TITLE: 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 and therefore 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 is a significant limitation as this is vital information for understanding the pace of adaptation—populations that can 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 sweeps are significantly more common than hard sweeps, 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* using 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.

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 (Maynard 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 (Maynard 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). In this way, selective sweep signatures can reveal loci underpinning adaptation as well as how quickly a population can adapt to new stressors, provided one can identify the type of sweep that occurred.

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).

Machine learning (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). However, this also means that the performance of these approaches is fundamentally tied to the demographic models used to generate the simulations because the classifier effectively "learns" the demographic assumptions that are provided. If the demographic history is misspecified, the classifier may misinterpret neutral patterns as selection or fail to identify true sweeps, meaning that accuracy is influenced by the chosen demographic model. There are, however, some methods that remain robust to demographic misspecification by focusing on spatial patterns of variation that signal selective sweeps, rather than relying on local levels of genetic variation (Schrider and Kern 2016). This is because demographic events typically leave signatures on a genome-wide scale, while strong selective sweeps may be expected to produce a localized perturbation in patterns of genetic diversity regardless of the population's demographic history. 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 (Schmidt et al. 2020; Gómez-Palacio et al. 2024; Kent et al. 2025; Lozada-Chávez et al. 2025). *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; Kolimenakis et al. 2021; Laporta et al. 2023). In addition to novel environmental stressors experienced during range expansion, *Ae. aegypti* has also encountered strong selective pressure in the form of insecticides (Love et al. 2023; Ware-Gilmore et al. 2023). Much of our understanding of the development of insecticide resistance in *Ae. aegypti* comes from studies in Asia and the Americas, whereas data from Africa remain comparatively limited. In some parts of the Americas, *Ae. aegypti* has experienced insecticide-based vector control for nearly a century while in Africa, *Ae. aegypti* has not been targeted by focused vector control strategies to the same extent as malaria vectors (Weetman et al. 2018; Wilson et al. 2020; Yaméogo et al. 2024). Nonetheless, increasing levels of insecticide resistance have now been documented across both Africa and the Americas, despite substantial variation in the intensity, duration, and documentation of insecticide use in these regions. This, coupled with their population sizes, genetic diversity levels, and capacity to develop resistance to insecticides in as little as five 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. 2025). However, efforts to detect such signals are complicated by the species' varied and complex demographic histories (Nielsen et al. 2005; Kent et al. 2025).

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). The movement of *Ae. aegypti* 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. 2025). 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* have relied on only a small number of genes (Bennett et al. 2016), mitochondrial DNA (Gonçalves da Silva et al. 2012), microsatellites, or reduced representation sequencing (Brown et al. 2014), and only recently have

147 whole genome-based resequencing and exome-based sequencing been performed (Crawford et al.
148 2017; Matthews et al. 2018; Kelly et al. 2021; Lozada-Chávez et al. 2025). This is largely due to
149 the size (1.3 Gb) and repetitive nature of its genome which poses a significant challenge (Matthews
150 et al. 2018).

151 Because whole-genome data for *Ae. aegypti* have only recently become available (Rose et
152 al. 2020; Kelly et al. 2021; Love et al. 2023), very little is known about recent positive selection
153 in this species. Currently, sweep scans in *Ae. aegypti* based on single-summary statistical
154 approaches have reported positive selection in genomic regions linked to human preference (Rose
155 et al. 2020), increased tolerance to egg desiccation (Venkataraman et al. 2022), and insecticide
156 resistance (Saavedra-Rodriguez et al. 2019; Love et al. 2023; Schmidt et al. 2024; Lozada-Chávez
157 et al. 2025). Among these, the best-characterized insecticide-resistance genes include the voltage-
158 gated sodium channel (*VGSC*), the acetylcholinesterase gene *ace-1*, and metabolic detoxification
159 genes such as cytochrome P450s and glutathione S-transferases (Fukami 1980; Mutero et al. 1994;
160 Williamson et al. 1996; Daborn et al. 2002). However, given their large population sizes, high
161 levels of genetic diversity, and growing evidence to suggest that IR-increasing alleles may spread
162 via soft sweeps in several insect species (Garud et al. 2015; Xue et al. 2021; Muralidhar and Veller
163 2022), it is likely that traditional methods may be missing informative signatures of selection.
164 Here, we leverage a robust machine learning approach (Kern and Schrider 2018) to identify
165 genomic targets of positive selection from environmental and anthropogenic stressors in *Ae.*
166 *aegypti*. We find that we are able to accurately detect both hard and soft sweeps in four globally
167 distributed population samples with complex population size histories. Importantly, we identify
168 novel insecticide resistance (IR) genes which will improve our understanding of the mechanisms
169 of IR evolution and uncover evidence that soft sweeps may play a central role in adaptation in *Ae.*
170 *aegypti*.
171

RESULTS

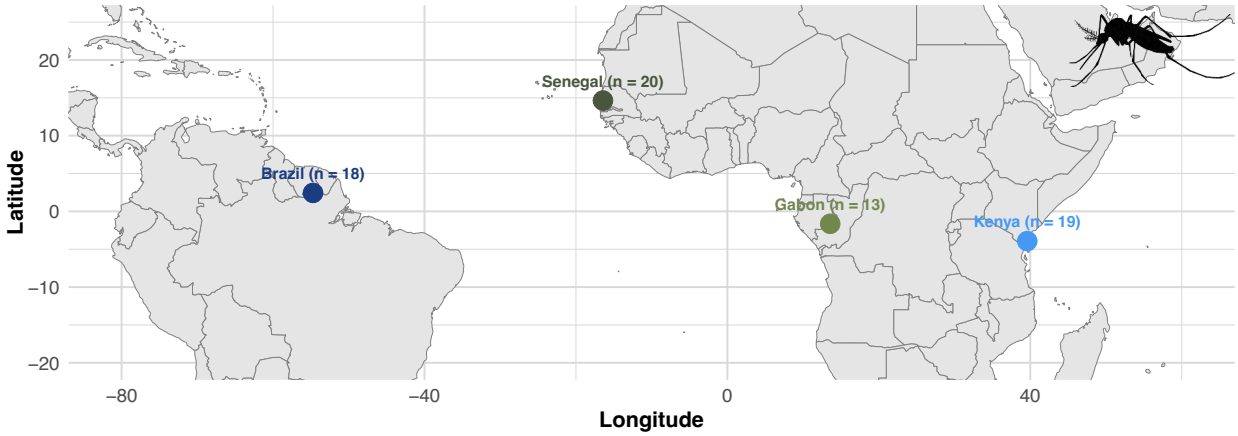

Figure 1. Map of sampling locations including the number of samples analyzed from each site. We have omitted the two sampling sites from Colombia as they were not included in the final analysis.

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; Figure 1). 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. (2025) (Supplemental Figure 1) that were estimated using SMC++ (Terhorst et al. 2017). SMC++ has been shown to infer accurate population size histories across broad time frames by combining the use of the SMC framework, which is most accurate in the deep past, with allele frequency information from a larger sample, which can improve estimation accuracy in the relatively recent past (Liu and Fu 2015; Patton et al. 2019; Liu and Fu 2020). Kent et al. further validated their N_e findings by running Stairway Plot 2 (Liu and Fu 2020) which showed similar patterns to those generated through SMC++. A generation mutation rate of 4.85 x 10^{-9} and a generation time of 0.067 years was used according to Rose et al. (2020). 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 specifically chose parameters designed to model recent and strong selective sweeps capable of substantially altering patterns of diversity.

We assessed the performance of our classifiers through Receiver Operating Characteristic curves (ROC), Precision Recall (PR) curves, and confusion matrices (Figure 2 and Supplemental Figure 2). 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). Classifier accuracy was evaluated using held-out simulated test datasets generated under the same prior parameter distributions used for training, meaning that the reported metrics reflect performance on data drawn from the same underlying assumptions as the training set. 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. The area under the ROC curve quantifies the classifier's ability to discriminate between classes across all thresholds, whereas the average precision summarizes performance across all recall values (equal to the area under the precision-recall curve).

