

RESEARCH ARTICLE

Evidence for serial founder events during the colonization of North America by the yellow fever mosquito, *Aedes aegypti*

Evlyn Pless^{1,2}  | Jeffrey R. Powell¹  | Krystal R. Seger³ | Brett Ellis³ | Andrea Gloria-Soria^{1,4} 

¹Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA

²Department of Anthropology, University of California, Davis, California, USA

³U.S. Virgin Islands Department of Health, Christiansted, VI, USA

⁴Department of Environmental Sciences, The Connecticut Agricultural Experiment Station, New Haven, Connecticut, USA

Correspondence

Evlyn Pless, Department of Anthropology, University of California, Davis, California, USA.

Email: espless@ucdavis.edu

Funding information

Center for Disease Control and Prevention, Grant/Award Number: U01CK000509; National Institutes of Health, Grant/Award Number: RO1-AI101112

Abstract

The *Aedes aegypti* mosquito first invaded the Americas about 500 years ago and today is a widely distributed invasive species and the primary vector for viruses causing dengue, chikungunya, Zika, and yellow fever. Here, we test the hypothesis that the North American colonization by *Ae. aegypti* occurred via a series of founder events. We present findings on genetic diversity, structure, and demographic history using data from 70 *Ae. aegypti* populations in North America that were genotyped at 12 microsatellite loci and/or ~20,000 single nucleotide polymorphisms, the largest genetic study of the region to date. We find evidence consistent with colonization driven by serial founder effect (SFE), with Florida as the putative source for a series of westward invasions. This scenario was supported by (1) a decrease in the genetic diversity of *Ae. aegypti* populations moving west, (2) a correlation between pairwise genetic and geographic distances, and (3) demographic analysis based on allele frequencies. A few *Ae. aegypti* populations on the west coast do not follow the general trend, likely due to a recent and distinct invasion history. We argue that SFE provides a helpful albeit simplified model for the movement of *Ae. aegypti* across North America, with outlier populations warranting further investigation.

KEYWORDS

Aedes aegypti, genetic structure, Invasive species, North America, population genetics, serial founder effect

TAXONOMY CLASSIFICATION

Population genetics

1 | INTRODUCTION

Range expansion is the process by which invasive species spread to new regions and environments. This process can be modeled as a continuous expanding wave, a series of founder events, or stochastic jumps (Peischl et al., 2016). Each type of range expansion leads to characteristic genetic patterns that are also influenced by

genetic drift and gene flow (Excoffier et al., 2009). Understanding the dynamics of range expansion for a given invasive species can shed light on their demographic history and may illuminate options for preventing future spread.

The *Aedes aegypti* mosquito (Linnaeus, 1762) is an invasive species that has successfully invaded tropical regions around the world and is increasingly reported in temperate regions (Kraemer

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

et al., 2015), with a rate of invasion that is expected to accelerate with climate change (Iwamura et al., 2020). The global distribution of this vector enabled the recent outbreaks of Zika and chikungunya in the Americas (Carlson et al., 2016; Leparco-Goffart et al., 2014), reemergence of yellow fever in Africa and South America (Hamlet et al., 2018), and a dramatic spread and increase of dengue cases around the world (Brady & Hay, 2020). Its widespread presence in Mexico and the Caribbean perpetuates endemic dengue, and its presence in the southern United States poses a public health threat.

Aedes aegypti likely arrived in the Americas in the 17th century aboard slave ships from Africa, where it rapidly spread throughout the continent, as evidenced by outbreaks of yellow fever and dengue; see the comprehensive review (Powell et al., 2018). Disease outbreaks ranged from the Caribbean to the North Atlantic and the southeast United States by the early 1800s (Carrigan, 1959; Moreno-Madriñán & Turell, 2018). Dengue made its way into the central-south United States in the 1850s (Chandler, 1956), and by the 1930s *Ae. aegypti* spanned much of Texas and coastal Mexico (Slosek, 1986). *Ae. aegypti* became more prevalent and widespread in New Mexico around 1994 (Merrill et al., 2005) and in California starting around 2013 (Metzger et al., 2017). Taken together, these lines of evidence paint a picture of the westward migration of the mosquito species.

When a small number of founders emigrate from a larger population, the new population will generally display a reduction in genetic variation relative to the original population that can last for many generations—a phenomenon called “founder effect” (Nei et al., 1975). The serial founder effect (SFE) (spread occurring through a series of these founder events) has been famously invoked to describe the movement of *Homo sapiens* out of Africa (Henn et al., 2012; Ramachandran et al., 2005). It has also been used to explain the global spread of the malaria parasite that accompanied humans out of Africa and the expansion of monarch butterflies across the Pacific (Pierce et al., 2014). Although numerous studies have examined *Ae. aegypti* invasions, migration, and structure using populations genetics from around the globe—including North America (Gloria-Soria et al., 2016; Kotsakiozi et al., 2018; Pless et al., 2017)—none have explicitly tested the hypothesis of colonization by SFE.

Here, we asked whether SFE explains the westward spread of *Ae. aegypti* across southern North America. These migrations were likely accomplished by a combination of active dispersal and passive human-mediated transport (e.g., trucks and ships) (Fonzi et al., 2015; Goncalves da Silva et al., 2012; Guagliardo et al., 2014; Medley et al., 2015). Given an active average lifetime dispersal of <200 m for *Ae. aegypti* (Honorio et al., 2003; Jasper et al., 2019; Reiter, 2007; Russell et al., 2005), we predict that short-distance invasions are more common for both active and passive dispersal due to a higher number of introduction events and higher propagule pressure (Sakai et al., 2001), which is likely to result in colonization via SFE.

To test if *Ae. aegypti* expansion across North America is consistent with a SFE moving westward, we evaluated our data against four expectations:

1. Source populations will have the highest number of alleles, with genetic diversity decreasing in proportion to the distance from the source (Hunley et al., 2012).
2. There will be a positive relationship between distance among groups and their genetic differentiation (Ramachandran et al., 2005).
3. Following each founding event, the daughter group will carry a subset of variation from the parental group (Ramachandran et al., 2005).
4. Demographic inference modeling will support SFE over other possible scenarios.

We inferred genetic structure across the region and tested these predictions using data from 12 microsatellite loci and single nucleotide polymorphism (SNP) array data from 70 North American *Ae. aegypti* populations. Establishing the extent to which *Ae. aegypti* has spread via SFE is important for preventing and detecting future invasions, as well as modifying vector control in response to the movement of pesticide-resistant alleles.

2 | MATERIALS AND METHODS

2.1 | Mosquito collection

Our analysis includes 70 *Ae. aegypti* populations across continental North America and the eastern Caribbean (Figure 1, Table 1 and Table S1). Microsatellite and SNP genotypes for most populations in this study have been reported in Evans et al. (2015), Gloria-Soria et al. (2014), Gloria-Soria et al. (2016), Kotsakiozi et al. (2017), Pless et al. (2017), Pless et al. (2020), Pless et al. (2021) and Saarman et al. (2017). New data presented here include microsatellite genotypes from (1) La Altagracia, Dominican Republic, (2) San Jose de Ocoa, Dominican Republic, (3) St. Croix, USVI, and (4) St. Thomas, USVI, and genome-wide SNP data for (1) St. Thomas, USVI, (2) Alamogordo, NM, USA, (3) Las Cruces, NM, USA, (4) Lubbock, TX, USA, and (5) Bexar, TX, USA. The new populations genotyped fill important regional gaps, particularly in the Caribbean and the central United States. The remaining gaps in samplings, such as those between the panhandle of Florida and New Orleans, are due to the absence of *Ae. aegypti* in recent years.

The mean sample size is 33.8 for microsatellites and 10.4 for SNPs. The year range of sampling collections is 2006–2018 (with >75% of the samples collected in 2014 or later). All mosquito samples were collected as adults or eggs from traps and were shipped as adults to Yale University for analysis. No more than six individuals were used from a single ovitrap to minimize sampling relatives.

