Received: 22 June 2018 | Revised: 6 February 2019 | Accepted: 12 February 2019

DOI: 10.1111/mec.15049

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Divergent mitochondrial lineages arose within a large, panmictic population of the Savannah sparrow (*Passerculus sandwichensis*)

Phred M. Benham | **Zachary A. Cheviron**

Division of Biological Sciences, University of Montana, Missoula, Montana

Correspondence
Phred M. Benham, Division of Biological Sciences, University of Montana, Missoula, MT.
Email: phbenham@berkeley.edu

Present Address
Phred M. Benham, Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, California

Funding information
University of Montana, Grant/Award Number: internal funds; Sigma Xi, Grant/Award Number: Grant-in-aid-of-research; University of Illinois, Grant/Award Number: various; American Museum of Natural History, Grant/Award Number: Frank M. Chapman Memorial Fund; Systematics Association, Grant/Award Number: Systematic Research Fund; Society for the Study of Evolution, Grant/Award Number: Rosemary Grant Award; Illinois Ornithological Society, Grant/Award Number: Student research award; Society for Integrative and Comparative Biology, Grant/Award Number: Grant-in-aid-of-research; American Ornithologists' Union, Grant/Award Number: student research award

Abstract

Unusual patterns of mtDNA diversity can reveal interesting aspects of a species' biology. However, making such inferences requires discerning among the many alternative scenarios that could underlie any given mtDNA pattern. Next-generation sequencing methods provide large, multilocus data sets with increased power to resolve unusual mtDNA patterns. A mtDNA-based phylogeography of the Savannah sparrow (*Passerculus sandwichensis*) previously identified two sympatric, but divergent (~2%) clades within the nominate subspecies group and a third clade that consisted of birds sampled from northwest Mexico. We revisited the phylogeography of this species using a population genomic data set to resolve the processes leading to the evolution of sympatric and divergent mtDNA lineages. We identified two genetic clusters in the genomic data set corresponding to (a) the nominate subspecies group and (b) northwestern Mexico birds. Following divergence, the nominate clade maintained a large, stable population, indicating that divergent mitochondrial lineages arose within a panmictic population. Simulations based on parameter estimates from this model further confirmed that this demographic history could produce observed levels of mtDNA diversity. Patterns of divergent, sympatric mtDNA lineages are frequently interpreted as admixture of historically isolated lineages. Our analyses reject this interpretation for Savannah sparrows and underscore the need for genomic data sets to resolve the evolutionary mechanisms behind anomalous, locus-specific patterns.

KEYWORDS
genotyping-by-sequencing, mitochondrial DNA, *Passerculus*, Passerellidae, phylogeography

1 | INTRODUCTION

Unusual phylogeographic patterns present rare opportunities to investigate the impact of certain ecological and evolutionary processes on population differentiation, yet to extract meaningful insights from these unusual patterns requires accurate inference of the demographic history that gave rise to these patterns. The strong reliance on mtDNA variation in phylogeographic studies has uncovered a litany of unusual mtDNA patterns. Prominent examples include mtDNA haplotype sharing among divergent species (e.g., Andersson, 1999; Weckstein, Zink, Blackwell-Rago, & Nelson, 2001; Beckman & Witt, 2015; Good, Vanderpool, Keeble, & Bi, 2015; Sloan, Havird, & Sharbrough, 2016), and deeply divergent, but sympatric mtDNA lineages within a single species (e.g., Giska, Sechi, & Babik, 2015; Hogner et al., 2012; Quinn, 1992; Spottiswoode, Stryjewski, Quader, Colebrook-Robjent, & Sorenson, 2011; Xiao et al., 2012). Unfortunately, in most cases, data are not available to tease apart the different demographic scenarios that could lead to the evolution

of observed mtDNA patterns, and as a result, inference of species history has relied heavily on speculative interpretations of these unusual mtDNA patterns. As genomic data sets become the norm for phylogeographic inference, new opportunities are arising to revisit these unexpected phylogeographic patterns. These opportunities will enable rigorous tests of competing hypotheses and a deeper understanding of the evolutionary mechanisms that generally underlie discordant phylogeographic patterns.

