TITLE: Signatures of soft selective sweeps predominate in the yellow fever mosquito *Aedes aegypti*

Remi N. Ketchum[1*], Daniel R. Matute[2], Daniel R. Schrider[1]

* Corresponding Author

Email: rketchum@ad.unc.edu

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27514-2916, United States of America

[2]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27514-2916, United States of America

ABSTRACT

The *Aedes aegypti* mosquito is a vector for human arboviruses and zoonotic diseases, such as yellow fever, dengue, Zika, and chikungunya, and as such poses a serious threat to public health. Understanding how *Ae. aegypti* adapts to environmental pressures—such as insecticides—is critical for developing effective mitigation strategies. However, most traditional methods for detecting recent positive selection search for signatures of classic "hard" selective sweeps, and to date no studies have examined soft sweeps in *Ae. aegypti.* This represents a significant limitation as this is vital information for understanding the pace at which an organism can adapt—populations that are able to immediately respond to new selective pressures are expected to adapt more often via standing variation or recurrent adaptive mutations (both of which may produce soft sweeps) than via *de novo* mutations (which produces hard sweeps). To this end, we used a machine learning method capable of detecting hard and soft sweeps to investigate positive selection in *Ae. aegypti* population samples from Africa and the Americas. Our results reveal that soft sweep signatures are significantly more common than hard sweeps in all population samples, including those that have experienced population bottlenecks, which may imply that this species can respond quickly to environmental stressors. This is a particularly concerning finding for vector control methods that aim to eradicate *Ae. aegypti* through the use of insecticides. We highlight genes under selection

that include both well-characterized and putatively novel insecticide resistance genes. These findings underscore the importance of using methods capable of detecting and distinguishing hard and soft sweeps, implicate soft sweeps as a major selective mode in *Ae. aegypti,* and highlight genes that may aid in the control of *Ae. aegypti* populations.

INTRODUCTION

A central goal of evolutionary biology is to uncover the genomic underpinnings of adaptation by identifying loci under positive selection. Researchers have made significant progress in this challenging task in a wide array of species. Some well-studied examples include: the hypoxia pathway gene *EPAS1* that was implicated in differences in hemoglobin concentrations at high altitude in humans (Huerta-Sánchez et al. 2014) or the *Ectodysplasin (EDA)* locus in threespine stickleback fish that facilitated their transition from marine to freshwater (Barrett et al. 2008). These discoveries, as well as many others, have been made possible by theoretical and methodological advances which have allowed for the detection and characterization of selective sweep signatures.

In addition to revealing the loci underpinning adaptation, selective sweep signatures can be informative about the rate of adaptation, provided one can identify the *type* of sweep that occurred. Traditionally, selective sweeps were thought to occur when a *de novo* beneficial mutation appears and then quickly rises in frequency until it becomes fixed in a population (Smith and Haigh 1974). This type of selective sweep is referred to as a hard sweep and is characterized by a lack of diversity (except that which is introduced by recombination or mutation during the sweep) around the vicinity of the selected site (Smith and Haigh 1974). A hard sweep may also result in increased linkage disequilibrium (LD) on either side of the sweep (Kelly 1997; Kim and Nielsen 2004), or a skew in allele frequency patterns (Fay and Wu 2000; Nielsen et al. 2005). In contrast, soft sweeps act on a mutation that is initially neutral (or weakly deleterious) and evolves under drift until a change in the selective environment causes the mutation to become beneficial and sweep to fixation (Orr and Betancourt 2001; Hermisson and Pennings 2005). They may also occur when a beneficial allele arises via mutation or migration during the selective phase of a sweep (Pennings and Hermisson 2006a; Pennings and Hermisson 2006b). In soft sweeps, the beneficial allele exists on multiple haplotypes that share a common ancestor prior to the onset of the sweep and so the

resulting skew in patterns of genetic diversity around the selected region may be both qualitatively different and less pronounced than that of a hard sweep (Prezeworski et al. 2005; Schrider et al. 2015). There is evidence that in at least some populations, adaptation proceeds mainly via soft sweeps (Garud et al. 2015; Schrider and Kern 2017; Xue et al. 2021). This in turn implies that these populations are able to adapt rapidly, as they need not wait for an adaptive mutation to arise (Karasov et al. 2010).

Sweep detection methods, which generally involve calculating single summary and/or test statistics (e.g., (Fay and Wu 2000; Kim and Nielsen 2004; Nielsen et al. 2005; Voight et al. 2006; Ferrer-Admetlla et al. 2014; Garud et al. 2015)), tend to be biased towards detecting hard sweeps because their more prominent genomic footprints are easier to detect (Teshima et al. 2006; Garud et al. 2015; Alachiotis and Pavlidis 2018; Weigand and Leese 2018). Even methods that are capable of detecting soft sweeps often lose potentially valuable information by reducing population genomic diversity to a single statistic (Schrider and Kern 2016). This endeavor is further complicated in populations with complex or unknown demographic histories because certain events, like bottlenecks, can mimic selective sweeps (Simonsen et al. 1995; Jensen et al. 2005; Nielsen et al. 2005). In the last few years, the field has made substantial progress in leveraging powerful machine learning algorithms that have been shown to outperform traditional methods (Flagel et al. 2019; Torada et al. 2019; Adrion et al. 2020; Caldas et al. 2022; Hejase et al. 2022; Mo and Siepel 2023; Whitehouse and Schrider 2023). ML approaches yield impressive discriminatory power as they either use a combination of several summary statistics as input (Pavlidis et al. 2010; Lin et al. 2011; Ronen et al. 2013; Schrider and Kern 2016; Alachiotis and Pavlidis 2018; Kern and Schrider 2018; Mughal et al. 2020; Arnab et al. 2023) or bypass this step entirely and train neural networks to work directly on genome alignments (Chan et al. 2018; Flagel et al. 2019; Torada et al. 2019; Adrion et al. 2020; Sanchez et al. 2021). Central to ML approaches are simulated training datasets which can be generated using estimated population-specific demographic histories. Doing so allows researchers to better approximate the distributions of patterns of polymorphisms produced by selection and neutrality even under more realistic and complex scenarios, thereby improving sweep-detection accuracy (Pybus Oliveras et al. 2015; Schrider and Kern 2016). Finally, some ML approaches have the added capacity to distinguish between hard and soft sweeps (Kern and Schrider 2018; Mughal and DeGiorgio 2019). This functionality is especially important in large, diverse populations where greater levels of standing

variation and/or a higher population-scaled mutation rate may mean that adaptation predominantly occurs through soft sweeps (Messer and Petrov 2013).

As a globally distributed vector species with large population sizes and high levels of genetic diversity, the mosquito *Aedes aeygpti* poses a major threat to public health (Kent et al. 2024). *Ae. aegypti* transmits yellow and dengue fever, Zika, and chikungunya, making a thorough understanding of natural selection in this species critical for guiding mitigation and vector control strategies. Due to a combination of climate change which is making regions more suitable for *Ae. aegypti,* increased globalization, and their impressive capacity as an invasive species, *Ae. aegypti* populations are expanding in parts of Europe, Central America, east Africa, the United States, and Canada (Ryan et al. 2019; Iwamura et al. 2020). In addition to novel environmental stressors experienced during range expansion, *Ae. aegypti* has also encountered strong selective pressure in the form of insecticides for the better part of the last century (Love et al. 2023; Ware-Gilmore et al. 2023). This, coupled with their population sizes, genetic diversity levels, capacity to develop resistance to insecticides in as little as 5 generations, suggests that signatures of positive selection may be prevalent (Martins et al. 2012; Matthews et al. 2018; Thornton et al. 2020; Kent et al. 2024). However, efforts to detect such signals are complicated by the species' varied and complex demographic histories.

Originally native to Africa, *Ae. aegypti* has, over the last four centuries, expanded its geographic range to include most of the world's tropical belt (Ryan et al. 2019; Rose et al. 2023). Its movement from Africa to the Americas involved multiple introductions, bottlenecks, and expansion events and was likely facilitated by the high ship volume to the Americas during the transatlantic slave trade (Tabachnick 1991; Brown et al. 2014; Powell et al. 2018). Population size estimates and nucleotide diversity in African *Ae. aegypti* population samples are approximately 10^5–10^6 and 0.0370, respectively (Rose et al. 2023; Kent et al. 2024). These estimates are generally larger than those in *Drosophila melanogaster*—a species known for its large population sizes and high levels of genetic diversity (Terhorst et al. 2017; Kapopoulou et al. 2018). Most of the genetic studies performed on *Ae. aegypti* has relied on only a small number of genes, mitochondrial DNA, microsatellites, or reduced representation sequencing, and only recently have whole genome-based resequencing and exome-based sequencing been performed (Gonçalves da Silva et al. 2012; Brown et al. 2014; Bennett et al. 2016; Crawford et al. 2017; Matthews et al. 2018; Kelly et al.

2021a; Lozada-Chávez et al. 2025). This is largely due to the size (1.3 Gb) and repetitive nature of its genome which poses a significant challenge (Matthews et al. 2018).

Because whole-genome data for *Ae. aegypti* have only recently become available (Rose et al. 2020; Kelly et al. 2021b; Love et al. 2023), very little is known about recent positive selection in this species. Currently, sweep scans in *Ae. aegpyti* based on single-summary statistical approaches have reported positive selection in genomic regions linked to human preference (Rose et al. 2020), increased tolerance to egg desiccation (Venkataraman et al. 2022), and insecticide resistance (Saavedra-Rodriguez et al. 2019; Love et al. 2023; Schmidt et al. 2024; Lozada-Chávez et al. 2025). However, given their large population sizes, high levels of genetic diversity, and growing evidence to suggest that IR-increasing alleles may spread via soft sweeps in several insect species (Garud et al. 2015; Avalos et al. 2017; Xue et al. 2021; Muralidhar and Veller 2022), it is likely that traditional methods may be missing informative signatures of selection. Here, we leverage a robust machine learning approach (Kern and Schrider 2018) to identify genomic targets of positive selection from environmental and anthropogenic stressors in *Ae. aegypti*. We find that we are able to accurately detect both hard and soft sweeps in four globally distributed population samples with complex population size histories. Importantly, we identify novel insecticide resistance (IR) genes which will improve our understanding of the mechanisms of IR evolution and uncover evidence that soft sweeps may play a central role in adaptation in *Ae. aegypti*.

RESULTS

Accurate detection of hard and soft selective sweeps in *Aedes aegypti*

To identify genomic targets of selective sweeps in the yellow-fever mosquito, *Aedes aegypti*, we used a machine learning approach, diploS/HIC (Kern and Schrider 2018), which seeks to discriminate between hard sweeps, soft sweeps, regions linked to hard or soft sweeps, and purely neutrally evolving regions. We used previously published genomic data from Love et al., 2023 that examined mosquito population samples from Brazil (Santarém), Colombia (Río Claro and Cali), Gabon (Franceville), Senegal (Ngoye), and Kenya (Kaya Bomu). For each population sample, we applied a diploS/HIC classifier that was trained on simulations following the population size trajectory estimates from Kent et al. 2024 that were estimated using SMC++ (Terhorst et al. 2017).