2.2 | DNA extraction and genotyping

The microsatellite dataset includes 2132 individuals from 63 populations genotyped at 12 loci (Brown et al., 2011; Slotman et al.,

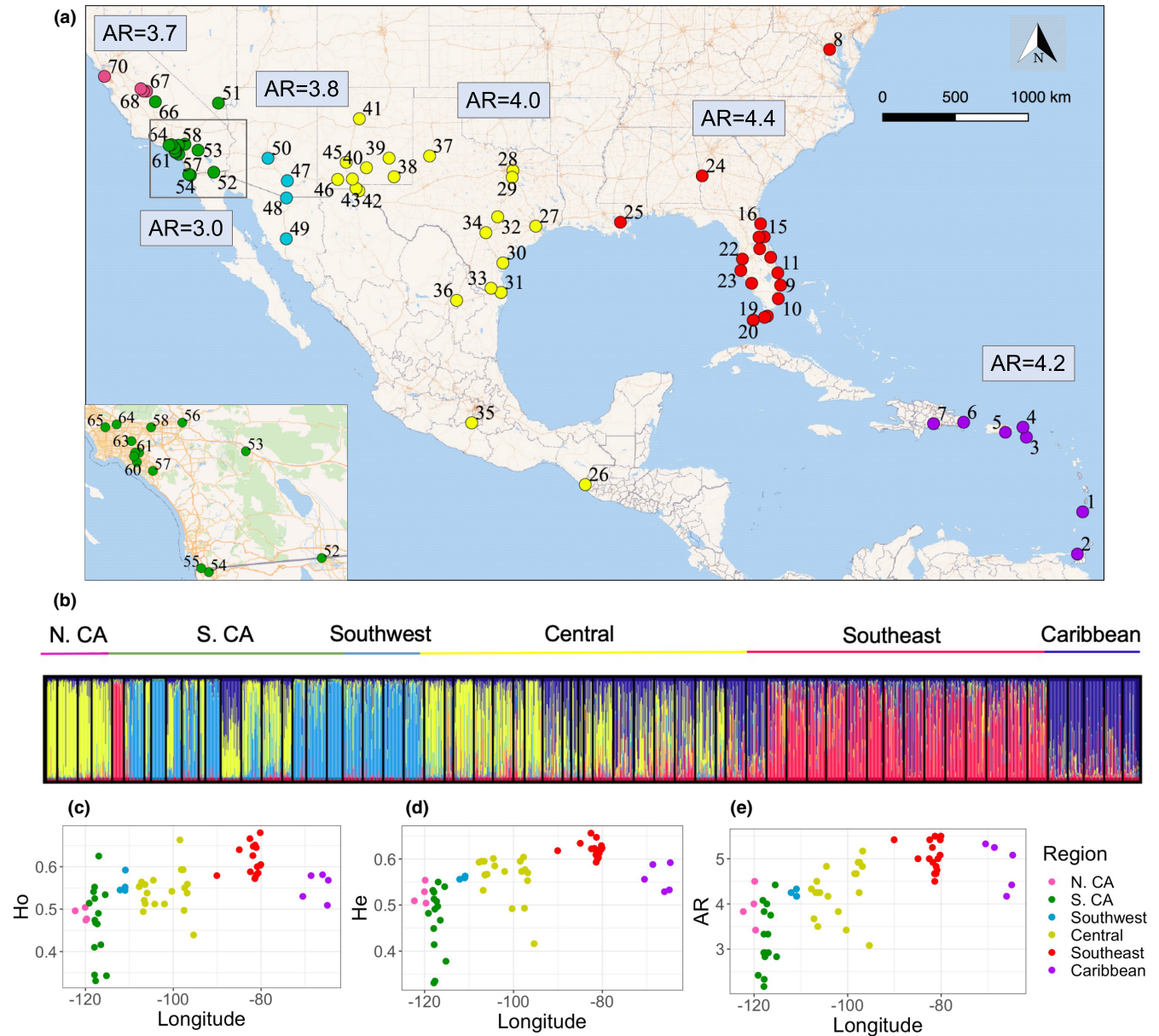


FIGURE 1 (a) Map showing locations for each *Aedes aegypti* sampling site included in this study, colored by regional group. The inset on the bottom left shows southern California. Mean allelic richness estimated from microsatellites by rarefaction of all sites within each region is shown in the boxes. (b) STRUCTURE plot of the complete microsatellite dataset with $K = 4$ number of clusters. Each column is an individual, and the heights of the color bars represent the proportion of ancestry that came from each of the four inferred clusters (yellow, light blue, red, and dark blue). The outlier (red) in southern California is Exeter County. The longitude of each site is plotted against its mean observed heterozygosity (c), expected heterozygosity (d), and allelic richness estimated by rarefaction using microsatellite data (e). Colors in the map (a), above the STRUCTURE plot (b), and in the diversity plots (c–e) are consistent and correspond to regional groups

2007). The SNP dataset includes 373 individuals from 36 populations genotyped at ~20,000 SNPs with the Axiom_aegypti array (Evans et al., 2015). Microsatellites are appropriate for unbiased genetic diversity estimates and demographic inference analysis due to their multiallelic, highly polymorphic nature, and complete allele frequency spectra. Additionally, their low cost facilitates the generation of large global and local reference databases. In contrast, the larger number of markers from the Axiom_aegypti array may provide higher fine-scale resolution for population structure, particularly when populations are of recent origin or when gene flow

is significant (Gloria-Soria et al., 2018). However, the ascertainment bias in the design of the SNP array is likely to affect measurements of genetic diversity and may also impact demographic analysis, which uses allele diversity spectrum to evaluate the likelihood of different scenarios. Therefore, the two sets of genetic markers provide complementary advantages and are useful for different analyses.

Whole genomic DNA was extracted from individual mosquitoes using the Qiagen DNeasy Blood and Tissue kit according to manufacturer instructions, including the optional RNase A step. Individuals were genotyped for 12 microsatellites as in Gloria-Soria et al. (2016).

TABLE 1 *Aedes aegypti* populations included in this study

Pop. number	Population	Abbrev.	Region name	Year	N ¹	N ²
1	St. Vincent	SV	Caribbean	2015	0	12
2	Trinidad	TRI	Caribbean	2014	0	12
3*	St. Croix, USVI	StX	Caribbean	2017	32	0
4**	St. Thomas, USVI	StT	Caribbean	2017	36	10
5	Patillas, Puerto Rico	PR	Caribbean	2014	40	12
6*	La Altagracia, DR	Alt	Caribbean	2018	31	0
7*	San Jose de Ocoa, DR	Oco	Caribbean	2018	39	0
8	Washington, D.C.	WashDC	Southeast	2014	0	11
9	Palm Beach County, FL, USA	PBC	Southeast	2013	40	12
10	Miami, FL, USA	Miami	Southeast	2011	40	8
11	Rio, FL, USA	FLO	Southeast	2014	40	0
12	Melbourne, FL, USA	Mel	Southeast	2014	40	12
13	Conch Key, FL, USA	Conch	Southeast	2006	40	0
14	Vaca Key, FL, USA	Vaca	Southeast	2009	40	0
15	Daytona Beach, FL, USA	DB	Southeast	2017	40	12
16	St. Augustine, FL, USA	StA	Southeast	2017	40	0
17	Orlando, FL, USA	Orl	Southeast	2014	32	11
18	Barberville, FL, USA	Bb	Southeast	2017	40	11
19	North Key West, FL	NK	Southeast	2013	0	11
20	Key West, FL, USA	KW16	Southeast	2016	40	12
21	Fort Myers, FL, USA	FM	Southeast	2014	37	12
22	Tampa, FL, USA	Tam	Southeast	2014	40	12
23	Sarasota, FL, USA	Sar	Southeast	2014	39	12
24	Muscogee, GA, USA	18 and 9	Southeast	2011	40	0
25	New Orleans LA, USA	NO and NO2	Southeast	2012	40	10
26	Tapachula Norte, CHP, MEX	TapaN	Central	2012	0	12
27	Houston, TX, USA	Houston and H11	Central	2011	19	8
28	Dallas, TX, USA	Dall	Central	2017	40	0
29	Ellis, TX, USA	El	Central	2017	40	0
30	Nueces, TX, USA	Nuec	Central	2017	40	0
31	Cameron, TX, USA	Cam	Central	2015	40	0
32	Travis, TX, USA	Tr	Central	2017	40	0
33	Hidalgo, TX, USA	Hid	Central	2017	40	0
34 ⁺	Bexar, TX, USA	Bex	Central	2017	18	8
35	Amacuzac, Morelos, Mexico	Amac16_P	Central	2014	0	12
36	Nuevo Leon, MEX	NL	Central	2017	10	0
37 ⁺	Lubbock, TX, USA	Lub	Central	2017	14	8
38	Carlsbad, NM, USA	Car	Central	2017	17	0
39	Roswell, NM, USA	Ros	Central	2017	39	0
40 ⁺	Alamagordo, NM, USA	Ala	Central	2017	35	4
41	Albuquerque, NM, USA	Alb	Central	2018	22	0
42	Juarez, MEX	Juar	Central	2017	40	0
43	Sunland Park, NM, USA	SP	Central	2017	37	0
44 ⁺	Las Cruces, NM, USA	LC18	Central	2018	40	6
45	Truth of Consequences, NM, USA	TC	Central	2017	17	0

TABLE 1 (Continued)

