2018

Molecular Population Structure for Feral Swine in the United States

Blake E. McCann
Theodore Roosevelt National Park, blake.mccann@gmail.com

Timothy J. Smyser
USDA APHIS Wildlife Services, Timothy.J.Smyser@aphis.usda.gov

Brandon S. Schmit
USDA APHIS Wildlife Services, Brandon.S.Schmit@aphis.usda.gov

Robert A. Newman
University of North Dakota, robert.newman@und.nodak.edu

Antoinette J. Piaggio
USDA APHIS Wildlife Services, Toni.J.Piaggio@aphis.usda.gov

See next page for additional authors

Authors
Blake E. McCann, Timothy J. Smyser, Brandon S. Schmit, Robert A. Newman, Antoinette J. Piaggio, Mathew J. Malek, Seth R. Swafford, Richard A. Sweitzer, and Rebecca B. Simmons

The Journal of Wildlife Management 82(4):821–832; 2018; DOI: 10.1002/jwmg.21452

Research Article

Molecular Population Structure for Feral Swine in the United States

BLAKE E. McCANN,[1] *Resource Management, Wildlife, Theodore Roosevelt National Park, 315 2nd Avenue, P.O. Box 7, Medora, ND 58645, USA*
TIMOTHY J. SMYSER, *U.S. Department of Agriculture, Wildlife Services, 4101 La Porte Avenue, Fort Collins, CO 80521, USA*
BRANDON S. SCHMIT, *U.S. Department of Agriculture, Wildlife Services, 4101 La Porte Avenue, Fort Collins, CO 80521, USA*
ROBERT A. NEWMAN, *Department of Biology, University of North Dakota, 10 Cornell Street, Stop 9019, Grand Forks, ND 58202, USA*
ANTOINETTE J. PIAGGIO, *U.S. Department of Agriculture, Wildlife Services, 4101 La Porte Avenue, Fort Collins, CO 80521, USA*
MATHEW J. MALEK, *University of South Dakota, Sanford School of Medicine, Department of General Surgery, 1400 West 22nd Street, Sioux Falls, SD 57105, USA*
SETH R. SWAFFORD, *U.S. Fish and Wildlife Service, 12595 MS Highway 149, Yazoo City, MS 39194, USA*
RICHARD A. SWEITZER, *The Great Basin Institute, 16750 Mt. Rose Highway, Reno, NV 89511, USA*
REBECCA B. SIMMONS, *Department of Biology, University of North Dakota, 10 Cornell Street, Stop 9019, Grand Forks, ND 58202, USA*

ABSTRACT Feral swine (*Sus scrofa*) have invaded most of the United States and continue to expand throughout North America. Given the ecological and economic threats posed by increasing feral swine abundance, it is imperative to develop an understanding of their patterns of natural range expansion and human-mediated introductions. Towards this goal, we used molecular markers to elucidate the genetic structure of feral swine populations throughout the United States and evaluated the association between historical introductions and contemporary patterns of genetic organization. We used STRUCTURE and discriminant analysis of principal components (DAPC) to delineate genetic clusters for 959 individuals genotyped at 88 single nucleotide polymorphism loci. We identified 10 and 12 genetic clusters for the 2 clustering approaches, respectively. We observed strong agreement in clusters across approaches, with both describing clusters having strong geographic association at regional levels reflecting past introduction and range expansion patterns. In addition, we evaluated patterns of isolation by distance to test for and estimate spatial scaling of population structure within western, central, and eastern regions of North America. We found contrasting spatial patterns of genetic relatedness among regions, suggesting differences in the invasion process, likely as a result of regional variation in landscape heterogeneity and the influence of human-mediated introductions. Our results indicate that molecular analyses of population genetic structure can provide reliable insights into the invasion processes of feral swine, thus providing a useful basis for management focused on minimizing continued range expansion by this problematic species. © 2018 The Wildlife Society.

KEY WORDS DNA, feral, genetic, pig, population, swine, United States, wild boar, wild pig.

The history of feral swine (*Sus scrofa*) introductions in the United States is complex, with populations descending from varied domestic and wild origins (Mayer and Brisbin 1991, McCann et al. 2014, Sweitzer et al. 2015). Briefly, swine were first introduced to the United States as early as 1200 with the arrival of Polynesians in the Hawaiian Islands (Kirch 1982, Mayer and Brisbin 1991, Linderholm et al. 2016). By the 1500s, populations were established on the United States mainland with pigs introduced throughout the southeast and California as a result of deliberate releases by Spanish explorers, a common practice of the time to establish harvestable populations for subsequent exploration parties (Towne and Wentworth 1950, Hanson and Karstad 1959, Mayer and Brisbin 1991). Following establishment in the wild, the genetic attributes of feral swine populations were continuously shaped by a variety of processes such as disparate selective pressures, range expansion, connectivity among established populations, and the introduction of novel diversity from released wild boar or escaped domestic pigs.

Wild boar have been released on multiple occasions from 1890 onward. Perhaps the most influential introduction of wild boar was associated with the establishment of a game preserve at Hooper Bald, in Graham County, North Carolina, USA in 1912. Following escape from the game preserve, wild boar began interbreeding with feral swine and

Received: 7 September 2016; Accepted: 19 December 2017

[1]*E-mail: blake_mccann@nps.gov*

expanded their range in western North Carolina and eastern Tennessee, USA, including what would later become Great Smoky Mountains National Park. The subsequent collection of individuals with wild boar phenotypes from the Hooper Bald area and release throughout Tennessee and California (1926), Florida (1960), West Virginia (1971), and South Carolina, USA (throughout 1970s) further elevated the importance of this wild boar introduction to the genetic composition of feral swine (Mayer and Brisbin 1991). Numerous other wild boar introductions have been documented that have shaped the phenotypic and genetic characteristics of local populations (Mayer and Brisbin 1991). The genetic characteristics of feral swine populations also have been influenced by gene flow from domestic pigs as a consequence of free-range livestock practices. From the time of European colonization through the enactment of closed-range livestock laws (varied by state but generally 1930s–1960s), pigs were routinely turned loose seasonally into wooded areas to fatten on mast and other naturally available food items (Mayer and Brisbin 1991). Many feral populations were either established or augmented by animals that escaped from these free-ranging herds.

Despite the varied pattern of swine introductions, populations generally remained restricted to areas of Hawaii, California, and the southeastern United States through the early 1980s. However, over the last 35 years, the range of feral swine has increased dramatically with populations now established in >40 states (Mayer and Brisbin 2009, Nolte and Anderson 2015). The expanding distribution and abundance of this invasive species has been accompanied by increasing economic and ecological costs. Economic consequences of feral swine in the United States are estimated broadly at $1.5 billion/year with costs associated with agricultural depredation, property damage, and the transmission of disease to humans, domestic animals, and wildlife (Pimentel 2007, Bevins et al. 2014, Nolte and Anderson 2015). Additionally, feral swine alter ecosystems with wallowing and rooting behaviors that modify and degrade wildlife habitats, denude forest understories, facilitate the spread of invasive flora, and decrease biodiversity (Bevins et al. 2014, Nolte and Anderson 2015). Further, feral swine can have direct negative effects on plant and animal species of conservation concern through foraging, predation, and nest depredation (Bevins et al. 2014, Nolte and Anderson 2015).