Our training simulations also included variation in mutation rates, recombination rates, and selective parameters such as the selection coefficient and the timing of the sweep (see Methods).

We assessed the performance of our classifiers through Receiver Operating Characteristic curves (ROC), Precision Recall (PR) curves, and confusion matrices (Figure 1). ROC curves measure true positive versus false positive rates across classification thresholds. An ideal classifier has a ROC curve that resembles a step-function with a true positive rate (or recall) as close to 1.0 as possible while achieving a false positive rate as close to 0 as possible. A PR curve measures the precision (or positive predictive value) against recall, with a good classifier being able to maintain high precision even at classification thresholds that yield high rates of recall. Here, our curves evaluate our models' performance on the binary task of distinguishing selective sweeps (whether hard or soft) from unselected regions (whether sweep-linked or fully neutrally evolving). The excellent quality of our classifiers is clearly discernible by the ROC and PR curves which respectively show an area under the curve (AUC) value greater or equal to 0.97 and an Average Precision (AP) value greater or equal to 0.95 in all population samples except for the two Colombia population samples.

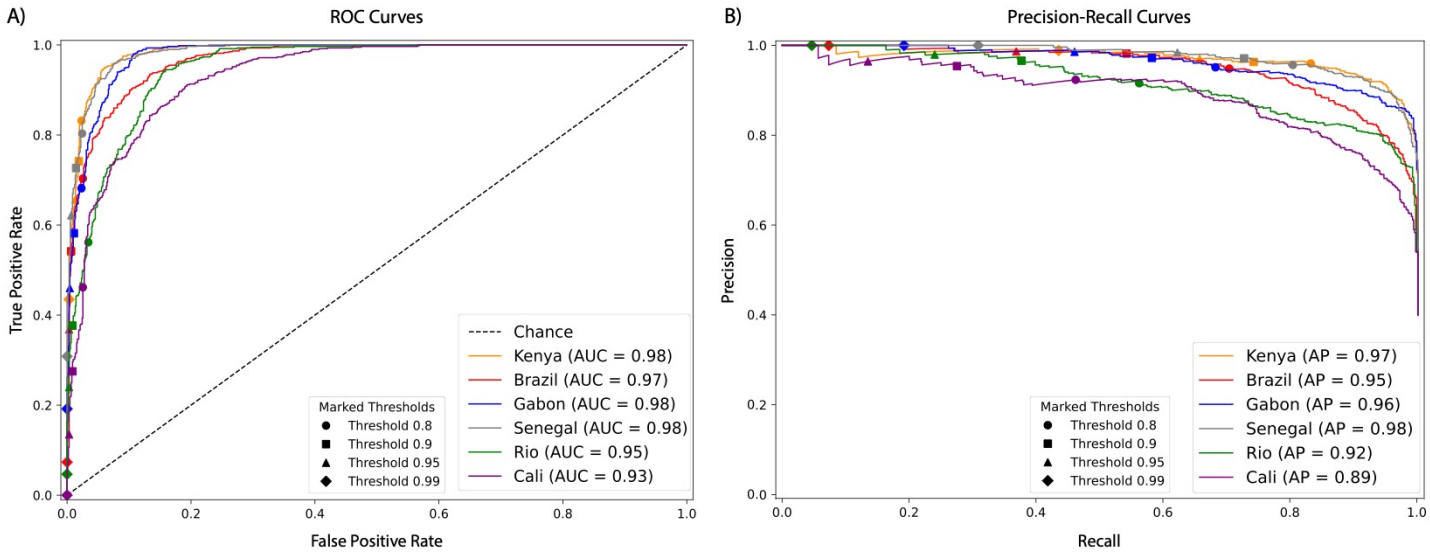

Figure 1. ROC curves and precision-recall curves summarizing the performance of each population samples' classifier. A) ROC curves showing the true and false positive rates for the binary classification task of distinguishing between selective sweeps (hard and soft) vs unselected regions (sweep-linked and neutral) with varying threshold cutoffs highlighted with different shapes. B) Precision-recall curves showing the classifiers performance at the same task (sweep vs unselected) with varying threshold cutoffs highlighted with different shapes. The precision is defined by the fraction of regions classified as sweeps that truly were sweeps and the recall is defined by the true positive rate. For the Cali classifier, there were no windows classified as a sweep with probability ≥0.99, so no marker is included on Cali's precision-recall curve for that threshold.

While ROC and PR curves are useful for summarizing the overall accuracy of a binary classifier, for multi-class inferential models like our trained diploS/HIC neural networks, these curves are limited because they collapse diploS/HIC's five classes down to two (selected and unselected regions), and thus they are not informative about the precise type of errors made within these two meta-classes. For example, if a hard sweep is correctly detected, how often is it misclassified as soft? Or, when neutrally evolving regions are misclassified as sweeps, are they generally classified as hard sweeps or soft sweeps? We therefore calculated confusion matrices, which show the fraction of test examples from each of the five classes that are correctly assigned to a given class. However, unlike the ROC and PR curves, the confusion matrices only show results from a single class membership probability threshold. We therefore made confusion matrices for

197 a range of thresholds for each population sample (see Methods). Note that the class membership
198 probabilities, which are produced by the softmax activation function from the final layer of the
199 diploS/HIC neural network, may not necessarily be well calibrated such that 80% of examples
200 classified as a hard sweep with 80% probability will be true hard sweeps.

201 In each population sample our ability to distinguish sweeps from neutral variation was
202 strong and improved even further when imposing sweep probability thresholds (e.g., < 7.5% of
203 neutral windows were misclassified as sweeping in each population sample with no sweep
204 probability cutoff versus < 4.2% in each population sample at the 0.80 cutoff; Supplemental
205 Figures 1-5). Our diploS/HIC classifiers had adequate power to distinguish hard from soft sweeps,
206 albeit with overall higher accuracy in African population samples than the South American
207 population samples (Supplemental Figures 1-5). For example, in Kenya with no sweep probability
208 cutoff, diploS/HIC correctly classified 88% of hard sweeps and 68% of soft sweeps (Supplemental
209 Figure 1); with a 0.80 threshold imposed this increased to 88% and 72%, respectively. This
210 improvement is observed because simulations classified as sweeps but with a combined sweep
211 probability lower than the classification threshold are treated as uncertain and are thus omitted
212 from the calculation, and those high-confidence predictions that remain are more likely to be
213 accurate. When applying cutoff values of 0.90, 0.95, and 0.99 to the Kenya classifier, accuracy
214 values for hard sweeps stayed relatively constant at 88%, 88%, and 90%, respectively, as did soft
215 sweeps at 71%, 69%, 66%. Conversely, for Cali (Supplemental Figure 1) with no cutoff,
216 diploS/HIC correctly classified 80% of hard sweeps and 33% of soft sweeps and increasing the
217 threshold to more stringent values did not dramatically improve these values (e.g., at 0.95, 77% of
218 hard sweeps and 37% of soft sweeps were recovered). At the highest threshold of 0.99, there were
219 no hard or soft sweep simulations for Cali passing the threshold, so the fraction of sweeps correctly
220 predicted could not be calculated. Performance for the other Colombian sample (Río Claro) was
221 similarly poor (Supplemental Figure 1-5), while performance for the Brazil, Gabon, and Senegal
222 classifiers was strong (obtaining similar false positive rates as for Kenya but slightly lower
223 accuracy overall). Given the comparatively poor performance of Colombia's classifiers, we chose
224 to focus our subsequent analyses on the other four population samples (Brazil, Gabon, Kenya, and
225 Senegal), for which our simulated test data suggested that we could detect and classify sweeps
226 with sufficient accuracy.

227 **Soft sweeps appear to predominate over hard sweeps in *Ae. aegypti***

228 We classified a total of 4598, 4599, 4599, and 4599 windows that passed our data filtering
229 cutoffs in the Brazil, Gabon, Senegal, and Kenya population samples, respectively. Of these
230 windows, 165 were classified as sweeps in Brazil, compared to 150 in Gabon, 79 in Senegal, and
231 363 in Kenya (Table 1). To construct higher-confidence sets of sweep candidates, we then imposed
232 increasing posterior probability cutoffs on these windows (Table 1). For example, when using a
233 cutoff of 0.95, there were 24 sweep windows in Brazil, 18 in Gabon, 17 in Senegal, and 65 in
234 Kenya. These windows comprise a set of 106 distinct candidate sweeps, some of which were
235 shared by more than one population sample: 1 (0.94%) was shared across all population samples,
236 1 (0.94%) was shared among the African population samples, and 93 (87.7%) were population-
237 specific sweeps (Figure 2). The remaining 11 (10.4%) sweeps were present in more than one
238 population but did not fit into the categories above.

239

240 **Table 1.** The number of hard and soft sweeps per population sample at different posterior
241 probability thresholds.

Pop Sample	No Cutoff		≥0.80		≥ 0.90		≥0.95		≥0.99	
	Hard	Soft (%)	Hard	Soft (%)	Hard	Soft (%)	Hard	Soft (%)	Hard	Soft (%)
Brazil	8	157 (95.2%)	4	75 (94.9%)	1	44 (97.8%)	1	23 (95.8%)	0	6 (100%)
Gabon	5	145 (96.7%)	0	66 (100%)	0	36 (100%)	0	18 (100%)	0	3 (100%)
Senegal	8	71 (89.9%)	5	37 (88.1%)	4	24 (85.7%)	2	15 (88.2%)	0	3 (100%)
Kenya	8	355 (97.8%)	5	149 (96.8%)	4	90 (95.7%)	4	61 (93.9%)	1	21 (95.5%)

242

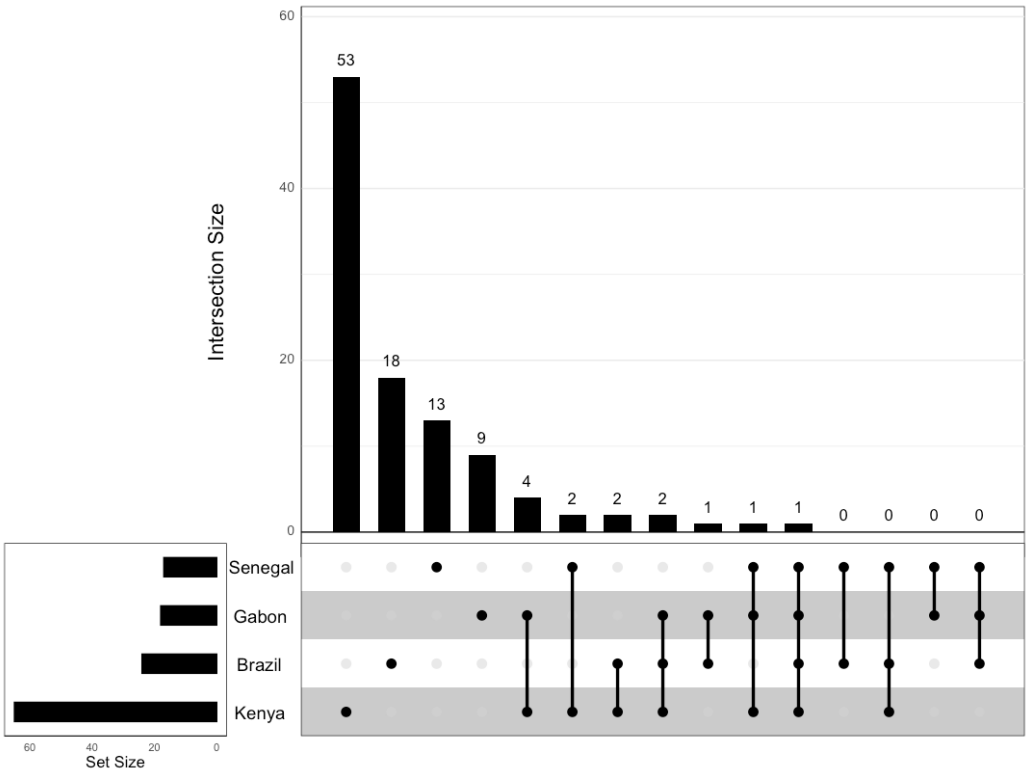

Figure 2. Upset plot showing the intersection of selective sweep locations across the four population samples in this study based on the posterior probability cutoff of a sweep of 0.95.