Pop. number	Population	Abbrev.	Region name	Year	N ¹	N ²
46	Deming, NM, USA	Dem	Central	2017	40	0
47	Tucson, AZ, USA	TJC2	Southwest	2012	40	12
48	Nogales, Son, MEX	Nog	Southwest	2013	40	9
49	Hermosillo, Son, MEX	Her	Southwest	2013	40	0
50	Maricopa County, AZ, USA	Az	Southwest	2013	39	0
51	Las Vegas, NV, USA	LV	Southern CA	2017	31	0
52	El Centro, CA, USA	EIC	Southern CA	2016	40	0
53	Coachella, CA, USA	Coa	Southern CA	2017	27	0
54	Tijuana, BCN, MEX	Tj	Southern CA	2013	20	10
55	San Diego, CA, USA	Cw and SY	Southern CA	2015	40	12
56	San Bernardino, CA, USA	SBern	Southern CA	2017	40	0
57	Mission Viejo, CA, USA	MV	Southern CA	2015	40	12
58	Montclair, CA, USA	Mc	Southern CA	2016	30	0
59	Orange, CA, USA	Or	Southern CA	2015	13	0
60	Santa Ana, CA, USA	SA17	Southern CA	2017	33	0
61	Anaheim, CA, USA	Ana_LC and Ana	Southern CA	2015	31	0
62	Garden Grove, CA, USA	GG	Southern CA	2015	29	12
63	La Habra, CA, USA	LH	Southern CA	2017	13	0
64	Rosemead, CA, USA	Ro	Southern CA	2017	40	0
65	Los Angeles, CA	GLA	Southern CA	2014	0	6
66	Exeter, CA, USA	Exe	Southern CA	2014	23	12
67	Clovis, CA, USA	Clovis and Cal	Northern CA	2013	40	6
68	Fresno, CA, USA	Fres	Northern CA	2015	27	12
69	Madera, CA, USA	MAD	Northern CA	2015	40	12
70	San Mateo, CA, USA	SM and SM2	Northern CA	2013	22	8

Note: Population number corresponding to Figure 1, population name, population abbreviation, region name, year sampled, the sample size for microsatellite data (N¹), and sample size for SNP data (N²). New microsatellite data are indicated with an asterisk (*) and new SNP data are indicated with a cross (†).

Abbreviations: DR, Dominican Republic; MEX, Mexico; USA, United States of America; USVI, United States Virgin Islands.

Four loci (A1, B2, B3, and A9) are trinucleotide repeats, and eight (AC2, CT2, AG2, AC4, AC1, AC5, AG1, and AG4) are di-nucleotide repeats. Any individuals that were genotyped at fewer than 10 loci were excluded from the analysis. We include only populations with a minimum of 10 individuals and arbitrarily selected 40 individuals from sites with more than 40 samples to control for unequal or low sample size (Puechmaile, 2016).

Individuals were genotyped using Axiom_aegypti, a high-throughput genotyping chip that has 50,000 probes (Evans et al., 2015). Genotyping was conducted by the Functional Genomics Core at the University of North Carolina, Chapel Hill. To prune the SNP dataset, we first excluded 2166 SNPs that failed a test of Mendelian inheritance (Evans et al., 2015). Since some analyses can be confounded by SNPs in linkage disequilibrium (Alexander et al., 2009), we excluded tightly linked SNPs with the plink command "--indep 75 kb 50 2" (Purcell, 2016; Purcell et al., 2007). We also excluded any SNPs that were genotyped in <90% of the individuals and those with a minor allele frequency of <5%, resulting in 20,003 SNPs remaining for analysis.

All microsatellite data are available in Dataset S1, and all SNP data are available in Dataset S2. Additionally, the data can be accessed on Dryad (<https://doi.org/10.5061/dryad.5x69p8d5j>) and VectorBase (VBP0000801).

2.3 | Inferring geographic regions based on genetic structure

All microsatellite loci were tested for within-population deviations from Hardy-Weinberg equilibrium and for linkage disequilibrium among loci pairs using the R package Genepop v. 1.1.4. with 5000 dememorizations, 500 batches, and 5000 iterations per batch for both tests (Raymond & Rousset, 1995). To correct for multiple testing, a Bonferroni correction was applied at the .05 α level of significance.

To establish regional groupings for subsequent analysis, we examined population structure using a number of methods. We performed principal component analysis (PCA) using the R package Adegenet v. 2.1.1. (Jombart, 2008) for the microsatellites.

We conducted 20 independent runs of STRUCTURE v. 2.3.4 for $K = 1-10$ (Pritchard et al., 2000) using the microsatellite data and 600,000 generations, with the first 100,000 discarded as burn-in. We visualized the STRUCTURE results using the programs Clumpak and DISTRUCT v.1.1 (Kopelman et al., 2015; Rosenberg, 2004), and we inferred the optimal value of K using relevant guidelines (Cullingham et al., 2020; Earl, 2012; Evanno et al., 2005). We repeated these analyses for the Caribbean microsatellite dataset (40 independent runs of STRUCTURE for $K = 1-6$), and for the Central and Southwest microsatellite datasets combined (20 independent runs of STRUCTURE for $K = 1-10$) to further explore the regional genetic structure of the two regions that include new populations. Additionally, we evaluated genetic structure within the Caribbean using a multivariate approach, Discriminant Analysis of Principal Components (DAPC), on the microsatellite data using the Adegenet package (Jombart, 2008).

A PCA for the complete SNP dataset was generated with Plink v.1.9 (Purcell, 2016; Purcell et al., 2007). Additionally, we ran 5 independent runs in fastSTRUCTURE 1.0 (Raj et al., 2014) using the SNP dataset for $K = 1-10$, and visualized the results using Clumpak and DISTRUCT v.1.1 (Kopelman et al., 2015; Rosenberg, 2004).

Based on the results from our genetic clustering analysis, we grouped the samples into six regions for further analyses (Table 1, Figure 1): Caribbean contains the Caribbean; Southeast contains Florida, Louisiana, Georgia, and Washington D.C.; Central contains eastern Texas, western Texas, New Mexico; Southwest contains Arizona; Southern CA contains southern California and Nevada; and Northern CA contains northern/central California.

2.4 | Prediction 1: Genetic diversity decreases toward the west

Observed heterozygosity (H_O), expected heterozygosity (H_E), the inbreeding coefficient (F_{IS}), and a number of private alleles (alleles found in no other population) were calculated from the microsatellite dataset for each population using GenAlEx v. 6.51 (Peakall & Smouse, 2006), and allelic richness was estimated by rarefaction ($N = 30$) using the software HP-Rare v. 1.0 (Kalinowski, 2005). These measurements were not calculated for the SNP dataset due to ascertainment bias in the design of the SNP array (Evans et al., 2015).

To assess if genetic diversity decreased westward, as expected from SFE, we calculated linear regressions for longitude versus the different genetic diversity metrics (observed heterozygosity, expected heterozygosity, and allelic richness) in R v. 4.0.2 (R Core Team, 2020). To determine whether Florida or the Caribbean was the more likely source of the expansion, we also calculated linear regressions for observed heterozygosity, expected heterozygosity, and allelic richness of each population versus (1) their distance to Florida and (2) their distance to the Caribbean.

We assessed regional genetic diversity by averaging the genetic diversity metrics across all the sites within each region. Since regional allelic richness and the number of private alleles could be

biased by the number of individuals in a region, we calculated the genetic diversity measures a second time after combining all individuals within each region and then randomly resampling them so that each region had the same number of individuals.

2.5 | Prediction 2: Positive relationship between geographic and genetic distance

Using both the microsatellite and SNP datasets, we calculated pairwise F_{ST} and evaluated significance with 1000 permutations using Arlequin 3.5 (Excoffier et al., 2005). We then tested for a relationship between pairwise genetic distance $\left(\frac{F_{ST}}{1-F_{ST}}\right)$ and geographic distance using a Mantel test with 9999 permutations, and repeated the test after excluding known new invasions, defined as populations that were first detected in 2013 or later (California, Las Vegas NV, and Albuquerque NM).

Because clustering and assignment methods may mistake continuous processes (e.g., isolation by distance, in which there is a positive relationship between geographic distance and gene flow between populations) for discrete processes, we implemented a method called conStruct v. 1.0.4 (Bradburd et al., 2018) using the SNP array data. This model-based clustering method uses isolation by distance when possible to explain genetic variation. We ran three independent runs of the program for $K = 1-4$ with 1000 iterations and the spatial model setting and another three independent runs with the non-spatial model setting. We assessed posterior probability, Markov chain Monte Carlo (MCMC) performance, and layer contributions to compare the independent runs and identify a putative optimal number of layers for each run. The program performs best when there are more loci than the number of samples, so it was not suitable to run with the microsatellite data.

2.6 | Prediction 3: Daughter groups nested within parental group

To test the hypothesis that daughter groups contain a subset of the allelic diversity from parent groups, we created the presence/absence matrices representing all alleles for each of the 12 microsatellites, where regional sites were columns, and the different alleles were rows. Microsatellites were used rather than SNPs because the latter are biallelic, and thus do not have the broad allele frequency spectra required for this type of analysis. To maximize the chance of finding a pattern if one existed, we focused on three regions that were most likely to display this pattern: Southeast, Central, and Southwest. To control for uneven sample size across different populations, we combined individuals within each region and resampled, so each region was represented by the same number of individuals before analysis. We calculated the nestedness metric based on the overlap and decreasing fill (NODF) for each matrix, in which higher scores indicate greater amounts of nestedness (Morrison, 2013),

using the R package “RInSp” (Zaccarelli et al., 2013). To compare our matrices and their NODF scores with the null expectation, we created five control model matrices by shuffling the elements in each row using the R package “picante,” specifically the randomizeMatrix tool, with 1000 iterations and the null model set as “richness” (Kembel et al., 2010).