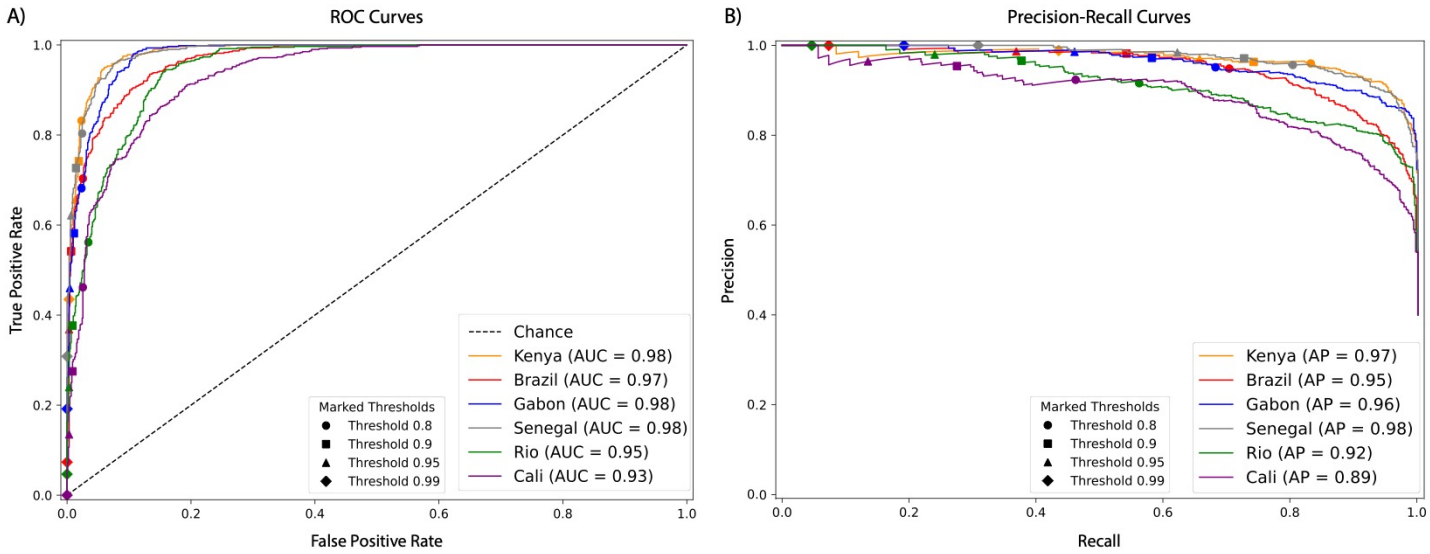

Figure 2. ROC curves and precision-recall curves summarizing the performance of each population samples' classifier. A) ROC curves showing the true and false positive rates for the binary classification task of distinguishing between selective sweeps (hard and soft) vs unselected regions (sweep-linked 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? To address these questions, we 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 a range of thresholds for each population sample (see Methods). Note that the class membership probabilities, which are produced by the softmax activation function

239 from the final layer of the diploS/HIC neural network, may not necessarily be well calibrated such
240 that 80% of examples classified as a hard sweep with 80% probability will be true hard sweeps.
241 In each population sample our ability to distinguish simulated sweeps from neutral
242 variation was strong and improved further when imposing sweep probability thresholds (e.g., <
243 7.5% of neutral windows were misclassified as sweeping in each population sample with no sweep
244 probability cutoff versus < 3.1% in each population sample at the 0.80 cutoff; Supplemental Figure
245 3). When neutrally evolving regions were misclassified, they tended to be misclassified as soft or
246 soft-linked, with a combined misclassification rate < 7.04% at the 0.80 cutoff across all population
247 samples except the Cali and Río Claro samples from Colombia, which showed lower accuracy (see
248 below). Neutrally evolving regions were never classified as hard sweeps at the 0.80 cutoff in any
249 population other than Río Claro, and the rate at which they were misclassified as soft sweeps (in
250 all populations except for those from Colombia) was < 1.7%. Our diploS/HIC classifiers also
251 demonstrated adequate power to discriminate between hard and soft sweeps. When sweep regions
252 were misclassified, they tended to be misclassified as the reciprocal sweep condition (i.e., soft
253 sweeps were misclassified as hard sweeps and vice versa). This same trend occurs in sweep-linked
254 regions. These results indicate that our classifiers were able to robustly distinguish sweeping
255 regions from neutrally evolving or sweep-linked regions.
256 Overall, the accuracy of our classifiers was higher in African population samples than the
257 South American population samples (Supplemental Figures 2-6). For example, in Kenya with no
258 sweep probability cutoff, diploS/HIC correctly classified 88% of hard sweeps and 68% of soft
259 sweeps (Supplemental Figure 2); with a 0.80 threshold imposed this increased to 88% and 72%,
260 respectively. This improvement is observed because simulations classified as sweeps, but with a
261 combined sweep probability lower than the classification threshold, are treated as uncertain and
262 are thus omitted from the calculation, and those high-confidence predictions that remain are more
263 likely to be accurate. When applying cutoff values of 0.90, 0.95, and 0.99 to the Kenya classifier,
264 accuracy values for hard sweeps stayed relatively constant at 88%, 88%, and 90%, respectively, as
265 did soft sweeps at 71%, 69%, 66%. Conversely, for Cali (Supplemental Figure 2) with no cutoff,
266 diploS/HIC correctly classified 80% of hard sweeps and 33% of soft sweeps and increasing the
267 threshold to more stringent values did not improve these values (e.g., at 0.90, 57% of hard sweeps
268 and 25% of soft sweeps were recovered). At the highest threshold of 0.99, there were no hard or
269 soft sweep simulations for Cali passing the threshold, so the fraction of sweeps correctly predicted

could not be calculated. Performance for the other Colombian sample (Río Claro) was similarly poor (Supplemental Figure 2-6). The poorer classifier performance obtained on simulated test data for the Colombian populations samples was expected due to their demographic histories which included stronger and more recent bottlenecks than the other population samples (Supplemental Figure 1). These bottlenecks reduce genetic diversity and distort genealogies in a manner that can resemble positive selection, making sweep detection inherently more difficult (Jensen et al. 2005; Schrider and Kern 2016). Given the performance of Colombia's classifiers, we chose to focus our subsequent analyses on the other four population samples (Brazil, Gabon, Kenya, and Senegal), for which our simulated test data suggested that we could detect and classify sweeps with sufficient accuracy.

Soft sweeps appear to predominate over hard sweeps in *Aedes aegypti*

We classified a total of 4598, 4599, 4599, and 4599 windows that passed our data filtering cutoffs in the Brazil, Gabon, Senegal, and Kenya population samples, respectively. Of these windows, 165 were classified as sweeps in Brazil, compared to 150 in Gabon, 79 in Senegal, and 363 in Kenya (Table 1 and Figure 3A). To construct higher-confidence sets of sweep candidates, we then imposed increasing posterior probability cutoffs on these windows (Table 1). For example, when using a cutoff of 0.95, there were 24 sweep windows in Brazil, 18 in Gabon, 17 in Senegal, and 65 in Kenya. These windows comprise a set of 106 distinct candidate sweeps, some of which were shared by more than one population sample: 1 (0.94%) was shared across all population samples, 1 (0.94%) was shared among the African population samples, and 93 (87.7%) were population-specific sweeps (Figure 3, Supplemental Figure 7). The remaining 11 (10.4%) sweeps were present in more than one population but did not fit into the categories above.

Table 1. The number of hard and soft sweeps per population sample at different posterior 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%)

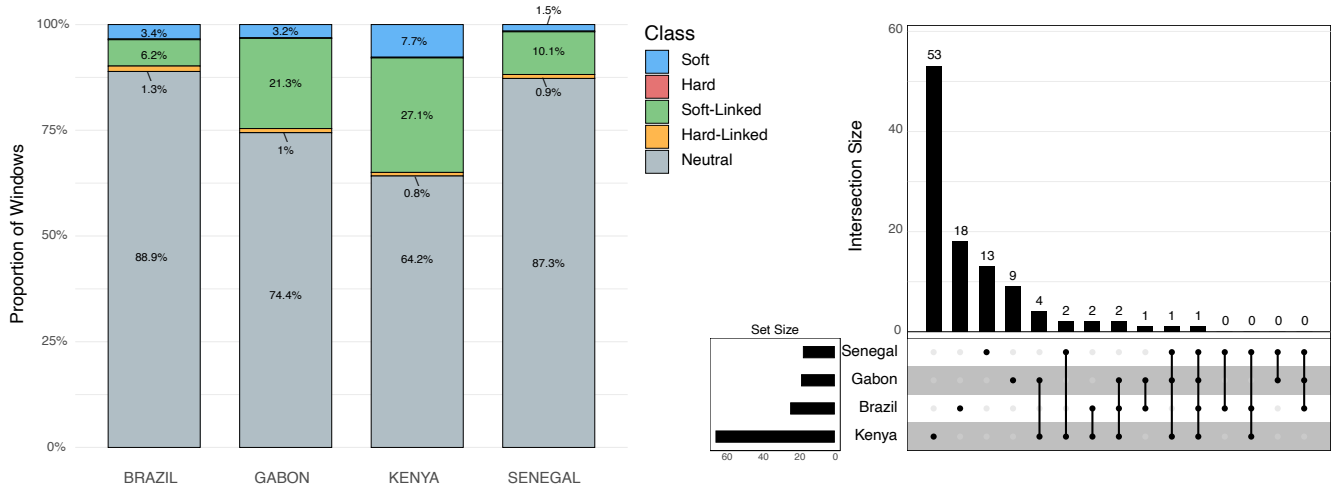

Figure 3. (Left) The proportions of each classification type per population sample with no threshold cutoff applied. diploS/HIC predicted 4,088 neutral, 286 soft-linked, 59 hard-linked, 157 soft, and 8 hard windows for Brazil. For Gabon, predictions included 3,423 neutral, 981 soft-linked, 45 hard-linked, 145 soft, and 5 hard windows. For Kenya, predictions included 2,952 neutral, 1,245 soft-linked, 39 hard-linked, 355 soft, and 8 hard windows. For Senegal, diploS/HIC predicted 4,013 neutral, 464 soft-linked, 43 hard-linked, 71 soft, and 8 hard windows. (Right) Upset plot showing the intersection of selective sweep locations across the four population samples in this study at a posterior probability cutoff 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 and Figure 3). 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, a false discovery rate (FDR)-adjusted *p*-value derived from procedure that is more powerful than the Benjamini–Hochberg FDR method (Storey 2002). We did this 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 (IR). 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

343 groups based on the nature and extent of evidence for their relevance to IR in the literature,
344 including: (1) functional similarity to known IR genes, (2) differential expression in response to
345 insecticide exposure, and (3) signatures of selection reported in other insect species under
346 insecticide pressure. It is important to note that while we can identify compelling candidate genes
347 within these sweep regions, there are often multiple genes in and around these windows, and we
348 thus cannot conclude with certainty which gene is the target of selection.