Across all population samples, the number of soft sweeps was substantially higher than the number of hard sweeps and this pattern did not change with more stringent posterior probability thresholds (Table 1). We further evaluated the robustness of this finding by applying a conservative approach using the error rates in our confusion matrices to estimate the proportion of sweeps that were soft after accounting for hard sweeps that may have been misclassified as soft, and to estimate potential false positive soft sweeps that are truly neutrally evolving—this approach may underestimate the fraction of soft sweeps because we ignore the possibility of soft sweeps being misclassified as hard (see Methods). Even after accounting for these estimated error rates, we found that soft sweeps predominate at every threshold and that the relative proportion of soft sweeps generally increases or remains stable with more stringent threshold cutoffs. Indeed, at every threshold and every population sample, more than 72% of all sweeps are predicted to be soft

(Supplemental Table 1). In all population samples other than Senegal, this (conservative) estimate was even higher: > 83.5% across all thresholds.

To further measure the degree of confidence in our sweep calls, we calculated *q*-values for each sweep window based on the combined posterior probabilities of hard and soft sweeps and our predicted false positive rates obtained from simulations for each population sample (Methods; see Supplemental Table 2 for the sweep candidates and their associated *q*-values). Notably, for those sweep windows with a combined posterior probability of ≥ 0.95, our *q*-value estimate could not be distinguished from zero (i.e., our neutral test simulations never achieved a sweep probability above this threshold) in all population samples except for Kenya (where sweeps passing the 0.95 threshold had an estimated *q*-value of ~0.4).

Insecticide Resistance Genes as Likely Targets of Selection

When examining candidate selective sweeps, we separately considered two sets of regions: shared sweeps (those with a posterior sweep probability of ≥0.95 in more than one population sample) and population-specific sweeps (those where only one population sample had a sweep probability ≥0.95). These two sets contained 13 and 93 sweep windows, respectively. Many of these candidate regions include genes with functions potentially linked to insecticide resistance. To better interpret their relevance, we categorized these genes into two groups: **well-characterized IR genes** and **putative or emerging IR candidates**. We assigned genes to these groups based on the nature and extent of evidence for their relevance to IR in the literature, including: (1) functional similarity to known IR genes, (2) differential expression in response to insecticide exposure, and (3) signatures of selection reported in other insect species under insecticide pressure. It is important to note that while we can identify compelling candidate genes within these sweep regions, there are often multiple genes in and around these windows, and we thus cannot conclude with certainty which gene is the target of selection.

Shared Sweep Windows Containing Putative or Emerging IR candidates

Of the 13 windows that were shared between 2 to 4 population samples, 3 windows contain candidate *Ae. aegypti* insecticide-resistance genes (see Supplemental Table 3 for the complete list). First, we found a high-confidence soft sweep shared in Kenya and Gabon (chr2: 452,250,001-452,500,000) that contains an ankyrin repeat domain-containing protein 29. This same region was

287 predicted to be a soft sweep in Brazil but its sweep probability of 0.94 did not meet our threshold
288 of 0.95. The ankyrin protein domain is a common protein-protein interaction motif that is present
289 in numerous structural proteins and is involved in cytoskeletal anchoring and mechanosensation.
290 Importantly, ankyrin proteins have been shown to directly interact with the well-established target
291 for insecticides, the voltage-gated sodium channel VGSC (Williamson et al. 1996). In a study on
292 pyrethroid resistant *Ae. aegypti* from Mexico, genes containing the ankyrin-domain were highly
293 associated with pyrethroid resistance (Campbell et al. 2019). Similarly, in pyrethroid-resistant
294 populations of the mosquito *Anopheles funestus* from Senegal, an ankyrin repeat domain protein
295 was one of the most highly overexpressed genes (Samb et al. 2016). In *Ae. aegypti*, an ankyrin
296 repeat domain-containing protein was found to be associated with pyrethroid resistance (Cosme et
297 al. 2022), and signatures of selection in two regions containing ankyrin genes (along with other IR
298 candidates) was also reported by Love et al., (2023). Similar patterns have also been observed in
299 other insect species (Kwiatkowska et al. 2013; Gouesbet et al. 2025), further highlighting the role
300 of these gene families in pyrethroid resistance.

301 The other two shared sweep windows containing putative IR genes were also classified as
302 sweeps in Kenya and Gabon. The first contained a gene that encodes for the RNA binding protein
303 Split ends (*spen;* chr2: 143,500,001-143,750,000) which provides a protective role against
304 cytotoxicity from the herbicide paraquat in *Drosophila* (Girard et al. 2020; Bresgen et al. 2023).
305 Although paraquat is an herbicide, it is known to cause oxidative stress in the mosquito *An.*
306 *gambiae* (Champion and Xu 2018; Tarimo et al. 2018). The second window contains *thioredoxin-*
307 *2* (chr2: 45,000,001-45,250,000), a key mitochondrial protein that regulates cellular redox and is
308 protective against oxidative stress, a common consequence of insecticide exposure, in *Drosophila*
309 (J. Svensson and Larsson 2007). Interestingly, the thioredoxin system recycles oxidized
310 glutathione to reduced glutathione, which can then be used by the well-established IR genes
311 glutathione S-transferases (GSTs) for detoxification (Tarimo et al. 2018). Other components of the
312 thioredoxin system, like thioredoxin peroxidase, have been shown to protect insects from
313 insecticide-induced oxidative injury (Zhao et al. 2022; Gao et al. 2025) and *thioredoxin* is listed as
314 of the candidate IR genes on the *Anopheles gambiae* 1000 Genomes Project Selection Atlas
315 (Clarkson et al. 2020).

316 *Population-Specific Sweep Windows Containing Well-Characterized IR Genes*

317 We classified 93 windows as a sweep with ≥ 95% probability in a single population sample.
318 Of these 93 windows, six contain genes with well-characterized IR-related functions. We describe
319 three of these genes below (See Supplemental Text for descriptions of the other three genes) and
320 explore their role in insecticide resistance.

321 One window classified as a soft sweep in Senegal, located at chr1:271,250,001-
322 271,500,000, contains three cytochrome P450 genes: *Cyp 6a8, Cyp 6a13, Cyp 6a14* (Figure 3).
323 Cytochrome P450 monooxygenases (CYPs) are critical for the metabolic detoxification and have
324 been implicated in insecticide resistance in many species, including mosquitoes (Balabanidou et
325 al. 2016; Rahman et al. 2021; Yang et al. 2021). *Cyp6a8* specifically has been shown to be
326 upregulated in response to *Piper nigrum* and DDT in *D. melanogaster* (Maitra et al. 1996; Jensen
327 et al. 2006). Notably, this sweeping window also contained several SweepFinder Composite
328 Likelihood Ratio (CLR) peaks (Nielsen 2005), with the highest value > 750. We also detected a
329 soft sweep in Kenya in a window that contains *cytochrome b5 reductase 4* (chr2:156,500,001-
330 156,750,000). Cytochrome P450's require cytochrome b5 reductase to function as an electron-
331 transfer intermediate and the latter is therefore likely to be involved in IR (Zhao et al. 2012).
332 Moreover, it has been shown to be upregulated in response to phenobarbital in the cotton bollworm,
333 *Helicoverpa armigera* (Zhao et al. 2012) and has been linked to cyantraniliprole resistance in the
334 tomato pest *Tuta absoluta* (Ullah et al. 2025).

335 We detected a soft sweep in Brazil in a window containing a glutathione S-transferase
336 (chr2:446,250,001-446,500,000), *GSTD7*, and that had a corresponding CLR peak with values >
337 460. Glutathione S-transferase epsilon clusters (GSTe) are another group of metabolic enzymes
338 that are well-known for their role in resistance to several classes of insecticides (Enayati et al.
339 2005), and other *GSTe* genes have previously been found to exhibit strong signatures of positive
340 selection (Love et al. 2023; Schmidt et al. 2024). *GSTD7*, specifically, was found to be under
341 selection in the fall armyworm, *Spodoptera frugiperda* (Tessnow et al. 2025). Moreover,
342 knockdown of GSTD7 resulted in higher mortality post-imidacloprid exposure in the silver
343 whitefly *Bemisia tabaci* (He et al. 2018), and *GSTD7* is highly expressed in *Drosophila suzukii*
344 after exposure to malathion (Hamby et al. 2013).

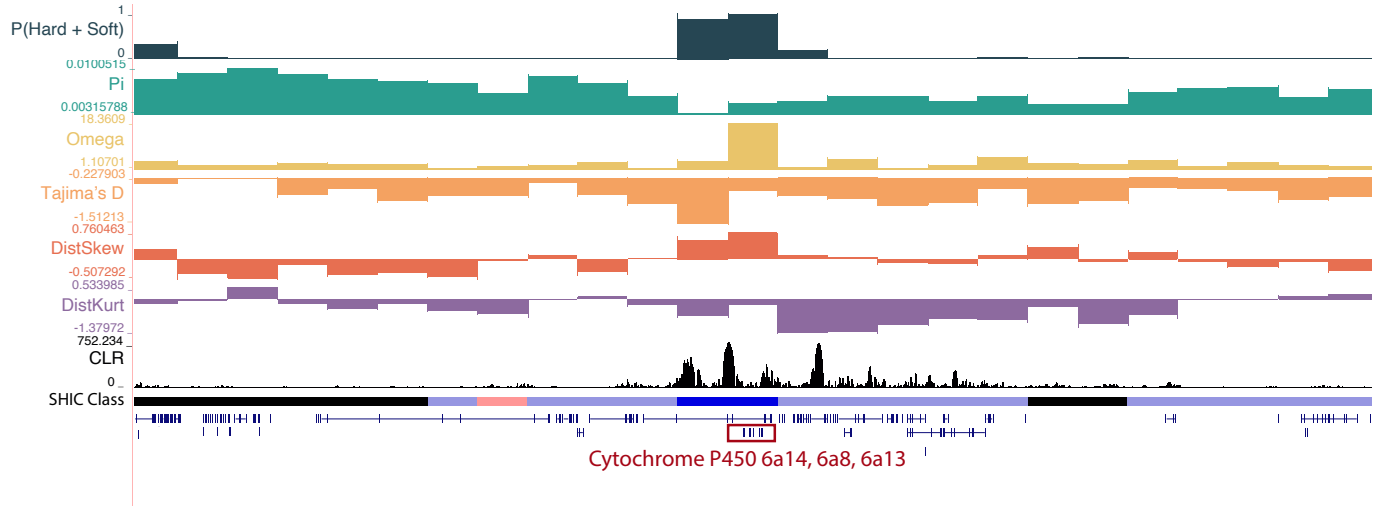

Figure 3. A soft sweep in Senegal at three cytochrome P450 genes: *Cyp6a14*, *Cyp6a8*, and *Cyp6a13*. The diploS/HIC classification track shows the class with the highest posterior probability, with soft sweeps as dark blue, soft sweep-linked regions as light blue, hard sweeps as red, hard sweep-linked as light red, and neutrally evolving regions in black. Above the diploS/HIC classifications are a subset of the summary statistics used by the classifier.

