

# An island 'endemic' born out of hybridization between introduced lineages

Jessie F. Salter  | Robb T. Brumfield  | Brant C. Faircloth 

Museum of Natural Science and  
Department of Biological Sciences,  
Louisiana State University, Louisiana,  
Baton Rouge, USA

## Correspondence

Jessie F. Salter, Ornithology Department,  
Natural History Museum of Los Angeles  
County, Los Angeles, CA, USA.  
Email: [jessie.salter@gmail.com](mailto:jessie.salter@gmail.com)

Brant C. Faircloth, Museum of Natural  
Science and Department of Biological  
Sciences, Louisiana State University,  
Baton Rouge, LA, USA.  
Email: [brant@faircloth-lab.org](mailto:brant@faircloth-lab.org)

## Funding information

American Museum of Natural History;  
Louisiana State University; National  
Science Foundation, Grant/Award  
Number: DBI-2109361 and DEB-1655624;  
Society of Systematic Biologists

Handling Editor: Angus Davison

## Abstract

Humans have profoundly impacted the distribution of plant and animal species over thousands of years. The most direct example of these effects is human-mediated movement of individuals, either through translocation of individuals within their range or through the introduction of species to new habitats. While human involvement may be suspected in species with obvious range disjunctions, it can be difficult to detect natural versus human-mediated dispersal events for populations at the edge of a species' range, and this uncertainty muddles how we understand the evolutionary history of populations and broad biogeographical patterns. Studies combining genetic data with archaeological, linguistic and historical evidence have confirmed prehistoric examples of human-mediated dispersal; however, it is unclear whether these methods can disentangle recent dispersal events, such as species translocated by European colonizers during the past 500 years. We use genomic DNA from historical museum specimens and historical records to evaluate three hypotheses regarding the timing and origin of Northern Bobwhites (*Colinus virginianus*) in Cuba, whose status as an endemic or introduced population has long been debated. We discovered that bobwhites from southern Mexico arrived in Cuba between the 12th and 16th centuries, followed by the subsequent introduction of bobwhites from the southeastern USA to Cuba between the 18th and 20th centuries. These dates suggest the introduction of bobwhites to Cuba was human-mediated and concomitant with Spanish colonial shipping routes between Veracruz, Mexico and Havana, Cuba during this period. Our results identify endemic Cuban bobwhites as a genetically distinct population born of hybridization between divergent, introduced lineages.

## KEYWORDS

Cuba, historical DNA, historical museum specimens, human history, introduced species, Northern Bobwhites

## 1 | INTRODUCTION

Translocation of plants and animals has been a pervasive practice throughout human history (Boivin et al., 2016), and genetic studies of the movement of domesticated and human-associated species have provided insights into population genetics (Larson & Burger, 2013), invasion biology (Tsutsui et al., 2000), phenotypic

evolution (Mathys & Lockwood, 2009) and disease ecology (Young et al., 2017). When combined with archaeological data, translocations have also shaped our understanding of human history. For example, genetic data from moth skins (*Lipinia noctua*) stowed away in prehistoric canoes provided strong support for the 'express train' hypothesis of human colonization in the Polynesian islands (Austin, 1999). Similarly, genetic data from modern and historical

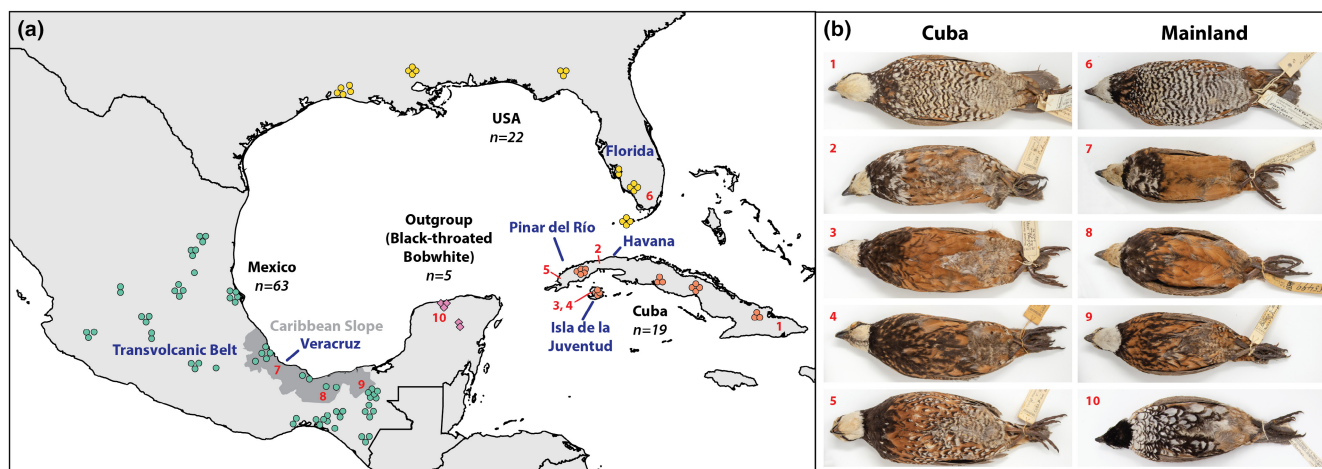
herbarium specimens of sweet potatoes (*Ipomoea batata*) supported the archaeological and linguistic evidence of pre-Columbian contact between Polynesians and South Americans (Roullier et al., 2013).

Evidence of human translocations is commonly detected on islands due to their unique biotic assemblages and because water is often a significant barrier to terrestrial dispersal (Hofman & Rick, 2018). For example, combined fossil, genetic and archaeological evidence suggested that northern common cuscuses (*Phalanger orientalis*) were introduced to several islands from New Guinea by humans ~20kya, and other marsupial distributions on islands throughout Australasia may have been shaped by similar prehistoric human translocations (Heinsohn, 2010; Summerhayes, 2007). For some species, the effects of human intervention on their distributions and demographic histories remain unclear, particularly when natural and human-mediated dispersal events occur within relatively recent timescales (e.g. the Holocene, 11.7kya-present) and near the edges of a species' native range (Heinsohn, 2003; Zeisset & Beebee, 2001). For example, these factors have limited our understanding of how Cuban tree frogs (*Osteopilus septentrionalis*) reached Florida: a short time frame for natural dispersal (<11.7kya) combined with 20th-century human-mediated introductions have confounded our understanding of whether they naturally colonized the state (Heinicke et al., 2011). Studying this latter class of natural versus human-mediated dispersal events across recent (11.7kya-present) timescales remains challenging, particularly when supporting fossils, genetic data, archaeological evidence or historical records are few.

Here, we combine historical literature and genomic DNA from historical museum specimens to investigate an island population of uncertain origin—the Cuban population of Northern Bobwhites (*Colinus virginianus*; hereafter bobwhites). Bobwhites are small,

sedentary quails native to southeastern Canada, the eastern United States, Mexico, Guatemala and Cuba (Figure 1). Along with Cuba, bobwhites have been documented on at least eight other islands in the Greater Antilles (Cory, 1885), and, except for Cuba, there is broad consensus that these populations were introduced by Europeans during the last 500 years due to the low dispersal capabilities of bobwhites (Smith et al., 1982; Stoddard, 1931) and the fact that bobwhites were not observed on these islands until relatively recently (Barbour, 1923; Bond, 1948; Johnsgard, 1988).

Bobwhites were first observed on mainland Cuba and Isla de la Juventud (an island 50km south of Pinar del Río province; Figure 1) by Europeans in 1839, and they were initially assumed to be introduced from populations in the eastern USA (D'Orbigny, 1839). Soon after, however, Cuban bobwhites were described as a distinct species, *C. cubanensis*, due to their smaller bills and their unique plumage pattern compared to bobwhites from mainland and other Caribbean island populations (Gould, 1850; Figure 1b.5). These morphological differences between Cuban and mainland bobwhites cast doubt on the initial hypothesis of recent human-mediated introduction given how little time had elapsed for phenotypic differences to accrue (Gould, 1850; Gundlach, 1876). However, German-Cuban naturalist Juan Gundlach published a conflicting account describing the introduction of bobwhites near Havana by a Spanish colonel during the late 18th century, although Gundlach did not know the source of the introduced population. Gundlach was also confused by the degree of morphological differentiation he observed between Cuban and mainland populations of bobwhites. Gundlach was able to integrate these competing ideas of recent introduction yet significant differentiation by proposing a theory that Cuban bobwhites were originally native to the savannas of Pinar del Río province in western



**FIGURE 1** Sampling localities and example plumages of Northern Bobwhites and Black-throated Bobwhites. (a) Sampling localities and sample number per locality for 104 Northern Bobwhites and 5 Black-throated Bobwhites. Markers are coloured by geographical population: USA (yellow), Mexico (teal), Cuba (orange) and Black-throated Bobwhites (outgroup; pink). Notable localities and geographical features mentioned throughout the text are labelled in blue; the Caribbean Slope of southern Mexico (including parts of Veracruz and Tabasco states) is shown in dark grey. Red numbers show collection localities of specimens in (b). (b) Example plumages of Northern Bobwhites (1–9) and Black-throated Bobwhites (10). Specimens 1–4 from Cuba have been paired with specimens 6–9 from mainland populations with similar plumages. Specimen 5 displays the unique *C. v. cubanensis* plumage described by Gould (1850). Specimen photographs were taken at the National Museum of Natural History, and these specific individuals are not represented in the genetic dataset.

Cuba. He added the contention that translocation of native, Pinar del Río individuals to Havana explained the eastward expansion of morphologically distinct Cuban bobwhites during the latter half of the 19th century (Gundlach, 1876).