The recent and rapid expansion of feral swine within the United States has been attributed to natural dispersal and, perhaps more importantly, the human-mediated establishment of new populations through the deliberate introduction of swine (Mayer and Brisbin 1991, Gipson et al. 1998, Waithman et al. 1999). Despite conventional efforts to track swine introductions, many populations have been either augmented or established through small, undocumented releases associated with the incidental escape of domestic pigs or deliberate introductions conducted by private citizens. Molecular tools can help elucidate the complex patterns of feral swine range expansion to identify mechanisms and sources for expanding populations across the United States. For example, McCann et al. (2014) evaluated patterns of mitochondrial haplotype diversity (mtDNA) throughout the United States. Their results demonstrated multiple genetic origins of feral swine populations and established the importance of human-mediated translocation in recent range expansions. By assessing mtDNA, McCann et al. (2014) successfully described the diversity of sources for swine introductions, yet the resolution of mtDNA for delineating population genetic structure was limited. Microsatellites markers similarly have been used to describe fine-scale patterns of population genetic structure and identify habitat attributes that facilitate or restrict gene flow; however, this work has typically been restricted to small spatial scales with a limited number of focal sampling locations (Delgado-Acevedo 2010, McCann 2012, Lopez et al. 2014, Hernández et al. 2018, Tabak et al. 2017). The ability to meaningfully extrapolate from focal studies to describe processes of biological invasion at a continental scale is limited by the diversity of landscapes occupied by feral swine, which differ with respect to community composition, resource availability, connectivity, and barriers to dispersal.

We sought to build upon previous molecular work by using single nucleotide polymorphism (SNP) nuclear markers to describe patterns of population genetic structure throughout the United States. Our specific objectives were to elucidate regional patterns of population structure and evaluate the association between genetic patterns, known introduction histories, and previously described range expansion. Given the diversity of swine lineages brought to North America and regional variation in animal husbandry practices and treatment of swine as a game species, we hypothesized that genetic structure would reflect historical introduction and range expansion processes of feral swine in western, southcentral, and southeastern regions of the continent and that unique lineages of wild boar could be detected and tracked subsequent to their introduction.

STUDY AREA

We collected feral swine tissue samples through cooperation with government and private organizations that were conducting legally sanctioned research, control, or eradication operations in 35 states throughout the United States (Fig. 1). Additionally, we cooperated with international partners to obtain samples from wild boar in Iran and Spain, which were selected to represent geographically and biologically distinct populations for comparison to feral swine.

METHODS

To minimize familial relatedness of individuals, we asked field personnel to collect samples from swine removed from as many different locations as possible and to avoid sampling individuals captured in the same location on the same day. As described previously (McCann et al. 2014, Sweitzer et al. 2015), blood samples were air dried on FTA cards (Whatman, Florham Park, NJ, USA) and stored at room temperature until processing. Other somatic tissues (skin, muscle, and bone) were stored at −20°C until processing. Hairs (~30 shafts with follicles) were pulled from pig carcasses with pliers, placed on

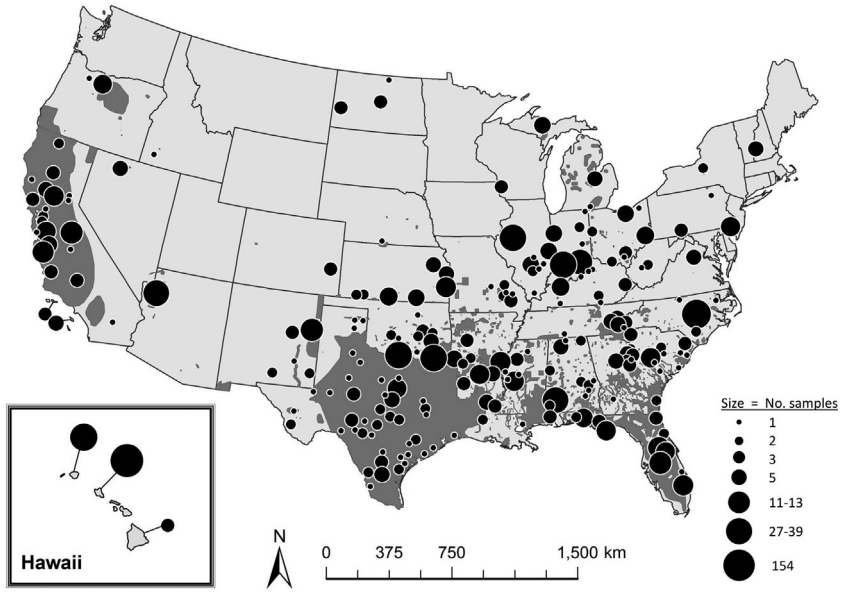

Figure 1. Distribution of 1,032 feral swine samples from 35 states in the United States collected during 1996–2013. Circles represent county centroids, with the exception of island locations, and overlapping locations where position was shifted to accommodate viewing. Circle size corresponds to number of samples from each area normalized with non-linear transformation. Shaded areas on the landscape represent approximated distribution of wild pigs in North America as described by state agencies and United States Department of Agriculture Wildlife Services (Southeastern Cooperative Wildlife Disease Study 2010). Sample locations in Spain ($n = 37$), Iran ($n = 10$), and those without known county of origin ($n = 9$) are not presented.

collection cards (GeneSeek, a Neogen Corporation, NE, USA), and stored at room temperature. This work was exempted from review by the University of North Dakota Institutional Animal Care and Use Committee because samples were collected ancillary to legally authorized management programs (e.g., agency cooperative agreements, management plans, depredation permits, and hunter harvest).

Data Extraction and Quality Control

We submitted all samples to an external laboratory (GeneSeek) for processing between March 2009 and September 2013. Genomic DNA was extracted with proprietary protocols, and specimens were genotyped at 96 SNP loci typically employed for differentiating parentage of swine (Table S1, available online in Supporting Information). Amplification proceeded for 701 samples using a single-base extension polymerase chain reaction (PCR) and nucleotide scoring with the MassARRAY® iPLEX Gold® assay (Sequinone, San Diego, CA, USA). The remaining 387 samples were processed as part of a concurrent genome-wide association study (B. S. Schmit, U.S. Department of Agriculture [USDA]/Animal and Plant Health Inspection Service [APHIS]/Wildlife Services, unpublished data) using the Porcine SNP60 bead chip v2 (Illumina, San Diego, CA; Ramos et al. 2009) containing the same loci of interest.

Although the assumptions of most genetic analyses are robust to missing data, the most reliable results are obtained when missing information is minimized (Wiens 1998, Zhen et al. 2012). Accordingly, we excluded all samples with <75% amplification across loci and then excluded all loci with <75% amplification across remaining individuals (Table S1, available online in Supporting Information). We duplicated population genetic structure analyses described below with threshold values of 90%. Observing similar patterns of genetic structure with both 75% and 90% thresholds, we opted to present the more inclusive dataset in the manuscript. We then tested for linkage disequilibrium across remaining loci using exact tests available in Program GENEPOP (Rousset 2008). We used default Markov chain parameters of 10,000 dememorization steps, 100 batches, and 5,000 iterations/batch, followed by Bonferroni adjustment of significance values to correct for family-wise error (Rice 1989). To differentiate linked loci from the effects of population structure on measures of linkage disequilibrium, we repeated this analysis on 8 subsets of swine with known or suspected biological or geographic differences from other animals in our dataset. Test subsets included wild boar from Spain ($n = 35$), wild boar from Iran ($n = 8$), hybrid (wild boar × feral) animals from Great Smoky Mountains National Park ($n = 14$; McCann et al. 2014), feral swine from 2 Hawaiian Islands (Oahu, $n = 101$; Kauai, $n = 31$), and feral swine from historical or isolated mainland populations in Sutter County, California ($n = 8$; Sweitzer et al. 2015), Fulton County, Illinois ($n = 25$; McCann et al. 2003), and Mohave County, Arizona ($n = 27$; Mayer and Brisbin 1991). For pairs of loci that were identified as linked, we used basic local alignment search tool (BLAST; Altschul et al. 1990) sweeps of flanking sequences on the current pig genome assembly (Sscrofa 10.2; http://www.ncbi.nlm.nih.gov/genome/guide/pig/), identified chromosome number and map locations, and identified the number of kilobases (kb) that separated loci on the same chromosome.