349

350 *Shared Sweep Windows Containing Putative or Emerging Insecticide Resistance Candidate Genes*

351 Of the 13 windows that were shared between two to four population samples, three
352 windows contain candidate *Ae. aegypti* insecticide-resistance genes (see Supplemental Table 3 for
353 the complete list). First, we found a high-confidence soft sweep shared in Kenya and Gabon (chr2:
354 452,250,001-452,500,000) that contains an ankyrin repeat domain-containing protein 29. This
355 same region was predicted to be a soft sweep in Brazil but its sweep probability of 0.94 did not
356 meet our threshold of 0.95. The ankyrin protein domain is a common protein-protein interaction
357 motif that is present in numerous structural proteins and is involved in cytoskeletal anchoring and
358 mechanosensation. Importantly, ankyrin proteins have been shown to directly interact with the
359 well-established target for insecticides, the voltage-gated sodium channel VGSC (Williamson et
360 al. 1996). In a study on pyrethroid resistant *Ae. aegypti* from Mexico, genes containing the ankyrin-
361 domain were highly associated with pyrethroid resistance (Campbell et al. 2019). Similarly, in
362 pyrethroid-resistant populations of the mosquito *Anopheles funestus* from Senegal, an ankyrin
363 repeat domain protein was one of the most highly overexpressed genes (Samb et al. 2016). In *Ae.*
364 *aegypti*, an ankyrin repeat domain-containing protein was found to be associated with pyrethroid
365 resistance (Cosme et al. 2022), and signatures of selection in two regions containing ankyrin genes
366 (along with other IR candidates) was also reported by Love et al., (2023). Similar patterns have
367 also been observed in other insect species (Kwiatkowska et al. 2013; Gouesbet et al. 2025), further
368 highlighting the role of these gene families in pyrethroid resistance.

369 The other two shared sweep windows containing putative IR genes were also classified as
370 sweeps in Kenya and Gabon. The first contained a gene that encodes for the RNA binding protein
371 Split ends (*spen;* chr2: 143,500,001-143,750,000) which provides a protective role against
372 cytotoxicity from the herbicide paraquat in *Drosophila* (Girard et al. 2020; Bresgen et al. 2023).

373 Although paraquat is an herbicide, it is known to cause oxidative stress in the mosquito *An.*
374 *gambiae* (Champion and Xu 2018; Tarimo et al. 2018). The second window contains *thioredoxin-*
375 *2* (chr2: 45,000,001-45,250,000), a key mitochondrial protein that regulates cellular redox and is
376 protective against oxidative stress, a common consequence of insecticide exposure, in *Drosophila*
377 (J. Svensson and Larsson 2007). Interestingly, the thioredoxin system recycles oxidized
378 glutathione to reduced glutathione, which can then be used by the well-established IR genes
379 glutathione S-transferases (GSTs) for detoxification (Tarimo et al. 2018). Other components of the
380 thioredoxin system, like thioredoxin peroxidase, have been shown to protect insects from
381 insecticide-induced oxidative injury (Zhao et al. 2022; Gao et al. 2025) and *thioredoxin* is listed as
382 of the candidate IR genes on the *Anopheles gambiae* 1000 Genomes Project Selection Atlas
383 (Clarkson et al. 2020).

384

385 *Population-Specific Sweep Windows Containing Well-Characterized Insecticide Resistance Genes*

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

390 One window classified as a soft sweep in Senegal, located at chr1:271,250,001-
391 271,500,000, contains three cytochrome P450 genes: *Cyp 6a8, Cyp 6a13, Cyp 6a14* (Figure 4).
392 Cytochrome P450 monooxygenases (CYPs) are critical for metabolic detoxification and have been
393 implicated in insecticide resistance in many species, including mosquitoes (Balabanidou et al.
394 2016; Rahman et al. 2021; Yang et al. 2021). *Cyp6a8* specifically has been shown to be upregulated
395 in response to *Piper nigrum* and DDT in *D. melanogaster* (Maitra et al. 1996; Jensen et al. 2006).
396 Notably, this sweeping window also contained several SweepFinder Composite Likelihood Ratio
397 (CLR) peaks (Nielsen 2005), with the highest value > 750. We also detected a soft sweep in Kenya
398 in a window that contains *cytochrome b5 reductase 4* (chr2:156,500,001-156,750,000).
399 Cytochrome P450's require cytochrome b5 reductase to function as an electron-transfer
400 intermediate and the latter is therefore likely to be involved in IR (Zhao et al. 2012). Moreover, it
401 has been shown to be upregulated in response to phenobarbital in the cotton bollworm,

Helicoverpa armigera (Zhao et al. 2012) and has been linked to cyantraniliprole resistance in the tomato pest *Tuta absoluta* (Ullah et al. 2025).

We detected a soft sweep in Brazil in a window containing glutathione S-transferase (chr2:446,250,001-446,500,000), *GSTD7*, and that had a corresponding CLR peak with values > 460 (Supplemental Figure 8). Glutathione S-transferase epsilon clusters (GSTe) are another group of metabolic enzymes that are well-known for their role in resistance to several classes of insecticides (Enayati et al. 2005), and other *GSTe* genes have previously been found to exhibit strong signatures of positive selection (Love et al. 2023; Schmidt et al. 2024). *GSTD7*, specifically, was found to be under selection in the fall armyworm, *Spodoptera frugiperda* (Tessnow et al. 2025). Moreover, knockdown of GSTD7 resulted in higher mortality post-imidacloprid exposure in the silver whitefly *Bemisia tabaci* (He et al. 2018), and *GSTD7* is highly expressed in *Drosophila suzukii* after exposure to malathion (Hamby et al. 2013).

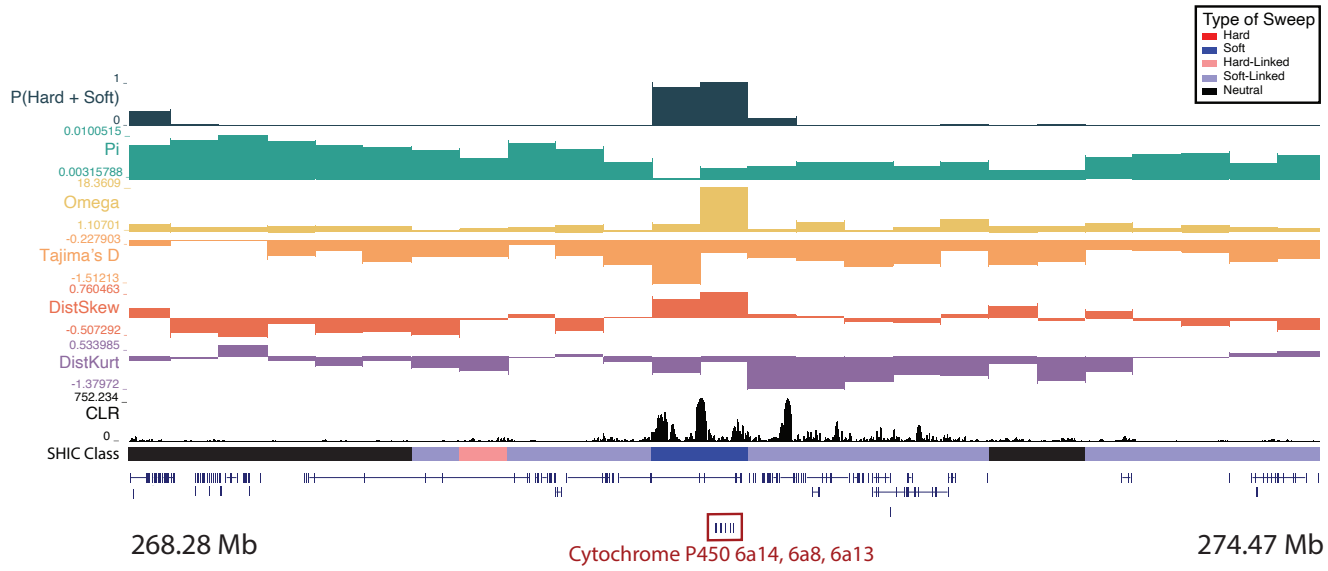

Figure 4. A soft sweep in Senegal at three cytochrome P450 genes: *Cyp6a14*, *Cyp6a8*, and *Cyp6a13* on chromosome one. The diploS/HIC classification track shows the class with the highest posterior probability, with soft sweeps as dark blue, soft-linked regions as light blue, hard sweeps as red, hard-linked 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.

Because we also wanted to explore two of the most well-known IR genes, our results next focus on two major selective sweeps previously reported in Schmidt et al. (2024), spanning the voltage-gated sodium channel gene (*VSSC/VGSC*) and a cluster of 15 glutathione S-transferase genes (*GST*)—regions likewise identified as sweeps in Love et al. (2023). Genome-browser visualizations of both regions are provided in Supplemental Figures 9 and 10. In Brazil, we detected a sweep signal in the *VGSC* region, although its combined posterior probability (0.91) fell slightly below our predefined 0.95 threshold. The magnitude of this sweep, visible as a multi-megabase Tajima's *D* valley, suggests that a larger diploS/HIC window size would be required for more confident classification. As this sweep is easily detectable without the aid of a machine-learning framework incorporating many summary statistics, we deemed it justified to employ a smaller window size so as to prioritize the detection of less expansive sweeps. We likewise detected a sweep signal in the *GST* gene cluster in Brazil, with one diploS/HIC window surpassing our threshold (posterior probability = 0.96). Although this sweep is also quite large, its signature is not as broad as that of *VGSC*, so our method was able to examine the entire sweep region and detect it with higher confidence.