To further complicate matters, translocations of Florida bobwhites (*C. v. floridanus*) to mainland Cuba began during the late 19th century (Chapman, 1892; Cory, 1885), and translocated individuals began hybridizing successfully with 'native' Cuban bobwhites despite differences in their natural habitats (Chapman, 1892). As of the early 20th century, few 'pure' Cuban bobwhites were thought to remain outside of Pinar del Río province and Isla de la Juventud (Figure 1) where there were no documented introductions; thus, these two localities have been suggested as the last refuges of 'pure' Cuban bobwhites (Barbour, 1923). However, based on the similarities in plumage between the 'pure' Cuban bobwhites from Isla de la Juventud and bobwhites from the Caribbean slope of southern Mexico (including the states of Veracruz and Tabasco; Figure 1), it has also been suggested that Cuban bobwhites were introduced from southern Mexico during Spain's colonization of both countries (Parkes, 1990).

Collectively, this prior work suggests the following hypotheses: (1) bobwhites were introduced to Cuba from the southeastern USA during the Spanish colonial period (D'Orbigny, 1839); (2) bobwhites were endemic to western Cuba prior to the arrival of Europeans (Gundlach, 1876); and (3) bobwhites were introduced to Cuba from southern Mexico during the Spanish colonial period (Parkes, 1990). Similar to other studies where natural and human-mediated dispersal events may have interacted over relatively recent timescales, molecular efforts to demystify the origins of Cuban bobwhites have produced equivocal results. For example, an analysis of two mitochondrial markers found shared haplotypes between eight bobwhites collected from central and eastern Cuba and individuals collected from Mexico and the United States. However, these data failed to identify any unique Cuban haplotypes; showed little resolution in the trees and haplotype networks among Cuban, Mexican and USA bobwhite populations; and showed discordance between haplogroups and population boundaries (Williford et al., 2016). A subsequent analysis using thousands of ultraconserved element (UCE) loci placed Cuban bobwhites sister to a clade of individuals collected from southern Mexico—suggesting the two populations share a common ancestor (Salter et al., 2022). Although this analysis only sampled a single individual from Cuba, the specimen was collected in Pinar del Río province—one of the potential remaining refuges of 'pure' Cuban bobwhites (Barbour, 1923; Gundlach, 1876). While these results suggest that Cuban bobwhites share genetic ancestry with multiple mainland bobwhite populations, the origin and timing of how bobwhites arrived to and dispersed across Cuba remain unclear.

To address these questions, we used a target capture approach appropriate for historical museum specimens (RADcap; Hoffberg et al., 2016) to collect thousands of restriction-site associated (RAD; Baird et al., 2008; Davey et al., 2011; Miller et al., 2007) loci from 109 bobwhite specimens sampled during the 1850s to the 1960s

throughout mainland Cuba, Isla de la Juventud and source populations in the United States and Mexico. Using phylogenetic, population genetic and demographic analyses in concert with historical records, we evaluated the three hypotheses that have been proposed to explain the origin of Cuban bobwhites.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling design and DNA extraction

We sampled 109 individuals, including 95 toepads from historical museum specimens collected between 1859 and 1966 (median collection year = 1941) and 14 tissue samples collected between 1985 and 2014 (median collection year = 2010; Table 1; Table S1). We included 104 Northern Bobwhites from Cuba ( $n=19$ ), the United States ( $n=22$ ) and Mexico ( $n=63$ ), as well as five Black-throated Bobwhites (*C. nigrogularis*) from the Yucatán Peninsula (Table 1; Table S1), which we used as outgroup individuals in some analyses (Salter et al., 2022). We extracted total DNA from tissues using a Qiagen DNeasy Blood & Tissue Kit following the manufacturer's instructions, and we extracted total DNA from toepads using a phenol-chloroform protocol (Tsai & Salter, 2018). Following extraction, we quantified samples with a Qubit Fluorometer (Life Technologies, Inc.).

### 2.2 | Library preparation, targeted enrichment and sequencing

For tissue samples, we prepared three 3RAD pools containing eight uniquely indexed individual libraries that were constructed in a way that allowed us to identify and remove PCR duplicates. For toepad samples, we prepared standard genomic libraries using the KAPA Hyper Prep library preparation kit (F. Hoffmann-La Roche AG) and custom indexes (Glenn et al., 2019). Prior to enrichment, we quantified libraries using a Qubit Fluorometer and combined toepad libraries into pools of eight individuals at equimolar ratios. We enriched all library pools using our custom RADcap bait set targeting 2351 neutral single nucleotide polymorphisms (SNPs) following the myBaits Hybridization Capture for Targeted NGS manual v4.01 and sequenced all pools using two partial lanes of 150-bp paired-end (PE150) sequencing on an Illumina NovaSeq 6000 (Novogene). Remaining samples on each lane had non-overlapping indexes. Additional details regarding identification of neutral SNPs, bait design, library preparation, targeted enrichment and sequencing are provided in the Supplemental Methods in Data S1 and Table S2.

### 2.3 | SNP calling and filtering

Because tissue and toepad libraries were prepared with different types of indexes, we performed slightly different steps to

TABLE 1 Summary of geographical sampling scheme, including the number of samples obtained from nine natural history museums.

Population	Museum code									Total samples
	DMNH	LACM	LSUMZ	MCZ	MLZ	SDNHM	UMMZ	WFVZ	YPM	
Outgroup (Black-throated Bobwhite)									5	5
Cuba				7					11	19
Northern Mexico			3		23				1	26
Southern Mexico		1	5		8	3	3	17		37
USA	3		11	8						22
Total	3	1	19	15	31	3	3	17	17	109

Note: Museum codes: DMNH, Delaware Museum of Natural History; LACM, Natural History Museum of Los Angeles County; LSUMZ, Louisiana State University, Museum of Natural Science; MCZ, Museum of Comparative Zoology, Harvard University; MLZ, Moore Laboratory of Zoology, Occidental College; SDNHM, San Diego Natural History Museum; WFVZ, Western Foundation of Vertebrate Zoology; UMMZ, University of Michigan, Museum of Zoology; YPM, Yale Peabody Museum of Natural History. For complete information on all samples used, including museum catalogue number, subspecies identification, collection locality, sample type, collection year, raw read pairs, average depth of coverage and the number of SNPs in the Set1a population structure dataset, see [Table S1](#).

demultiplex sequencing reads, trim reads for adapters and low-quality bases and remove PCR duplicates. The full procedure is described in the Supplemental Methods in Data S1. We aligned reads to the bobwhite reference genome (Salter et al., 2019) using bwa v0.7.17 (Li & Durbin, 2009) and samtools v1.10 (Li et al., 2009), and we called SNPs using a parallel implementation of the Best Practices for Variant Discovery (Salter & Faircloth, 2021; Van der Auwera & O'Connor, 2020) in GATK v4.1.9.0, including base quality score recalibration (McKenna et al., 2010).

Because the analyses we implemented assume sites are putatively neutral and unlinked (Frichot et al., 2014; Kamm et al., 2020), we used VCFTools to filter SNPs in the recalibrated VCF file to the single locus targeted during bait design. We also used VCFTools to exclude: loci determined to be indels, sites with less than a minimum depth of 10, sites with more than 10% missing data and sites out of Hardy-Weinberg equilibrium due to excess heterozygosity, which can indicate genotyping errors (Chen et al., 2017). Because missing data can bias population structure assignment (Yi & Latch, 2021), we also removed individuals missing more than 25% of calls at the remaining sites.

Because particular filters applied to a set of variable SNPs can affect different types of analyses in unexpected ways (Chattopadhyay et al., 2014), we produced two types of VCF files: Set1 for population genetic summary statistics, genetic structure and phylogenetic analyses; and Set2 for demographic analysis. To produce the first set of files (Set1), we used VCFTools to remove loci having a minor allele count (MAC) less than three (Linck & Battey, 2019), and we output files: (Set1a) including all bobwhite samples+the outgroup individuals (ingroup+outgroup); and (Set1b) including only the Cuban, the United States and Mexican populations of bobwhites (ingroup only). The inclusion of the outgroup individuals in the Set1a data file allowed us to test whether there was any allele sharing between Black-throated Bobwhites and Cuban bobwhites because some populations of Black-throated Bobwhites on the Yucatán Peninsula are physically closer to Cuba

than populations of Northern Bobwhites in either Florida or the Caribbean slope of southern Mexico. To produce the second file for demographic analyses (Set2), we removed the MAC filtering because rare alleles can be important for inferring demographic history (Marth et al., 2004), and we included all bobwhites+the outgroup individuals so that we could include an external constraint on the common ancestor of the ingroup and outgroup (see below). Additional details on base quality score recalibration, SNP calling and filtering procedures are provided in the Supplemental Methods in Data S1.

## 2.4 | Patterns of population genetic diversity

To assess population structure and potential admixture among populations, we calculated ancestry coefficients in sNMF (Frichot et al., 2014) using Set1a and Set1b VCF files. To further characterize the patterns of genetic clustering in our dataset, we performed a discriminant analysis of principal components (DAPC) using VCFR v1.12.0 (Knaus & Grünwald, 2017) and adegenet v2.1.3 (Jombart, 2008) in R v4.0.3 (R Core Team, 2020). Finally, we performed a principal components analysis (PCA) using SNPrelate (Zheng et al., 2012) in R v4.0.3 (R Core Team, 2020) because PCA does not require a priori selection of the number of clusters present in the data.