Population Inference

We used 2 independent approaches to estimate population genetic structure from the genetic diversity found among

individuals in the sample. Our first approach used the Bayesian clustering method implemented in STRUCTURE version 2.3.4 to identify the presence of genetic clusters (K; Pritchard et al. 2000). The underlying premise of this approach is that discrete genetic clusters can be identified that conform to Hardy–Weinberg genotype frequencies when genetic structure is evaluated with loci in linkage equilibrium. The second approach, discriminant analysis of principal components (DAPC; Jombart et al. 2010) is a purely statistical clustering method maximizing between group variance in principal components (i.e., a set of orthogonal linear combinations) of individual multilocus genotypes using discriminant analysis, while minimizing within group variance. When genetic groupings are not known *a priori*, clusters are identified by k-means clustering on the principal components using model selection to identify the best supported number of clusters (K) that minimizes within group variance. Both methods depend on probabilistic identification of discrete groups within multivariate datasets, but Bayesian clustering is based on underlying assumptions arising from a fundamental evolutionary model, whereas DAPC is simply based on maximizing among group variance and minimizing within group variance in allelic composition. To further explore the patterns of genetic structure, we also conducted a hierarchical STRUCTURE analysis (in which we initially describe the top level of hierarchical structure and then describe substructure within initial clusters; Evanno et al. 2005, Janes et al. 2017) and a Bayesian analysis of population structure (BAPS; Corander and Marttinen 2006, Corander et al. 2008). These analyses corroborated the patterns described with STRUCTURE and DAPC; therefore, we restricted the presentation of detailed results from additional analyses to the supplemental materials.

Anderson and Dunham (2008) demonstrated that the inclusion of related individuals in a sample can inflate the number of genetic clusters identified by STRUCTURE. This may be of particular concern for our data given that many samples were collected secondarily to ongoing control efforts and pigs often organize as matrilineal social groups (sounders; Gabor et al. 1999) that could be culled simultaneously. Accordingly, we conducted 2 independent but identical STRUCTURE analyses to first, identify plausible genetic delineations within which we would identify closely related individuals (full siblings and parent offspring; Anderson and Dunham 2008) and second, describe regional patterns of genetic structure with a dataset in which closely related individuals had been removed. For both STRUCTURE analyses, we evaluated K over a range from 1 to 65 using the admixture model with correlated allele frequencies, a 50,000 iteration burn-in period, and 100,000 Markov chain Monte Carlo repetitions with 10 independent iterations for each value of K (Pritchard et al. 2000, Falush et al. 2003). Initial exploratory analysis demonstrated an asymptotic pattern in the mean natural log of the probability of the data given the number of clusters (ln $Pr[X|K]$; Pritchard et al. 2000) with little increase in value for $K > 38$. Therefore, we selected a maximum *ad hoc* K value of 65 for the subsequent iterations to ensure that we considered all plausible partitions of genetic clusters.

From the preliminary STRUCTURE analysis for the identification of closely related individuals, we selected $K = 49$ because this value represented the highest ln $Pr(X|K)$ and represented a localized minima in standard deviation among iterations. Given that the objective of this analysis was to eliminate the effect of sampling family groups on the estimated number of genetic clusters, we were not concerned with overestimating K because closely related individuals should be assigned to the same genetic cluster. We then used the software package RELATED (Pew et al. 2015) implemented in R (R Development Core Team 2017) to calculate Queller and Goodnight's (1989) estimate of the coefficient of relatedness (r) for all pairs of individuals within clusters. Next, we used RELATED to conduct simulation models across 11 of the 49 clusters (selected to represent the range in number of individuals assigned to clusters; range = 5–72) to identify a threshold in relatedness estimates that would allow removal of closely related individuals (expected $r \geq 0.5$; consistent with full siblings, parent offspring, or inbred individuals) while retaining dyads with relatedness values consistent with half siblings (expected $r = 0.25$), as recommended by Anderson and Dunham (2008). Using allele frequencies calculated for each of the 11 genetic clusters, we simulated 3×100 dyads that represented half siblings, full siblings, and parent-offspring relationships and calculated respective Queller and Goodnight (1989) relatedness values. We then compared the distribution of r values for half siblings versus full siblings and parent-offspring dyads to identify a mean value across the 11 simulated populations that would minimize bias associated with false exclusion of half siblings or false inclusion of full siblings or parent-offspring dyads. For dyads within the observed data with r estimates that exceeded the specified threshold, we summed each individual's pair-wise relatedness values across all other individuals within the genetic cluster and pruned the individual with the higher relatedness sum. In this manner, we strategically reduced relatedness within genetic clusters while minimizing the number of individuals removed from the sample. Following the removal of all closely related individuals, we repeated the STRUCTURE analysis to identify the most informative population genetic structure based on the mean natural log of the probability of the data (ln $Pr[X|K]$; Pritchard et al. 2000, Evanno et al. 2005), the strength of individual assignments to genetic clusters (Q_{max}; Pritchard et al. 2000, Puechmaille 2016), and the expectation that true genetic clusters should demonstrate geographic cohesion (Puechmaille 2016).

We analyzed the same reduced data set with the alternative statistical clustering method, DAPC. We used the k-means clustering algorithm in the adegenet package (version 2.1.0; Jombart 2008) for Program R (R Development Core Team 2017) to describe genetic structure among unrelated individuals (as described above). Following the analysis recommendations of Jombart and Collins (2017), we first conducted a principal components analysis to distill the variation in the genotypic data into orthogonal principal components. Carrying forward all of the described principal components, we then used Bayesian Information Criterion

(BIC) to evaluate among competing values of K across the range of 1–65. We then conducted a discriminant analysis with a reduced number of principal components on the clusters that had been identified in the principal components analysis. Following Jombart and Collins (2017), we determined the optimal number of principal component axes to retain based on the a-score, which measures the balance between discriminatory power and over-fitting. We retained 14 principal components and used all discriminant functions for the assignment of individuals into clusters. For the purposes of illustrating the alignment of clusters between analysis approaches, and to visualize the patterns of similarities and differences in genetic composition of groups, we also conducted a discriminant analysis on the clusters defined by STRUCTURE and plotted clusters on primary discriminant function axes.

Isolation by Distance

We conducted isolation by distance analyses to identify spatial patterns of connectivity throughout the contiguous United States and independently within regions representing unique historical invasions. We limited the global analysis to those samples from contiguous states for which we had precise geographic coordinates for sampling locations ($n = 570$). For regional analyses, we divided samples into 3 regions: western (CA and contiguous states west of NM; $n = 87$), central (TX and contiguous states west of the Mississippi River and east of AZ; $n = 180$), and eastern (FL and contiguous states east of the Mississippi River; $n = 303$). We used GENEPOP to estimate pairwise genetic distances among individuals using Rousset's a (Rousset 2000). We used Mantel tests (Mantel 1967) with 999 permutations available in PASSAGE (Rosenberg and Anderson 2011) to test for correlation between Rousset's a and geographic distance both across the entire sample and within region described above.

To obtain a more detailed assessment of spatial effects on genetic identity, we also tested for spatial autocorrelation in PASSAGE using defined distance class maxima of 10, 25, 50, 75, 100, 150, 200, 250, 300, 400, and 500 km, representing both conventional and interpretable thresholds for managers. To facilitate comparisons of regional differences in geographic structure, we also fit linear regressions of pairwise Rousset's a to $\log_{10}$(geographic distance), the theoretical relationship between pairwise Rousset's a and geographic distance under a drift-gene flow model (Rousset 2000). For this analysis, we excluded pairwise values of individuals sampled from the same locations (geographic distance = 0) to avoid bias in parameter estimation (Rousset 2000), but we evaluated regional differences in Rousset's a for co-located individuals (geographic distance = 0) in a separate analysis of variance. To avoid confounding the effects of variation in the spatial extent of regions with differences among regions in patterns of isolation by distance, we also limited the analysis to pairwise geographic distances not exceeding the maximum separation distance in the smallest region (greatest pairwise geographic distance in Western region = 1,260 km).

RESULTS

We sampled 1,041 feral swine from 35 states and 47 Eurasian wild boar from Spain and Iran. After excluding individuals and loci with amplification rates <75%, our dataset consisted of 959 individuals genotyped at 88 loci with an average genotyping success rate of 96%. Retained samples represented 205 counties from 34 states ($n = 916$), 3 provinces in Spain ($n = 35$), and southwestern Iran ($n = 8$; Fig. 1).