2.7 | Prediction 4: Demographic inference modeling supports SFE

We performed demographic history analysis using DIYABC-RF (Random Forest) v.1.1.1-beta (Collin et al., 2021) on the microsatellite dataset, to avoid possible effects of ascertainment bias on allele frequency spectra derived from the SNP chip selection process. The DIYABC-RF approach enables efficient discrimination among scenarios and estimation of the posterior probabilities with a lower computational burden than classic approximate Bayesian computation approaches (Collin et al., 2021). The program applies supervised machine learning methods to population genetic data for statistical inference through the use of a training set. The training set includes a given number of datasets simulated under different evolutionary scenarios using parameter values drawn from prior distributions. Priors were set to be as wide as possible within reason given known colonization events (Table S2). Each resulting dataset is then summarized with a set of descriptive statistics. These summary statistics describe genetic variation within populations, between pairs or triplets of populations, averaged over loci. The RF algorithm then chooses the best scenario from the simulated datasets and characterizes the posterior distribution of parameters of interest under a specific scenario and assesses the performance of RF-based inferences in terms of prediction through the computation of error and accuracy measurements.

Specifically, we compared SFE scenario—in which Southern CA split from Southwest, which split from Central, which split from the Southeast—against a scenario where each region diverged from the Southeast independently (Figure 2). Each region was represented

by 158 randomly selected samples, and we performed the analysis in duplicate, using a second random draw from each region. We excluded Exeter from Southern California since it is a known outlier that likely has a different origin than the other sites in Southern California (see Results). To ensure that pooling across populations did not bias the outcome, we also repeated the analysis twice replacing the regions with single, arbitrarily selected populations from each region (Table S2). We ran DIYABC-RF using a training set of 200,000 simulated datasets. The constructed RF for both model choice and parameter estimation used 1000 trees, and the best scenario was selected based on linear discriminant analysis and partial least squares regression analysis, in each case. Full parameters and priors are listed in Table S2.

3 | RESULTS

3.1 | Inferring geographic regions based on genetic structure

We first confirmed that our microsatellites behave as independent, neutral loci by performing 756 Hardy–Weinberg equilibrium (HWE) tests (12 microsatellite loci X 63 populations) and 4158 linkage disequilibrium pairwise tests on loci within populations. After applying a Bonferroni correction ($p = .05/\text{number of tests}$), 13 loci-by-populations (1.7%) were out of HWE, and there were not enough data to determine the p-values for 18 tests. Similarly, 57 loci pairs within populations (1.4%) showed significant evidence of being in linkage disequilibrium after a Bonferroni correction, and there were not enough data to determine the p-values for 127 tests. These observations, 1.7% and 1.4%, are below the “significance” cutoff of 5%.

To test our hypothesis that *Ae. aegypti* primarily colonized North America through a series of westward founder effects, it is useful to represent our 70 populations as major regions defined by their geography and population structure. We explored genetic population structure using both microsatellites and SNP markers and various methods. Using Bayesian clustering with the CV error

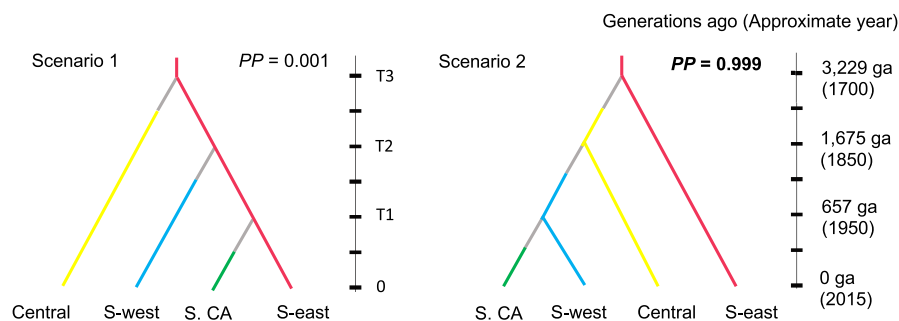


FIGURE 2 Scenarios tested by DIYABC-RF using microsatellite data. The two trees depict scenarios of *Ae. aegypti* colonization of continental North America. Scenario 1 shows independent invasions from the Southeast with bottlenecks after each lineage division and Scenario 2 shows a pattern of serial founder effect moving west from the Southeast, also with bottlenecks. Posterior probabilities (PP) are shown for each scenario. Time 0 corresponds to the present (when the samples were collected). Colors for each lineage correspond to Figure 1, and changes in effective population size (bottlenecks) are shown in gray. Posterior probabilities and divergence times correspond to a single run of the model using random draws from each region (Table S2b); results are similar for the other three independent runs (Table S2)

method (Evanno et al., 2005) for the microsatellite data, we found two primary clusters which generally split the west half of North America from the east half, with a large transition zone through New Mexico and Texas (Figure S1a). At $K = 4$, the four ancestries correspond generally to 1: Northern California + Central, 2: Southwest, 3: Southeast, and 4: Caribbean (Figure 1b). Southern CA has sites that are outliers (e.g., Exeter) or appear to show admixture (e.g., Mission Viejo), consistent with the complex, recent invasions known to characterize the region (Pless et al., 2020). Bayesian clustering using the SNP data reinforced the results from the microsatellites and added additional resolution in southern California, for example showing that Garden Grove and Mission Viejo each have distinct genetic signatures (Figure 3).

PCA using SNPs showed a similar pattern. The Southeast and Caribbean are grouped together on the right of the graph, along with Exeter, a known outlier that may represent a third invasion into California (Figure S2). Southern California and Southwest clustered in the top left, Northern CA was in the bottom right, and Central was in the middle. PCA using the microsatellites showed a similar pattern but with a higher level of overlap among all the populations (not shown). The consistency in the clustering results suggests the identified regions could represent major “steps” during SFE colonization and that subsequent analyses could be performed using these groupings.

We further explored the genetic structure within the Caribbean and the Southwest, since we have new populations in these regions compared to previous publications. The CV error method (Evanno et al., 2005) suggested $K = 5$ for the Caribbean microsatellite data (Figure S1e), but $K = 3$ made the most biological sense, showing moderate differentiation among the Dominican Republic (DR), United States Virgin Islands (USVI), and Puerto Rico (PR) (Figure S1d). DAPC showed a connection between some of the samples from Ocoa, DR, and PR (Figure S3). In the Southwest and Central, Bayesian clustering at $K = 2$ found a strong divide between Arizona and New Mexico (Figure S1b); at $K = 3$ western NM (through Alamogordo) separated from more eastern sites in New Mexico and Texas (Figure S1c).

3.2 | Prediction 1: Genetic diversity decreases toward the west

The microsatellite data showed that, overall, the highest genetic diversity was found in the Southeast, followed by the Caribbean, Central, Southwest, Northern CA, and Southern CA (R^2 for longitude vs. allelic richness = 0.32, $p < 10^{-8}$) (Figure 1, Table 2). After combining and standardizing the number of individuals per region, we found that Southeast and Southern California had the highest number of private alleles (alleles found in no other region), followed by the Caribbean and Central (Table 2). This is consistent with the Southeast having the oldest population and Southern California being founded multiple times from different regions. There was a stronger correlation between heterozygosity versus distance from

Florida ($R^2 = .53$, $p < 10^{-11}$) than between heterozygosity and distance from the Caribbean ($R^2 = .32$, $p < 10^{-11}$).

3.3 | Prediction 2: Positive relationship between geographic and genetic distance

Considering the microsatellite data, the F_{ST} (a measure of genetic differentiation due to genetic structure) values between all population pairs ($N = 1953$) were 0.13 ± 0.070 (mean \pm SD), and all values were significantly different than zero at the 0.05 level after correction for multiple tests (Table S4). Geographic distance was a significant predictor of genetic distance (calculated as $\frac{F_{ST}}{(1-F_{ST})}$) (Mantel $R = .14$, $p = .024$) (Figure 4a). After removing known new invasions from the dataset (detected in 2013 or later), the strength of the relationship increased (Mantel $R = .52$, $p < 10^{-4}$) (Figure 4b). Similarly, the pairwise F_{ST} values generated from the SNP data ($N = 630$) were 0.10 ± 0.045 (mean \pm SD), and all values were significantly different than zero except for the pairs between Madera and San Mateo (Table S5). Using the SNP data, there was also a significant correlation between geographic and genetic distance (Mantel $R = 0.23$, $p = .0038$) (Figure 4c), and the strength of the correlation increased when new invasions were removed (Mantel $R = .62$, $p < 10^{-4}$) (Figure 4d).

To investigate the effect of continuous processes, namely isolation by distance, on clustering, we ran conStruct for $K = 1-4$ using the SNP dataset. We focus on $K = 2$ because this model consistently produced well-behaved MCMC posterior probability plots, and the first two layers contributed most of the variance in the majority of the runs (Table S6). The three independent non-spatial models produced very consistent results at $K = 2$, showing a clear structure, especially between Florida + Exeter and California (Figure 5 and Figure S4a). After controlling for the effect of isolation by distance, the three spatial runs showed significantly less genetic structure (Figure 5 and Figure S4b), although one run continued to show Exeter as a clear outlier (Figure S4c). In all cases, the posterior probability across MCMC runs was higher for the spatial models than the non-spatial models. The non-spatial model showed additional structure at $K = 3$ and $K = 4$ (e.g., differences between the Caribbean and continental North America), but again the spatial model at these higher K levels showed less structure.