To describe patterns of differentiation and genetic diversity within and among bobwhites (ingroup only), we calculated population genetic summary statistics (observed and expected heterozygosity, allelic richness, inbreeding coefficients, nucleotide diversity and  $F_{ST}$  between each pair of populations) using the Set1b VCF file. We also calculated the number of private alleles in each population using a custom Python (Van Rossum & Drake, 2009) script, which can be found on GitHub (Faircloth, 2021). Additional details on analyses of genetic structure and population genetic summary statistics are in the Supplemental Methods in Data S1.

## 2.5 | Phylogenetic analyses

To investigate the evolutionary relationships among populations, we used the Set1b VCF file to estimate species trees using SNAPP v1.5.1 (Bryant et al., 2012) implemented in Beast v2.6.3 (Bouckaert et al., 2014). We chose SNAPP because it was designed to model the coalescent process using allele frequencies from biallelic SNP data and because it produces output that enables users to visualize conflicting signals that may arise due to allele sharing among multiple populations, which we expected to exist in our dataset based on the possibility of multiple introductions into Cuba. Because SNAPP is computationally intensive, precluding us from estimating a species tree for our entire dataset, we used two types of subsampling schemes to provide resolution at broad and fine geographical scales: (1) five replicate subsamples containing five randomly sampled individuals from each of the four geographical populations identified by our sNMF analyses (USA, northern Mexico, southern Mexico and Cuba; 20 individuals per analysis; Table S3); and (2) locality-specific subsets for each of the five, distinct Cuban localities that included all individuals from that location (range 3–6 individuals) as well as two randomly selected individuals from each of the 12 mainland populations representing northern Mexico, southern Mexico and the United States (total 27–30 individuals per analysis; Table S4).

We ran each SNAPP analysis with default parameters for 2 million iterations, sampling every 1000 iterations and discarding the first 10% of sampled iterations as burn-in. We used Tracer v1.7.2 (Rambaut et al., 2018) to visualize estimated parameters and to ensure that the effective sample size (ESS) of each was greater than 200 (Drummond et al., 2006). We visualized the resulting species trees and posterior distribution of gene trees using DensiTree v2.2.7 (Bouckaert, 2010; Bouckaert & Heled, 2014). Additional details on random sampling protocols and SNAPP analyses are provided in the Supplemental Methods in Data S1.

## 2.6 | Demographic analyses

Inferring evolutionary relationships and visualizing allele sharing among populations provide one mechanism for understanding population origins, but these types of analyses do not typically allow objective comparisons of alternative evolutionary scenarios, including comparisons of the timing of evolutionary events and/or the source(s) and direction of migrants to specific populations. This is particularly true in the absence of suitable fossil calibrations, which is currently the case for Cuban bobwhites (Orihuela, 2019). To perform these types of analyses, we used momi2 (Kamm et al., 2020) with the site frequency spectrum derived from the Set2 VCF file.

Because the historical record contradicts a hypothesis of a single introduction of bobwhites to Cuba (Barbour, 1923; Chapman, 1892), we tested two categories of models given our data: (1) single source models with a second dispersal event and (2) multiple source models. We designed the single source models to test scenarios where one of the mainland populations (the ancestor of northern + southern

Mexico, northern Mexico, southern Mexico or USA) founded the Cuban population, followed by a single dispersal event from the same founding population to Cuba. This design is consistent with either natural or human-mediated dispersal scenarios from a single geographical origin. We designed the multiple source models to test scenarios where one of the mainland populations founded the Cuban population, followed by a single dispersal event from a different mainland population during the last 500 years. This design is consistent with either natural or human-mediated dispersal scenarios establishing the Cuban population followed by human-mediated introductions during the last 500 years, as noted in several historical sources (Barbour, 1923; Chapman, 1892; Gundlach, 1876). We constrained models in the multiple source category to always include the United States, either as the founding population or as the source of more recent dispersal to Cuba, because SNAPP analyses produced topologies suggesting allele sharing between bobwhite populations in the United States and Cuba. For example, we did not model a scenario in which northern Mexico bobwhites founded the Cuban population, followed by recent dispersal of bobwhites from southern Mexico because this was inconsistent with the SNAPP results. To allow tests of hypotheses regarding the timing of bobwhite arrival on Cuba, we incorporated five different temporal scenarios to all models in both categories for the founding of the Cuban population—specifically, that the Cuban bobwhite population was founded: (1) within 0.5 kya (since European arrival on Cuba; De las Casas, 1877); (2) between 0.5 and 5.0 kya (since Indigenous arrival on Cuba; Allaire, 1999); (3) between 5.0 and 11.7 kya (since the Pleistocene–Holocene Transition; Cohen et al., 2013); (4) between 11.7 and 23.0 kya (since the last glacial maximum; Cohen et al., 2013); or (5) between 23.0 and 140.0 kya (since the penultimate glacial maximum; Colleoni et al., 2016). In total, this produced 45 separate models (Figure S2; Table S5).

Before running momi2, we set the effective size of the population ancestral to bobwhites and their sister species, Black-throated Bobwhites to  $3.35 \times 10^5$ , the mean value estimated across multiple G-PhoCs (Gronau et al., 2011) runs (additional details on G-PhoCs analyses are in the Supplemental Methods in Data S1); we specified a generation time of 1.22 years, which was the median value estimated from multiple radiotelemetry and survivorship studies of wild bobwhites (Halley et al., 2014); and we specified a mutation rate of  $1.91 \times 10^{-9}$  sites per year, which was estimated from chickens (Nam et al., 2010). Finally, we constrained the divergence time between Northern Bobwhites and Black-throated Bobwhites to have occurred since 1.563 Ma, which was the older bound of the 95% highest posterior density interval estimated for their divergence in a time-calibrated phylogeny (Hosner et al., 2015).

We ran each of the 45 models 10 times and computed corrected Akaike information criterion (AICc) scores (Hurvich & Tsai, 1989, 1993) for all models in each set of the 10 model runs. We then used AIC-based model comparison (Burnham & Anderson, 2002) to rank and compare models (Table S6). For each of the three models that fell within the confidence set (Royall, 1997) among any of the 10 runs, we completed an additional 90 runs. We selected the best (highest)

log-likelihood value for each model out of 100 (10+90) runs, and we used the best log-likelihood obtained for each of the three top models to compute final AICc scores, delta AICc values and Akaike weights following Burnham and Anderson (2002). Finally, for the two (out of three) models in the final confidence set, we ran 100 bootstrap replicates in *mom2* and computed the 95% confidence intervals for the parameter estimates produced by each model.

### 3 | RESULTS

#### 3.1 | Sequence data from tissue and toepad samples

Sequence data collected for both tissue and toepad samples were comparable, although we observed some differences between sample types and greater variance in all metrics among sequence data collected from toepads (Table 2). Specifically, we collected an average of 1,629,289 reads (range 206,636–2,483,216, 95% confidence interval (CI)  $\pm 377,407$ ) from tissue samples prepared as 3RAD libraries, and 1,338,505 reads (range 269,012–5,980,941, 95% CI  $\pm 158,242$ ) from toepad samples prepared as standard genomic libraries. The main difference between the two sample types was the efficacy of target capture: 95.7% of reads collected from tissue samples were on-target, whereas 24.1% of reads were on-target for toepad samples. We also discarded substantially more duplicate reads from toepad samples (24.8%) than from tissues (4.96%). After removing duplicates, the average depth of coverage across targeted SNPs was lower in tissues (range 35–72, mean  $47 \pm 4$  95% CI), although the variance was greater for toepads (range 19–209, mean  $66 \pm 6$  95% CI).

After filtering, the Set1 VCF files used for population genetic and phylogenetic analyses contained 1258 SNPs (Set1a, ingroup only) and 1267 SNPs (Set1b, ingroup+outgroup), and the Set2 VCF file used for demographic modelling contained 2228 SNPs.

#### 3.2 | Population structure and genetic diversity

sNMF runs for individuals from the United States, Mexico and Cuba identified four populations as the best-fitting value of K, a result that is broadly concordant with sampling region. All Cuban and USA samples were assigned to their own populations, respectively, and the Mexican samples were assigned to either a northern or a southern population, consistent with phylogenetic breaks observed in UCE data for this species that correspond to the Transvolcanic

Belt (Marshall & Liebherr, 2000; Morrone, 2010; Salter et al., 2022; Figure 2a). sNMF also showed that the four populations were well differentiated with little evidence of admixture in most individuals. The exceptions to this general trend were admixture between the northern and southern Mexico populations evident in samples collected from the Caribbean slope of southern Mexico and admixture between the northern Mexico population and US individuals collected in Louisiana and Georgia. Notably, individuals from the Caribbean slope of southern Mexico shared alleles with Cuban individuals (Figure 2a). The DAPC results also identified four populations as best fitting the data, and population assignments were identical to the results from sNMF (Figure 2b). PCA analyses, which did not require a priori selection of the number of clusters in the data, also showed four well-differentiated populations and identical population assignments relative to the sNMF and DAPC results (Figure 2c). When we included Black-throated Bobwhites (*C. nigrogularis*) as the outgroup in similar analyses, the best-fitting K-value suggested by sNMF increased to five populations, with little change in the admixture proportions estimated among the bobwhite samples or in the clustering results from either the DAPC or PCA (Figure S1). These results suggest Black-throated Bobwhite populations from the Yucatán Peninsula were not the source of Cuban bobwhites.