Tests of linkage disequilibrium were statistically significant for 267 pairwise relationships between loci for the full dataset. Conversely, no linkage disequilibrium was detected within subsets of individuals sampled in Great Smoky Mountains National Park, Illinois, Iran, Spain, or California. Seven pairwise relationships were significant within subsets representing Arizona ($n = 1$), Kauai ($n = 2$), and Oahu ($n = 4$), with 1 set of loci demonstrating significant linkage on both sampled Hawaiian Islands. Among these 7 pairs of loci, 5 were on the same chromosome, whereas 2 were on different chromosomes; linkage of the 2 loci on different chromosomes is most likely attributable to population structure (Table S2, available online in Supporting Information). Among the 5 pairs of loci on the same chromosome, the minimum and mean distances between loci was 1,846 kb and 28,653 kb, respectively, and correlation between adjacent loci would be expected to be minimal ($r^2 < 0.20$; Badke et al. 2012). Collectively, the lack of linkage relationships among 5 population subsets, map distances >1,000 kb separating loci, and inconsistencies in linkage estimates among population subsets suggest that population structure rather than physical linkage was the cause of the linkage disequilibrium detected in both the full dataset and among the subset of sampling locations. Accordingly, we retained the complete complement of 88 SNPs for analysis of population structure.

Based on simulations conducted in RELATED, we excluded an individual from each dyad with $r > 0.39$ (range of threshold values among 11 simulated populations = 0.38–0.40), providing an expected false exclusion rate for half siblings and false inclusion rate for full siblings or parent-offspring dyads of 14% (range 11–16%). Applying this threshold, we pruned 178 individuals from the dataset while retaining 781 individuals for population genetic analysis.

Bayesian Genetic Clustering Analysis

Evaluating population genetic structure among 781 unrelated individuals, we deemed the genetic clusters produced with K ranging from 4 to 10 to be the most biologically informative based on the criteria of mean ln $Pr[X|K]$, the distribution of Q_{max} values, and geographic cohesion of the delineated clusters (Puechmaille 2016; Fig. 2). Specifically, K values from 4 to 10 produced geographically cohesive clusters with smaller, geographically cohesive clusters emerging from more inclusive clusters with incremental increases in K. The model with $K = 34$ produced the highest mean ln $Pr[X|K]$ with comparatively low variance among independent runs (Fig. S1, available online in Supporting Information). For values of K between 10 and 34, we observed a pattern in which small geographically cohesive clusters emerged with increasing values of K, yet these small cohesive clusters were

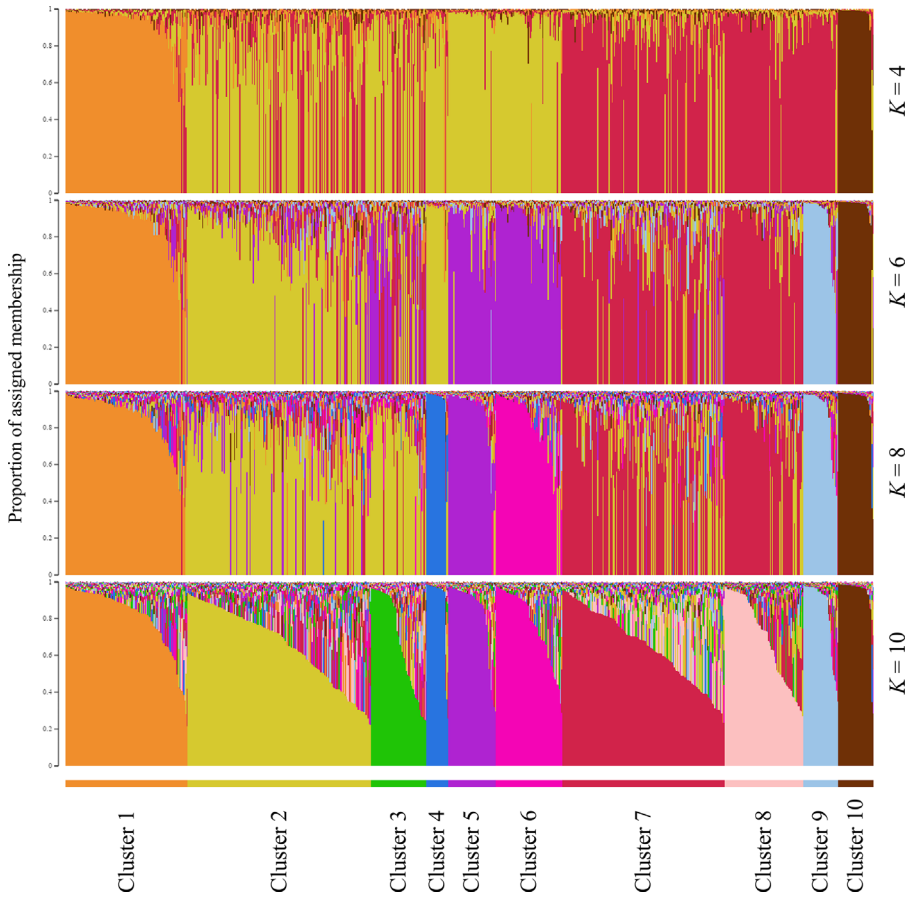

Figure 2. Bar plot generated with STRUCTURE PLOT (Ramasamy et al. 2014) representing proportional assignment of 781 feral swine (sampled across the United States from 1996 to 2013) and wild boar (sampled in Spain and Iran from 2010 to 2013) for $K = 4$, 6, 8, and 10 genetic clusters delineated with STRUCTURE, with numeric cluster identifiers for $K = 10$ presented below the plot.

accompanied by widespread clusters that extensively overlapped geographically with similarly widespread clusters. Accordingly, we deemed 10 genetic clusters to offer the finest level of genetic partitioning that could be resolved with the discriminatory power of the marker set at this spatial scale.

With $K = 10$, 49% (386/781) individuals assigned strongly ($Q_{max} > 0.8$) to a genetic cluster with an average Q_{max} of 73% (Fig. 2). The 10 clusters were associated primarily (i.e., individuals assigning at $Q_{max} \geq 0.8$) with the general regions of 1) Hawaii and 3 continental states; 2) southeastern states and portions of California; 3) Illinois, Texas, Florida, New Hampshire, and Iranian wild boar; 4) Mohave County, Arizona; 5) Johnston County, North Carolina; 6) Great Smoky Mountains National Park, Northern California, and Nevada; 7) southcentral states (primarily west of the Mississippi River); 8) Oklahoma, Arkansas, and Kansas; 9) Spanish wild boar; and 10) Jackson and Lawrence Counties Indiana (Fig. 3; Table S3, available online in Supporting Information).

Discriminant Analysis of Principal Component Genetic Clustering

Evaluation of k-means clustering with BIC demonstrated $K = 12$ was the best model; however, models with K values from 11 to 16 were competitive ($\Delta BIC < 2$; Fig. S2, available in online Supporting Information). For the purposes of interpreting genetic clusters defined with DAPC, we selected a value of $K = 12$ in accord with the suggestions of Jombart and Collins (2017) because it explained the structural patterns within the genetic data with the fewest number of parameters. Geographic patterns between the 2 clustering approaches were very similar, with DAPC splitting swine sampled in the Hawaiian archipelago (cluster 1 from STRUCTURE analysis) into 2 clusters corresponding with Oahu and Kauai. Similarly, DAPC split STRUCTURE cluster 2 (southeastern states and portions of CA) into 2 clusters (Fig. 4; Figs. S3 and S4, available in online Supporting Information). Remaining clusters showed strong patterns of geographic congruence between the clustering approaches (Table S4, available online in Supporting Information). The genetic relationships among clusters revealed by discriminant analyses are also very similar for the 2 methods of analysis (Fig. 4; Figs. S3 and S4, available in online Supporting Information).