Population-Specific Sweeps Windows with Putative or Emerging IR Candidates

Of these 93 windows that were classified as high-confidence sweeps, 13 contain genes with putative or emerging IR-related functions. We describe seven of these genes below and explore their role in insecticide resistance (see Supplemental Table 4 for the complete list). The remaining six genes and associated role in IR are included in the Supplementary Text.

We found three high-confidence windows that were classified as soft sweeps in Kenya which contained *arrestin* (chr2:246,250,001-246,500,000) and overlapped with the *sodium/potassium/calcium exchanger NCKX30C* (chr2:288,750,001-289,000,000; Figure 4). In vertebrates, arrestins regulate the signaling and trafficking of G-protein-coupled receptors (Gurevich and Gurevich 2006), and in the mosquito *Culex pipiens pallens*, *arrestin* expression was higher in deltamethrin-resistant strains and knockdown via siRNA resulted in decreased viability after deltamethrin treatment (Sun et al. 2012). *NCKX30C* is likely involved in the maintenance of calcium homeostasis (Haug-Collet et al. 1999; Webel et al. 2002) and was found to be differentially regulated in insecticide resistant strains of *An. gambiae* (Vontas et al. 2005). The last window classified as a hard sweep in Kenya contains *mitochondrial protoheme IX farnesyltransferase* (chr2:262,250,001-262,500,000), or *COX10*. *COX10* encodes for an enzyme that catalyzes the

367 farnesylation of heme, which plays a crucial role in cytochrome c oxidase (COX) function (Diaz
368 et al. 2006). COX genes have been linked to insecticide resistance in a number of species (Steele
369 et al. 2018), including *D. melanogaster* (Song and Scharf 2009), the housefly *Musca domestica*
370 (Sacktor 1951), the German cockroach *Blattella germanica* (Pridgeon and Liu 2003), and *Ae.*
371 *aegypti* (Pridgeon et al. 2009).

372 In Brazil, we found sweeping windows that contain or overlap the *muscle calcium channel*
373 *subunit alpha-1* (chr2:225,750,001-226,000,000), *potassium voltage-gated channel subfamily*
374 *KQT member 1* (chr2:144,750,001-145,000,000), *neuroligin-1* (chr3:351,750,001-352,000,000),
375 and *thioredoxin-related transmembrane protein 1* (chr3:12,250,001-12,500,000). Of these four
376 windows, three are classified as soft sweeps while the second window is classified as a hard sweep.
377 Further, the first, third, and fourth window all contain CLR peaks with values greater than 475,
378 280, and 900, respectively. The second window was also initially classified as a soft sweep in
379 Kenya, although the posterior probability (0.83) did not meet our threshold. Although the role of
380 both the *muscle calcium channel subunit alpha-1* and the *potassium voltage-channel subfamily*
381 *KQT member 1* in IR is not well understood, both mechano-susceptible and voltage-gated channels
382 are confirmed targets of various insecticide and pesticide classes. *Neuroligin-1* is a synaptic cell-
383 adhesion molecule (Song et al. 1999) that is highly associated with pyrethroid resistance in *Ae.*
384 *aegypti* (Campbell et al. 2019), likely a result of insecticides disrupting excitatory synapses.
385 Similar to the *thioredoxin-2* sweep window described above, *thioredoxin-related transmembrane*
386 *protein 1* (Trx1) also plays a role in protection against oxidative stress in insects (Zhang et al.
387 2015) and has been shown to be involved in antioxidant defense in *Apis cerana cerana* (Yao et al.
388 2014).

389 In the Supplementary Text, we discuss additional candidate sweep regions containing the
390 following putative IR genes: *acyl-CoA synthetase family member 4*, *fatty acyl-CoA reductase 1,*
391 *membrane-associated progesterone receptor component 1, D(2) dopamine receptor*
392 *A, transcription factor grauzone,* and *NADH dehydrogenase 1 beta subcomplex subunit 5,*

393

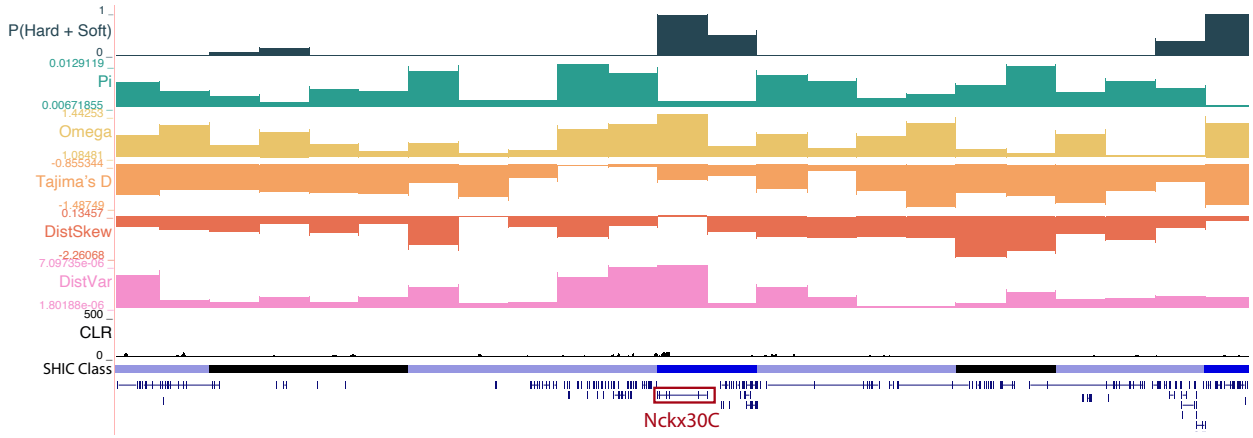

Figure 4. A soft sweep in Kenya at the *sodium/potassium/calcium exchanger NCKX30C*. The diploS/HIC classification track shows the class with the highest posterior probability, with soft sweeps as dark blue, soft sweep-linked regions as light blue, hard sweeps as red, hard sweep-linked as light red, and neutrally evolving regions in black. Above the diploS/HIC classifications are a subset of the summary statistics used by the classifier.

DISCUSSION

Natural selection leaves conspicuous footprints in the genome, and our ability to accurately detect these signatures has significant implications for public health, agriculture, and conservation. While early methods focused heavily on detecting hard selective sweeps, growing evidence supports the widespread occurrence of soft sweeps, particularly in large, diverse populations. Soft sweeps leave a more subtle signature than hard sweeps and are therefore more difficult to detect with traditional approaches. Machine learning has emerged as a powerful tool in this space, well-suited to detecting high-dimensional signals of selection and capable of accounting for complex demographic histories that confound inferences. This is an important advancement for species like *Aedes aegypti*—a major disease vector with a global impact on public health—which poses a significant challenge to evolutionary analyses due to its high degree of genetic variation and complex demographic history. Here we leverage the machine learning tool diploS/HIC (Kern and Schrider 2018) and recently published demographic history estimates (Kent et al. 2024) to uncover the targets of recent positive selection in *Ae. aegypti* and provide evidence that soft sweeps may play a central role in adaptive evolution in this important vector species.

This work contributes to a growing body of literature that challenges the traditional view that selective sweeps mainly act on a single-origin *de novo* mutation. Indeed, our work mirrors the

418 findings of similar studies performed on humans (Schrider and Kern 2017), *D. melanogaster*
419 (Garud et al. 2015), *Anopheles gambiae* (Weedall et al. 2020; Xue et al. 2021), and HIV (Feder et
420 al. 2016), that implicate soft selective sweeps as the dominant mode of adaptation (although these
421 studies have proved controversial; see (Harris et al. 2018; Schrider and Kern 2018; Feder et al.
422 2021; Garud et al. 2021; Johri et al. 2022)). In fact, in all population samples, and at every sweep
423 probability threshold, the percentage of sweeps that were soft was always > 85%. When we applied
424 a conservative approach designed to account for the potential impact of false positives and of soft
425 sweeps being misclassified as hard (while ignoring the possibility of misclassification in the
426 opposite direction), the percentage of soft sweeps was still > 72% in all population samples and
427 thresholds. This suggests that selection in *Ae. aegypti* can act on standing genetic variation instead
428 of waiting for novel mutations to arise, and *Ae. aegypti* populations are therefore able to rapidly
429 respond to new selective pressures like insecticides. Soft sweeps even predominated in Brazil—a
430 population sample that has experienced a protracted bottleneck associated with the introduction
431 from Africa to the Americas (Rose et al. 2023; Kent et al. 2024). This suggests that even
432 bottlenecked populations of *Ae. aegypti* may have sufficient standing variation (or a high enough
433 population-scaled mutation rate; see below) to facilitate rapid adaptation. Out of every population
434 sample, Kenya had substantially more high-confidence soft selective sweeps predicted than any
435 other population sample (61 in Kenya, compared to 23, 18, 15 in Brazil, Gabon and Senegal,
436 respectively, when considering sweeps with a posterior probability ≥ 0.95). This is in line with the
437 findings presented in Kent et al. (2025) where Kenya was inferred to have a larger effective
438 population size than the other population samples and therefore would be expected to experience
439 efficacious natural selection and more soft sweeps.

440 Our findings are predicated on modeling soft sweeps as selection on standing variation,
441 which is plausible given the high levels of genetic diversity observed. Alternatively, selection on
442 recurrent *de novo* mutations can also produce sweep signatures similar to that of a soft sweep—
443 albeit typically less pronounced–with multiple haplotypes carrying the beneficial allele (Pennings
444 and Hermisson 2006b). Such sweeps are especially likely in species where the number of new
445 mutations entering the population each generation is large, either due to large population size
446 and/or a high spontaneous mutation rate (Pennings and Hermisson 2006b; Karasov et al. 2010;
447 Garud et al. 2015). As a result, our machine learning models may detect signals of both selection
448 on standing variation and on recurrent mutations, both of which are a sign of a population that

449 need not wait for long periods for a *de novo* beneficial mutation to arise following a change in the
450 selective environment.