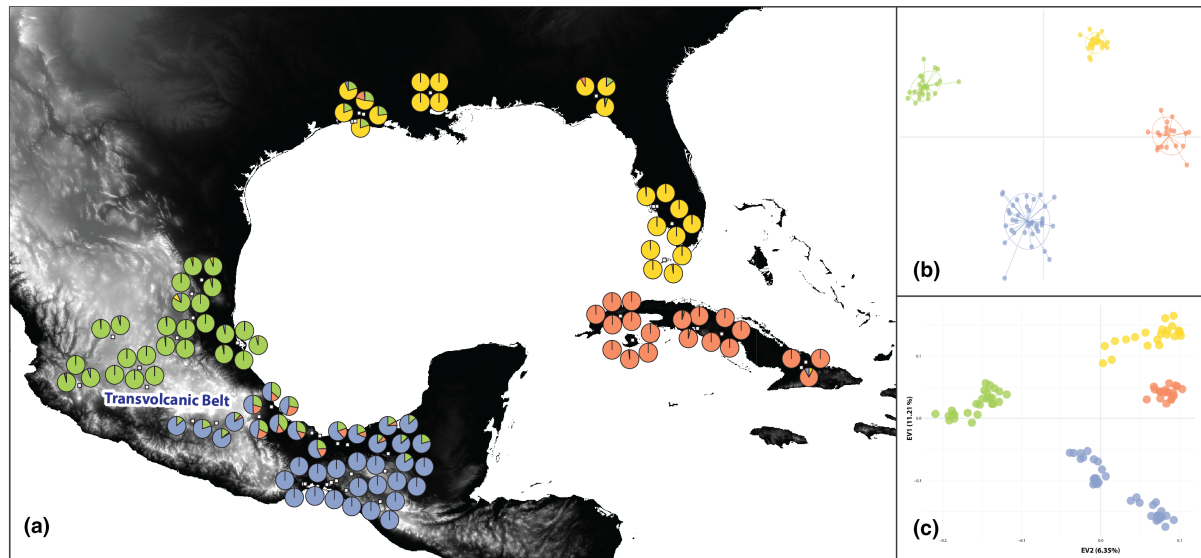
Across all populations, mean nucleotide diversity was 0.277 (0.268–0.285 95% CI), and observed heterozygosity was lower than expected ( $t=21.845$ ,  $df=5067$ ,  $p\text{-value}=2.2e-16$ ; Table 3). When we considered populations separately, nucleotide diversity and observed and expected heterozygosity were highest among US bobwhites and lowest among southern Mexico bobwhites (Table 3). All four populations had positive inbreeding coefficients, ranging from 0.050 in the Cuban population to 0.155 in southern Mexico (Table 3).  $F_{ST}$  values were moderate between populations (Table 4), with the Cuban population showing the least differentiation compared to the United States, followed by comparisons of Cuban individuals to individuals collected from the northern Mexico and southern Mexico populations (Table 4). We identified private alleles in all four populations, including seven alleles unique to the Cuban population (Table 4). When we expanded the analysis to look for alleles unique to pairs of populations, we found one private allele shared between Cuba and northern Mexico, 11 shared between Cuba and southern Mexico, and 47 shared between Cuba and the United States (Table 4).

#### 3.3 | Phylogenetic analyses

SNAPP analyses using a broad geographical subsampling scheme, in which we randomly sampled individuals from each of the four

TABLE 2 Sequencing summary statistics for tissues and toepads.

Sample type	No. of samples	Mean raw reads $\pm$ 95% CI	Reads on target (%)	Duplicate reads (%)	Mean depth of coverage $\pm$ 95% CI
Tissues	14	1,629,289 $\pm$ 377,407	95.7	4.96	47 $\pm$ 4
Toepads	95	1,338,505 $\pm$ 158,242	24.1	24.8	66 $\pm$ 6



**FIGURE 2** Population structure of Northern Bobwhites from Cuba, the United States and Mexico. (a) sNMF results for the best-fitting  $K$ -value of four populations. Pie charts show admixture proportions for each individual plotted by sampling locality (points have been jittered around exact coordinates to allow easier viewing). (b) DAPC results for best-fit four population clusters: USA (yellow), northern Mexico (green), southern Mexico (blue) and Cuba (orange). (c) PCA results for the same set of individuals. Clusters have been coloured as in (a, b) to reflect population assignment from sNMF and DAPC.

**TABLE 3** Population genetic summary statistics.

Population	Sample size	$H_O$	$H_S$	Allelic richness	$\pi$	$F_{IS}$
Cuba	19	0.223	0.237	1.679	0.236	0.050
N. Mex	26	0.232	0.249	1.785	0.249	0.052
S. Mex	37	0.174	0.214	1.693	0.213	0.155
USA	22	0.260	0.281	1.893	0.280	0.062

Note: For each population, the table shows the number of individuals, observed ( $H_O$ ) and expected ( $H_S$ ) heterozygosity, allelic richness, nucleotide diversity ( $\pi$ ), and inbreeding coefficient ( $F_{IS}$ ). All statistics were calculated from the Set1b dataset.

Abbreviations: N. Mex, northern Mexico; S. Mex, southern Mexico.

**TABLE 4**  $F_{ST}$  between population pairs (lower half in white), private alleles within each population (black diagonal), and private alleles within pairs of populations (upper half in grey).

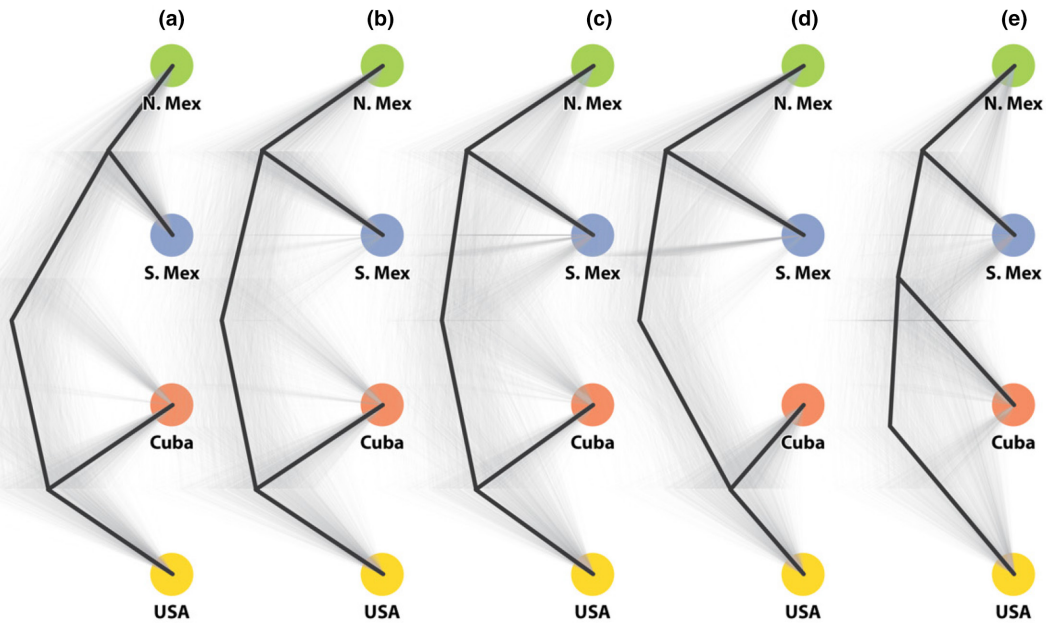
Population	Cuba	N. Mexico	S. Mexico	USA
Cuba	7	1	11	47
N. Mex	0.155	6	57	74
S. Mex	0.184	0.167	24	24
USA	0.126	0.135	0.226	59

Abbreviations: N. Mex, northern Mexico; S. Mex, southern Mexico.

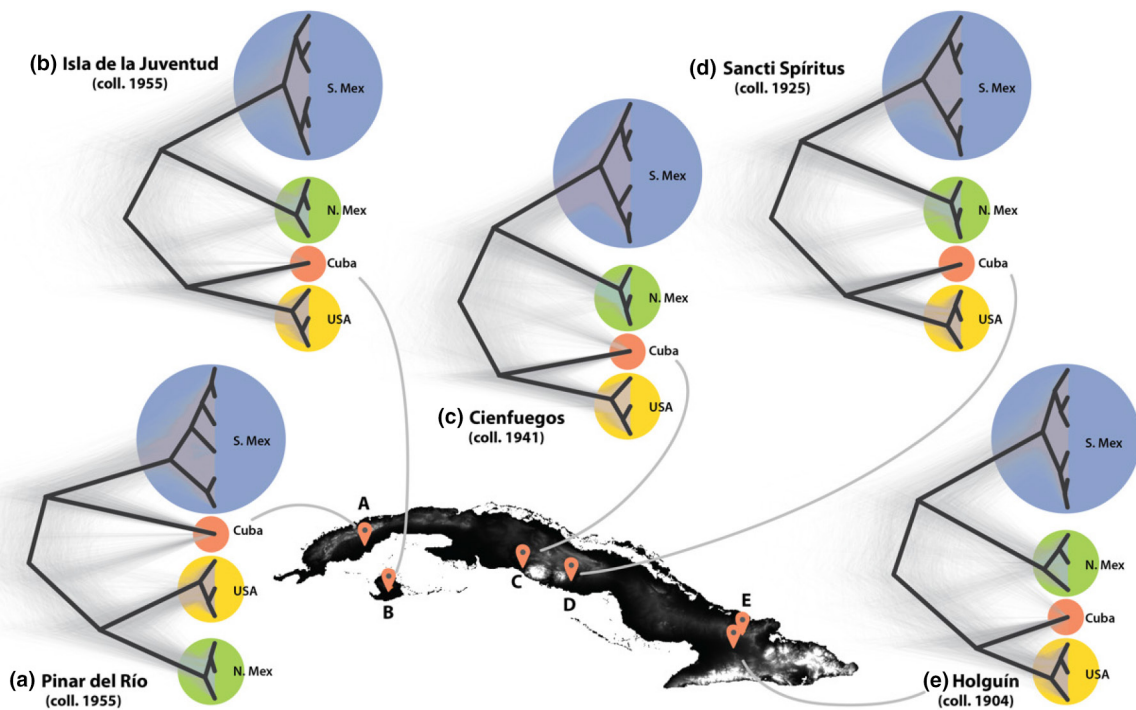
populations identified by sNMF (Cuba, northern Mexico, southern Mexico and the United States), produced different species tree topologies (Figure 3). In four of the five analyses, the consensus topology resolved the Cuban population as sister to the US population, with Cuba+USA resolved as sister to a clade consisting of both Mexican populations (Figure 3a–d). However, the remaining analysis resolved the Cuban population as sister to the two Mexican populations (Figure 3e). Across all five analyses, the posterior distribution

of trees showed considerable variation in topology, including inconsistencies that suggested allele sharing between the Cuban population and multiple mainland populations (Figure 3a–d).