Spatial Effects on Genetic Structure

Individual pairwise genetic and geographic distances were significantly correlated throughout the contiguous United States and within regional subsets ($P < 0.001$ for all tests), with Mantel correlations of 0.159 overall and 0.628, 0.375,

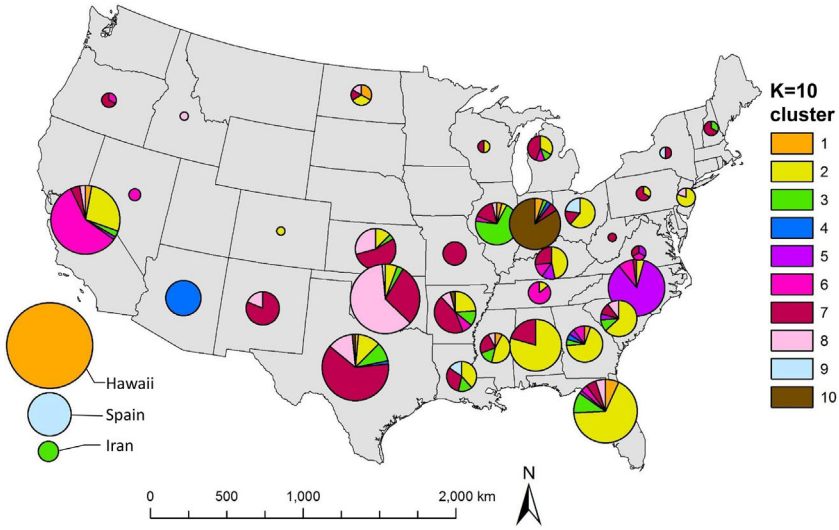

Figure 3. Geographic distribution of individuals across 10 genetic clusters representing feral swine sampled across the United States from 1996 to 2013. Chart size corresponds to number of samples from each state, placed at centroids of state polygons, with colors corresponding to assigned clusters.

and 0.137 for western, central, and eastern subsets, respectively. Mantel correlations were significant ($P < 0.001$) across all distance classes for western and central subsets, and all but the 400–500-km distance class for the eastern subset (Fig. 5); however, the magnitude of the Mantel correlations exhibited a highly nonlinear relationship with geographic distance. The strength of correlation was greatest among local (i.e., 10 and 25 km maxima) distance classes and then diminished abruptly, followed by a more gradual decline with farther geographic separation.

Regional differences in the spatial pattern of pairwise individual genetic distances were also evident. Among co-located individuals (those sampled at the same location and not included in subsequent regression analyses), the greatest Rousset's *a* was observed in the eastern region (0.502), followed by the western (0.490), and the central (0.457) regions. Similarly, at the shortest geographic distance included in the regression analyses, the estimated mean Rousset's *a* was greatest in the east (*y*-intercept = 0.411, SE = 0.005) with lower genetic distances observed among proximate individuals for the western and central regions (western *y*-intercept = 0.238 SE = 0.012; central *y*-intercept = 0.220, SE = 0.008). The influence of increasing geographic separation on genetic distance was greatest in the western region (slope = 0.074, SE = 0.002; $R^2 = 0.346$), followed by the central region (slope = 0.063, SE = 0.002; $R^2 = 0.103$), with a relatively flat relationship between genetic and geographic distance in the eastern region (slope = 0.039, SE = 0.001, $R^2 = 0.05$). Within all 3 regions, we identified individuals that were genetically similar (low Rousset's *a*) but geographically separated by hundreds of kilometers (Fig. 6), indicative of the effect of anthropogenic translocation.

DISCUSSION

The observed patterns of genetic structure illustrate the complexity of processes influencing feral swine populations throughout the United States. Evaluations of population genetic structure and isolation by distance demonstrate that the genetic attributes of feral swine have been shaped by both natural range expansion from long-established feral populations (e.g., southcentral and southeastern genetic clusters) and the human facilitated establishment of new populations. With initial introductions to southcentral and southeastern states beginning nearly 500 years ago, these regions are each defined by clusters that span broad geographic areas with a putative contact zone near the Mississippi River (Fig. 3; clusters 2 and 7). The genetic association between swine in the southeast and portions of California (Fig. 3) is consistent with historical accounts of free-ranging pig populations being seeded by Spanish exploration parties and also accompanying Spanish settlements. The association of contemporary genetic patterns with the historical record supports a hypothesis of a shared Spanish ancestry; however, additional genetic analyses are needed to test this hypothesis by comparing the genetic attributes of these feral populations to historical Spanish breeds.

Similar to McCann et al. (2014), we observed a genetic association between disjunct populations within Great Smoky Mountains National Park and northern California (cluster 6 from STRUCTURE analysis). The introduction of wild boar into Hooper Bald, North Carolina, subsequent geographic expansion, collection and introduction of descendants to Monterey County, California, and ensuing serial introductions throughout northern California is documented (Mayer and Brisbin 1991, Waithman et al. 1999). At present, this lineage is dominant across northern California with genetically similar individuals sampled in Nevada. The persistence of this Hooper Bald genetic signal, nearly 100 years post-introduction, demonstrates the unique influence of wild boar introgression on the molecular profile of feral swine (Fig 3; Mayer and Brisbin 1991). The geographic expansion of pigs from this unique lineage may be due to the preferential translocation of animals with wild boar phenotypes given their desirability to hunters (Mayer

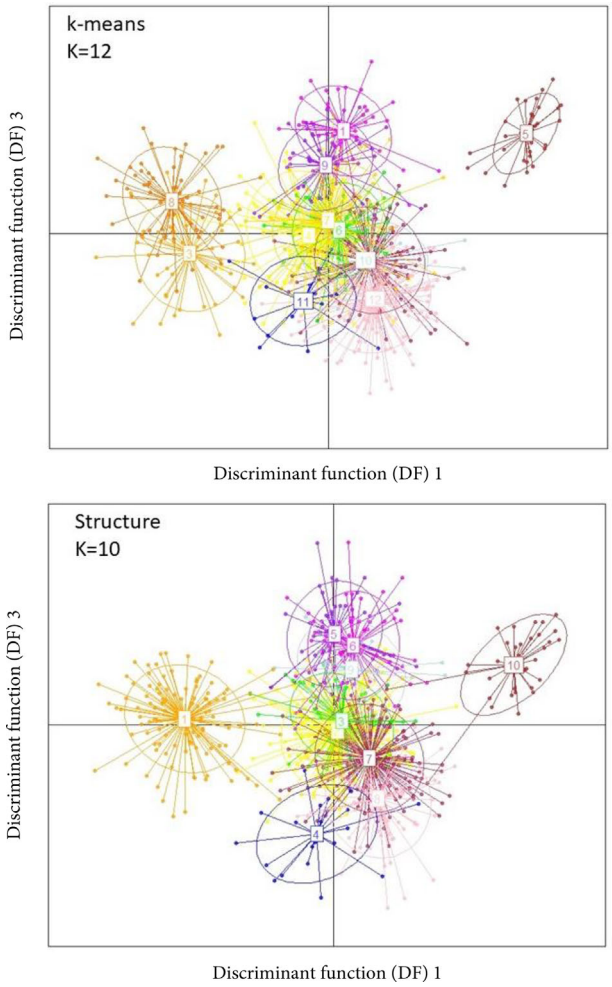

Figure 4. Comparison of discriminant analysis of best supported clusters of pig single nucleotide polymorphism genotypes based on k-means clustering using discriminant analysis of principal components (top) and STRUCTURE clusters (bottom) inferred from 88 single nucleotide polymorphism genotypes for 781 feral swine sampled throughout the United States and wild boar from Spain and Iran, 1996–2013. Plots of discriminant functions (DF) 3 versus 1 best spread clusters for visual presentation. Dots are individual pigs, ellipses are 95% of cluster memberships. Colors and cluster numbers match those used in the STRUCTURE bar plot.

and Brisbin 1991, Waithman et al. 1999, Caudell et al. 2013). Alternately, the prevalence of this group within northern California may lend credence to the hypothesis by Waithman et al. (1999) that individuals with wild boar genetic ancestry are more ecologically adaptable, making them more efficient natural dispersers in California than other conspecifics that have descended directly from domestic pigs.