One such discordant pattern is the presence of broadly sympatric, but deeply divergent (>2% sequence divergence) mtDNA clades within an otherwise panmictic species. Although this phylogeographic pattern is relatively rare in animal studies, it has been found across many different classes, including birds (e.g., Block, Goodman, Hackett, Bates, & Raherilalao, 2015; Hogner et al., 2012; Quinn, 1992; Spottiswoode et al., 2011), mammals (Hoelzer, Dittus, Ashley, & Melnick, 1994) and invertebrates (Giska et al., 2015; Xiao et al., 2012). Traditionally, this pattern has been submitted as evidence for the merger of historically isolated lineages (e.g., Avise, 2000; Hogner et al., 2012; Quinn, 1992); an interpretation based on theoretical considerations given the primarily uniparental inheritance, severely reduced recombination and relatively small effective population sizes (on average 4× smaller than the nuclear genome) characterizing the mtDNA genome. These characteristics result in rapid fixation of divergent haplotypes in isolation, with divergent haplotypes potentially persisting upon secondary contact despite the erosion of genetic structure across the nuclear genome (Block et al., 2015). A recent and well-supported example can be found in a Madagascar bird species, the spectacled tetraka (*Xanthomixis zosterops*). Although this species does not exhibit any evidence of population structure in the nuclear genome, it harbours several divergent and sympatric mitochondrial haplotypes (~5% sequence divergence) (Block et al., 2015). Importantly, the authors also found evidence of associations between mtDNA haplogroups and lineages of host-specific chewing lice, corroborating the hypothesis that this species experienced a history of isolation followed by recent merger. Although this represents one empirical example supporting a history of recent admixture between historically isolated mtDNA lineages, many studies lack sufficient corroborating evidence to exclude other potential explanations (e.g., Giska et al., 2015; Hogner et al., 2012; Quinn, 1992).

Historical isolation is not always necessary to generate divergent mtDNA lineages within a population. First, the reduced recombination rate of the mitochondrial genome can result in presence of multiple mtDNA haplotypes that arose within a large panmictic population and transiently persist through stochastic lineage sorting processes (Hudson & Turelli, 2003; Slatkin & Hudson, 1991). Second, the thirteen protein-coding genes found in the animal mitochondrial genome encode subunits of key proteins within the electron transport chain, which may be a frequent target of natural selection (Ballard & Whitlock, 2004). Experimental evolution studies have shown that selection on mitochondrial function can lead to sympatric, but divergent mtDNA haplotypes through processes such as negative frequency-dependent selection (e.g., Kazancioğlu & Arnqvist, 2014). Third, uniparental inheritance of the mitochondrial genome can also lead to discordant mitochondrial and autosomal genealogies in cases where the sexes differ in life history attributes (e.g., Spottiswoode et al., 2011). Finally, the mitochondrial genome has been found to be highly susceptible to introgression among populations relative to elements of the nuclear genome, resulting in discordance between mitochondrial and nuclear structure (Bachtrog, Thornton, Clark, & Andolfatto, 2006; Currat, Ruedi, Petit, & Excoffier, 2008; Good et al., 2015; Mims, Hulsey, Fitzpatrick, & Streelman, 2010; Sloan et al., 2016). Independent data sets to discern among these alternative hypotheses are lacking in most instances where sympatric, divergent mtDNA patterns have been documented. Demographic analysis of population genomic data sets thus presents a promising opportunity to explicitly test these competing hypotheses.

The Savannah sparrow (*Passerculus sandwichensis*) is one example where previous mtDNA phylogeographic analyses uncovered a pattern of divergent and sympatric mitochondrial lineages, but the initial study lacked any corroborating evidence to understand how this pattern arose (Zink et al., 2005). The Savannah sparrow is one of the most widespread North American songbirds with populations found in a range of environments spanning tundra, prairie, meadow and salt marsh habitats (Wheelwright & Rising, 2008). The 18 subspecies currently recognized (Dickinson & Christidis, 2014) are often classified into several subspecies groups that have been treated as distinct species by some authors (e.g., Rising, 2017). The subspecies groups include: (a) the largely migratory nominate group found across most of the breeding range; (b) the *beldingi* group that is resident along the Pacific coast of southern California and the Baja peninsula; (c) the resident *rostratus* group of coastal Sonora, Mexico; and (d) *P. s. sanctorum* of Isla San Benito off the west coast of Baja California (Figure 1). Despite this ecological and taxonomic diversity, the previous mtDNA survey found little evidence for significant geographic structure across the species' distribution; however, the nominate subspecies group contained two broadly sympatric mtDNA clades (A & B) that exhibited approximately 2% sequence divergence (Figure 2; Zink et al., 2005). A third clade, sister to one of the nominate clades (B), included salt marsh residents from the *beldingi*, *rostratus* and *sanctorum* subspecies groups of northwestern Mexico. The authors did not find evidence for selection shaping patterns of haplotype diversity within either of the two clades based on MacDonald–Kreitman tests, and it is unlikely that Savannah sparrows exhibit sex-based differences in life history (Wheelwright & Rising, 2008). This leaves three potential explanations for the observed sympatry of mtDNA clades A and B in the nominate Savannah sparrows: (a) the introgression of mitochondrial haplotypes exceeds the rate of nuclear introgression between two historically isolated populations; (b) the two independently sorting haplotype groups arose in a large, panmictic population; and (c) complete admixture of two historically isolated populations gave rise to the existence of divergent sympatric mtDNA clades. Hypotheses one and three are distinguished by the presence or absence of differential introgression of mitochondrial and nuclear loci. Hypothesis two differs from the others by not invoking a period of historical isolation.