SNAPP analyses with a finer population-level subsampling scheme, which we used to understand the complicated history of introductions across Cuba, suggested that there was a geographical pattern in the discordance we observed in the first subsampling scheme: individuals collected in Pinar del Río, the westernmost population in Cuba, were resolved as sister to the southern Mexico population, with northern Mexico and US populations forming the sister clade to this group (Figure 4a). However, individuals collected from three populations in the central and eastern provinces of Cuba, as well as individuals collected from Isla de la Juventud, were resolved as sister to the US population, with northern and southern Mexico populations forming a subsequent sister clade (Figure 4b–e). As in the first subsampling scheme and despite the differences in topology, the posterior distribution of trees in all analyses included inconsistencies that suggested allele sharing between the Cuban population and multiple mainland populations.



**FIGURE 3** Evolutionary relationships of Northern Bobwhite populations from Cuba, the United States and Mexico. Species trees from five independent SNAPP analyses using five randomly sampled individuals per population (a–e). Consensus topology is depicted by black lines; alternate topologies are depicted using grey lines. The samples used for each analysis are listed in [Table S3](#).



**FIGURE 4** Cuban bobwhite populations have heterogeneous evolutionary histories. Species trees from five independent SNAPP analyses using 3–5 samples per Cuban population and two randomly sampled individuals from all subpopulations ([Table S1](#)) of each mainland population (a–e). The year the Cuban samples were collected is listed below the locality name. Consensus topology is depicted by black lines; alternate topologies are depicted using grey lines. The samples used for each analysis are listed in [Table S4](#).

### 3.4 | Demographic analyses

To allow objective comparisons of historically and biologically informed alternative evolutionary scenarios for the arrival of

bobwhites to Cuba, we tested 45 demographic models. In the initial comparison of all 45 models, only three models fell within the candidate set. After running each of these three models 100 times to search the likelihood space, only two models remained



in the confidence set with Akaike weights ( $w_i$ ) of 0.503 and 0.493 (Tables 5 and 6). Both models support a scenario in which the Cuban bobwhite population was founded by bobwhites from southern Mexico, followed by introduction of bobwhites from the United States (Figure 5; Table 6). The model receiving greater support ( $w_i=0.503$ ) constrained the initial colonization from southern Mexico between 0.5 and 5.0 kya (estimated at  $855 \pm 217$  years ago or ca. 1165 CE), with the introduction from the United States occurring  $301 \pm 26$  years ago (ca. 1720 CE). The second-best model ( $w_i=0.436$ ) constrained the timing of colonization to within the past 500 years and estimated slightly younger ages for founding and migration events (colonization from southern Mexico  $467 \pm 22$  years ago, introduction from the United States  $158 \pm 25$  years ago). Both models estimated small effective population sizes for Cuban bobwhites (393–752 individuals), consistent with a scenario of a small

founding population. Both models estimated the pulse probability from the United States at ~58%, meaning 58% of the effective population size of Cuban bobwhites are migrants from the United States (Kamm et al., 2020). Remaining parameter estimates were concordant between the two models (Table 6). Both models estimated: (1) the divergence between Northern Bobwhites and Black-throated Bobwhites occurred ca. 1.563 Ma, which was equal to the constraint we imposed for this event based on results from a time-calibrated phylogeny (Hosner et al., 2015) and (2) simultaneous divergence of the three mainland lineages ca. 334 kya. The effective population size estimates were similar for the three mainland bobwhite lineages, ranging from ~920,000 for the United States to ~1.1 million for southern Mexico and ~1.3 million for northern Mexico. Effective population size of Black-throated Bobwhites was estimated at ~2.2 million (Table 6).

TABLE 5 Results of demographic model selection.

Model	Model description	K	Likelihood score	AICc	$\Delta_i$	$w_i$
m27	Southern Mexico founder (0.5–5.0 kya), subsequent migration from US (<0.5 kya)	11	-19,238.203	38,501.128	0.000	0.503
m26	Southern Mexico founder (<0.5 kya), subsequent migration from US (<0.5 kya)	11	-19,238.225	38,501.171	0.043	0.493
m41	US founder (<0.5 kya), subsequent migration from Southern Mexico (<0.5 kya)	11	-19,243.078	38,510.878	9.750	0.004

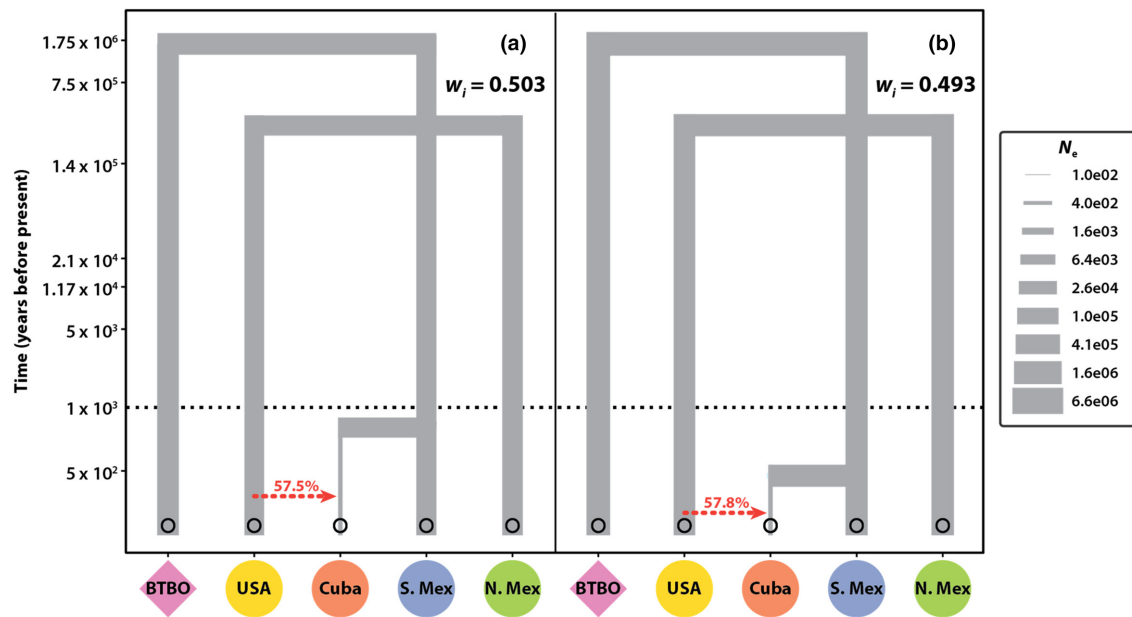
Note: Each model was run 100 times and the highest likelihood score for each was used for AIC model selection. The number of parameters (K), corrected AIC scores (AIC<sub>c</sub>), difference between model AIC<sub>c</sub> and the lowest AIC<sub>c</sub> ( $\Delta_i$ ), and Akaike weights ( $w_i$ ) are given. Model numbers correspond to the graphical depictions in Figure S2.

TABLE 6 Parameter estimates for the two best-supported demographic models  $\pm$  95% confidence intervals.

Parameter	Southern Mexico founder (0.5–5 kya), subsequent migration from USA (<0.5 kya)	Southern Mexico founder (<0.5 kya), subsequent migration from USA (<0.5 kya)
Akaike weight	0.503	0.493
Age of migration from USA	$301 \pm 26$ (upper limit 500)	$158 \pm 25$ (upper limit 467)
Migration pulse probability	$57.50 \pm 0.01\%$	$57.82 \pm 0.01\%$
Divergence age between Cuba and southern Mexico populations	$855 \pm 217$ (constrained 500–5000)	$467 \pm 22$ (upper limit 500)
Divergence age between northern and southern Mexico populations	$333,786 \pm 11,434$	$334,693 \pm 18,884$
Divergence age between USA and ancestral Mexico population	$333,786 \pm 15,362$	$334,693 \pm 17,133$
Divergence age between Black-throated and ancestral Northern Bobwhite population	$1,562,873 \pm 27,289$ (upper limit 1.563 M)	$1,562,902 \pm 28,510$ (upper limit 1.563 M)
Cuba $N_e$	$752 \pm 68$	$393 \pm 49$
Northern Mexico $N_e$	$1,290,940 \pm 55,260$	$1,310,580 \pm 80,307$
Southern Mexico $N_e$	$1,108,260 \pm 21,675$	$1,098,556 \pm 25,430$
USA $N_e$	$916,239 \pm 65,384$	$927,960 \pm 94,878$
Black-throated Bobwhite $N_e$	$2,213,379 \pm 426,636$	$2,262,612 \pm 340,677$

Note: Divergence ages are given in years.

Abbreviation:  $N_e$ , effective population size.