451 On the other hand, there are several phenomena that may result in signatures of soft sweeps
452 even in the absence of rapid adaptation. First, patterns of genetic diversity in a genomic region
453 flanking a hard sweep can mirror that of a soft sweep (the "soft shoulder" effect; (Schrider et al.
454 2015)). However, this is an unlikely explanation for our results given that we detected very few
455 hard sweeps in our population samples, and diploS/HIC is designed to account for this shoulder
456 effect by examining polymorphisms across a larger window containing the sweep (Schrider and
457 Kern 2016). Second, allelic gene conversion can "soften" hard selective sweeps by transferring the
458 beneficial mutation onto multiple haplotypes (Jones and Wakeley 2008; Schrider et al. 2015;
459 Schrider 2023). This is more common in large populations where there is more time for gene
460 conversion events to occur (Schrider 2023). Third, the dearth of hard sweeps across all population
461 samples may result from the inherent difficulty in accounting for spatial population structure when
462 detecting selective sweeps. In low-dispersal scenarios (which are common in species with wide
463 geographic ranges) an adaptive mutation cannot rise in frequency as rapidly as it would in a
464 panmictic population. As a result, hard selective sweeps can become enriched in intermediate-
465 frequency variants which can cause them to resemble soft sweeps (Chotai et al. 2024). Thus, the
466 expectation for large populations like those examined here might be that we would detect more
467 soft sweeps than hard sweeps even without selection acting on standing genetic variation or
468 recurrent mutations. Nonetheless, there is direct evidence that *Ae. aegypti* can adapt rapidly to
469 insecticides (Martins et al. 2012), and Lozada-Chávez et al. (2025) recently reported evidence of
470 selection in out-of-Africa population samples on variants that are segregating in Africa. Thus, true
471 soft sweeps may therefore represent the most parsimonious explanation for the patterns we
472 observe.

473 While the interpretation of soft sweep signatures is not straightforward, the above
474 arguments suggest that it is essential for selection scans in large/diverse populations to consider
475 soft sweeps in order to detect a larger fraction of the targets of recent positive selection. Previous
476 studies have detected selection on several IR-related genes, including the voltage-gated sodium
477 channel gene (*VGSC*), glutathione S-transferases (GSTs), *ace-1,* carboxylesterases, and many
478 cytochrome P450 genes (Love et al. 2023; Schmidt et al. 2024). While these studies have provided

479 insight into how mosquitoes develop resistance to insecticides (e.g., through metabolic
480 detoxification or variation in the insecticide target location (Kliot and Ghanim 2012; Yahouédo et
481 al. 2017)) they did not consider the possibility of soft sweeps, and thus may have missed additional
482 targets of selection. Here we identified both shared and population-specific sweep windows that
483 were almost entirely classified as soft sweeps (96.8% and 93.5%, respectively). Within these
484 windows, we identified 22 genes with either well-characterized or putative roles in IR. To the best
485 of our knowledge, several of our highlighted IR candidates (e.g., *ankyrin repeat domain-*
486 *containing protein 29*, *split ends*, *sodium/potassium/calcium exchanger NCKX30C, muscle*
487 *calcium channel subunit alpha-1*, or *potassium voltage-gated channel subfamily KQT member 1*),
488 have not been reported as IR genes in *Ae. aegypti* and thus represent novel targets for functional
489 validation. Interestingly, of our 13 shared selective sweep windows, only two of them had
490 corresponding CLR peaks (defined by a CLR value > 200) and in our population-specific dataset,
491 only 22% had corresponding CLR peaks. These results make sense given that CLR is based on a
492 hard selective sweep model and the majority of our selective sweeps were classified as soft, further
493 highlighting the need for methods specifically designed to detect the signatures of soft selective
494 sweeps in order to more fully appreciate the landscape of adaptive evolution.

495 Although we were able to identify several novel IR candidate regions by searching for
496 selective sweeps, we note that IR is a polygenic trait. One might therefore expect polygenic
497 selection, in which the selected phenotype moves towards its optimum through the combined effect
498 of small shifts in the frequencies of a large number of small-effect alleles (Berg and Coop 2014;
499 Stephan 2016; Höllinger et al. 2019; Hayward and Sella 2022), to be the primary mode of
500 adaptation in *Ae. aegypti*. Indeed, polygenic adaptation has been implicated in the evolution of
501 complex traits, including insecticide resistance (Chen et al. 2023; Hobbs et al. 2023). However,
502 selective sweeps may still occur during a polygenic shift to a new fitness optimum (Thornton
503 2019), which may explain our success in finding sweep candidates at both known and potentially
504 novel IR loci, even though diploS/HIC would be underpowered to detect more subtle allele
505 frequency shifts that may also be contributing to adaptation for IR and other traits in *Ae. aegypti*.

506 While our examination of candidate sweep regions primarily focuses on genes associated
507 with IR, we note that other traits may also be subject to recent positive selection. Given *Ae.*
508 *aegypti's* broad and expanding geographic distribution, robust invasive capacity, and the impact of

509 ongoing climate change, it is highly likely some of the sweeps reported here are unrelated to IR.
510 This is particularly relevant for the population samples from Brazil and Senegal, which are human
511 specialists unlike the Kenya and Gabon samples which exhibit the ancestral generalist feeding
512 behavior. The Senegal and Brazil samples thus have experienced additional selective pressures
513 associated with domestication (e.g., adapting to urban environments and acquiring a preference
514 for feeding on human blood; (Rose et al. 2020; Lozada-Chávez et al. 2025)). Although we do not
515 analyze non-IR genes in this work, we provide a comprehensive list of all sweep calls with ≥0.95
516 confidence in any population sample in Supplemental Tables 3 and 4. Our focus on IR-related
517 genes reflects both the urgency of addressing developing IR, as well as the rich body of existing
518 literature on these genes, which helps facilitate the challenging task of interpreting sweep signals.

519 In summary, in this study we leveraged whole-genome data from multiple *Ae. aegypti*
520 population samples together with a powerful machine learning tool to enrich our understanding of
521 the genomic targets and mode of positive selection in this important vector species. We discovered
522 several novel insecticide resistance candidate genes, suggesting that current surveillance efforts
523 may be missing variants/loci that contribute substantially to segregating resistance. More broadly,
524 our results suggest that insect vectors, particularly those with large, diverse populations, can
525 rapidly adapt to insecticides and other control strategies, perhaps through selection on standing
526 genetic variation, recurrent originations of adaptive alleles, and/or polygenic selection. This in turn
527 may imply the presence of many additional adaptive variants that have been missed by previous
528 scans for selection in other insect pests.

529

530 **MATERIALS AND METHODS**

531 **Sampling, Sequencing, and Variant Calling**

532 In this study, we used a previously curated whole genome sequencing dataset of 104
533 individuals from (Love et al. 2023). This dataset includes 18 samples from Santarém, Brazil; 13
534 from Franceville, Gabon; 19 from Kaya Bomu, Kenya; and 20 from Ngoye, Senegal, all taken
535 from Rose et al. (2020), as well as 10 and 24 samples from Cali and Río Claro, Colombia,
536 respectively, sequenced in Love et al. (2023). Detailed methods for DNA extractions, sequencing,
537 and variant calling and filtering are available in (Love et al. 2023). Briefly, reads from all

individuals were aligned to the AaegL5 reference genome (cite; NCBI accession GCF_002204515.2) using bwa-mem2 v. 2.1 (Vasimuddin et al. 2019), indels were removed and SNPs were filtered following GATK's best practices (Koboldt 2020). Repetitive regions, non-biallelic SNPs, and non-uniquely mappable regions were removed and genotypes with qualities less than 20 were masked.

Using Classifiers to Detect Selective Sweeps

To detect selective sweeps, we used diploS/HIC, a supervised machine learning approach that has been shown to be robust to nonequilibrium demography and has the capacity to distinguish between hard and soft sweeps in unphased data (Schrider and Kern 2016; Kern and Schrider 2018). In short, diploS/HIC classifies individual genomic windows into five categories (hard sweep, soft sweep, hard-linked, soft-linked, or neutral) based on a vector of windowed and transformed summary statistics that are input into a convolutional neural network (CNN). While this classification approach allows for robust inference based on many features jointly, it requires training datasets that consist of examples known to belong to each class. Therefore, we generated the training data using discoal (Kern and Schrider 2016). DiploS/HIC then subdivides each simulated region into a number of equally sized, adjacent subwindows; we used the default value of 11 subwindows. For hard and soft sweeps, we selected the location of selection from a uniform distribution within each of the 11 subwindows. Specifically, we conducted 3,000 simulations where selection occurred uniformly within the leftmost subwindow, another 3,000 for the second subwindow, and so on for all 11 subwindows.

We sought to train each population sample's classifier on simulations under demographic models that provide a reasonable fit to that population sample's genomic data. To this end, we used Kent et al (2025)'s population size histories, which were estimated using SMC++ for each population sample examined here. To account for variation in mutation and recombination rates across the genome, we allowed these parameters to vary across our training replicates. We set our mutation rate (μ) to vary uniformly from 8.82×10^{-10} to 8.82×10^{-9}, giving a mean value of 4.85×10^{-9} which matches the estimate from Rose et al. (2023). We drew the recombination rate (r) from an exponential distribution with mean 4.85×10^{-9} (with values greater than $3*r$ not allowed), choosing this mean value because the average recombination rate in *Ae. aegypti* appears to be similar to our mean value of μ (Matthews et al. 2018; Chen et al. 2022); note that because this is a truncated

568 exponential distribution the true mean value of r will be somewhat lower than our mean parameter
569 value.

570 For the simulations involving sweeps, there are several additional parameters whose
571 distributions are unknown, like the strength and timing of selection. For these, we drew from a
572 wide distribution to ensure that the range of parameters seen during training encompasses those
573 likely encountered during downstream inference. Specifically, we drew the selection coefficient
574 from a log-uniform distribution ranging from 0.005 to 0.05 and the time of fixation of the beneficial
575 allele, from $U(0, 0.001)$. For soft and soft-linked sweeps, we simulated selection on a previously
576 neutral standing variant, drawing the frequency of the previously neutral allele at the onset of
577 selection, (f_0), from $U(0, 0.05)$. In total, we generated 3,000 simulated replicates for each class for
578 each population. Because our goal was to identify strong selective sweeps that affect a wide stretch
579 of the chromosome, we sought to simulate 2.75 Mb regions, but discoal was unable to simulate
580 such regions due to memory constraints. We therefore simulated 550 kb regions and decreased the
581 selection coefficient to $U(0.001, 0.01)$, resulting in simulated sweeps with the same ratio of s/r as
582 we would have obtained from 2.75 Mb simulations with $s \sim U(0.005, 0.05)$.