Wild boar have similarly influenced the genetic profiles of feral swine populations beyond North Carolina, Tennessee, California, and Nevada. Though not strongly assigned, our cluster 6 included animals sampled in 7 other states, of which Georgia and Florida are known recipients of Hooper Bald descendants (Mayer and Brisbin 2009). One animal from this group was sampled from Southern Illinois, a region where wild boar phenotypes have been previously reported (McCann et al. 2003). Cluster 3 included our outgroup of

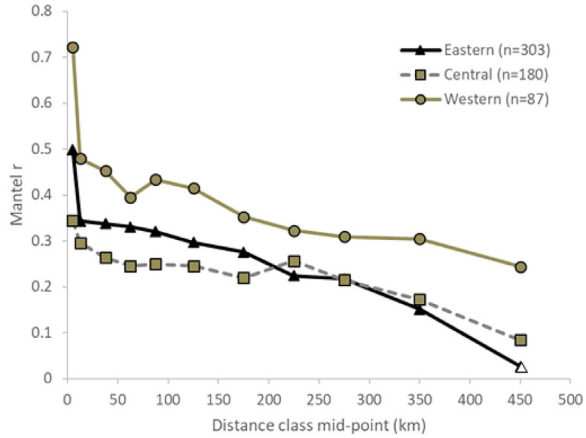

Figure 5. Mantel correlograms for 3 genotypic subsets of feral swine across distance class maxima of 10, 25, 50, 75, 100, 150, 200, 250, 300, 400, and 500 km. Regional subsets evaluated were western (CA and contiguous states west of NM), central (TX and contiguous states west of the Mississippi River and east of AZ), and eastern (FL and contiguous states east of the Mississippi River) from samples collected 1996–2013. Correlations not significant at $P \leq 0.01$ are denoted by hollow markers. Distance classes are presented at midpoint of estimate for each range on the *x*-axis.

Iranian wild boar and 1 animal from Sullivan County, New Hampshire, USA near Corbin's Park, a known Eurasian wild boar introduction site (Mayer and Brisbin 1991). These associations could indicate the possibility of domestic pig-wild boar hybrid ancestry for animals sampled in Illinois, Texas, Florida, and 9 other states with wild boar lineages separate from that of Hooper Bald (Table S3, available online in Supporting Information). Further, pelage characteristics indicative of hybrid wild boar (i.e., striped young, wild-grizzled pelage in adults) have been observed among swine collected in 12 states (B. E. McCann, Theodore Roosevelt National Park, unpublished data). Therefore, introduction histories, coupled with field observations and genetic relationships, suggest that wild boar contributions to

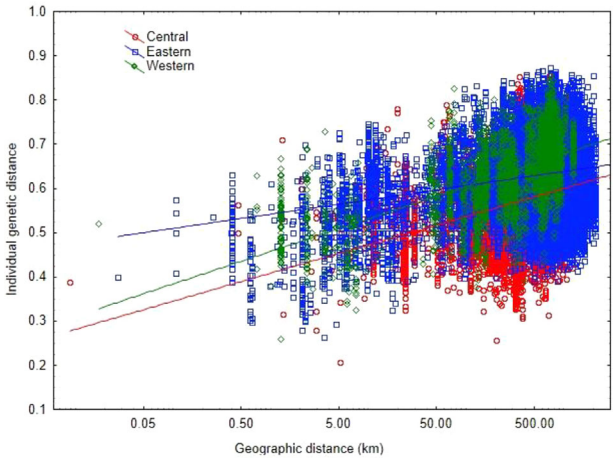

Figure 6. Scatterplot of pairwise Rousset's *a* versus pairwise Euclidean distance for feral swine in the United States, 1996–2013, with known geographic coordinates for collection sites ranging to 500 km. Lines are region-specific regressions of Rousset's *a* on log10[geographic distance].

feral swine populations are pervasive in North America (Caudell et al. 2013).

The genetic uniqueness of pigs on the Hawaiian Islands represents another distinct historical influence. Linderholm et al. (2016) characterized Hawaiian feral swine as a genetic mix of Pacific Clade swine, introduced with Polynesian colonization, and European domestic pigs, which have interbred with established Polynesian populations from the time of first European contact in 1778 through the present. The Pacific Clade of pigs is a derived state from the East Asian Clade. There is a long history of genetic isolation between Asian and European Clades with the European Clade diverging from the Asian Clade an estimated 0.8–1.6 million years ago (Groenen et al. 2012). Thus, we would expect Hawaiian feral swine, with admixed Pacific ancestry, to be genetically distinct from mainland feral swine populations, which are believed to descend primarily from domestic pigs and wild boar of European origin (Mayer and Brisbin 1991). However, a variety of European domestic pig breeds were crossed with Asian breeds starting in the 1700s. Many of the improved or commercial breeds propagated in the United States are of mixed lineages, which could have served as sources for Asian genetics in feral populations (Jones 1998, McCann et al. 2014). Further, a limited number of feral swine have been culled with phenotypic characteristics similar to that of pot-bellied pigs (Caudell et al. 2013), a breed that originated in Vietnam and was first imported to North America in 1985 (Tynes 1999). Thus, a plausible hypothesis for the genetic assignment of pigs sampled in the contiguous United States to the Hawaiian population is that these individuals descended from Asian breeds, which, in our sample, would be most closely related to Hawaiian feral swine.

Our $K = 10$ clustering result described with STRUCTURE revealed distinct genetic groupings in Arizona, Indiana, and North Carolina that were associated with limited geographic distributions and in some cases more recent invasions. The strong genetic differentiation of feral pigs in Indiana and Arizona could be explained by geographic isolation or introductions from novel genetic sources (Fig. 3). Indeed, these 2 groups were more distantly related to most feral swine groupings in the United States than were Spanish wild boar (Figs. 3 and 4). The geographically isolated Arizona population was established before 1900 with the escape of domestic pigs from a nearby ranch in Needles, California (Mayer and Brisbin 1991). Despite their duration as a free-living population, the pigs in Mohave County, Arizona have retained flopped ears and a prominent forehead (M. W. Lutman, USDA/APHIS/Wildlife Services, personal communication), phenotypic traits common in many domestic breeds. Combining oral history with microsatellite data, Caudell et al. (2013) reported that the feral swine population in southern Indiana was most likely founded by individuals collected in Louisiana that possessed high wild boar ancestry. The genetic association between the Indiana cluster and the southcentral cluster (which includes LA) provides some support for the hypothesis presented by Caudell et al. (2013). However, given the genetic uniqueness of the Indiana population relative to others sampled, it is possible that the true source was not included in this analysis or that extreme founder effect, as evidenced by extremely low levels of observed heterozygosity, contributed to the uniqueness of this group. We observed similar yet less extreme patterns for populations in Johnston County, North Carolina. In all 3 instances, many sampled individuals had to be excluded from analysis because of close relatedness, possibly indicating founder effect and inbreeding associated with isolation (Table S3, available online in Supporting Information).

Geographically cohesive and biologically plausible populations were apparent in evaluations of STRUCTURE results for values of $K > 10$. For example, wild boar sampled in Iran first emerged as a distinct genetic cluster at $K = 11$. At $K = 34$, the value of K that produced the maximum ln $Pr[X|K]$, 19 groups corresponded primarily with county- or state-level geographic boundaries, including reference populations of Great Smoky Mountains National Park; Fulton County, Illinois; Kauai and Honolulu counties, Hawaii; Sutter County, California; and Mojave County, Arizona, along with novel groupings (Table S3, available in online Supplemental Information). A parallel exploratory analysis of the 781 genotypes in Program BAPS 6.0 (Corander and Marttinen 2006, Corander et al. 2008) under default admixture clustering settings returned $K = 44$ as the optimal partition, with many similar fine-scale genetic associations identified (Table S3, available in online Supplemental Information). The suggestion of fine-scale genetic partitioning present within the described clusters is consistent with recent findings by Tabak et al. (2017) who identified 21 genetic clusters for feral swine in California alone using a panel of 43 microsatellite loci. Indeed, given the convoluted history of feral pig invasion in North America, broad-scale structure described at $K = 10$ almost certainly exists in parallel with finer-scale structure, but the latter is complex and variable in geographic cohesiveness and consequently more difficult to discern with confidence.