**FIGURE 5** Parameter estimates of the best supported demographic models (a, b). Y-axis is log-scaled above the dotted line. Timing and direction of migration is shown with red-dashed arrows; percentages represent the percent of the Cuban effective population that are migrants from the United States. BTBO, Black-throated Bobwhites; S. Mex, southern Mexico; N. Mex, northern Mexico. Branches are scaled for effective population size ( $N_e$ ); scale shown at right.

## 4 | DISCUSSION

### 4.1 | Cuban bobwhites are a unique population born through hybridization of divergent lineages

Together, our data suggest that a small number of bobwhites from the Caribbean slope of southern Mexico colonized Cuba, followed by significant introgression from bobwhites introduced to Cuba from the southeastern USA. The conflicting signals of ancestry present in the SNAPP analyses suggest that Cuban bobwhites are the result of interacting evolutionary lineages, consistent with introductions from mainland source populations in southern Mexico and the United States (Figures 3 and 4). Our demographic model results are unequivocal that the Cuban bobwhite population was founded by individuals from southern Mexico, followed by substantial introgression with bobwhites introduced from the southeastern USA (Table 6; Figure 5). Models supporting an alternate order of introduction received no support (Table 5).

Although the sNMF results showed little evidence of allele sharing between Cuban bobwhites and individuals in northern Mexico and the United States, the nine individuals that we sampled from the Caribbean slope of southern Mexico showed a considerable degree of allele sharing with the Cuban population (11%–26% population assignment; Figure 2). This finding is consistent with a scenario in which individuals from this region of southern Mexico founded the Cuban population, followed by subsequent founder effects and genetic drift in the Cuban population which can significantly alter allele frequencies without much effect on heterozygosity (Allendorf, 1986). Despite the small effective population sizes for Cuban bobwhites estimated by the demographic models (Figure 5, Table 6), we observed higher

heterozygosity and allelic diversity among Cuban bobwhites than in the much larger southern Mexico population (Table 3), consistent with a founding population bottleneck followed by substantial movement of individuals from the United States into Cuba, similar to patterns of genetic diversity in recently introduced populations of Caribbean anoles (Kolbe et al., 2004, 2007) and frogs (Heinicke et al., 2011). Taken together, these results reject the hypothesis that Cuban bobwhites were originally introduced from the southeastern USA, although our demographic model parameter estimates suggest that the effective size of the Cuban bobwhite population comprises more individuals from the United States than from southern Mexico (Table 6).

Conversely, SNAPP results support conjecture (Barbour, 1923; Gundlach, 1876) that Pinar del Río province is a reservoir of ‘native’ Cuban bobwhites and potentially the site of their founding population (Figure 4). Pinar del Río was the only population for which SNAPP inferred a species tree that resolved Cuban bobwhites as sister to bobwhites collected in southern Mexico, rather than the United States (Figure 4a), unlike the three populations in central and eastern Cuba where 19th-century introductions from the southeastern USA have been documented (Chapman, 1892). However, our SNAPP results challenge the idea that Isla de la Juventud remains a population of ‘native’ Cuban bobwhites that have not experienced subsequent introgression with bobwhites from the United States. Individuals from this locality were resolved as sister to the US population rather than sister to the southern Mexico population (Figure 4b). That said, given the small size of the island (2200 km<sup>2</sup>) and our small sample size ( $n=5$ ), a single additional individual could conceivably alter this result.

In effect, the Cuban bobwhites we sampled (collected between 1859 and 1966) are advanced generation hybrids between two

well-differentiated lineages of Northern Bobwhites that have been evolving somewhat independently for >300,000 years (Figure 5, Table 6). This conclusion explains observations of diverse plumage phenotypes in Cuban bobwhites, with specimens collected from around Cuba and Isla de la Juventud displaying plumage typical of Florida bobwhites (Ridgway, 1894), Caribbean slope bobwhites (Parkes, 1990), apparent Cuban-Floridian hybrids (Chapman, 1892) and the distinct *Colinus cubanensis* plumage described by Gould (1850) (Figure 1b).

#### 4.2 | The timing of bobwhite arrival on Cuba suggests human-mediated introduction

Although both demographic models in our confidence set agree on the source and order of introductions to Cuba, the small difference in their Akaike weights suggests they are equivocal regarding the precise timing of these events. The best-weighted model suggests bobwhites arrived in Cuba from southern Mexico 638–1072 years ago (ca. 948–1382 CE), but this scenario is only 1.02 times more likely given our data than the model that estimated the arrival of bobwhites 445–489 years ago (ca. 1531–1575 CE). This equivocacy may be the result of how we constrained our models. In both candidate models, the divergence time between Northern Bobwhites and Black-throated Bobwhites was estimated to be slightly younger than the older limit that we allowed based on a previous time-calibrated phylogeny. We constrained this split to the older 95% HPD from that analysis, which is ~400,000 years older than the mean divergence time of 1.133 Ma (Hosner et al., 2015)—suggesting that without this constraint, the divergence time estimates from our dataset could be considerably older than that of previous studies. Our choice of mutation rate and generation time could also have influenced these estimates, so we caution over-interpreting the precise dates obtained from demographic model runs. Ultimately, the equivocacy between the two models in the confidence set constitutes a difference of a few hundred years and resolving this exact difference may go beyond the degree of resolution we can expect from this type of molecular data.

The demographic model results unambiguously indicate that bobwhites first arrived in Cuba very recently, well after the earliest documented human settlement. The older end of our divergence time estimates suggests that bobwhites were already present in Cuba at the time of European settlement beginning in 1510 (Table 6); however, this timing does not preclude human involvement in the introduction of bobwhites to Cuba. The oldest archaeological evidence of human settlement on Cuba dates to 3100 BCE, with subsequent waves of migrants arriving from the Lesser Antilles during the following 2000 years (Allaire, 1999). The indigenous practice of moving wildlife, particularly animals used for food, has been well documented in the Caribbean (Newsom & Wing, 2004; Wing, 2001), including recent molecular and fossil evidence that Lucayan people shaped the distribution of endemic rodents (*Geocapromys ingrahami*) throughout the Bahamian archipelago (Oswald et al., 2020), although

the extent of this practice by indigenous peoples inhabiting Cuba is unclear. There is some speculation that the Taíno people of Cuba reached south Florida during the late 15th century prior to European arrival, and that they may have made contact with the Yucatán Peninsula, although archaeological evidence for these conclusions is lacking (reviewed in Chard, 1950). The older time frame we inferred for the original introduction of bobwhites to Cuba is consistent with the confusing scenario proposed by Gundlach: that bobwhites were already present on the island prior to Spanish colonization and that bobwhites were introduced near Havana by a Spanish colonel during the late 1700s (Gundlach, 1876).

The younger divergence time estimates from our demographic models place the introduction of bobwhites to Cuba shortly after European colonization. Cuba was one of the islands visited by Christopher Columbus in 1492, and the first Spanish settlers arrived during 1510, establishing permanent settlements in present-day Havana by 1515 (De las Casas, 1877; Monzote, 2009). Between 1575 and 1779, the main trans-Atlantic shipping artery used by Spanish fleets travelled between Veracruz and Havana before departing across the Atlantic (Figure 1), with as many as 100 ships making this journey each year by the end of the 16th century (reviewed in Lugo-Fernández et al., 2007). This history is remarkably consistent with our estimates, which indicated the introduction of bobwhites occurred during the mid-16th century (Figure 4, Table 6). The historical Veracruz-Havana connection also supports the signals of shared ancestry in our sNMF analyses between bobwhites in Cuba and the Caribbean slope of southern Mexico (Figure 2) and the previously reported sister relationship between Cuban bobwhites and a clade including Caribbean slope populations (Salter et al., 2022). These results are consistent with the hypothesis that bobwhites were introduced to Cuba from the southern Caribbean slope of Mexico prior to the 19th century (Parkes, 1990). Furthermore, the younger dates we estimated suggest that bobwhites from the southeastern USA arrived in Cuba ca. 1860, which is supported by the collection of Cuban bobwhites with Florida-like plumage during the following decades (Chapman, 1892).

Based on our results, we cannot reject either Gundlach's hypothesis that bobwhites were already present in Cuba prior to European arrival or Parkes' hypothesis that bobwhites were brought to Cuba from Mexico by the Spanish, in part because they are not mutually exclusive. The range of introduction dates estimated by our demographic analyses could support Gundlach's theory that there was already a population of bobwhites in Cuba when the Spanish began introducing them, while the younger dates are consistent with the hypothesis that the Spanish were wholly responsible for the introduction of bobwhites. Gundlach did not specify any sources of bobwhites introduced to Cuba, whereas Parkes was unequivocal that Cuban bobwhites were introduced from the Caribbean slope of Mexico. The same shipping vessels travelling between Veracruz and Havana also stopped in Florida (Lugo-Fernández et al., 2007), where the Spanish established colonies beginning in 1568, so a Floridian origin of Cuban bobwhites introduced by the Spanish is plausible within this timeframe. However, our results clearly support southern Mexico as the original source of the Cuban bobwhite population.