583 As described above, in our real data we masked genomic positions that had genotype
584 qualities below 20. To incorporate heterogeneity in data quality into our training/test data, for each
585 simulated window, we randomly selected a corresponding window from our empirical dataset and
586 masked the same sites in the simulated window that had been masked in the empirical window.
587 This ensured that our masking procedure affected our simulated data in the same way as our real
588 data. Feature vectors were then calculated for each simulation replicate. These features vectors
589 measure the spatial patterns of a number of population genetic summary statistics, including π
590 (Tajima 1983), θ_ω (Watterson 1975), Tajima's D (Tajima 1989), the number of distinct haplotypes,
591 average haplotype homozygosity (also referred to as H_1 by Garud et al. 2015), H_{12} and H_2/H_1
592 (Garud et al. 2015), Z_{ns} (Kelly 1997), and the maximum value of Kim and Nielsen's $\boldsymbol{\omega}$ (Kim and
593 Nielsen 2004); note that the latter two statistics, which measure linkage disequilibrium, were based
594 on calculations of Rogers and Huff's estimator of LD for unphased data (Rogers and Huff 2009).
595 In addition to these commonly used statistics, diploS/HIC also includes in its feature vector
596 estimates of the variance, skewness, and kurtosis of the distribution of the number of pairwise
597 differences for all pairs of individuals in the sample; these summaries are useful for detecting hard

598 and soft sweeps and discriminating between them (Kern and Schrider 2018). DiploS/HIC
599 subdivides the simulated region into 11 adjacent windows and calculates each of these statistics in
600 each window, dividing that value by the sum of values for that statistic across all 11 subwindows.
601 If the smallest of these 11 values for a given statistic is less than 0, the value for this subwindow
602 is increased to zero, and the value for each other subwindow is increased by the same amount (i.e.,
603 the absolute value of the smallest subwindow is added to each subwindow in such cases).

604 After computing summary statistics, for each population sample we used diploS/HIC's
605 makeTrainingSets command to combine our simulations into a training set of 13,500 examples:
606 2,700 each for the hard sweep class (consisting of simulations where a hard sweep occurred in the
607 central subwindow), the hard-linked class (where a hard sweep occurred in any subwindow but the
608 central one), the soft sweep class (where a soft sweep occurred in the central subwindow), the soft-
609 linked class (where a soft sweep occurred in any other subwindow), and the neutral class (where
610 no sweep occurred). Similarly, we constructed an independent test set of 1,500 examples, 300 of
611 each of the five classes. We then used diploS/HIC to train a CNN classifier for each population,
612 holding out 10% of the training simulations as a validation set used for early stopping during
613 training—if 5 consecutive epochs of training yielded no improvement (defined as a decrease in the
614 categorical cross entropy loss function of at least 0.001), then training terminated and the best-
615 performing set of neural network weights up to that point was used for the final classifier. We then
616 applied each population sample's classifier to its corresponding test set and evaluated performance
617 using confusion matrices, precision recall curves (PR), and receiver operating characteristic (ROC)
618 curves.

619 After training our classifiers, we then calculated feature vectors on the genomic data from
620 the corresponding *Ae. aegypti* population samples, again using 2.75 Mb windows (each subdivided
621 into 11 adjacent 250 kb subwindows). We next applied our classifiers to the genomic data, and we
622 used diploS/HIC's posterior class membership probability estimates in order to experiment with
623 four different thresholds: 0.80, 0.90, 0.95, and 0.99. For a given threshold, we required the sum of
624 the windows' hard and soft sweep posterior probabilities to be greater or equal to the threshold
625 before labelling the window as a sweep, and all windows whose highest-probability class was hard
626 or soft but for whom the sweep probability threshold was not met were treated as uncertain and

627 thus were not assigned to a class. We plotted the intersections of selective sweep locations for all
628 populations for a given threshold using the UpsetR package (Conway et al. 2017).

629 **Accounting for the Impact of Classification Errors on the Relative Numbers of Hard and**
630 **Soft Sweeps**

631 To investigate how false positive rates may impact our ratio of hard sweep to soft sweep
632 calls, we calculated the expected number of false discoveries obtained at a given sweep probability
633 threshold for a specified population sample by using the *q*-value obtained at that threshold. We
634 then made the conservative assumption that all false sweep discoveries are classified as soft sweeps
635 and reduced our expected number of soft sweeps accordingly. We then examined the confusion
636 matrix for the corresponding probability threshold to obtain the fraction of hard sweeps that are
637 misclassified as soft, further reducing our total number of expected soft sweeps—during this step
638 we conservatively ignored the possibility of soft sweeps being misclassified as hard.

639 **Identifying Novel High-Confidence Sweep Candidates**

640 To generate a list of high-confidence regions that are sweeping in one or more population
641 samples, we imposed sweep probability cutoffs as described above. We also calculated false
642 discovery *q*-values for each sweep candidate as follows: first, for each sweep in each population
643 sample, we obtained the sweep probability score from diploS/HIC, treating it as our current
644 threshold, and then counted the numbers of both real and simulated sweeps that were equal to or
645 exceeded this threshold. We then obtained the false discovery rate (FDR) by comparing the
646 estimated number of neutral simulations misclassified as a sweep to the total number of predicted
647 sweeps. Once we obtained an FDR for each candidate sweep and its corresponding threshold, *q*-
648 values were calculated as described in (Storey and Siegmund 2001). This procedure was done
649 separately for each population sample.

650 Because diploS/HIC is designed to detect the changes in diversity summaries at varying
651 recombination distances away from a sweep, it cannot produce accurate predictions in non-
652 recombining regions of the genome. We therefore calculated weighted average recombination rates
653 in each of our predicted windows (using data from (Matthews et al. 2018)) and removed any
654 regions where the recombination rate was equal to zero. Finally, to exclude regions of low
655 confidence, we removed regions where the fraction of sites that were masked according to the data

filtering criteria described above exceeded 0.85. Our set of high-confidence, shared sweep candidates was generated by ensuring that the combined posterior probabilities of a sweep was ≥0.95 in all populations that contained the sweep. Similarly, the set of high-confidence population-specific sweeps consisted of windows where a sweep was predicted with ≥0.95 probability in a single population sample. Visualization of sweep candidates was performed using the UCSC Genome Browser (Kent et al. 2002; Perez et al. 2025), along with custom tracks highlighting several population genetic summary statistics calculated by diploS/HIC and CLR scores (Nielsen 2005; DeGiorgio et al. 2016) generated in Love et al., 2023. When searching for potential targets of selection, we focused on genes that were either contained entirely within or partially overlapped the sweeping window, while noting that it is possible that in some cases selection may be acting on a regulatory region impacting the expression of a gene lying outside of this window.

ACKNOWLEDGEMENTS

We would like to thank members of the Schrider lab for useful discussion. This work was funded by NIH awards R01HG010774 and R35GM138286.

682 **REFERENCES**

683 Adrion JR, Galloway JG, Kern AD. 2020. Predicting the landscape of recombination using deep
684 learning. *Molecular Biology and Evolution* 37:1790–1808.

685 Alachiotis N, Pavlidis P. 2018. RAiSD detects positive selection based on multiple signatures of
686 a selective sweep and SNP vectors. *Communications Biology* 1:79.

687 Arnab SP, Amin MR, DeGiorgio M. 2023. Uncovering footprints of natural selection through
688 spectral analysis of genomic summary statistics. *Molecular Biology and Evolution*
689 40:msad157.

690 Avalos A, Pan H, Li C, Acevedo-Gonzalez JP, Rendon G, Fields CJ, Brown PJ, Giray T,
691 Robinson GE, Hudson ME. 2017. A soft selective sweep during rapid evolution of gentle
692 behaviour in an Africanized honeybee. *Nature Communications* 8:1550.

693 Balabanidou V, Kampouraki A, MacLean M, Blomquist GJ, Tittiger C, Juárez MP, Mijailovsky
694 SJ, Chalepakis G, Anthousi A, Lynd A. 2016. Cytochrome P450 associated with
695 insecticide resistance catalyzes cuticular hydrocarbon production in Anopheles gambiae.
696 *Proceedings of the National Academy of Sciences* 113:9268–9273.

697 Barrett RD, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine
698 stickleback. *Science* 322:255–257.

699 Bennett KL, Shija F, Linton Y-M, Misinzo G, Kaddumukasa M, Djouaka R, Anyaele O, Harris
700 A, Irish S, Hlaing T. 2016. Historical environmental change in Africa drives divergence
701 and admixture of Aedes aegypti mosquitoes: a precursor to successful worldwide
702 colonization? *Molecular Ecology* 25:4337–4354.

703 Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genetics*
704 10:e1004412.

705 Bresgen N, Kovacs M, Lahnsteiner A, Felder TK, Rinnerthaler M. 2023. The janus-faced role of
706 lipid droplets in aging: insights from the cellular perspective. *Biomolecules* 13:912.

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

710 Caldas IV, Clark AG, Messer PW. 2022. Inference of selective sweep parameters through
711 supervised learning. *bioRxiv*:2022.07. 19.500702.

712 Campbell CL, Saavedra-Rodriguez K, Kubik TD, Lenhart A, Lozano-Fuentes S, Black IV WC.
713 2019. Vgsc-interacting proteins are genetically associated with pyrethroid resistance in
714 Aedes aegypti. *PLoS One* 14:e0211497.

715 Champion CJ, Xu J. 2018. Redox state affects fecundity and insecticide susceptibility in
716 Anopheles gambiae. *Scientific Reports* 8:13054.

717 Chan J, Perrone V, Spence J, Jenkins P, Mathieson S, Song Y. 2018. A likelihood-free inference
718 framework for population genetic data using exchangeable neural networks. *Advances in*
719 *Neural Information Processing Systems* 31.

720 Chen C, Compton A, Nikolouli K, Wang A, Aryan A, Sharma A, Qi Y, Dellinger C, Hempel M,
721 Potters M. 2022. Marker-assisted mapping enables forward genetic analysis in Aedes
722 aegypti, an arboviral vector with vast recombination deserts. *Genetics* 222:iyac140.

723 Chen YH, Cohen ZP, Bueno EM, Christensen BM, Schoville SD. 2023. Rapid evolution of
724 insecticide resistance in the Colorado potato beetle, Leptinotarsa decemlineata. *Current*
725 *Opinion in Insect Science* 55:101000.

726 Chotai M, Wei X, Messer PW. 2024. Signatures of selective sweeps in continuous-space
727 populations. *bioRxiv*.

728 Clarkson CS, Miles A, Harding NJ, Lucas ER, Battey CJ, Amaya-Romero JE, Kern AD,
729 Fontaine MC, Donnelly MJ, Lawniczak MK. 2020. Genome variation and population
730 structure among 1142 mosquitoes of the African malaria vector species Anopheles
731 gambiae and Anopheles coluzzii. *Genome Research* 30:1533–1546.

732 Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of
733 intersecting sets and their properties. *Bioinformatics* 33:2938–2940.

734 Cosme LV, Lima JBP, Powell JR, Martins AJ. 2022. Genome-wide association study reveals new
735 loci associated with pyrethroid resistance in Aedes aegypti. *Frontiers in Genetics*
736 13:867231.

737 Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, Surendran SN, Black WC,
738 Pain A, Jiggins FM. 2017. Population genomics reveals that an anthropophilic population
739 of Aedes aegypti mosquitoes in West Africa recently gave rise to American and Asian
740 populations of this major disease vector. *BMC Biology* 15:1–16.

741 DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. 2016. SweepFinder2: increased
742 sensitivity, robustness and flexibility. *Bioinformatics* 32:1895–1897.

743 Diaz F, Fukui H, Garcia S, Moraes CT. 2006. Cytochrome c oxidase is required for the
744 assembly/stability of respiratory complex I in mouse fibroblasts. *Molecular and cellular*
745 *Biology* 26:4872–4881.

746 Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and insecticide
747 resistance. *Insect Molecular Biology* 14:3–8.