3.4 | Prediction 3: Daughter groups nested within parental group

To evaluate if each putative daughter group contained a subset of the diversity in the presumed parent group, we calculated NODF for each microsatellite allele for the three regions where we expected the larger effect: Southeast, Central, and the Southwest. However, nine loci had <4 alleles, an insufficient number to detect nestedness patterns (Table 3). Of the three remaining loci, one (AG2) showed

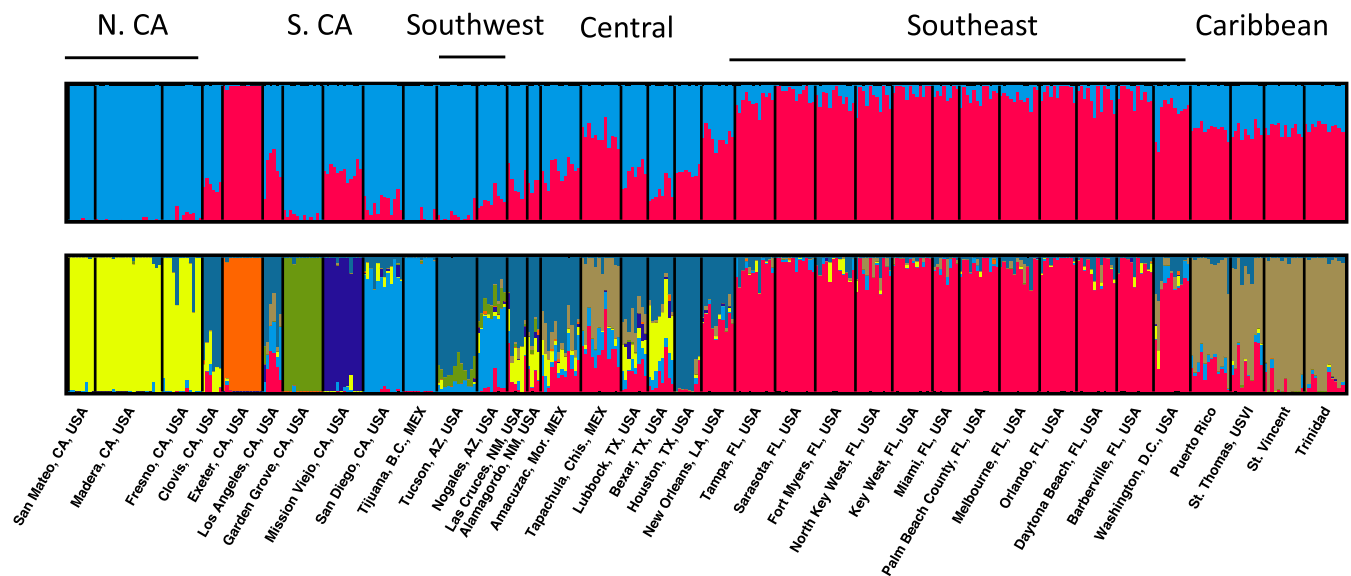


FIGURE 3 Population structure in North America generated by fastStructure using the SNP dataset. Populations are arranged according to their longitude (west on the left). Each column is an individual, each color represents an inferred group, and the height of the color bar shows the proportion of ancestry that came from the K inferred groups ($K = 2$ on the top panel and $K = 8$ on the bottom panel)

TABLE 2 Genetic diversity of each region calculated with microsatellites

Region	$H_o \pm SD^a$	H_o^b	$H_e \pm SD^a$	H_e^b	$AR \pm SD^a$	AR^b	Private alleles ^b
Caribbean	0.55 ± 0.032	0.56	0.56 ± 0.030	0.60	4.21 ± 0.32	6.65	4
Southeast	0.62 ± 0.035	0.62	0.62 ± 0.016	0.65	4.42 ± 0.22	6.89	8
Central	0.54 ± 0.051	0.55	0.56 ± 0.051	0.61	3.97 ± 0.44	6.31	3
Southwest	0.55 ± 0.029	0.57	0.57 ± 0.023	0.60	3.83 ± 0.018	5.23	1
Southern CA	0.47 ± 0.086	0.48	0.47 ± 0.072	0.58	2.99 ± 0.53	6.07	6
Northern CA	0.49 ± 0.015	0.49	0.52 ± 0.023	0.57	3.69 ± 0.37	4.82	0

Abbreviations: H_o , observed heterozygosity; H_e , expected heterozygosity; AR, allelic richness estimated by rarefaction ($n = 30$); Private alleles, number of unique alleles found only in that region.

^aMean of populations within each region.

^bIndividuals in each region combined and resampled to standardized size before analyses.

evidence of nestedness (NODF score $>1SD$ above the neutral NODF scores), and the other two were within $1SD$ of the neutral model (Table 3). Additional analyses using more regions or individual populations rather than regions did not show significant nestedness.

3.5 | Prediction 4: Demographic inference modeling supports SFE

Using approximate Bayesian computation combined with supervised machine learning, we compared two invasion scenarios to test whether the data were more consistent with SFE or independent invasions from the Southeast (Figure 2). SFE had a higher posterior probability (PP > 0.922 across all four independent runs) than independent invasions from the Southeast (PP < 0.07 across all four independent runs). Under the SFE model and assuming ten generations/year, we estimated that Southern CA split from Southwest

~ 39 – 66 years ago, Southwest split from Central ~ 166 – 207 years ago, and Central split from Southeast ~ 322 – 345 years ago (Table S2). We ran the model twice using randomly selected samples from each region and twice using single populations from each region; results were similar across the four runs (Table S2).

4 | DISCUSSION

4.1 | Regional genetic structure

In this study, we tested whether SFE described the westward spread of *Ae. aegypti* across North America by evaluating the genetic diversity and population structure from a rich dataset of 70 populations. We first established relevant regions for analysis based on genetic structure; these are the hypothesized “stepping-stone” regions for the westward spread of *Ae. aegypti*. We identified six genetic regions

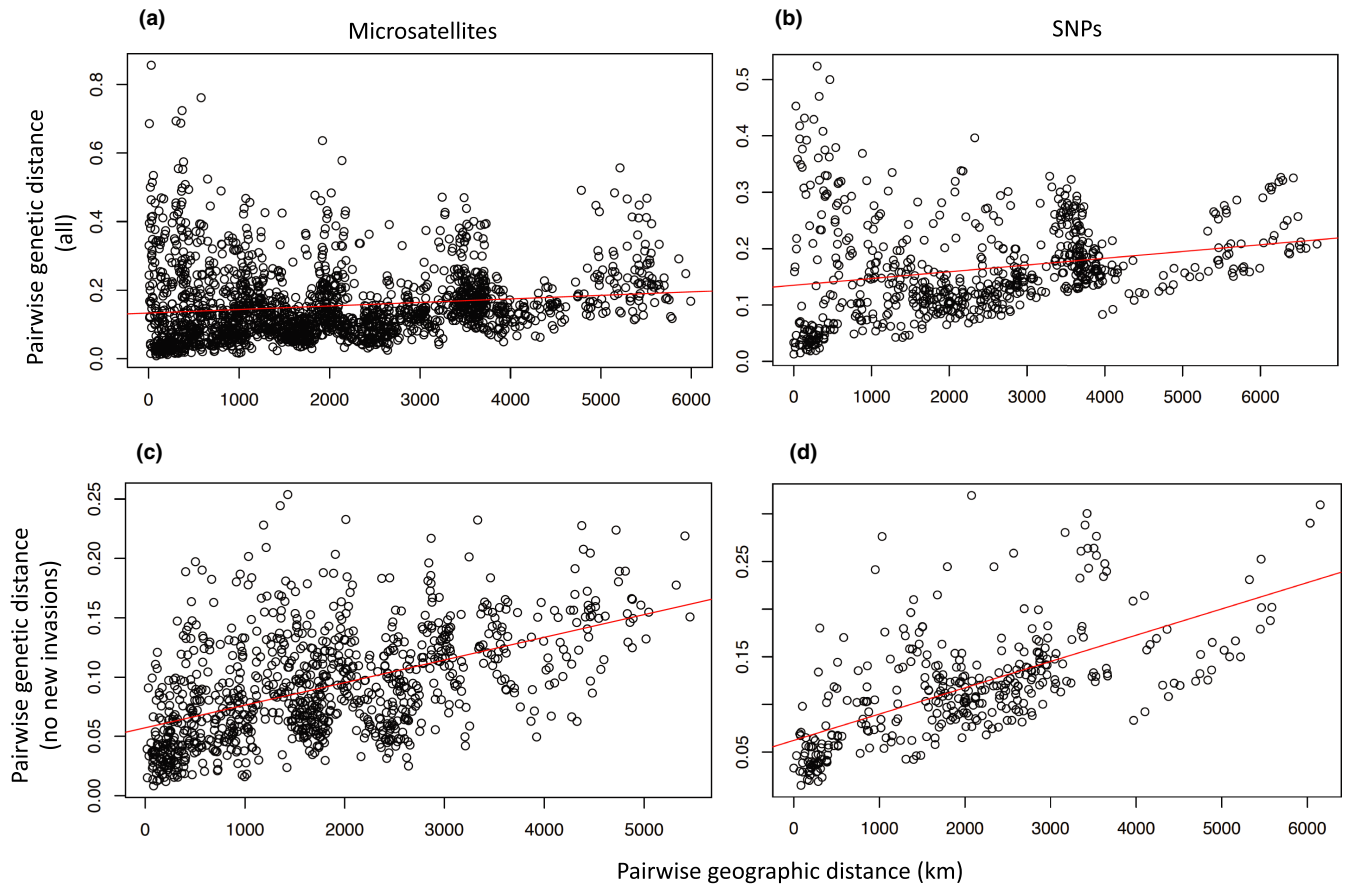


FIGURE 4 Pairwise genetic distance (linearized F_{ST}) for pairs of populations as a function of the geographic distance (km) between them for (a) all microsatellite data, (b) all SNP data, (c) all microsatellite data excluding known new invasions (California, Las Vegas NV, and Albuquerque NM), and (d) all SNP data excluding known new invasions