In the absence of explicit documentation that humans (either indigenous or Europeans) brought the first bobwhites to Cuba from southern Mexico, our data cannot exclude a scenario in which bobwhites arrived on Cuba through some kind of natural dispersal; however, following established criteria for distinguishing translocated versus naturally occurring populations of animals (Heinsohn, 2010), we contend that it is extremely unlikely, given both the dispersal capability of bobwhites and the evidence from our genetic data. Bobwhites are largely sedentary and rarely disperse over distances greater than a few kilometres (Smith et al., 1982; Stoddard, 1931). The shortest distances between Cuba and mainland North America are over 200km to either south Florida (~225 km) or the Yucatán Peninsula (~210km), and the distance between the Caribbean slope and Cuba is even greater (~1290km). Furthermore, many sedentary birds are averse to dispersal over water (Moore et al., 2008), making a natural dispersal scenario improbable. The only other odontophorid quails found on islands are Catalina California Quails (*Callipepla californica catalinensis*), which are considered endemic to Catalina Island off the coast of southern California; notably, this population is suspected to have been introduced by Paleolndians following their arrival to the island ~12 kya (Johnson, 1972; Zink et al., 1987). Finally, our estimate of a recent introduction is consistent with the lack of fossil evidence of bobwhites (or any galliformes) on Cuba during the Quaternary, despite a diverse avian fossil record (Orihuela, 2019; Zelenkov & Belichenko, 2022) and the identification of several bobwhite fossils from Florida (Holman, 1961). In summary, the population genetic and demographic evidence we report here, combined with historical records and observations, supports the conclusion that bobwhites arrived in Cuba through human-mediated introduction within the recent past.

## 5 | CONCLUSIONS

In light of our results, it is worth reconsidering Gould's taxonomic description of Cuban bobwhites (Gould, 1850). Gould's recognition of Cuban bobwhites as distinct from all other mainland populations of bobwhites (including *C.v. pectoralis* of Veracruz, which he also described) is largely supported by our data. Despite their recent arrival on the island, the combination of genetic drift, multiple introductions and small population size have given Cuban bobwhites a distinct allelic profile that is well differentiated from the mainland bobwhite populations we sampled (Figure 4, Tables S3–S4). The unique plumage phenotypes observed in Cuban bobwhites, which inspired Gould's taxonomic description, also suggest a compelling avenue of future study: as hybrids between two phenotypically divergent lineages, Cuban bobwhites offer a unique opportunity to study the genetic basis of plumage traits that have given rise to the remarkable phenotypic diversity within bobwhites, which have more subspecies (described by male plumage) than 99% of all other birds (Dickinson & Remsen Jr, 2013).

The combination of historical museum specimens and a wealth of historical records provided a unique opportunity to test demographic

hypotheses in this system, but our results have implications beyond bobwhites. Many animal distributions have been shaped by human activity over thousands of years, indirectly through landscape and habitat changes as well as through direct introduction and translocation of animals. Yet, in some cases, the extent of human involvement or our ability to detect human effects remains unclear. Our results demonstrate that genomic data from historical museum specimens are sufficiently powerful to infer complex population histories over very recent timescales, and similar approaches could be applied to study other, ambiguous animal distributions, particularly on islands.

## AUTHOR CONTRIBUTIONS

J.F.S., R.T.B. and B.C.F. designed the study. J.F.S. collected and analysed the data with input from R.T.B. and B.C.F. J.F.S., R.T.B. and B.C.F. wrote the paper. J.F.S., R.T.B. and B.C.F. contributed funds.

## ACKNOWLEDGEMENTS

We thank Angus Davison, Town Peterson and two anonymous reviewers for their comments which improved this manuscript. We also thank the following museums and collections staff that provided loans for this project, including samples that were used to design the custom RADcap baits: Delaware Museum of Natural History (DMNH—Jean Woods); University of Kansas Biodiversity Research Institute (KU—Mark Robbins); Natural History Museum of Los Angeles County (LACM—Kimball Garrett); Louisiana State University Museum of Natural Science (LSU—Donna Dittmann, Van Remsen, & Fred Sheldon); Harvard Museum of Comparative Zoology (MCZ—Jeremiah Trimble); Moore Laboratory of Zoology (MLZ—James Maley & John McCormack); Royal Ontario Museum (ROM—Santiago Claramunt); San Diego Museum of Natural History (SDNHM—Phil Unitt); University of Michigan Museum of Zoology (UMMZ—Janet Hinshaw); National Museum of Natural History (USNM—Christopher Milensky); University of Washington Burke Museum (UWBM—Sharon Birks); Western Foundation of Vertebrate Zoology (WFVZ—René Corado); and the Yale Peabody Museum (YPM—Kristof Zyskowski). We would additionally like to thank Christopher Milensky and the National Museum of Natural History for providing access to specimens for photography. Thanks also to Pete Hosner, who provided the calibration dates for the demographic analyses, and John Carroll, who directed us to Ken Parkes' work on Cuban bobwhites. Discussions with Anna Hiller, Glauca Del Rio and Jessica Oswald provided helpful feedback through the development of this project. Finally, we thank Sheila Rodriguez Machado and Zenaida Navarro Martínez for their help translating the manuscript to Spanish.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## FUNDING INFORMATION

This research was supported by funds from the American Museum of Natural History Chapman Award, the Society of Systematic Biology Graduate Student Research Award, and the Louisiana

State University Museum of Natural Science Ornithology Student Support Fund to J.F.S. J.F.S. was also supported by a Carrie Lynn Yoder Superior Graduate Student Scholarship from the Louisiana State University Department of Biological Sciences and NSF DBI-2109361. Startup funds from LSU to B.C.F. and funds from NSF DEB-1655624 to B.C.F. and R.T.B. also supported this work. Portions of this research were conducted with high-performance computational resources provided by Louisiana State University (<http://www.hpc.lsu.edu>). Funders had no role in creating or approving manuscript content.

#### DATA AVAILABILITY STATEMENT

Raw sequencing reads are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive as part of BioProject PRJNA875956. Custom computer code, DNA alignments, analysis inputs and analysis outputs are available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.tjq2bw31>).

#### BENEFITS SHARING

Benefits from these data and results are derived from the sharing of both on public databases, including the steps and computer code used in the analyses. Additionally, the data associated with specimens loaned from the institutions outlined in Table S1 will be made available to those institutions upon publication, so that they may be integrated to the appropriate specimen database(s). Finally, a Spanish translation of this manuscript is included in Dryad (<https://doi.org/10.5061/dryad.tjq2bw31>).

#### ORCID

Jessie F. Salter  <https://orcid.org/0000-0002-8975-5914>

Robb T. Brumfield  <https://orcid.org/0000-0003-2307-0688>

Brant C. Faircloth  <https://orcid.org/0000-0002-1943-0217>

#### REFERENCES

- Allaire, L. (1999). Archaeology of the Caribbean region. In F. Salomon & S. B. Schwartz (Eds.), *The Cambridge history of the native peoples of the Americas*. Cambridge University Press.
- Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, 5, 181–190.
- Austin, C. C. (1999). Lizards took express train to Polynesia. *Nature*, 397, 113–114.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, e3376.
- Barbour, T. (1923). *The birds of Cuba*. Memoirs of the Nuttall Ornithological Club.
- Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., Denham, T., & Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 6388–6396.
- Bond, J. (1948). Origin of the bird fauna of the West Indies. *The Wilson Bulletin*, 60, 207–229.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10, e1003537.
- Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, 26, 1372–1373.
- Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing trees through the Forest. *bioRxiv*, 11. <https://doi.org/10.1101/012401>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29, 1917–1932.
- Burnham, K. P., & Anderson, D. R. (2002). *A practical information-theoretic approach. Model Selection and Multimodel Inference*, 2.
- Chapman, F. M. (1892). *Notes on birds and mammals observed near Trinidad, Cuba: With remarks on the origin of west Indian bird-life*. Order of the Trustees, American Museum of Natural History.
- Chard, C. S. (1950). Pre-Columbian trade between north and South America. *Papers. Kroeber Anthropological Society*, 1, 1–27.
- Chattopadhyay, B., Garg, K. M., & Ramakrishnan, U. (2014). Effect of diversity and missing data on genetic assignment with RAD-Seq markers. *BMC Research Notes*, 7, 841.
- Chen, B., Cole, J. W., & Grond-Ginsbach, C. (2017). Departure from hardy Weinberg equilibrium and genotyping error. *Frontiers in Genetics*, 8, 167.
- Cohen, K. M., Finney, S. C., Gibbard, P. L., & Fan, J.-X. (2013). The ICS international chronostratigraphic chart. *Episodes*, 36, 199–204.
- Colleoni, F., Wekerle, C., Näslund, J.-O., Brandefelt, J., & Masina, S. (2016). Constraint on the penultimate glacial maximum northern hemisphere ice topography ( $\approx 140$  kyrs BP). *Quaternary Science Reviews*, 137, 97–112.
- Cory, C. B. (1885). *A list of the birds of the West Indies, including the Bahama Islands and the greater and Lesser Antilles, excepting the islands of Tobago and Trinidad*. Estes & Lauriat.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics*, 12, 499–510.
- De las Casas, B. (1877). *Historia de las Indias*. Imprenta y litografía de I. Paz.
- Dickinson, E. C., & Remsen, J. V., Jr. (Eds.). (2013). *The Howard & Moore Complete Checklist of the birds of the world. Aves Press*.
- D'Orbigny, A. (1839). Ornithologie. In R. de la Sagra (Ed.), pp. III–336 *Historia física, política y natural de la isla de Cuba*. Artus Bertrand.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, e88.
- Faircloth, B. C. (2021). *Private-alleles: Compute private alleles in one population relative to another*. Github.
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196, 973–983.
- Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S., Garcia-De Leon, F. J., Del Rio Portilla, M. A., Reed, K. D., Anderson, J. L., Meece, J. K., Aggrey, S. E., Rekaya, R., Alabady, M., Belanger, M., ... Faircloth, B. C. (2019). Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ*, 7, e7755.
- Gould, J. (1850). *A monograph of the Odontophorinæ, or, Partridges of America*. John Gould, F. R. S. Preface [a monograph of the Odontophorinæ, or, partridges of America].
- Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, 43, 1031–1034.
- Gundlach, J. C. (1876). *Contribucion á la ornitologia cubana. Imp. "La Antilla" de N. Cacho-Negrete*. Habana.