748 Fay JC, Wu C-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–
749 1413.

750 Feder AF, Pennings PS, Petrov DA. 2021. The clarifying role of time series data in the
751 population genetics of HIV. *PLoS Genetics* 17:e1009050.

752 Feder AF, Rhee S-Y, Holmes SP, Shafer RW, Petrov DA, Pennings PS. 2016. More effective
753 drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *Elife*
754 5:e10670.

755 Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or
756 hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*
757 31:1275–1291.

758 Flagel L, Brandvain Y, Schrider DR. 2019. The unreasonable effectiveness of convolutional
759 neural networks in population genetic inference. *Molecular Biology and Evolution*
760 36:220–238.

761 Gao Y, Liu D-Y, Deng Q, Feng W-L, Yu J-M, Li M-Y, Liu S. 2025. Characterization of the
762 thioredoxin peroxidase gene and its role in lambda-cyhalothrin tolerance in Agrotis
763 ipsilon (Lepidoptera: Noctuidae). *Journal of Asia-Pacific Entomology* 28:102403.

764 Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North
765 American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genetics*
766 11:e1005004.

767 Garud NR, Messer PW, Petrov DA. 2021. Detection of hard and soft selective sweeps from
768 Drosophila melanogaster population genomic data. *PLoS Genetics* 17:e1009373.

769 Girard V, Goubard V, Querenet M, Seugnet L, Pays L, Nataf S, Dufourd E, Cluet D, Mollereau
770 B, Davoust N. 2020. Spen modulates lipid droplet content in adult Drosophila glial cells
771 and protects against paraquat toxicity. *Scientific Reports* 10:20023.

772 Gonçalves da Silva A, Cunha IC, Santos WS, Luz SL, Ribolla PE, Abad-Franch F. 2012. Gene
773 flow networks among American Aedes aegypti populations. *Evolutionary Applications*
774 5:664–676.

775 Gouesbet G, Renault D, Derocles SA, Colinet H. 2025. Strong resistance to β-cyfluthrin in a
776 strain of the beetle Alphitobius diaperinus: a de novo transcriptome analysis. *Insect*
777 *Science* 32:209–226.

778 Gurevich VV, Gurevich EV. 2006. The structural basis of arrestin-mediated regulation of G-
779 protein-coupled receptors. *Pharmacology & Therapeutics* 110:465–502.

780 Hamby KA, Kwok RS, Zalom FG, Chiu JC. 2013. Integrating circadian activity and gene
781 expression profiles to predict chronotoxicity of Drosophila suzukii response to
782 insecticides. *PloS One* 8:e68472.

783 Harris RB, Sackman A, Jensen JD. 2018. On the unfounded enthusiasm for soft selective sweeps
784 II: examining recent evidence from humans, flies, and viruses. *PLoS Genetics*
785 14:e1007859.

786 Haug-Collet K, Pearson B, Webel R, Szerencsei RT, Winkfein RJ, Schnetkamp PPM, Colley NJ.
787 1999. Cloning and characterization of a potassium-dependent sodium/calcium exchanger
788 in Drosophila. *Journal of Cell Biology* 147:659–670.

789 Hayward LK, Sella G. 2022. Polygenic adaptation after a sudden change in environment. *Elife*
790 11:e66697.

791 He C, Xie W, Yang X, Wang S-L, Wu Q-J, Zhang Y-J. 2018. Identification of glutathione S-
792 transferases in Bemisia tabaci (Hemiptera: Aleyrodidae) and evidence that GSTd7 helps
793 explain the difference in insecticide susceptibility between B. tabaci Middle East-Minor
794 Asia 1 and Mediterranean. *Insect Molecular Biology* 27:22–35.

795 Hejase HA, Mo Z, Campagna L, Siepel A. 2022. A deep-learning approach for inference of
796 selective sweeps from the ancestral recombination graph. *Molecular Biology and*
797 *Evolution* 39:msab332.

798 Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from
799 standing genetic variation. *Genetics* 169:2335–2352.

800 Hobbs NP, Weetman D, Hastings IM. 2023. Insecticide resistance management strategies for
801 public health control of mosquitoes exhibiting polygenic resistance: A comparison of
802 sequences, rotations, and mixtures. *Evolutionary Applications* 16:936–959.

803 Höllinger I, Pennings PS, Hermisson J. 2019. Polygenic adaptation: From sweeps to subtle
804 frequency shifts. *PLoS Genetics* 15:e1008035.

805 Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang YU, Yi X, He M,
806 Somel M. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-
807 like DNA. *Nature* 512:194–197.

808 Iwamura T, Guzman-Holst A, Murray KA. 2020. Accelerating invasion potential of disease
809 vector Aedes aegypti under climate change. *Nature Communications* 11:2130.

810 J. Svensson M, Larsson J. 2007. Thioredoxin-2 affects lifespan and oxidative stress in
811 Drosophila. *Hereditas* 144:25–32.

812 Jensen HR, Scott IM, Sims S, Trudeau VL, Arnason JT. 2006. Gene expression profiles of
813 Drosophila melanogaster exposed to an insecticidal extract of Piper nigrum. *Journal of*
814 *Agricultural and Food Chemistry* 54:1289–1295.

815 Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing between
816 selective sweeps and demography using DNA polymorphism data. *Genetics* 170:1401–
817 1410.

818 Johri P, Stephan W, Jensen JD. 2022. Soft selective sweeps: addressing new definitions,
819 evaluating competing models, and interpreting empirical outliers. *PLoS Genetics*
820 18:e1010022.

821 Jones DA, Wakeley J. 2008. The influence of gene conversion on linkage disequilibrium around
822 a selective sweep. *Genetics* 180:1251–1259.

823 Kapopoulou A, Pfeifer SP, Jensen JD, Laurent S. 2018. The demographic history of African
824 Drosophila melanogaster. *Genome Biology and Evolution* 10:2338–2342.

825 Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in Drosophila is not limited
826 by mutation at single sites. *PLoS Genetics* 6:e1000924.

827 Kelly ET, Mack LK, Campos M, Grippin C, Chen T-Y, Romero-Weaver AL, Kosinski KJ, Brisco
828 KK, Collier TC, Buckner EA. 2021a. Evidence of local extinction and reintroduction of
829 Aedes aegypti in Exeter, California. *Frontiers in Tropical Diseases* 2:703873.

830 Kelly ET, Mack LK, Campos M, Grippin C, Chen T-Y, Romero-Weaver AL, Kosinski KJ, Brisco
831 KK, Collier TC, Buckner EA. 2021b. Evidence of local extinction and reintroduction of
832 Aedes aegypti in Exeter, California. *Frontiers in Tropical Diseases* 2:703873.

833 Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* 146:1197–1206.

834 Kent TV, Schrider DR, Matute DR. 2024. Demographic history and the efficacy of selection in
835 the globally invasive mosquito Aedes aegypti. *bioRxiv*.

836 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The
837 human genome browser at UCSC. *Genome Research* 12:996–1006.

838 Kern AD, Schrider DR. 2016. Discoal: flexible coalescent simulations with selection.
839 *Bioinformatics* 32:3839–3841.

840 Kern AD, Schrider DR. 2018. diploS/HIC: an updated approach to classifying selective sweeps.
841 *G3: Genes, Genomes, Genetics* 8:1959–1970.

842 Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics*
843 167:1513–1524.

844 Kliot A, Ghanim M. 2012. Fitness costs associated with insecticide resistance. *Pest Management*
845 *Science* 68:1431–1437.

846 Koboldt DC. 2020. Best practices for variant calling in clinical sequencing. *Genome Medicine*
847 12:91.

848 Kwiatkowska RM, Platt N, Poupardin R, Irving H, Dabire RK, Mitchell S, Jones CM, Diabaté A,
849 Ranson H, Wondji CS. 2013. Dissecting the mechanisms responsible for the multiple
850 insecticide resistance phenotype in Anopheles gambiae ss, M form, from Vallee du Kou,
851 Burkina Faso. *Gene* 519:98–106.

852 Lin K, Li H, Schlotterer C, Futschik A. 2011. Distinguishing positive selection from neutral
853 evolution: boosting the performance of summary statistics. *Genetics* 187:229–244.

Love RR, Sikder JR, Vivero RJ, Matute DR, Schrider DR. 2023. Strong positive selection in Aedes aegypti and the rapid evolution of insecticide resistance. *Molecular Biology and Evolution* 40:msad072.

Lozada-Chávez AN, Lozada-Chávez I, Alfano N, Palatini U, Sogliani D, Elfekih S, Degefa T, Sharakhova MV, Badolo A, Sriwichai P. 2025. Adaptive genomic signatures of globally invasive populations of the yellow fever mosquito Aedes aegypti. *Nature Ecology & Evolution*:1–20.

Maitra S, Dombrowski SM, Waters LC, Ganguly R. 1996. Three second chromosome-linked clustered Cyp6 genes show differential constitutive and barbital-induced expression in DDT-resistant and susceptible strains of Drosophila melanogaster. *Gene* 180:165–171.

Martins AJ, Ribeiro CD e M, Bellinato DF, Peixoto AA, Valle D, Lima JBP. 2012. Effect of insecticide resistance on development, longevity and reproduction of field or laboratory selected Aedes aegypti populations. *PloS One* 7:e31889.

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

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* 28:659–669.

Mo Z, Siepel A. 2023. Domain-adaptive neural networks improve supervised machine learning based on simulated population genetic data. *PLoS Genetics* 19:e1011032.

Mughal MR, DeGiorgio M. 2019. Localizing and classifying adaptive targets with trend filtered regression. *Molecular Biology and Evolution* 36:252–270.

Mughal MR, Koch H, Huang J, Chiaromonte F, DeGiorgio M. 2020. Learning the properties of adaptive regions with functional data analysis. *PLoS Genetics* 16:e1008896.

Muralidhar P, Veller C. 2022. Dominance shifts increase the likelihood of soft selective sweeps. *Evolution* 76:966–984.

Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Research* 15:1566–1575.

Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. *Genetics* 157:875–884.

Pavlidis P, Jensen JD, Stephan W. 2010. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics* 185:907–922.

887 Pennings PS, Hermisson J. 2006a. Soft sweeps III: the signature of positive selection from
888 recurrent mutation. *PLoS Genetics* 2:e186.

889 Pennings PS, Hermisson J. 2006b. Soft sweeps II—molecular population genetics of adaptation
890 from recurrent mutation or migration. *Molecular Biology and Evolution* 23:1076–1084.

891 Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN,
892 Hinrichs AS, Lee CM. 2025. The UCSC Genome Browser database: 2025 update.
893 *Nucleic Acids Research* 53:D1243–D1249.

894 Powell JR, Gloria-Soria A, Kotsakiozi P. 2018. Recent history of Aedes aegypti: Vector
895 genomics and epidemiology records. *Bioscience* 68:854–860.

896 Prezeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic
897 variation. *Evolution* 59:2312–2323.

898 Pridgeon JW, Becnel JJ, Clark GG, Linthicum KJ. 2009. Permethrin induces overexpression of
899 cytochrome c oxidase subunit 3 in Aedes aegypti. *Journal of Medical Entomology*
900 46:810–819.