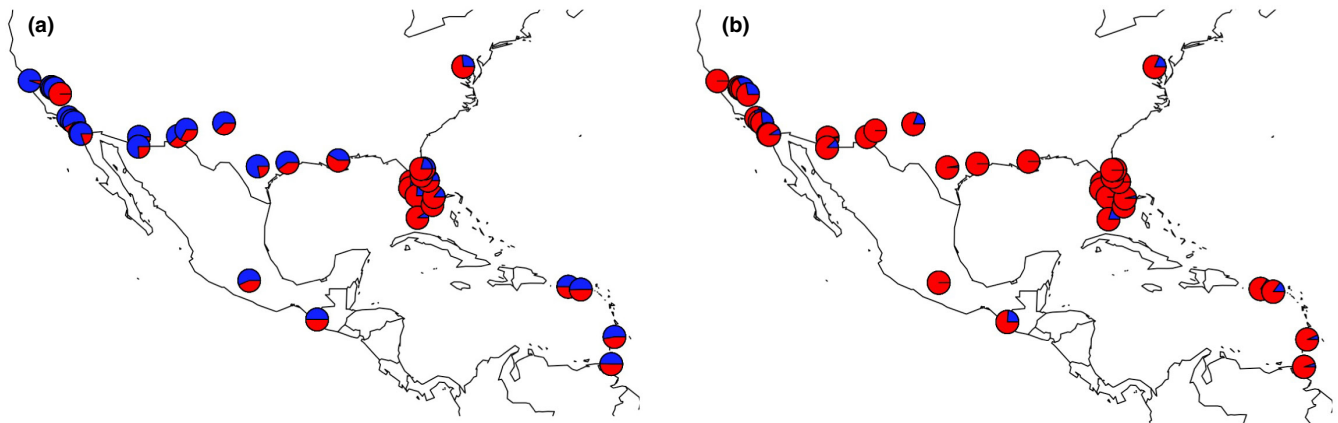


FIGURE 5 North America population structure ($K = 2$) generated in conStruct with the SNP dataset using (a) a non-spatial model and (b) a spatial model that first attributes genetic variation to isolation by distance where possible

based on both microsatellites and SNPs that are consistent with previous work (Gloria-Soria et al., 2016; Kotsakiozi et al., 2018). Broadly, we detected two primary clusters in the microsatellite dataset which split the eastern populations from the western ones with a large admixture zone through the Southwest and Central regions (Figure 2a). At higher levels of K , additional structure emerged (Figure 2), and we found genetic differentiation by geography with some outliers,

specifically Exeter CA, which clusters with Florida and likely represents the third invasion into the state (Figure S1). Southern California has unusual genetic patterns (e.g., high genetic differentiation and a high number of private alleles despite low diversity), as discussed elsewhere (Pless et al., 2020). The stability of the regions we identified is indicated by older allozyme data from parts of the same distribution using collections from the 1970s and 1980s (Wallis et al.,

TABLE 3 Test of allelic nestedness for Southeast, Central, and Southwest using microsatellite data

Microsatellite	NODF	Neutral \pm SD	No. alleles total	No. alleles not fixed
AG2*	64.7	61.9 \pm 1.5	20	9
AC5	49.5	49.4 \pm 1.0	16	5
AC2	65.3	58.3 \pm 7.8	6	4
AC1	70.8	64.5 \pm 4.1	7	3
AG1	56.7	67.3 \pm 3.8	7	3
B2	65.3	62.2 \pm 6.8	6	3
AG5	45.2	45.2 \pm 0.0	8	2
A1	41.7	51.7 \pm 9.1	7	2
B3	46.2	45.2 \pm 0.0	5	1
CT2	66.7	66.7 \pm 0.0	3	1

Note: Asterisk (*) indicates nestedness metric (NODF) significantly greater than neutral score.

Abbreviations: Microsatellite, each microsatellite locus; NODF, nestedness metric based on the overlap and decreasing fill; Neutral \pm SD, generated by randomly shuffling the rows of the nestedness matrix; No. alleles total, number of alleles for given microsatellite; No. alleles not fixed, number of alleles that vary across the three regions.

1983). In particular, the break between the Southeast and Central around the border of Texas and Louisiana was clearly identified in these early studies.

We further explored the intra-region structure for the Central and Caribbean, since these are the regions where we added new genetic data. Bayesian clustering using microsatellites found two primary ancestries in the Southwest and Central regions combined, clearly separating the five most eastern sites (AZ and MX) from the other sites (NM, TX, and MX) (Figure S2b). At $K = 3$, an additional divide emerged between Alamogordo NM and Roswell NM (Figure S2c). Despite some geographic structuring in the Caribbean (Figures S2d and S4), most analyses show high homogeneity among these sites, consistent with significant human-mediated gene flow among the islands as previously reported (Wallis et al., 1984). Although *Ae. aegypti* was likely established in the Caribbean and Florida at similar times in history (1600–1700s), the lower genetic diversity and higher amount of genetic structure in the Caribbean may reflect the accumulation of genetic drift due to smaller population sizes and lower migration between the islands than between sites in Florida and/or population bottlenecks from eradication efforts on the islands (Dick et al., 2012; Sherpa et al., 2018).

4.2 | Genetic predictions of SFE

The genetic data supported three of our predictions of SFE (#1, #2, and #4), with the Southeast as the putative source for westward spread across North America. The highest numbers of private alleles were found in Florida and the Caribbean, and genetic diversity

largely decreased in populations westward (Tables 1 and 2, Figure 1). Distance to Florida was strongly correlated with heterozygosity; distance to the Caribbean was correlated to heterozygosity, but the signal was weaker. Florida likely harbors the oldest and most stable populations of *Ae. aegypti* in continental North America today because (1) southern Florida is hospitable year-round for *Ae. aegypti* (Johnson et al., 2017), (2) vector control never fully eliminated *Ae. aegypti* from the region (Slosek, 1986; Soper, 1965), and (3) *Ae. albopictus* never fully displaced *Ae. aegypti* from the region (Lounibos et al., 2016).

In line with our second expectation of SFE, geographic distance was a strong predictor of genetic distance, especially after removing recent invasions on the west coast (including the highly differentiated southern California sites, and the northern California sites, which were likely founded by long-distance human-mediated movement) (Pless et al., 2017) (Figure 4). We also found that some of the genetic structures in North America could be explained by a continuous process like isolation by distance (Bradburd et al., 2018). Accounting for genetic variance due to isolation by distance resulted in less genetic clustering (Figure 5 and Figure S4).

Our third expectation, that daughter populations would contain a subset of the diversity contained by their parental group, could not be properly tested. The overall low genetic diversity of *Ae. aegypti* in the Americas, relative to that of its native Africa (Gloria-Soria et al., 2016), resulted in a limited number of alleles per microsatellite and frequent presence of rare alleles that prevented the detection of diversity nestedness in the dataset. Exploring different regional categories and using higher sample sizes could help clarify these results. We do find support for the SFE model in the demographic analysis (prediction 4). Specifically, westward spread via a series of founder effects was better supported than independent invasions from the Southeast.

To the best of our knowledge, the SFE model has not been tested for any mosquito species, and has only been tested on a few insects more generally, see Pierce et al. (2014). We argue this model is useful in thinking about the expansion of *Ae. aegypti* out of Africa and across continents such as North America. Like isolation by distance, SFE can serve as a null model for *Ae. aegypti* genetic structure in North America, and outliers from the model warrant additional research and explanation (e.g., California being founded by multiple invasions, some from across the country). Our study also has some limitations that should be addressed in future work. While the linear decline of heterozygosity is indicative of SFE, it can also be explained by extensive admixture from an early branching lineage into later branching lineages (Pickrell & Reich, 2014). The increase in genetic distance with an increase in geographic distance can be caused by the budding pattern of migration under SFE or solely by isolation by distance (Wright, 1943). More extensive sequencing and modeling would be useful for testing these different demographic scenarios.