Population-Specific Sweep Windows with Putative or Emerging Insecticide Resistance Candidate Genes

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; Supplemental Figure 11) and overlapped with the *sodium/potassium/calcium exchanger Nckx30C* (chr2:288,750,001-289,000,000; Figure 5). 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 and contains *mitochondrial protoheme IX farnesyltransferase* (chr2:262,250,001-262,500,000), or *COX10*. *COX10* encodes for an enzyme that catalyzes the farnesylation of heme, which plays a crucial role in cytochrome c oxidase (COX) function (Diaz et al. 2006). COX genes have been linked to insecticide resistance in a number of species (Steele et al. 2018), including *D. melanogaster* (Song and Scharf 2009), the housefly *Musca domestica* (Sacktor 1951), the German cockroach *Blattella germanica* (Pridgeon and Liu 2003), and *Ae. aegypti* (Pridgeon et al. 2009).

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

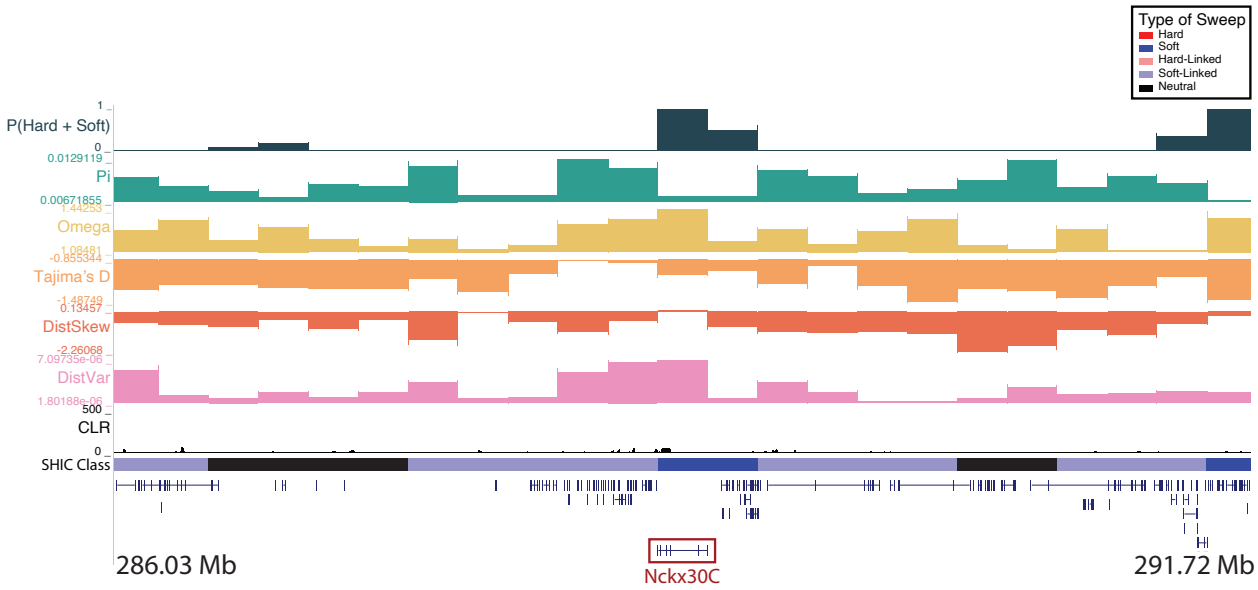

Figure 5. A soft sweep in Kenya at the *sodium/potassium/calcium exchanger Nckx30C* on chromosome two. The diploS/HIC classification track shows the class with the highest posterior probability, with soft sweeps as dark blue, soft-linked regions as light blue, hard sweeps as red, hard-linked as light red, and neutrally evolving regions in black. Above the diploS/HIC classifications are a subset of the summary statistics used by the classifier.

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

No Evidence of Enrichment of Insecticide Resistance Genes or GO Terms in Sweep Windows

To test whether genes previously implicated in insecticide resistance were overrepresented in our high-confidence sweep windows (posterior probability cutoff ≥ 0.95), we used a list of 157 *Ae. aegypti* genes previously implicated in insecticide resistance (Supplemental Table 5). Importantly, this list was generated prior to identifying and annotating candidate sweep windows with diploS/HIC so that the enrichment results would provide an unbiased assessment of whether known IR loci disproportionately overlapped our inferred sweep regions. Across all four populations, we found no statistically significant evidence of known IR genes among sweep regions. In Gabon and Kenya, the overlap between sweep windows and IR genes was equal to or

below that expected by chance. In Brazil and Senegal, enrichment was somewhat elevated by not statistically significant (Supplemental Table 6). We also tested for significantly enriched GO terms within our sweep windows (posterior probability cutoff ≥ 0.80). Although several GO categories showed nominally significant enrichment (P-value < 0.05), none of these categories survived FDR correction (q > 0.05 in all cases; Supplemental Table 7).

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 (Wilson et al. 2014). 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. 2025) 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 findings of similar studies performed on humans (Schrider and Kern 2017), *D. melanogaster* (Garud et al. 2015), *Anopheles gambiae* (Weedall et al. 2020; Xue et al. 2021), and HIV (Feder et al. 2016), that implicate soft selective sweeps as the dominant mode of adaptation (although these studies have proved controversial; see (Harris et al. 2018; Schrider and Kern 2018; Feder et al. 2021; Garud et al. 2021; Johri et al. 2022)). In fact, in all population samples, and at every sweep probability threshold, the percentage of sweeps that were soft was always > 85%. When we applied a conservative approach designed to account for the potential impact of false positives and of soft

526 sweeps being misclassified as hard (while ignoring the possibility of misclassification in the
527 opposite direction), the percentage of soft sweeps was still > 72% in all population samples and
528 thresholds. This suggests that selection in *Ae. aegypti* can act on standing genetic variation instead
529 of waiting for novel mutations to arise, and *Ae. aegypti* populations are therefore able to rapidly
530 respond to new selective pressures like insecticides. Soft sweeps even predominated in Brazil—a
531 population sample that has experienced a protracted bottleneck associated with the introduction
532 from Africa to the Americas (Rose et al. 2023; Kent et al. 2025). This suggests that even
533 bottlenecked populations of *Ae. aegypti* may have sufficient standing variation (or a high enough
534 population-scaled mutation rate; see below) to facilitate rapid adaptation. Out of every population
535 sample, Kenya had substantially more high-confidence soft selective sweeps predicted than any
536 other population sample (61 in Kenya, compared to 23, 18, 15 in Brazil, Gabon and Senegal,
537 respectively, when considering sweeps with a posterior probability ≥ 0.95). This is in line with the
538 findings presented in Kent et al. (2025) where Kenya was inferred to have a larger effective
539 population size than the other population samples and therefore would be expected to experience
540 efficacious natural selection and more soft sweeps.

541 Our classification approach modeled soft sweeps as occurring via selection on standing
542 variation, which is plausible given the high levels of genetic diversity observed. Alternatively,
543 selection on recurrent *de novo* mutations can also produce sweep signatures similar to that of a soft
544 sweep on standing variation—albeit typically less pronounced–with multiple haplotypes carrying
545 the beneficial allele (Pennings and Hermisson 2006b). Such sweeps are especially likely in species
546 where the number of new mutations entering the population each generation is large, either due to
547 large population size and/or a high spontaneous mutation rate (Pennings and Hermisson 2006b;
548 Karasov et al. 2010; Garud et al. 2015). Given that soft sweeps on recurrent mutations can look
549 very similar to soft sweeps on standing variation in terms of the values of numerous summary
550 statistics (Schrider et al. 2015), our machine learning models may detect signals of both selection
551 on standing variation and on recurrent mutations. However, both of these sweeps are indicative of
552 a population that need not wait for long periods for a *de novo* beneficial mutation to arise following
553 a change in the selective environment.