901 Pridgeon JW, Liu N. 2003. Overexpression of the cytochrome c oxidase subunit I gene
902 associated with a pyrethroid resistant strain of German cockroaches, Blattella germanica
903 (L.). *Insect Biochemistry and Molecular Biology* 33:1043–1048.

904 Pybus Oliveras M, Luisi P, Dall'Olio GM, Uzkudun M, Laayouni H, Bertranpetit J, Engelken J.
905 2015. Hierarchical boosting: a machine-learning framework to detect and classify hard
906 selective sweeps in human populations. *Bioinformatics. 2015 Dec 15; 31 (24): 3946-52*.

907 Rahman RU, Souza B, Uddin I, Carrara L, Brito LP, Costa MM, Mahmood MA, Khan S, Lima
908 JBP, Martins AJ. 2021. Insecticide resistance and underlying targets-site and metabolic
909 mechanisms in Aedes aegypti and Aedes albopictus from Lahore, Pakistan. *Scientific*
910 *Reports* 11:4555.

911 Rogers AR, Huff C. 2009. Linkage disequilibrium between loci with unknown phase. *Genetics*
912 182:839–844.

913 Ronen R, Udpa N, Halperin E, Bafna V. 2013. Learning natural selection from the site frequency
914 spectrum. *Genetics* 195:181–193.

915 Rose NH, Badolo A, Sylla M, Akorli J, Otoo S, Gloria-Soria A, Powell JR, White BJ, Crawford
916 JE, McBride CS. 2023. Dating the origin and spread of specialization on human hosts in
917 Aedes aegypti mosquitoes. *Elife* 12:e83524.

918 Rose NH, Sylla M, Badolo A, Lutomiah J, Ayala D, Aribodor OB, Ibe N, Akorli J, Otoo S,
919 Mutebi J-P. 2020. Climate and urbanization drive mosquito preference for humans.
920 *Current Biology* 30:3570-3579. e6.

Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. 2019. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Neglected Tropical Diseases* 13:e0007213.

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

Sacktor B. 1951. Some aspects of respiratory metabolism during metamorphosis of normal and DDT-resistant house flies, Musca domestica L. *The Biological Bulletin* 100:229–243.

Samb B, Konate L, Irving H, Riveron JM, Dia I, Faye O, Wondji CS. 2016. Investigating molecular basis of lambda-cyhalothrin resistance in an Anopheles funestus population from Senegal. *Parasites & vectors* 9:1–10.

Sanchez T, Cury J, Charpiat G, Jay F. 2021. Deep learning for population size history inference: design, comparison and combination with approximate Bayesian computation. *Molecular Ecology Resources* 21:2645–2660.

Schmidt TL, Endersby-Harshman NM, van Rooyen AR, Katusele M, Vinit R, Robinson LJ, Laman M, Karl S, Hoffmann AA. 2024. Global, asynchronous partial sweeps at multiple insecticide resistance genes in Aedes mosquitoes. *Nature Communications* 15:6251.

Schrider DR. 2023. Allelic gene conversion softens selective sweeps. *bioRxiv*.

Schrider DR, Kern AD. 2016. S/HIC: robust identification of soft and hard sweeps using machine learning. *PLoS Genetics* 12:e1005928.

Schrider DR, Kern AD. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Molecular Biology and Evolution* 34:1863–1877.

Schrider DR, Kern AD. 2018. On the well-founded enthusiasm for soft sweeps in humans: a reply to Harris, Sackman, and Jensen. Zenodo

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* 200:267–284.

Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141:413–429.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genetics Research* 23:23–35.

Song C, Scharf ME. 2009. Mitochondrial impacts of insecticidal formate esters in insecticide-resistant and insecticide-susceptible Drosophila melanogaster. *Pest Management Science: formerly Pesticide Science* 65:697–703.

Song J-Y, Ichtchenko K, Südhof TC, Brose N. 1999. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. *Proceedings of the National Academy of Sciences* 96:1100–1105.

Steele LD, Coates BS, Seong KM, Valero MC, Mittapalli O, Sun W, Clark J, Pittendrigh BR. 2018. Variation in mitochondria-derived transcript levels associated with DDT resistance in the 91-R strain of Drosophila melanogaster (Diptera: Drosophilidae). *Journal of Insect Science* 18:1.

Stephan W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology* 25:79–88.

Storey JD, Siegmund D. 2001. Approximate p-values for local sequence alignments: numerical studies. *Journal of Computational Biology* 8:549–556.

Sun Y, Zou P, Yu X-Y, Chen C, Yu J, Shi L-N, Hong S-C, Zhou D, Chang X-L, Wang W-J. 2012. Functional characterization of an arrestin gene on insecticide resistance of Culex pipiens pallens. *Parasites & Vectors* 5:1–11.

Tabachnick WJ. 1991. Evolutionary genetics and arthropod-borne disease: the yellow fever mosquito. *American Entomologist* 37:14–26.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tarimo BB, Law HCH, Tao D, Pastrana-Mena R, Kanzok SM, Buza JJ, Dinglasan RR. 2018. Paraquat-Mediated Oxidative Stress in Anopheles gambiae Mosquitoes Is Regulated by An Endoplasmic Reticulum (ER) Stress Response. *Proteomes* 6:47.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics* 49:303–309.

Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Research* 16:702–712.

Tessnow AE, Nagoshi RN, Meagher RL, Gilligan TM, Sadd BM, Carrière Y, Davis HN, Fleischer SJ, Richers K, Palumbo JC. 2025. Genomic patterns of strain-specific genetic structure, linkage, and selection across fall armyworm populations. *BMC Genomics* 26:116.

Thornton J, Gomes B, Ayres C, Reimer L. 2020. Insecticide resistance selection and reversal in two strains of Aedes aegypti. *Wellcome Open Research* 5:183.

988 Thornton KR. 2019. Polygenic adaptation to an environmental shift: temporal dynamics of
989 variation under Gaussian stabilizing selection and additive effects on a single trait.
990 *Genetics* 213:1513–1530.

991 Torada L, Lorenzon L, Beddis A, Isildak U, Pattini L, Mathieson S, Fumagalli M. 2019.
992 ImaGene: a convolutional neural network to quantify natural selection from genomic
993 data. *BMC Bioinformatics* 20:337.

994 Ullah F, Gul H, Panda RM, Murtaza G, Zhang Z, Huang J, Li X, Desneux N, Lu Y. 2025.
995 Nanocarrier-mediated RNAi of CYP9E2 and CYB5R enhance susceptibility of invasive
996 tomato pest, Tuta absoluta to cyantraniliprole. *Frontiers in Plant Science* 16:1573634.

997 Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-
998 MEM for multicore systems. In: 2019 IEEE international parallel and distributed
999 processing symposium (IPDPS). IEEE. p. 314–324.

1000 Venkataraman K, Shai N, Lakhiani P, Zylka S, Zhao J, Herre M, Zeng J, Neal LA, Molina H,
1001 Zhao L. 2022. Rapidly evolving genes underlie Aedes aegypti mosquito reproductive
1002 resilience during drought. *bioRxiv*:2022.03. 01.482582.

1003 Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the
1004 human genome. *PLoS Biology* 4:e72.

1005 Vontas J, Blass C, Koutsos AC, David J-P, Kafatos FC, Louis C, Hemingway J, Christophides
1006 GK, Ranson H. 2005. Gene expression in insecticide resistant and susceptible Anopheles
1007 gambiae strains constitutively or after insecticide exposure. *Insect Molecular Biology*
1008 14:509–521.

1009 Ware-Gilmore F, Novelo M, Sgrò CM, Hall MD, McGraw EA. 2023. Assessing the role of
1010 family level variation and heat shock gene expression in the thermal stress response of
1011 the mosquito Aedes aegypti. *Philosophical Transactions of the Royal Society B*
1012 378:20220011.

1013 Watterson GA. 1975. On the number of segregating sites in genetical models without
1014 recombination. *Theoretical Population Biology* 7:256–276.

1015 Webel R, Haug-Collet K, Pearson B, Szerencsei RT, Winkfein RJ, Schnetkamp PPM, Colley NJ.
1016 2002. Potassium-Dependent Sodium-Calcium Exchange through the Eye of the Fly.
1017 *Annals of the New York Academy of Sciences* 976:300–314.

1018 Weedall GD, Riveron JM, Hearn J, Irving H, Kamdem C, Fouet C, White BJ, Wondji CS. 2020.
1019 An Africa-wide genomic evolution of insecticide resistance in the malaria vector
1020 Anopheles funestus involves selective sweeps, copy number variations, gene conversion
1021 and transposons. *PLoS Genetics* 16:e1008822.

1022 Weigand H, Leese F. 2018. Detecting signatures of positive selection in non-model species using
1023 genomic data. *Zoological journal of the Linnean Society* 184:528–583.

1024 Whitehouse LS, Schrider DR. 2023. Timesweeper: accurately identifying selective sweeps using
1025 population genomic time series. *Genetics* 224:iyad084.

1026 Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL. 1996. Identification of mutations
1027 in the housefly para-type sodium channel gene associated with knockdown resistance
1028 (kdr) to pyrethroid insecticides. *Molecular and General Genetics MGG* 252:51–60.

1029 Xue AT, Schrider DR, Kern AD. 2021. Discovery of ongoing selective sweeps within Anopheles
1030 mosquito populations using deep learning. *Molecular Biology and Evolution* 38:1168–
1031 1183.

1032 Yahouédo GA, Chandre F, Rossignol M, Ginibre C, Balabanidou V, Mendez NGA, Pigeon O,
1033 Vontas J, Cornelie S. 2017. Contributions of cuticle permeability and enzyme
1034 detoxification to pyrethroid resistance in the major malaria vector Anopheles gambiae.
1035 *Scientific Reports* 7:11091.

1036 Yang T, Li T, Feng X, Li M, Liu S, Liu N. 2021. Multiple cytochrome P450 genes: conferring
1037 high levels of permethrin resistance in mosquitoes, Culex quinquefasciatus. *Scientific*
1038 *Reports* 11:9041.

1039 Yao P, Chen X, Yan Y, Liu F, Zhang Y, Guo X, Xu B. 2014. Glutaredoxin 1, glutaredoxin 2,
1040 thioredoxin 1, and thioredoxin peroxidase 3 play important roles in antioxidant defense in
1041 Apis cerana cerana. *Free Radical Biology and Medicine* 68:335–346.

1042 Zhang S, Li Z, Nian X, Wu F, Shen Z, Zhang B, Zhang Q, Liu X. 2015. Sequence analysis,
1043 expression profiles and function of thioredoxin 2 and thioredoxin reductase 1 in
1044 resistance to nucleopolyhedrovirus in Helicoverpa armigera. *Scientific Reports* 5:15531.

1045 Zhao C, Tang T, Liu J, Feng X, Qiu L. 2012. Identification and expression analysis of NADH-
1046 cytochrome b5 reductase gene in the cotton bollworm, Helicoverpa armigera. *Gene*
1047 511:96–102.

1048 Zhao L, Cao Y, Wang D-D, Chen N, Li S-G, Liu S, Li M-Y. 2022. A thioredoxin peroxidase
1049 protects Pieris rapae from oxidative stress induced by chlorantraniliprole exposure.
1050 *Archives of Insect Biochemistry and Physiology* 111:e21964.

1051