Clearly, *Ae. aegypti* can and does make stochastic long-distance jumps, for example, from the U.S. southeast to the Netherlands (Brown et al., 2011) and to California (Gloria-Soria et al., 2014). However, the demonstrated relevance of SFE and of environmental

variables (Pless et al., 2021) on *Ae. aegypti* gene flow suggests short-distance migration is important. Given the importance of short-distance migration, controlling *Ae. aegypti* in one area should help protect areas around it from becoming infested. These short-distance migrations occur across the state and country lines (e.g., the Mexico samples cluster with U.S. samples at the same longitude, not with each other), highlighting the importance of international cooperation to prevent further invasions and control vector-borne diseases.

AUTHOR CONTRIBUTIONS

Evlyn Pless: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Methodology (equal); Writing – original draft (lead); Writing – review & editing (equal). **Jeffrey Powell:** Conceptualization (equal); Funding acquisition (lead); Methodology (equal); Supervision (equal); Writing – review & editing (equal). **Krystal Seger:** Resources (equal); Writing – review & editing (equal). **Brett Ellis:** Resources (equal); Writing – review & editing (supporting). **Andrea Gloria-Soria:** Conceptualization (supporting); Data curation (supporting); Formal analysis (supporting); Methodology (equal); Supervision (equal); Writing – original draft (supporting); Writing – review & editing (equal).

ACKNOWLEDGMENTS

We thank our sample contributors: Kathryn Hanley and the NMSU-NMDoH collaborative SWARM project (New Mexico); Bethany Bolling, Alexander Wilson-Fallon, and the Texas Department of State Health Services (Texas); Antonio De la Mora Covarrubias (Juarez); Angel Solís and Yennis Ferreras (Dominican Republic); David Delgado, Aubrey Drummond II, and the U.S. Virgin Islands Department of Health (USVI). Thanks to Casey Dunn, Kim Medley, and Muyang Lu for helpful discussion and ideas. Some of the data collection was supported by NIH grant RO1-AI101112. AG-S was supported by the Cooperative Agreement U01CK000509, funded by the Center for Disease Control and Prevention. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the Center for Disease Control and Prevention or the Department of Health and Human Services.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Genetic data (microsatellite calls and SNP array dataset) are included in the supplementary materials and are archived at Dryad (<https://doi.org/10.5061/dryad.5x69p8d5j>) and VectorBase (VBPO000801).

ORCID

Evlyn Pless  <https://orcid.org/0000-0002-6402-3397>

Jeffrey R. Powell  <https://orcid.org/0000-0002-7651-5153>

Andrea Gloria-Soria  <https://orcid.org/0000-0002-5401-3988>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1), 33–52. <https://doi.org/10.1534/genetics.118.301333>
- Brady, O. J., & Hay, S. I. (2020). The global expansion of Dengue: How *Aedes Aegypti* mosquitoes enabled the first pandemic arbovirus. *Annual Review of Entomology*, 65, 191–208.
- Brown, J. E., Scholte, E.-J., Dik, M., Den Hartog, W., Beeuwkes, J., & Powell, J. R. (2011). *Aedes aegypti* mosquitoes imported into the Netherlands, 2010. *Emerging Infectious Diseases*, 17(12), 2335–2337.
- Carlson, C. J., Dougherty, E. R., & Getz, W. (2016). An ecological assessment of the pandemic threat of Zika virus. *PLOS Neglected Tropical Diseases*, 10(8), e0004968. <https://doi.org/10.1371/journal.pntd.0004968>
- Carrigan, J. A. (1959). Yellow fever in New Orleans, 1853: Abstractions and realities. *The Journal of Southern History*, 25(3), 339–355. <https://doi.org/10.2307/2954767>
- Chandler, A. (1956). History of *Aedes aegypti* control work in Texas. *Mosquito News*, 16, 58–63.
- Collin, F.-D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J.-M., & Estoup, A. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Molecular Ecology Resources*, 21(8), 2598–2613. <https://doi.org/10.1111/1755-0998.13413>
- Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., & Janes, J. K. (2020). Confidently identifying the correct K value using the ΔK method: When does K= 2? *Molecular Ecology*, 29, 862–869.
- Dick, O. B., San Martín, J. L., Montoya, R. H., del Diego, J., Zambrano, B., & Dayan, G. H. (2012). The history of dengue outbreaks in the Americas. *The American Journal of Tropical Medicine and Hygiene*, 87(4), 584–593. <https://doi.org/10.4269/ajtmh.2012.11-0770>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Evans, B. R., Gloria-Soria, A., Hou, L., McBride, C., Bonizzoni, M., Zhao, H., & Powell, J. R. (2015). A multipurpose, high-throughput single-nucleotide polymorphism chip for the dengue and yellow fever mosquito, *Aedes aegypti*. *G3 (Bethesda)*, 5(5), 711–718.
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>
- Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, 47–50. 117693430500100003.
- Fonzi, E., Higa, Y., Bertuso, A. G., Futami, K., & Minakawa, N. (2015). Human-mediated marine dispersal influences the population structure of *Aedes aegypti* in the Philippine Archipelago. *PLOS Neglected Tropical Diseases*, 9(6), e0003829. <https://doi.org/10.1371/journal.pntd.0003829>
- Gloria-Soria, A., Ayala, D., Bheecarry, A., Calderon-Arguedas, O., Chadee, D. D., Chiappero, M., & Kamal, H. A. (2016). Global genetic diversity of *Aedes aegypti*. *Molecular Ecology*, 25, 5377–5395.
- Gloria-Soria, A., Brown, J. E., Kramer, V., Yoshimizu, M. H., & Powell, J. R. (2014). Origin of the dengue fever mosquito, *Aedes aegypti*, in