554 On the other hand, there are several phenomena that may result in signatures of soft sweeps
555 even in the absence of rapid adaptation. First, patterns of genetic diversity in a genomic region
556 flanking a hard sweep are expected to mirror that of a soft sweep (the "soft shoulder" effect;

557 (Schrider et al. 2015)). However, this is an unlikely explanation for our results given that we
558 detected very few hard sweeps in our population samples, and diploS/HIC is designed to account
559 for this shoulder effect by examining polymorphisms across a larger window containing the sweep
560 (Schrider and Kern 2016). Second, allelic gene conversion can "soften" hard selective sweeps by
561 transferring the beneficial mutation onto multiple haplotypes (Jones and Wakeley 2008; Schrider
562 et al. 2015; Schrider 2023). This is more common in large populations where there is more time
563 for gene conversion events to occur (Schrider 2023). Third, the dearth of hard sweeps across all
564 population samples may result from the inherent difficulty in accounting for spatial population
565 structure when detecting selective sweeps. In low-dispersal scenarios (which are common in
566 species with wide geographic ranges) an adaptive mutation cannot rise in frequency as rapidly as
567 it would in a panmictic population. As a result, hard selective sweeps can become enriched in
568 intermediate-frequency variants which can cause them to resemble soft sweeps (Chotai et al.
569 2024). Thus, the expectation for large populations like those examined here might be that we would
570 detect more soft sweeps than hard sweeps even without selection acting on standing genetic
571 variation or recurrent mutations. Finally, we note that demographic model misspecification can
572 diminish accuracy in discriminating between hard and soft selective sweeps. To mitigate this effect,
573 we have trained our classifiers using demographic models estimated from the same population
574 samples examined here, and for which two different demographic estimation methods yielded
575 qualitatively similar results (Kent et al. 2025). Nonetheless, there is direct evidence that *Ae. aegypti*
576 can adapt rapidly to insecticides and Martins et al. (2012) and Lozada-Chávez et al. (2025) recently
577 reported evidence of selection in out-of-Africa population samples on variants that are segregating
578 in Africa. Thus, true soft sweeps may therefore represent the most parsimonious explanation for
579 the patterns we observe, and at least some of these may be due to selection on standing genetic
580 variation.

581 While the interpretation of soft sweep signatures is not straightforward, the above
582 arguments suggest that it is essential for selection scans in large/diverse populations to consider
583 soft sweeps in order to detect a larger fraction of the targets of recent positive selection. Previous
584 studies have detected selection on several IR-related genes, including the voltage-gated sodium
585 channel gene (*VGSC*), glutathione S-transferases (GSTs), *ace-1,* carboxylesterases, and many
586 cytochrome P450 genes (Love et al. 2023; Schmidt et al. 2024). While these studies have provided
587 insight into how mosquitoes develop resistance to insecticides (e.g., through metabolic

588 detoxification or variation in the insecticide target location (Kliot and Ghanim 2012; Yahouédo et
589 al. 2017) they did not consider the possibility of soft sweeps, and thus may have missed additional
590 targets of selection. Here we identified both shared and population-specific sweep windows that
591 were almost entirely classified as soft sweeps (96.8% and 93.6%, respectively). Within these
592 windows, we identified 22 genes with either well-characterized or putative roles in IR. To the best
593 of our knowledge, several of our highlighted IR candidates (e.g., *ankyrin repeat domain-*
594 *containing protein 29*, *split ends*, *sodium/potassium/calcium exchanger Nckx30C, muscle calcium*
595 *channel subunit alpha-1*, or *potassium voltage-gated channel subfamily KQT member 1*), have not
596 been reported as IR genes in *Ae. aegypti* and thus represent novel targets for functional validation.
597 Interestingly, of our 13 shared selective sweep windows, only two of them had corresponding CLR
598 peaks (defined by a CLR value > 200) and in our population-specific dataset, only 22% had
599 corresponding CLR peaks. These results make sense given that CLR is based on a hard selective
600 sweep model and the majority of our selective sweeps were classified as soft, further highlighting
601 the need for methods specifically designed to detect the signatures of soft selective sweeps in order
602 to more fully appreciate the landscape of adaptive evolution.

603 Although we were able to identify several novel IR candidate regions by searching for
604 selective sweeps, we note that IR is a polygenic trait. One might therefore expect polygenic
605 selection, in which the selected phenotype moves towards its optimum through the combined effect
606 of small shifts in the frequencies of a large number of small-effect alleles (Berg and Coop 2014;
607 Stephan 2016; Höllinger et al. 2019; Hayward and Sella 2022), to be the primary mode of
608 adaptation in *Ae. aegypti*. Indeed, polygenic adaptation has been implicated in the evolution of
609 complex traits, including insecticide resistance (Chen et al. 2023; Hobbs et al. 2023). However,
610 selective sweeps may still occur during a polygenic shift to a new fitness optimum (Thornton
611 2019). Indeed, this may explain our success in finding sweep candidates at both known and
612 potentially novel IR loci, even though diploS/HIC would be underpowered to detect more subtle
613 allele frequency shifts that may also be contributing to adaptation for IR and other traits in *Ae.*
614 *aegypti*.

615 While our examination of candidate sweep regions primarily focuses on genes associated
616 with IR, we note that other traits may also be subject to recent positive selection. Given *Ae.*
617 *aegypti's* broad and expanding geographic distribution, robust invasive capacity, and the impact of
618 ongoing climate change, it is highly likely some of the sweeps reported here are unrelated to IR.

619 This is particularly relevant for the population samples from Brazil and Senegal, which are human
620 specialists unlike the Kenya and Gabon samples which exhibit the ancestral generalist feeding
621 behavior. The Senegal and Brazil samples thus have experienced additional selective pressures
622 associated with domestication (e.g., adapting to urban environments and acquiring a preference
623 for feeding on human blood; (Rose et al. 2020; Lozada-Chávez et al. 2025)). These reasons also
624 likely underscore why our enrichment analysis did not yield significant results for previously
625 known IR genes. It is also important to note that the list of 157 IR genes that we used to test for
626 enrichment within sweep windows only represents a subset of the genes likely involved in IR and
627 is therefore not comprehensive. Thus, we cannot rule out the possibility that an examination of a
628 more complete set of IR genes in *Ae. aegypti* would reveal significant enrichment within sweep
629 windows. Nonetheless, we concentrate our discussion on sweep windows containing selection on
630 IR genes for two main reasons: 1) selection on these genes may impact mosquito control efforts,
631 and 2) focusing on candidate regions with IR genes may help facilitate the difficult task of
632 identifying the targets of selection: the fact that exceptionally strong signatures of selection have
633 previously been observed near IR genes, suggests that, in sweep regions spanning multiple genes,
634 it is reasonable to prioritize any IR genes as the most likely target of selection. Finally, although
635 we do not analyze non-IR genes in this work, we provide a comprehensive list of all sweep calls
636 with ≥ 0.95 confidence in any population sample in Supplemental Tables 3 and 4.
637 In this study we leveraged whole-genome data from multiple *Ae. aegypti* population
638 samples together with a powerful machine learning tool to enrich our understanding of the genomic
639 targets and mode of positive selection in this important vector species. In summary, our findings
640 suggest that the rapid adaptation of *Ae. aegypti* to insecticides and other control strategies may, in
641 part, be driven by soft selective sweeps. The persistence of high genetic diversity in out-of-Africa
642 populations suggests that these population samples are likely to retain substantial adaptive
643 potential, enabling rapid responses to future selective pressures. This underscores the importance
644 of considering modes of adaptation beyond classic hard sweeps, as approaches that primarily target
645 hard sweep signatures may overlook many adaptive variants. Our work highlights the need for
646 systematic methods to detect soft sweeps and a more complete understanding of how standing
647 genetic variation and polygenic selection contribute to adaptive evolution in insect vectors.
648

649 **METHODS**

Sampling, Sequencing, and Variant Calling

In this study, we used a previously curated whole genome sequencing dataset of 104 individuals from (Love et al. 2023). This dataset includes 18 samples from Santarém, Brazil; 13 from Franceville, Gabon; 19 from Kaya Bomu, Kenya; and 20 from Ngoye, Senegal, all taken from Rose et al. (2020), as well as 10 and 24 samples from Cali and Río Claro, Colombia, respectively, sequenced in Love et al. (2023). Detailed methods for DNA extractions, sequencing, and variant calling and filtering are available in (Love et al. 2023). Briefly, reads from all individuals were aligned to the AaegL5 reference genome (Matthews et al. 2018; 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 exponential distribution the true mean value of r will be somewhat lower than our mean parameter value.

For the simulations involving sweeps, there are several additional parameters whose distributions are unknown, like the strength and timing of selection. For these, we drew from a wide distribution to ensure that the range of parameters seen during training encompasses those likely encountered during downstream inference. Specifically, we drew the selection coefficient from a log-uniform distribution ranging from 0.005 to 0.05 and the time of fixation of the beneficial allele, from $U(0, 0.001)$. For soft and soft-linked sweeps, we simulated selection on a previously neutral standing variant, drawing the frequency of the previously neutral allele at the onset of selection, (f_0), from $U(0, 0.05)$. In total, we generated 3,000 simulated replicates for each class for each population. Because our goal was to identify strong selective sweeps that affect a wide stretch of the chromosome, we sought to simulate 2.75 Mb regions, but the version of discoal available at the time these simulations were performed was unable to simulate such regions due to memory constraints. We therefore simulated 550 kb regions and decreased the selection coefficient to $U(0.001, 0.01)$, resulting in simulated sweeps with the same ratio of s/r as we would have obtained from 2.75 Mb simulations with $s \sim U(0.005, 0.05)$. More recently, discoal has been updated to be more memory efficient. This allows simulation of 2.75 Mb genomic regions. These simulations are very time consuming, so rather than regenerating training and test data for each population sample, we decided to directly assess the extent to which our approach of using smaller windows with smaller selection coefficients impacts our classification accuracy. To this end, we generated an additional test set of simulations for the Kenyan population sample using 2.75 Mb regions under the same demographic and selective parameters (i.e., $U[0, 0.05]$) . We then applied the original diploS/HIC classifier that was trained on 550 kb simulations to these newly generated 2.75 Mb

711 simulations and evaluated classifier performance using confusion matrices (Supplemental Figure
712 13). The resulting confusion matrices show that the classifier actually performs slightly better on
713 the larger-window simulations compared to its performance on the original 550 kb simulated
714 windows. This is probably a result of the increased number of polymorphisms in the larger
715 windows, which decreases variance in our summary statistic calculations. In any case, these results
716 indicate that the classifier generalizes well to the larger window size and that the change in window
717 scale does not negatively affect classification accuracy.