Comparison of Q_{max} values, bar plots, and cluster assignments across values of K reveals a general hierarchy of nested relationships (Fig. 2), and further corroborates the assertion that substructure is likely present within groupings inferred at $K = 10$. We began to elucidate the hierarchical organization of genetic structure with additional analyses presented in the supplemental materials. However, a conservative assessment of genetic structure was better supported by our data given the realized sample sizes and the genetic resolution provided by the marker set, as some clusters at high values of K had extensive geographic overlap and comparatively low Q_{max} values that were difficult to interpret biologically. Sympatric distributions for genetically differentiated clusters could be explained in part by rampant translocation or multiple introductions from the same domestic lineages, which are both known pathways of invasion (Mayer and Brisbin 1991, McCann et al. 2014, Tabak et al. 2017). However, pilot data collected on a sampling of United States feral swine populations with high density SNP arrays (Ramos et al. 2009) indicate that greater geographic cohesion, consistent with our interpretation of $K = 10$, should be

expected for this invasive species (B. S. Schmit, unpublished data).

Other authors have demonstrated that, with intensive genetic sampling, fine-scale genetic structure can be resolved for feral swine (Delgado-Acevedo 2010, Lopez et al. 2014, Tabak et al. 2017, Hernández et al. 2018). With this analysis, we expressly sought to distribute our sampling effort across the entirety of the invaded range in the United States, a design not best suited to resolve fine-scale genetic difference. Rather, our analysis demonstrates that feral swine populations are hierarchically organized and characterized by regional patterns of genetic structure. Though genetic units are not spatially discrete, distinct genetic clusters are apparent. For example, our group associated with Oklahoma, Kansas, and Arkansas (cluster 8), is closely related to but genetically distinct from the southcentral group (cluster 7), first emerging at $K = 10$ (Figs. 2 and 3). In addition to the STRUCTURE and DAPC results detailed above, patterns of genetic organization were largely corroborated by additional hierarchical STRUCTURE and Program BAPS analyses (including those employing spatial or location priors). In all cases, similar patterns of population structure and geographic distributions were apparent, indicating a consistent molecular signal (results detailed in supplemental materials). However, regardless of analysis technique, some individuals did not fit well statistically, biologically, or geographically with their group of assignment. This may be due to a lack of reference individuals locally, the influx of novel genetics, or panmixis within some invaded areas. Far greater sample sizes and genetic resolution will be needed to fully elucidate fine-scale genetic partitioning across the extent of the invaded range within the United States.

Isolation by distance detected within regions and throughout North America supports findings of molecular population structure, as genetic relationships are expected to scale with geography. Differences in slope and strength of relationships across regions indicate a landscape likely experiencing multiple stages of invasion, differing spatial use responses of pigs to local habitat and land use regimes, and varied levels of human-mediated gene flow. Local genetic structure may also be attributable to female philopatry and the long-term stability of matriarchal social groups (Gabor et al. 1999, Podgórski et al. 2014).

Regardless of region, isolation by distance patterns appears to reach an asymptote at 50–100 km, after which genetic diversity is no longer strongly structured by geographic connectivity (Fig. 6). Mantel distance classifications show stronger genetic correlations at fine spatial scales and a decay in genetic similarity with increasing spatial separation. The sharp decrease in genetic correlation with geographic distance from 0–25 km, particularly in the western and eastern regions, is consistent with influences of limited dispersal and matriarchal lineages in structuring swine populations within a historically invaded range and a topographically complex landscape (Gabor et al. 1999; Fig. 5). The weaker Mantel correlations of the central region, coupled with high genetic similarity at short and long geographic distances, indicate that frequent translocations have diluted isolation by distance patterns here more than elsewhere, which generally fits recent range expansion patterns in this region (Gipson et al. 1998; Figs. 5 and 6). Finally, the non-significant correlations (Fig. 5) and highly variable genetic distances (Fig. 6) for the eastern region at the 401–500 km distance class likely reflect the higher diversity of origins for feral pigs previously reported for this part of North America (McCann et al. 2014).

Overall, isolation by distance patterns indicate a very heterogeneous genetic landscape for feral swine in the United States, influenced by multiple natural and anthropogenic factors operating across spatial scales, with translocation by humans being a key factor in long-range dispersal. Moreover, Rousset's *a* is a complementary approach to discerning genetic structure that does not depend on an *a priori*, and possibly erroneous, delineation of populations or inferred clusters. Rather, it is a lens through which to observe patterns with respect to associations such as geographic separation or any other presumptive structuring factor (e.g., landscape features). Detection of genetic structure, starting with Rousset's *a*, has thus provided an independent perspective to understand scaling of effects of geographic separation regardless of *K* populations inferred.

Our analysis has successfully provided a regional perspective on the genetic structure of feral swine in the United States that complements prior work describing the diversity of introduction sources with mtDNA and fine-scale genetic processes with nuclear microsatellites. However, additional genetics research is needed to fully describe the diversity of population process demonstrated by feral swine, given the great breadth of land cover types invaded, how these processes might deviate between saturated and vacant landscapes, and the phenotypic (and presumed ecological) variation observed among feral swine with morphotypes that range from that of wild boar to domestic pig. Further, our results have demonstrated that human-mediated gene flow and introduction from novel genetic sources has continued to shape feral swine distribution and genetic patterns over recent years. Additional fine-scale analyses will allow for more precise estimates of the frequency of such translocations and their importance for range expansion versus natural dispersal.

MANAGEMENT IMPLICATIONS

Molecular techniques will allow managers to target discrete genetic units, evaluate removal efficacy, and track gene flow to elucidate patterns of natural and anthropogenic dispersal of feral pigs. Our identification of molecular population structure linked with geography confirms that discrete breeding units can be identified and targeted for removal, as isolation of populations is key to successful eradication. Where animals are detected in areas thought to be recently eradicated of swine, genotypes may be evaluated to determine whether they represent previously undetected individuals or newly translocated animals. Hierarchical genetic structure described here provides a first glimpse at the utility of nuclear genetic markers for describing patterns of gene flow and identifying sources of new populations. For

instance, we may infer that California was the source for animals collected in Humboldt County, Nevada (Fig 3; cluster 6). Therefore, managers should be encouraged to continue using genetic tools to inform feral swine control and elimination strategies.

ACKNOWLEDGMENTS

Thanks to the USDA Wildlife Disease Program and field agents for collection of pig tissue samples from throughout the United States. Thanks to the National Park Service, Institute for Wildlife Studies, and many other state and private organizations for collection of samples that made this work possible. Thanks to M. Karimi and C. Gortázar for providing Eurasian wild boar samples. We thank North Dakota Experimental Program to Stimulate Competitive Research, University of North Dakota Department of Biology, University of North Dakota Graduate School, and USDA for funding of this research.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Anderson, E. C., and K. K. Dunham. 2008. The influence of family groups on inferences made with the program Structure. Molecular Ecology Resources 8:1219–1229.

Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, and J. P. Steibel. 2012. Estimation of linkage disequilibrium in four US pig breeds. BMC Genomics 13:24.

Bevins, S. N., K. Pedersen, M. W. Lutman, T. Gidlewski, and T. J. DeLiberto. 2014. Consequences associated with the recent range expansion of nonnative feral swine. BioScience 64:291–299.

Caudell, J. N., B. E. McCann, S. E. Backs, B. S. Schmit, R. A. Newman, R. A. Sweitzer, and R. B. Simmons. 2013. Identification of putative origins of introduced pigs in Indiana using nuclear microsatellite markers and oral history. Proceedings of the 15th Wildlife Damage Management Conference, Clemson, South Carolina, USA.

Corander, J., and P Marttinen. 2006. Bayesian identification of admixture events using multilocus molecular markers. Molecular Ecology 15:2833–2843.