- California. *PLoS Neglected Tropical Diseases*, 8(7), e3029. <https://doi.org/10.1371/journal.pntd.0003029>
- Gloria-Soria, A., Lima, A., Lovin, D. D., Cunningham, J. M., Severson, D. W., & Powell, J. R. (2018). Origin of a high-latitude population of *Aedes aegypti* in Washington, DC. *The American journal of tropical medicine and hygiene*, 98(2), 445.
- Goncalves da Silva, A., Cunha, I. C., Santos, W. S., Luz, S. L., Ribolla, P. E., & Abad-Franch, F. (2012). Gene flow networks among American *Aedes aegypti* populations. *Evolutionary Applications*, 5(7), 664–676.
- Guagliardo, S. A., Barboza, J. L., Morrison, A. C., Astete, H., Vazquez-Prokopec, G., & Kitron, U. (2014). Patterns of geographic expansion of *Aedes aegypti* in the Peruvian Amazon. *PLoS Neglected Tropical Diseases*, 8(8), e3033. <https://doi.org/10.1371/journal.pntd.0003033>
- Hamlet, A., Jean, K., Perea, W., Yactayo, S., Biey, J., Van Kerkhove, M., Ferguson, N., & Garske, T. (2018). The seasonal influence of climate and environment on yellow fever transmission across Africa. *PLOS Neglected Tropical Diseases*, 12(3). <https://doi.org/10.1371/journal.pntd.0006284>
- Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), 17758–17764. <https://doi.org/10.1073/pnas.1212380109>
- Honorio, N. A., Silva, W. D., Leite, P. J., Goncalves, J. M., Lounibos, L. P., & Lourenco-de-Oliveira, R. (2003). Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. *Memorias do Instituto Oswaldo Cruz*, 98(2), 191–198. <https://doi.org/10.1590/S0074-02762003000200005>
- Hunley, K., Bowern, C., & Healy, M. (2012). Rejection of a serial founder effects model of genetic and linguistic coevolution. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736), 2281–2288.
- Iwamura, T., Guzman-Holst, A., & Murray, K. A. (2020). Accelerating invasion potential of disease vector *Aedes aegypti* under climate change. *Nature Communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-020-16010-4>
- Jasper, M., Schmidt, T. L., Ahmad, N. W., Sinkins, S. P., & Hoffmann, A. A. (2019). A genomic approach to inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito. *Molecular Ecology Resources*, 19(5), 1254–1264. <https://doi.org/10.1111/1755-0998.13043>
- Johnson, T. L., Haque, U., Monaghan, A. J., Eisen, L., Hahn, M. B., Hayden, M. H., Savage, H. M., McAllister, J., Mutebi, J.-P., & Eisen, R. J. (2017). Modeling the environmental suitability for *Aedes* (*Stegomyia*) *aegypti* and *Aedes* (*Stegomyia*) *albopictus* (Diptera: Culicidae) in the contiguous United States. *Journal of Medical Entomology*, 54(6), 1605–1614. <https://doi.org/10.1093/jme/tjx163>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kalinowski, S. T. (2005). hp-rare 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, 5(1), 187–189.
- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11), 1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191.
- Kotsakiozi, P., Gloria-Soria, A., Caccone, A., Evans, B., Schama, R., Martins, A. J., & Powell, J. R. (2017). Tracking the return of *Aedes aegypti* to Brazil, the major vector of the dengue, chikungunya and Zika viruses. *PLoS neglected tropical diseases*, 11(7), e0005653.
- Kotsakiozi, P., Gloria-Soria, A., Schaffner, F., Robert, V., & Powell, J. R. (2018). *Aedes aegypti* in the Black Sea: Recent introduction or ancient remnant? *Parasites & Vectors*, 11, 1–3. <https://doi.org/10.1186/s13071-018-2933-2>
- Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A. Q. N., Shearer, F. M., Barker, C. M., Moore, C. G., Carvalho, R. G., Coelho, G. E., Van Bortel, W., Hendrickx, G., Schaffner, F., Elyazar, I. R. F., Teng, H.-J., Brady, O. J., Messina, J. P., Pigott, D. M., Scott, T. W., Smith, D. L., ... Hay, S. I. (2015). The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife*, 4, e08347. <https://doi.org/10.7554/eLife.08347>
- Leparco-Goffart, I., Nougaiere, A., Cassadou, S., Prat, C., & De Lamballerie, X. (2014). Chikungunya in the Americas. *The Lancet*, 383(9916), 514. [https://doi.org/10.1016/S0140-6736\(14\)60185-9](https://doi.org/10.1016/S0140-6736(14)60185-9)
- Lounibos, L. P., Bargielowski, I., Carrasquilla, M. C., & Nishimura, N. (2016). Coexistence of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in peninsular Florida two decades after competitive displacements. *Journal of Medical Entomology*, 53(6), 1385–1390.
- Medley, K. A., Jenkins, D. G., & Hoffman, E. A. (2015). Human-aided and natural dispersal drive gene flow across the range of an invasive mosquito. *Molecular Ecology*, 24(2), 284–295. <https://doi.org/10.1111/mec.12925>
- Merrill, S. A., Ramberg, F. B., & Hagedorn, H. H. (2005). Phylogeography and population structure of *Aedes aegypti* in Arizona. *The American Journal of Tropical Medicine and Hygiene*, 72(3), 304–310. <https://doi.org/10.4269/ajtmh.2005.72.304>
- Metzger, M. E., Hardstone Yoshimizu, M., Padgett, K. A., Hu, R., & Kramer, V. L. (2017). Detection and establishment of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) mosquitoes in California, 2011–2015. *Journal of Medical Entomology*, 54(3), 533–543. <https://doi.org/10.1093/jme/tjw237>
- Moreno-Madriñán, M. J., & Turell, M. (2018). History of mosquito-borne diseases in the United States and implications for new pathogens. *Emerging Infectious Diseases*, 24(5), 821. <https://doi.org/10.3201/eid2405.171609>
- Morrison, L. W. (2013). Nestedness in insular floras: Spatiotemporal variation and underlying mechanisms. *Journal of Plant Ecology*, 6(6), 480–488. <https://doi.org/10.1093/jpe/rtt002>
- Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29(1), 1–10. <https://doi.org/10.1111/j.1558-5646.1975.tb00807.x>
- Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), 288–295.
- Peischl, S., Dupanloup, I., Bosshard, L., & Excoffier, L. (2016). Genetic surfing in human populations: From genes to genomes. *Current Opinion in Genetics & Development*, 41, 53–61. <https://doi.org/10.1016/j.gde.2016.08.003>
- Pickrell, J. K., & Reich, D. (2014). Toward a new history and geography of human genes informed by ancient DNA. *Trends in Genetics*, 30(9), 377–389. <https://doi.org/10.1016/j.tig.2014.07.007>
- Pierce, A. A., Zalucki, M. P., Bangura, M., Udawatta, M., Kronforst, M. R., Altizer, S., Haeger, J. F., & de Roode, J. C. (2014). Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797), 20142230. <https://doi.org/10.1098/rspb.2014.2230>
- Pless, E., Gloria-Soria, A., Evans, B. R., Kramer, V., Bolling, B. G., Tabachnick, W. J., & Powell, J. R. (2017). Multiple introductions of the dengue vector, *Aedes aegypti*, into California. *PLOS Neglected Tropical Diseases*, 11(8), e0005718. <https://doi.org/10.1371/journal.pntd.0005718>
- Pless, E., Hopperstad, K. A., Ledesma, N., Dixon, D., Henke, J. A., & Powell, J. R. (2020). Sunshine versus gold: The effect of population age on genetic structure of an invasive mosquito. *Ecology and Evolution*, 10(18), 9588–9599. <https://doi.org/10.1002/ece3.6661>

- Pless, E., Saarman, N. P., Powell, J. R., Caccone, A., & Amatulli, G. (2021). A machine-learning approach to map landscape connectivity in *Aedes aegypti* with genetic and environmental data. *Proceedings of the National Academy of Sciences of the United States of America*, 118(9), e2003201118.
- Powell, J. R., Gloria-Soria, A., & Kotsakiozi, P. (2018). Recent history of *Aedes aegypti*: Vector genomics and epidemiology records. *BioScience*, 68(11), 854–860.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16(3), 608–627.
- Purcell, S. (2016). PLINK [1.9]. <http://pngu.mgh.harvard.edu/purcell/plink/>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- R Core Team (2020). *R: A language and environment for statistical computing (Version 4.0.2)*. R Foundation for Statistical Computing. <https://R-project.org/>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences*, 102(44), 15942–15947. <https://doi.org/10.1073/pnas.0507611102>
- Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86(3), 248–249.
- Reiter, P. (2007). Oviposition, dispersal, and survival in *Aedes aegypti*: Implications for the efficacy of control strategies. *Vector Borne Zoonotic Disease*, 7(2), 261–273.
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Russell, R. C., Webb, C. E., Williams, C. R., & Ritchie, S. A. (2005). Mark-release-recapture study to measure dispersal of the mosquito *Aedes aegypti* in Cairns, Queensland, Australia. *Medical and Veterinary Entomology*, 19(4), 451–457. <https://doi.org/10.1111/j.1365-2915.2005.00589.x>
- Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., Baughman, S., Cabin, R. J., Cohen, J. E., Ellstrand, N. C., McCauley, D. E., O'Neil, P., Parker, I. M., Thompson, J. N., & Weller, S. G. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics*, 32(1), 305–332. <https://doi.org/10.1146/annurev.ecolsys.32.081501.114037>
- Saarman, N. P., Gloria-Soria, A., Anderson, E. C., Evans, B. R., Pless, E., Cosme, L. V., Gonzalez-Acosta, C., Kamgang, B., Wesson, D. M., & Powell, J. R. (2017). Effective population sizes of a major vector of human diseases, *Aedes aegypti*. *Evolutionary Applications*, 10(10), 1031–1039.
- Sherpa, S., Rioux, D., Goindin, D., Fouque, F., François, O., & Despres, L. (2018). At the origin of a worldwide invasion: Unraveling the genetic makeup of the Caribbean bridgehead populations of the dengue vector *Aedes aegypti*. *Genome Biology and Evolution*, 10(1), 56–71. <https://doi.org/10.1093/gbe/evx267>
- Slosek, J. (1986). *Aedes aegypti* mosquitoes in the Americas: A review of their interactions with the human population. *Social Science & Medicine*, 23(3), 249–257. [https://doi.org/10.1016/0277-9536\(86\)90345-X](https://doi.org/10.1016/0277-9536(86)90345-X)
- Slotman, M. A., Kelly, N. B., Harrington, L. C., Kitthawee, S., Jones, J. W., Scott, T. W., Caccone, A., & Powell, J. R. (2007). Polymorphic microsatellite markers for studies of *Aedes aegypti* (Diptera: Culicidae), the vector of dengue and yellow fever. *Molecular Ecology Notes*, 7(1), 168–171.
- Soper, F. L. (1965). The 1964 status of *Aedes aegypti* eradication and yellow fever in the Americas. *The American Journal of Tropical Medicine and Hygiene*, 14(6), 887–891. <https://doi.org/10.4269/ajtmh.1965.14.887>
- Wallis, G. P., Tabachnick, W. J., & Powell, J. R. (1983). Macrogeographic genetic variation in a human commensal: *Aedes aegypti*, the yellow fever mosquito. *Genetics Research*, 41(3), 241–258.
- Wallis, G. P., Tabachnick, W. J., & Powell, J. R. (1984). Genetic heterogeneity among Caribbean populations of *Aedes aegypti*. *The American Journal of Tropical Medicine and Hygiene*, 33(3), 492–498. <https://doi.org/10.4269/ajtmh.1984.33.492>
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114. <https://doi.org/10.1093/genetics/28.2.114>
- Zaccarelli, N., Bolnick, D. I., & Mancinelli, G. (2013). RInSp: An r package for the analysis of individual specialization in resource use. *Methods in Ecology and Evolution*, 4(11), 1018–1023.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Pless, E., Powell, J. R., Seger, K. R., Ellis, B., & Gloria-Soria, A. (2022). Evidence for serial founder events during the colonization of North America by the yellow fever mosquito, *Aedes aegypti*. *Ecology and Evolution*, 12, e8896. <https://doi.org/10.1002/ece3.8896>