718 As described above, in our real data we masked genomic positions that had genotype
719 qualities below 20. To incorporate heterogeneity in data quality into our training/test data, for each
720 simulated window, we randomly selected a corresponding window from our empirical dataset and
721 masked the same sites in the simulated window that had been masked in the empirical window.
722 This ensured that our masking procedure affected our simulated data in the same way as our real
723 data. Feature vectors were then calculated for each simulation replicate. These features vectors
724 measure the spatial patterns of a number of population genetic summary statistics, including π
725 (Tajima 1983), θ_ω (Watterson 1975), Tajima's *D* (Tajima 1989), the number of distinct haplotypes,
726 average haplotype homozygosity (also referred to as H_1 by Garud et al. 2015), H_{12} and H_2/H_1
727 (Garud et al. 2015), Z_{ns} (Kelly 1997), and the maximum value of Kim and Nielsen's ω (Kim and
728 Nielsen 2004); note that the latter two statistics, which measure linkage disequilibrium, were based
729 on calculations of Rogers and Huff's estimator of LD for unphased data (Rogers and Huff 2009).
730 In addition to these commonly used statistics, diploS/HIC also includes in its feature vector
731 estimates of the variance, skewness, and kurtosis of the distribution of the number of pairwise
732 differences for all pairs of individuals in the sample; these summaries are useful for detecting hard
733 and soft sweeps and discriminating between them (Kern and Schrider 2018). diploS/HIC
734 subdivides the simulated region into 11 adjacent windows and calculates each of these statistics in
735 each window, dividing that value by the sum of values for that statistic across all 11 subwindows.
736 If the smallest of these 11 values for a given statistic is less than 0, the value for this subwindow
737 is increased to zero, and the value for each other subwindow is increased by the same amount (i.e.,
738 the absolute value of the smallest subwindow is added to each subwindow in such cases).

739 After computing summary statistics, for each population sample we used diploS/HIC's
740 makeTrainingSets command to combine our simulations into a training set of 13,500 examples:
741 2,700 each for the hard sweep class (consisting of simulations where a hard sweep occurred in the

central subwindow), the hard-linked class (where a hard sweep occurred in any subwindow but the central one), the soft sweep class (where a soft sweep occurred in the central subwindow), the soft-linked class (where a soft sweep occurred in any other subwindow), and the neutral class (where no sweep occurred). Similarly, we constructed an independent test set of 1,500 examples, 300 of each of the five classes. We then used diploS/HIC to train a CNN classifier for each population, holding out 10% of the training simulations as a validation set used for early stopping during training—if 5 consecutive epochs of training yielded no improvement (defined as a decrease in the categorical cross entropy loss function of at least 0.001), then training terminated and the best-performing set of neural network weights up to that point was used for the final classifier. We then applied each population sample's classifier to its corresponding test set and evaluated performance using confusion matrices, precision recall curves (PR), and receiver operating characteristic (ROC) curves.

After training our classifiers, we then calculated feature vectors on the genomic data from the corresponding *Ae. aegypti* population samples, again using 2.75 Mb windows (each subdivided into 11 adjacent 250 kb subwindows); this was done across the three chromosomes of the *Ae. aegypti* reference genome, while the smaller scaffolds were ignored. We next applied our classifiers to the genomic data and classified a total of 4,738 windows in each of our four population samples. Because diploS/HIC is designed to detect the changes in diversity summaries at varying recombination distances away from a sweep, it cannot produce accurate predictions in non-recombining regions of the genome. We therefore calculated weighted average recombination rates in each of our predicted windows (using data from (Matthews et al. 2018)) and removed any regions where the recombination rate was equal to zero. This resulted in 4,631 windows classified in each of our population samples. Finally, to exclude regions of low confidence, we removed regions where the fraction of sites that were masked according to the data filtering criteria described above exceeded 0.85, which resulted in 4,599 windows in each population sample except for Brazil which contained 4,598 windows. These filtered windows were used for all downstream analyses. We then used diploS/HIC's posterior class membership probability estimates in order to experiment with four different thresholds: 0.80, 0.90, 0.95, and 0.99. For a given threshold, we required the sum of the windows' hard and soft sweep posterior probabilities to be greater or equal to the threshold before labelling the window as a sweep, and all windows whose highest-probability class was hard or soft but for whom the sweep probability threshold was not met were

773 treated as uncertain and thus were not assigned to a class. We plotted the intersections of selective
774 sweep locations for all populations for a given threshold using the UpsetR package (Conway et al.
775 2017).
776

777 **Accounting for the Impact of Classification Errors on the Relative Numbers of Hard and**
778 **Soft Sweeps**

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

787

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

789 To generate a list of high-confidence regions that are sweeping in one or more population
790 samples, we imposed sweep probability cutoffs as described above (See section: ***Using Classifiers***
791 ***to Detect Selective Sweeps***). We also calculated false discovery *q*-values for each sweep candidate
792 as follows: first, for each sweep in each population sample, we obtained the sweep probability
793 score from diploS/HIC, treating it as our current threshold, and then counted the numbers of both
794 real and simulated sweeps that were equal to or exceeded this threshold. We then obtained the false
795 discovery rate (FDR) by comparing the estimated number of neutral simulations misclassified as
796 a sweep to the total number of predicted sweeps. Once we obtained an FDR for each candidate
797 sweep and its corresponding threshold, *q*-values were calculated as described in (Storey 2002).
798 This procedure was done separately for each population sample.
799 Our set of high-confidence, shared sweep candidates was generated by ensuring that the
800 combined posterior probabilities of a sweep was ≥ 0.95 in all populations that contained the sweep.
801 Similarly, the set of high-confidence population-specific sweeps consisted of windows where a
802 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.

Enrichment Analysis

We first performed an enrichment analysis to test whether genes previously implicated in IR were overrepresented among genomic regions under selection. This was done using a list of 157 *Ae. aegypti* genes previously implicated in insecticide resistance, compiled by Love et al. (2023). For each population, we only considered sweep windows with a posterior probability ≥ 0.95. To test for enrichment, we performed a permutation-based analysis where we permuted the association between the sweep statistic values and their genomic positions 10,000 times. For each permutation, we recalculated the number of IR genes that overlapped with sweep windows. Gene windows were extended by 1 kb on either side to account for uncertainty around sweep locations and the locations of nearby regulatory elements. One-sided *p*-values were calculated as the proportion of permutations in which the number of overlapping IR genes was greater than or equal to that observed in the real data. Fold enrichment was calculated as the ratio between the observed and mean permutated number of IR genes overlapping sweep windows.

Next, we performed an additional enrichment test (https://github.com/SchriderLab/permEnrichmentTest) following that of Love et al. (2023) to look for overrepresented GO terms among genes overlapping selective sweep windows with a posterior probability ≥ 0.8. Importantly, this method tests for significantly enriched GO terms after controlling for linkage between genes. It does this by concatenating the three chromosomes into a circular sequence and rotating all sweep scores around the circle by a random interval. We ran this enrichment analysis using 10,000 permutations of the sweep window posterior probabilities and then asked which GO terms were enriched in the selective sweep set compared to the permuted set. Finally, genes were extended by 1 kb in either direction.

ACKNOWLEDGEMENTS

We would like to thank members of the Schrider lab for useful discussion. This work was funded by NIH awards R01HG010774 and R35GM138286. DRM was funded by NIH award R35GM148244.

DATA AVAILABILITY

All sequencing data used in this paper are publicly available (see Love et al. 2023). All code used for the analyses presented in this paper can be found at https://github.com/SchriderLab/Detecting_Selection_Aedes_Aegypti_2026.

858 **REFERENCES**

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

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

863 Arnab SP, Amin MR, DeGiorgio M. 2023. Uncovering footprints of natural selection through
864 spectral analysis of genomic summary statistics. *Molecular Biology and Evolution*
865 40:msad157.

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

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

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

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

879 Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genetics*
880 10:e1004412.

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

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

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

888 Campbell CL, Saavedra-Rodriguez K, Kubik TD, Lenhart A, Lozano-Fuentes S, Black IV WC.
889 2019. Vgsc-interacting proteins are genetically associated with pyrethroid resistance in
890 Aedes aegypti. *PLoS One* 14:e0211497.

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

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

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

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

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

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

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

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

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P. 2002. A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* 297:2253–2256.

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

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

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

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

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

927 Feder AF, Rhee S-Y, Holmes SP, Shafer RW, Petrov DA, Pennings PS. 2016. More effective
928 drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *Elife*
929 5:e10670.

930 Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or
931 hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*
932 31:1275–1291.

933 Flagel L, Brandvain Y, Schrider DR. 2019. The unreasonable effectiveness of convolutional
934 neural networks in population genetic inference. *Molecular Biology and Evolution*
935 36:220–238.

936 Fukami J-I. 1980. Metabolism of several insecticides by glutathion S-transferase. *Pharmacology*
937 *& therapeutics* 10:473–514.

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

941 Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North
942 American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genetics*
943 11:e1005004.