Corander, J., P. Marttinen, J. Siren, and J. Tang. 2008. Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. BMC Bioinformatics 9:539.

Delgado-Acevedo, J. H. 2010. Feral pig management in southern Texas: a landscape genetics approach. Dissertation, Texas A&M University, Kingsville, USA.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611–2620.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Gabor, T. M., E. C. Hellgren, R. A. Van Den Bussche, and N. J. Silvy. 1999. Demography, sociospatial behaviour and genetics of feral pigs (*Sus scrofa*) in a semi-arid environment. Journal of the Zoological Society of London 247:311–322.

Gipson, P. S., B. Hlavachick, and T. Berger. 1998. Range expansion by wild hogs across the central United States. Wildlife Society Bulletin 26:274–278.

Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H.-J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. F. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S. Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S.-H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. M. A. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapio, B. Dibbits, N. Drou, Z.-Q. Du, K. Eversole, J. Fadista, S. Fairley, T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. R. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z.-L. Hu, S. J. Humphray, T. Hunt, H. Hornshøj, J.-T. Jeon, P. Jern, M. Jones, J. Jurka, H. Kanamori, R. Kapetanovic, J. Kim, J.-H. Kim, K.-W. Kim, T.-H. Kim, G. Larson, K. Lee, K.-T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. Truong Nguyen, P. Ni, S.-J. Oh, S. Onteru, F. Panitz, E.-W. Park, H.-S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle, B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher, and L. B. Schook. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491:393–398.

Hanson, R. P., and L. Karstad. 1959. Feral swine in the southeastern United States. Journal of Wildlife Management 23:63–74.

Hernández, F. A., B. M. Parker, C. L. Pylant, T. J. Smyser, A. J. Piaggio, S. L. Lance, M. P. Milleson, J. D. Austin, and S. M. Wisely. 2018. Invasion ecology of wild pigs (*Sus scrofa*) in Florida, USA: the role of humans in the expansion and colonization of an invasive wild ungulate. Biological Invasions. In press. https://doi.org/10.1007/s10530-018-1667-6.

Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, and R. L. Andrew. 2017. The K=2 conundrum. Molecular Ecology 26:3594–3602.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94.

Jombart, T., and C. Collins. 2017. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.1.0. Imperial College, London, United Kingdom.

Jones, G. F. 1998. Genetic aspects of domestication, common breeds and their origin. Pages 17–50 *in* M. F. Rothschild and A. Ruvinsky, editors. The genetics of the pig. CAB International, Wallingford, Oxdordshire, United Kingdom.

Kirch, P. V. 1982. The impact of the prehistoric Polynesians on the Hawaiian ecosystem. Pacific Science 36:1–14.

Linderholm, A., D. Spencer, V. Battista, L. Frantz, R. Barnett, R. C. Fleischer, H. F. James, D. Duffy, J. P. Sparks, D. R. Clements, L. Andersson, K. Dobney, J. A. Leonard, and G. Larson. 2016. A novel *MC1R* allele for black coat colour reveals the Polynesian ancestry and hybridization patterns of Hawaiian feral pigs. Royal Society Open Science 3:160304.

Lopez, J., D. Hurwood, B. Dryden, and S. Fuller. 2014. Feral pig populations are structured at fine spatial scales in tropical Queensland Australia. PLoS One 9(3):e91657.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–220.

Mayer, J. J., and I. L. Brisbin. 1991. Wild pigs in the United States: their history, comparative morphology, and current status. University of Georgia Press, Athens, USA.

Mayer, J. J., and I. L. Brisbin Jr. 2009. Wild pigs: biology, damage, control techniques and management. Savannah River National Laboratory, Aiken, South Carolina, USA.

McCann, B. E. 2012. Genetic relationships of wild pigs (*Sus scrofa*) in the United States: geographic origins and genotypic distribution of the species with implications for management. Dissertation, University of North Dakota, Grand Forks, USA.

McCann, B. E., D. K. Davie, and G. A. Feldhamer. 2003. Distribution, habitat use, and morphotypes of feral hogs (*Sus scrofa*) in Illinois. Transactions of the Illinois State Academy of Science 96:301–311.

McCann, B. E., M. J. Malek, R. A. Newman, B. S. Schmit, S. R. Swafford, R. A. Sweitzer, and R. B. Simmons. 2014. Mitochondrial diversity supports multiple origins for invasive pigs. Journal of Wildlife Management. 78:202–213.

Nolte, D. L., and W. S. Anderson. 2015. Collaborating to halt feral swine damage. Wildlife Professional 9:56–59.

Pew, J., P. H. Muir, J. Wang, and T. R. Frasier. 2015. related: an R package for analyzing pairwise relatedness from codominant molecular markers. Molecular Ecology Resources 15:557–561.
Pimentel, D. 2007. Environmental and economic costs of vertebrate species invasions into the United States. G. W. Witmer, W. C. Pitt, and K. A. Fagerstone, editors. Managing vertebrate invasive species: proceedings of an international symposium. USDA/APHIS/WS, National Wildlife Research Center, Fort Collins, Colorado, USA.
Podgórski, T., D. Lusseau, M. Scandura, L., Sönnichsen, and B. Jędrzejewska. 2014. Long-lasting, kin-directed female interactions in a spatially structured wild boar social network. PLoS ONE 9:e99875.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
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 16:608–627.
Queller, D. C., and K. F. Goodnight. 1989. Estimating relatedness using molecular markers. Evolution 43:258–275.
R Development Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.
Ramasamy, R. K., S. Ramasamy, B. B. Bindroo, and V. G. Naik. 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3:431.
Ramos, A. M., R. P. M. A. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, J. E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M. S. Hansen, J. Hedegaard, Z-L. Hu, H. H. Kerstens, A. S. Law, H-J. Megens, D. Milan, D. J. Nonneman, G. A. Rohrer, M. F. Rothschild, T. P. L. Smith, R. D. Schnabel, C. P. Van Tassel, J. F. Taylor, R. T. Wiedmann, L. B. Schook, and M. A. M. Groenen. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS ONE 4:e6524.
Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.
Rosenberg, M. S., and C. D. Anderson. 2011. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods in Ecology and Evolution 2:229–232.
Rousset, F. 2000. Genetic differentiation between individuals. Journal of Evolutionary Biology 13:58–62.
Rousset, F. 2008. GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. Molecular Ecological Resources 8:103–106.
Southeastern Cooperative Wildlife Disease Study. 2010. 2010 National Feral Swine Mapping System. Southeastern Cooperative Wildlife Disease Study, University of Georgia, Athens, USA. http://swine.vet.uga.edu/nfsms/. Accessed 1 Sep 2010.
Sweitzer, R. A., B. E. McCann, R. E. Loggins, and R. B. Simmons. 2015. Mitochondrial DNA perspectives on the introduction and range expansion of wild pigs in California. California Fish and Game 101:131–145.
Tabak, M. A., A. J. Piaggio, R. S. Miller, R. A. Sweitzer, and H. B. Ernest. 2017. Anthropogenic factors predict movement of an invasive species. Ecosphere 8(6)e01844.
Towne, C. W., and E. N. Wentworth. 1950. Pigs from cave to cornbelt. University of Oklahoma Press, Norman, USA.
Tynes, V. V. 1999. Potbellied pig husbandry and nutrition. Veterinary Clinics of North America: Exotic Animal Practice 2:193–207.
Waithman, J. D., R. A. Sweitzer, D. Van Vuren, J. D. Drew, A. J. Brinkhaus, I. A. Gardner, and W. M. Boyce. 1999. Range expansion, population sizes, and management of wild pigs in California. Journal of Wildlife Management 63:298–308.
Wiens, J. J. 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? Systematic Biology 47:625–640.
Zhen, L., V. Gopal, X. Li, J. M. Davis, and G. Casella. 2012. Simultaneous SNP identification in association studies with missing data. Annals of Applied Statistics 6:432–456.

Associate Editor: Zach Olson.

SUPPORTING INFORMATION

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