- Halley, Y. A., Dowd, S. E., Decker, J. E., Seabury, P. M., Bhattarai, E., Johnson, C. D., Rollins, D., Tizard, I. R., Brightsmith, D. J., Peterson, M. J., Taylor, J. F., & Seabury, C. M. (2014). A draft de novo genome assembly for the northern bobwhite (*Colinus virginianus*) reveals evidence for a rapid decline in effective population size beginning in the late Pleistocene. *PLoS ONE*, *9*, e90240.
- Heinicke, M. P., Diaz, L. M., & Hedges, S. B. (2011). Origin of invasive Florida frogs traced to Cuba. *Biology Letters*, *7*, 407–410.
- Heinsohn, T. (2003). Animal translocation: Long-term human influences on the vertebrate zoogeography of Australasia (natural dispersal versus ethnoploresy). *The Australian Zoologist*, *32*, 351–376.
- Heinsohn, T. E. (2010). Marsupials as introduced species: Long-term anthropogenic expansion of the marsupial frontier and its implications for zoogeographic interpretation. In S. G. Haberle, J. Stevenson, & M. Prebble (Eds.), *Altered Ecologies: Fire, Climate and Human Influence on Terrestrial Landscapes* (pp. 133–176). ANU Press. <http://www.jstor.org/stable/j.ctt24h8rj.11>
- Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: Sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, *16*, 1264–1278.
- Hofman, C. A., & Rick, T. C. (2018). Ancient biological invasions and Island ecosystems: Tracking translocations of wild plants and animals. *Journal of Archaeological Research*, *26*, 65–115.
- Holman, J. A. (1961). *Osteology of living and fossil New World quails (Aves, Galliformes)*. University of Florida.
- Hosner, P. A., Braun, E. L., & Kimball, R. T. (2015). Land connectivity changes and global cooling shaped the colonization history and diversification of New World quail (Aves: Galliformes: Odontophoridae). *Journal of Biogeography*, *42*, 1883–1895.
- Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.
- Hurvich, C. M., & Tsai, C.-L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis/A Journal Sponsored by the Bernoulli Society for Mathematical Statistics and Probability*, *14*, 271–279.
- Johnsgard, P. A. (1988). *Quails, partridges, and francolins of the world*. Oxford University Press.
- Johnson, N. K. (1972). Origin and differentiation of the avifauna of the Channel Islands, California. *The Condor*, *74*, 295–315.
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405.
- Kamm, J., Terhorst, J., Durbin, R., & Song, Y. S. (2020). Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association*, *115*, 1472–1487.
- Knaus, B. J., & Grünwald, N. J. (2017). Vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*, 44–53.
- Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, *431*, 177–181.
- Kolbe, J. J., Glor, R. E., Rodríguez Schettino, L., Lara, A. C., Larson, A., & Losos, J. B. (2007). Multiple sources, admixture, and genetic variation in introduced anolis lizard populations. *Conservation Biology: The Journal of the Society for Conservation Biology*, *21*, 1612–1625.
- Larson, G., & Burger, J. (2013). A population genetics view of animal domestication. *Trends in Genetics: TIG*, *29*, 197–205.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*, 639–647.
- Lugo-Fernández, A., Ball, D. A., Gravois, M., Horrell, C., & Irion, J. B. (2007). Analysis of the Gulf of Mexico's Veracruz-Havana route of La Flota de la Nueva España. *Journal of Maritime Archaeology*, *2*, 24–47.
- Marshall, C. J., & Liebherr, J. K. (2000). Cladistic biogeography of the Mexican transition zone. *Journal of Biogeography*, *27*, 203–216.
- Marth, G. T., Czabarka, E., Murvai, J., & Sherry, S. T. (2004). The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics*, *166*, 351–372.
- Mathys, B. A., & Lockwood, J. L. (2009). Rapid evolution of great kiskadees on Bermuda: An assessment of the ability of the Island rule to predict the direction of contemporary evolution in exotic vertebrates. *Journal of Biogeography*, *36*, 2204–2211.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*, 240–248.
- Monzote, R. F. (2009). *From rainforest to cane field in Cuba: An environmental history since 1492*. Univ of North Carolina Press.
- Moore, R. P., Robinson, W. D., Lovette, I. J., & Robinson, T. R. (2008). Experimental evidence for extreme dispersal limitation in tropical forest birds. *Ecology Letters*, *11*, 960–968.
- Morrone, J. J. (2010). Fundamental biogeographic patterns across the Mexican transition zone: An evolutionary approach. *Ecography*, *33*, 355–361.
- Nam, K., Mugal, C., Nabholz, B., Schielzeth, H., Wolf, J. B., Backström, N., Künstner, A., Balakrishnan, C. N., Heger, A., Ponting, C. P., Clayton, D. F., & Ellegren, H. (2010). Molecular evolution of genes in avian genomes. *Genome Biology*, *11*, R68.
- Newsom, L. A., & Wing, E. S. (2004). *On land and sea: Native American uses of biological resources in the West Indies*. University of Alabama Press.
- Orihuela, J. (2019). An annotated list of late quaternary extinct birds of Cuba. *Ornitologia Neotropical*, *30*, 57–67.
- Oswald, J. A., Allen, J. M., LeFebvre, M. J., Stucky, B. J., Folk, R. A., Albury, N. A., Morgan, G. S., Guralnick, R. P., & Steadman, D. W. (2020). Ancient DNA and high-resolution chronometry reveal a long-term human role in the historical diversity and biogeography of the Bahamian hutia. *Scientific Reports*, *10*, 1373.
- Parkes, K. C. (1990). *The origin of the Cuban bobwhite "Colinus virginianus cubanensis" (Gray)*. Annual Meeting of the Society of Caribbean Ornithology.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian Phylogenetics using tracer 1.7. *Systematic Biology*, *67*, 901–904.
- Ridgway, R. (1894). *Colinus virginianus cubanensis* not a Florida bird. *The Auk*, *11*, 324–325.
- Roullier, C., Benoit, L., McKey, D. B., & Lebot, V. (2013). Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 2205–2210.
- Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. CRC Press.
- Salter, J. F., & Faircloth, B. C. (2021). *Running GATK in parallel*.
- Salter, J. F., Hosner, P. A., Tsai, W. L. E., McCormack, J. E., Braun, E. L., Kimball, R. T., Brumfield, R. T., & Faircloth, B. C. (2022). Historical specimens and the limits of subspecies phylogenomics in the

- New World quails (Odontophoridae). *Molecular Phylogenetics and Evolution*, 175, 107559.
- Salter, J. F., Johnson, O., Stafford, N. J., 3rd, Herrin, W. F., Jr., Schilling, D., Cedotal, C., Brumfield, R. T., & Faircloth, B. C. (2019). A highly contiguous reference genome for northern bobwhite (*Colinus virginianus*). *G3 (Bethesda)*, 9, 3929–3932.
- Smith, G. F., Kellogg, F. E., Davidson, W. R., & Mack Martin, W. (1982). A 10-year study of bobwhite quail movement patterns. *National Quail Symposium Proceedings*, 2, 6.
- Stoddard, H. L. (1931). *The bobwhite quail*. C. Scribner's Sons.
- Summerhayes, G. R. (2007). Island Melanesian pasts: a view from archaeology. In J. S. Friedlaender (Ed.), *Population Genetics, Linguistics, and Culture History in the Southwest Pacific*, (Chap. 2, pp. 10–35). Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780195300307.003.0002>
- Tsai, W. L. E., & Salter, J. F. (2018). *Phenol chloroform extraction (for toe-pads)*. Moore Laboratory of Zoology, Occidental College; Louisiana State University.
- Tsutsui, N. D., Suarez, A. V., Holway, D. A., & Case, T. J. (2000). Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 5948–5953.
- Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: Using Docker, GATK, and WDL in Terra*.
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual: (Python Documentation Manual Part 2)*. CreateSpace Independent Publishing Platform.
- Williford, D., Deyoung, R. W., Honeycutt, R. L., Brennan, L. A., & Hernández, F. (2016). Phylogeography of the bobwhite (*Colinus*) quails. *Wildlife Monographs*, 193, 1–49.
- Wing, E. S. (2001). Native American use of animals in the Caribbean. In C. A. Woods & F. E. Sergile (Eds.), *Biogeography of the West Indies* (pp. 481–518). CRC Press.
- Yi, X., & Latch, E. K. (2021). Nonrandom missing data can bias PCA inference of population genetic structure. *Molecular Ecology Resources*, 22, 602–611.
- Young, H. S., Parker, I. M., Gilbert, G. S., Sofia Guerra, A., & Nunn, C. L. (2017). Introduced species, disease ecology, and biodiversity–disease relationships. *Trends in Ecology & Evolution*, 32, 41–54.
- Zeisset, I., & Beebee, T. J. (2001). Determination of biogeographical range: An application of molecular phylogeography to the European pool frog *Rana lessonae*. *Proceedings. Biological Sciences/the Royal Society*, 268, 933–938.
- Zelenkov, N. V., & Belichenko, E. S. (2022). Dynamics of the late quaternary avifauna of Western Cuba (based on material from El Abrón cave). *Doklady Biological Sciences: Proceedings of the Academy of Sciences of the USSR, Biological Sciences Sections/Translated from Russian*, 503, 54–57.
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328.
- Zink, R. M., Lott, D. F., & Anderson, D. W. (1987). Genetic variation, population structure, and evolution of California quail. *The Condor*, 89, 395–405.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Salter, J. F., Brumfield, R. T., & Faircloth, B. C. (2023). An island 'endemic' born out of hybridization between introduced lineages. *Molecular Ecology*, 00, 1–15. <https://doi.org/10.1111/mec.16990>