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

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

949 Gómez-Palacio A, Morinaga G, Turner PE, Micieli MV, Elnour M-AB, Salim B, Surendran SN,
950 Ramasamy R, Powell JR, Soghigian J. 2024. Robustness in population-structure and
951 demographic-inference results derived from the *Aedes aegypti* genotyping chip and
952 whole-genome sequencing data. *G3: Genes, Genomes, Genetics* 14:jkae082.

953 Gonçalves da Silva A, Cunha IC, Santos WS, Luz SL, Ribolla PE, Abad-Franch F. 2012. Gene
954 flow networks among American *Aedes aegypti* populations. *Evolutionary Applications*
955 5:664–676.

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

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

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

964 Harris RB, Sackman A, Jensen JD. 2018. On the unfounded enthusiasm for soft selective sweeps
965 II: examining recent evidence from humans, flies, and viruses. *PLoS Genetics*
966 14:e1007859.

967 Haug-Collet K, Pearson B, Webel R, Szerencsei RT, Winkfein RJ, Schnetkamp PPM, Colley NJ.
968 1999. Cloning and characterization of a potassium-dependent sodium/calcium exchanger
969 in Drosophila. *Journal of Cell Biology* 147:659–670.

970 Hayward LK, Sella G. 2022. Polygenic adaptation after a sudden change in environment. *Elife*
971 11:e66697.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kent TV, Schrider DR, Matute DR. 2025. Demographic history, genetic load, and the efficacy of selection in the globally invasive mosquito *Aedes aegypti*. *Genome Biology and Evolution* 17:evaf066.

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

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

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

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

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

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

Kolimenakis A, Heinz S, Wilson ML, Winkler V, Yakob L, Michaelakis A, Papachristos D, Richardson C, Horstick O. 2021. The role of urbanisation in the spread of *Aedes*

1029 mosquitoes and the diseases they transmit—A systematic review. *PLoS neglected tropical*
1030 *diseases* 15:e0009631.

1031 Kwiatkowska RM, Platt N, Poupardin R, Irving H, Dabire RK, Mitchell S, Jones CM, Diabaté A,
1032 Ranson H, Wondji CS. 2013. Dissecting the mechanisms responsible for the multiple
1033 insecticide resistance phenotype in *Anopheles gambiae* ss, M form, from Vallee du Kou,
1034 Burkina Faso. *Gene* 519:98–106.

1035 Laporta GZ, Potter AM, Oliveira JF, Bourke BP, Pecor DB, Linton Y-M. 2023. Global
1036 distribution of Aedes aegypti and Aedes albopictus in a climate change scenario of
1037 regional rivalry. *Insects* 14:49.

1038 Lin K, Li H, Schlotterer C, Futschik A. 2011. Distinguishing positive selection from neutral
1039 evolution: boosting the performance of summary statistics. *Genetics* 187:229–244.

1040 Liu X, Fu Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nature*
1041 *Genetics* 47:555–559.

1042 Liu X, Fu Y-X. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency
1043 spectra. *Genome Biology* 21:280.

1044 Love RR, Sikder JR, Vivero RJ, Matute DR, Schrider DR. 2023. Strong positive selection in
1045 Aedes aegypti and the rapid evolution of insecticide resistance. *Molecular Biology and*
1046 *Evolution* 40:msad072.

1047 Lozada-Chávez AN, Lozada-Chávez I, Alfano N, Palatini U, Sogliani D, Elfekih S, Degefa T,
1048 Sharakhova MV, Badolo A, Sriwichai P. 2025. Adaptive genomic signatures of globally
1049 invasive populations of the yellow fever mosquito *Aedes aegypti. Nature Ecology &*
1050 *Evolution*:1–20.

1051 Maitra S, Dombrowski SM, Waters LC, Ganguly R. 1996. Three second chromosome-linked
1052 clustered Cyp6 genes show differential constitutive and barbital-induced expression in
1053 DDT-resistant and susceptible strains of *Drosophila melanogaster. Gene* 180:165–171.

1054 Martins AJ, Ribeiro CD e M, Bellinato DF, Peixoto AA, Valle D, Lima JBP. 2012. Effect of
1055 insecticide resistance on development, longevity and reproduction of field or laboratory
1056 selected *Aedes aegypti* populations. *PloS One* 7:e31889.

1057 Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ,
1058 Herre M, Redmond SN, Rose NH. 2018. Improved reference genome of *Aedes aegypti*
1059 informs arbovirus vector control. *Nature* 563:501–507.

1060 Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genetics*
1061 *Research* 23:23–35.

1062 Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps.
1063 *Trends in Ecology & Evolution* 28:659–669.

1064 Mo Z, Siepel A. 2023. Domain-adaptive neural networks improve supervised machine learning
1065 based on simulated population genetic data. *PLoS Genetics* 19:e1011032.

1066 Mughal MR, DeGiorgio M. 2019. Localizing and classifying adaptive targets with trend filtered
1067 regression. *Molecular Biology and Evolution* 36:252–270.

1068 Mughal MR, Koch H, Huang J, Chiaromonte F, DeGiorgio M. 2020. Learning the properties of
1069 adaptive regions with functional data analysis. *PLoS Genetics* 16:e1008896.

1070 Muralidhar P, Veller C. 2022. Dominance shifts increase the likelihood of soft selective sweeps.
1071 *Evolution* 76:966–984.

1072 Mutero A, Pralavorio M, Bride J-M, Fournier D. 1994. Resistance-associated point mutations in
1073 insecticide-insensitive acetylcholinesterase. *Proceedings of the National Academy of*
1074 *Sciences* 91:5922–5926.

1075 Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218.

1076 Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for
1077 selective sweeps using SNP data. *Genome Research* 15:1566–1575.

1078 Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic
1079 variation. *Genetics* 157:875–884.

1080 Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK, Ruiz-Aravena M,
1081 Ryder O, McCallum HI, Jones ME. 2019. Contemporary demographic reconstruction
1082 methods are robust to genome assembly quality: a case study in Tasmanian devils.
1083 *Molecular Biology and Evolution* 36:2906–2921.

1084 Pavlidis P, Jensen JD, Stephan W. 2010. Searching for footprints of positive selection in whole-
1085 genome SNP data from nonequilibrium populations. *Genetics* 185:907–922.

1086 Pennings PS, Hermisson J. 2006a. Soft sweeps III: the signature of positive selection from
1087 recurrent mutation. *PLoS Genetics* 2:e186.

1088 Pennings PS, Hermisson J. 2006b. Soft sweeps II—molecular population genetics of adaptation
1089 from recurrent mutation or migration. *Molecular Biology and Evolution* 23:1076–1084.

1090 Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN,
1091 Hinrichs AS, Lee CM. 2025. The UCSC Genome Browser database: 2025 update.
1092 *Nucleic Acids Research* 53:D1243–D1249.

1093 Powell JR, Gloria-Soria A, Kotsakiozi P. 2018. Recent history of Aedes aegypti: Vector
1094 genomics and epidemiology records. *Bioscience* 68:854–860.

1095 Prezeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic
1096 variation. *Evolution* 59:2312–2323.

1097 Pridgeon JW, Becnel JJ, Clark GG, Linthicum KJ. 2009. Permethrin induces overexpression of
1098 cytochrome c oxidase subunit 3 in *Aedes aegypti. Journal of Medical Entomology*
1099 46:810–819.

1100 Pridgeon JW, Liu N. 2003. Overexpression of the cytochrome c oxidase subunit I gene
1101 associated with a pyrethroid resistant strain of German cockroaches, Blattella germanica
1102 (L.). *Insect Biochemistry and Molecular Biology* 33:1043–1048.

1103 Pybus Oliveras M, Luisi P, Dall'Olio GM, Uzkudun M, Laayouni H, Bertranpetit J, Engelken J.
1104 2015. Hierarchical boosting: a machine-learning framework to detect and classify hard
1105 selective sweeps in human populations. *Bioinformatics. 2015 Dec 15; 31 (24): 3946-52*.

1106 Rahman RU, Souza B, Uddin I, Carrara L, Brito LP, Costa MM, Mahmood MA, Khan S, Lima
1107 JBP, Martins AJ. 2021. Insecticide resistance and underlying targets-site and metabolic
1108 mechanisms in *Aedes aegypti* and *Aedes albopictus* from Lahore, Pakistan. *Scientific*
1109 *Reports* 11:4555.

1110 Rogers AR, Huff C. 2009. Linkage disequilibrium between loci with unknown phase. *Genetics*
1111 182:839–844.

1112 Ronen R, Udpa N, Halperin E, Bafna V. 2013. Learning natural selection from the site frequency
1113 spectrum. *Genetics* 195:181–193.

1114 Rose NH, Badolo A, Sylla M, Akorli J, Otoo S, Gloria-Soria A, Powell JR, White BJ, Crawford
1115 JE, McBride CS. 2023. Dating the origin and spread of specialization on human hosts in
1116 *Aedes aegypti* mosquitoes. *Elife* 12:e83524.

1117 Rose NH, Sylla M, Badolo A, Lutomiah J, Ayala D, Aribodor OB, Ibe N, Akorli J, Otoo S,
1118 Mutebi J-P. 2020. Climate and urbanization drive mosquito preference for humans.
1119 *Current Biology* 30:3570-3579. e6.

1120 Ryan SJ, Carlson CJ, Mordecai EA, Johnson LR. 2019. Global expansion and redistribution of
1121 Aedes-borne virus transmission risk with climate change. *PLoS Neglected Tropical*
1122 *Diseases* 13:e0007213.

1123 Saavedra-Rodriguez K, Campbell CL, Lenhart A, Penilla P, Lozano-Fuentes S, Black IV WC.
1124 2019. Exome-wide association of deltamethrin resistance in *Aedes aegypti* from Mexico.
1125 *Insect Molecular Biology* 28:591–604.

1126 Sacktor B. 1951. Some aspects of respiratory metabolism during metamorphosis of normal and
1127 DDT-resistant house flies, *Musca domestica* L. *The Biological Bulletin* 100:229–243.

1128 Samb B, Konate L, Irving H, Riveron JM, Dia I, Faye O, Wondji CS. 2016. Investigating
1129 molecular basis of lambda-cyhalothrin resistance in an *Anopheles funestus* population
1130 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, 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 Neglected Tropical Diseases* 14:e0008463.

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.

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. 2002. A Direct Approach to False Discovery Rates. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64:479–498.

1166 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.
1167 Functional characterization of an arrestin gene on insecticide resistance of *Culex pipiens*
1168 *pallens*. *Parasites & Vectors* 5:1–11.

1169 Tabachnick WJ. 1991. Evolutionary genetics and arthropod-borne disease: the yellow fever
1170 mosquito. *American Entomologist* 37:14–26.

1171 Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*
1172 105:437–460.

1173 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA
1174 polymorphism. *Genetics* 123:585–595.

1175 Tarimo BB, Law HCH, Tao D, Pastrana-Mena R, Kanzok SM, Buza JJ, Dinglasan RR. 2018.
1176 Paraquat-Mediated Oxidative Stress in *Anopheles gambiae* Mosquitoes Is Regulated by
1177 An Endoplasmic Reticulum (ER) Stress Response. *Proteomes* 6:47.

1178 Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from
1179 hundreds of unphased whole genomes. *Nature Genetics* 49:303–309.

1180 Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for
1181 selective sweeps? *Genome Research* 16:702–712.

1182 Tessnow AE, Nagoshi RN, Meagher RL, Gilligan TM, Sadd BM, Carrière Y, Davis HN,
1183 Fleischer SJ, Richers K, Palumbo JC. 2025. Genomic patterns of strain-specific genetic
1184 structure, linkage, and selection across fall armyworm populations. *BMC Genomics*
1185 26:116.

1186 Thornton J, Gomes B, Ayres C, Reimer L. 2020. Insecticide resistance selection and reversal in
1187 two strains of *Aedes aegypti*. *Wellcome Open Research* 5:183.

1188 Thornton KR. 2019. Polygenic adaptation to an environmental shift: temporal dynamics of
1189 variation under Gaussian stabilizing selection and additive effects on a single trait.
1190 *Genetics* 213:1513–1530.

1191 Torada L, Lorenzon L, Beddis A, Isildak U, Pattini L, Mathieson S, Fumagalli M. 2019.
1192 ImaGene: a convolutional neural network to quantify natural selection from genomic
1193 data. *BMC Bioinformatics* 20:337.

1194 Ullah F, Gul H, Panda RM, Murtaza G, Zhang Z, Huang J, Li X, Desneux N, Lu Y. 2025.
1195 Nanocarrier-mediated RNAi of CYP9E2 and CYB5R enhance susceptibility of invasive
1196 tomato pest, *Tuta absoluta* to cyantraniliprole. *Frontiers in Plant Science* 16:1573634.

1197 Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-
1198 MEM for multicore systems. In: 2019 IEEE international parallel and distributed
1199 processing symposium (IPDPS). IEEE. p. 314–324.

1200 Venkataraman K, Shai N, Lakhiani P, Zylka S, Zhao J, Herre M, Zeng J, Neal LA, Molina H,
1201 Zhao L. 2022. Rapidly evolving genes underlie *Aedes aegypti* mosquito reproductive
1202 resilience during drought. *bioRxiv*:2022.03. 01.482582.

1203 Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the
1204 human genome. *PLoS Biology* 4:e72.

1205 Vontas J, Blass C, Koutsos AC, David J-P, Kafatos FC, Louis C, Hemingway J, Christophides
1206 GK, Ranson H. 2005. Gene expression in insecticide resistant and susceptible *Anopheles*
1207 *gambiae* strains constitutively or after insecticide exposure. *Insect Molecular Biology*
1208 14:509–521.

1209 Ware-Gilmore F, Novelo M, Sgrò CM, Hall MD, McGraw EA. 2023. Assessing the role of
1210 family level variation and heat shock gene expression in the thermal stress response of
1211 the mosquito *Aedes aegypti*. *Philosophical Transactions of the Royal Society B*
1212 378:20220011.

1213 Watterson GA. 1975. On the number of segregating sites in genetical models without
1214 recombination. *Theoretical Population Biology* 7:256–276.

1215 Webel R, Haug-Collet K, Pearson B, Szerencsei RT, Winkfein RJ, Schnetkamp PPM, Colley NJ.
1216 2002. Potassium-Dependent Sodium-Calcium Exchange through the Eye of the Fly.
1217 *Annals of the New York Academy of Sciences* 976:300–314.

1218 Weedall GD, Riveron JM, Hearn J, Irving H, Kamdem C, Fouet C, White BJ, Wondji CS. 2020.
1219 An Africa-wide genomic evolution of insecticide resistance in the malaria vector
1220 *Anopheles funestus* involves selective sweeps, copy number variations, gene conversion
1221 and transposons. *PLoS Genetics* 16:e1008822.

1222 Weetman D, Kamgang B, Badolo A, Moyes CL, Shearer FM, Coulibaly M, Pinto J, Lambrechts
1223 L, McCall PJ. 2018. *Aedes* mosquitoes and *Aedes*-borne arboviruses in Africa: current
1224 and future threats. *International Journal of Environmental Research and Public Health*
1225 15:220.

1226 Weigand H, Leese F. 2018. Detecting signatures of positive selection in non-model species using
1227 genomic data. *Zoological journal of the Linnean Society* 184:528–583.

1228 Whitehouse LS, Schrider DR. 2023. Timesweeper: accurately identifying selective sweeps using
1229 population genomic time series. *Genetics* 224:iyad084.

1230 Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL. 1996. Identification of mutations
1231 in the housefly para-type sodium channel gene associated with knockdown resistance
1232 (kdr) to pyrethroid insecticides. *Molecular and General Genetics MGG* 252:51–60.

1233 Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, Torr SJ, Lindsay SW. 2020. The
1234 importance of vector control for the control and elimination of vector-borne diseases.
1235 *PLoS Neglected Tropical Diseases* 14:e0007831.

1236 Wilson BA, Petrov DA, Messer PW. 2014. Soft selective sweeps in complex demographic
1237 scenarios. *Genetics* 198:669–684.

1238 Xue AT, Schrider DR, Kern AD. 2021. Discovery of ongoing selective sweeps within *Anopheles*
1239 mosquito populations using deep learning. *Molecular Biology and Evolution* 38:1168–
1240 1183.

1241 Yahouédo GA, Chandre F, Rossignol M, Ginibre C, Balabanidou V, Mendez NGA, Pigeon O,
1242 Vontas J, Cornelie S. 2017. Contributions of cuticle permeability and enzyme
1243 detoxification to pyrethroid resistance in the major malaria vector *Anopheles gambiae.*
1244 *Scientific Reports* 7:11091.

1245 Yaméogo F, Sombié A, Oté M, Saiki E, Sakurai T, Wangrawa DW, McCall PJ, Weetman D,
1246 Kanuka H, Badolo A. 2024. Three years of insecticide resistance evolution and associated
1247 mechanisms in *Aedes aegypti* populations of Ouagadougou, Burkina Faso. *PLoS*
1248 *Neglected Tropical Diseases* 18:e0012138.

1249 Yang T, Li T, Feng X, Li M, Liu S, Liu N. 2021. Multiple cytochrome P450 genes: conferring
1250 high levels of permethrin resistance in mosquitoes, *Culex quinquefasciatus*. *Scientific*
1251 *Reports* 11:9041.

1252 Yao P, Chen X, Yan Y, Liu F, Zhang Y, Guo X, Xu B. 2014. Glutaredoxin 1, glutaredoxin 2,
1253 thioredoxin 1, and thioredoxin peroxidase 3 play important roles in antioxidant defense in
1254 *Apis cerana cerana*. *Free Radical Biology and Medicine* 68:335–346.

1255 Zhang S, Li Z, Nian X, Wu F, Shen Z, Zhang B, Zhang Q, Liu X. 2015. Sequence analysis,
1256 expression profiles and function of thioredoxin 2 and thioredoxin reductase 1 in
1257 resistance to nucleopolyhedrovirus in *Helicoverpa armigera*. *Scientific Reports* 5:15531.

1258 Zhao C, Tang T, Liu J, Feng X, Qiu L. 2012. Identification and expression analysis of NADH-
1259 cytochrome b5 reductase gene in the cotton bollworm, *Helicoverpa armigera*. *Gene*
1260 511:96–102.

1261 Zhao L, Cao Y, Wang D-D, Chen N, Li S-G, Liu S, Li M-Y. 2022. A thioredoxin peroxidase
1262 protects Pieris rapae from oxidative stress induced by chlorantraniliprole exposure.
1263 *Archives of Insect Biochemistry and Physiology* 111:e21964.

1264