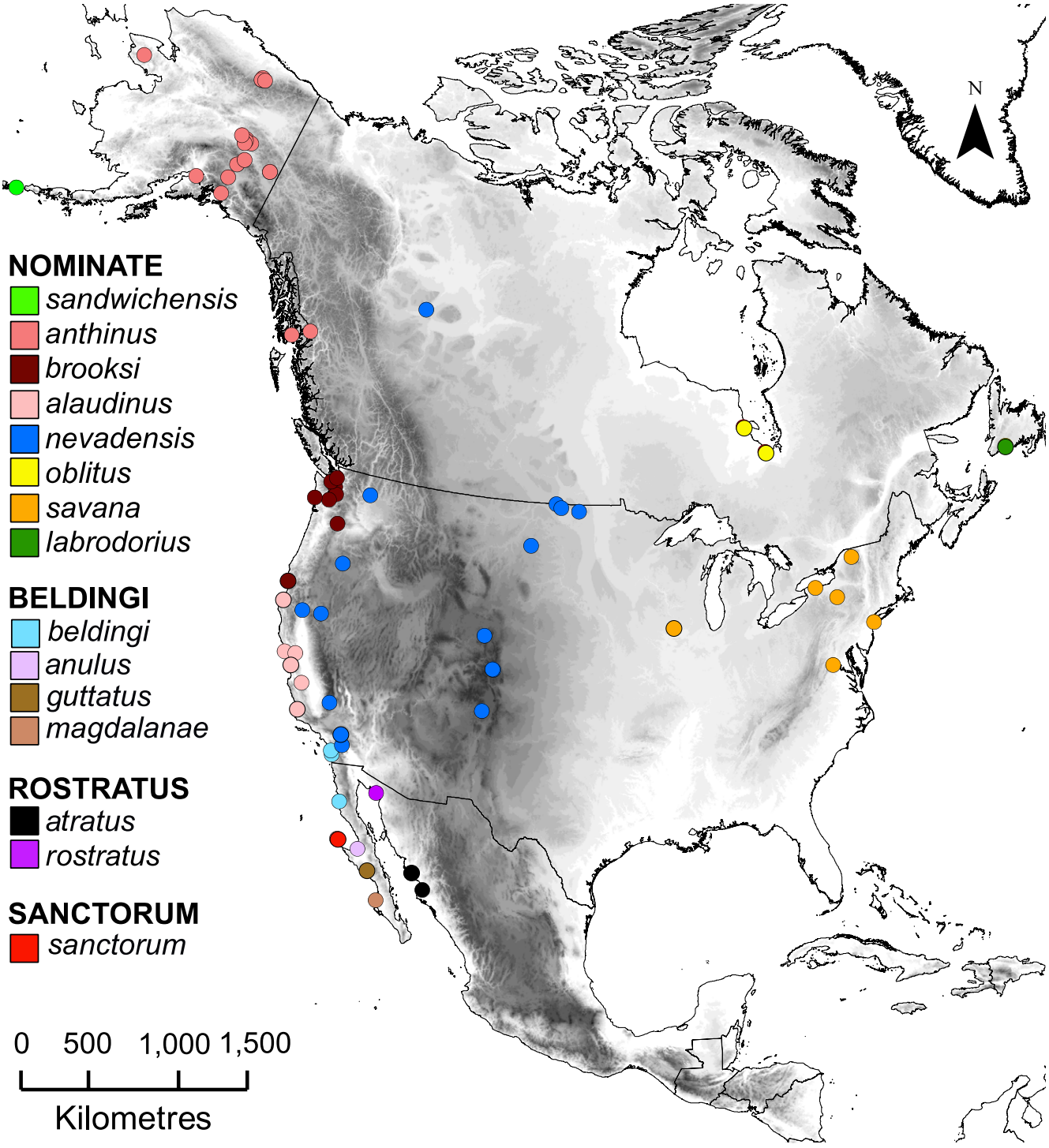

FIGURE 1 Savannah sparrow sampling map for this study. Each circle represents the geographic locality for all samples coloured by putative subspecies. The subspecies are clustered into four frequently recognized subspecies groups (Wheelwright & Rising, 2008) [Colour figure can be viewed at wileyonlinelibrary.com]

To tease apart these three hypotheses, we generated a large data set of SNPs distributed across the nuclear genome from individuals that were sampled widely across the geographic range of the Savannah sparrows. We analysed this data set using several methods to assess patterns of geographic structure, and infer the historical demography of Savannah sparrows. To distinguish the three competing hypotheses, we made the following predictions. The first hypothesis of extensive, bidirectional mtDNA introgression with reduced nuclear introgression would be supported by evidence of nuclear population structure and lower levels of historical introgression in nuclear loci among populations within the nominate subspecies group relative to mtDNA introgression.

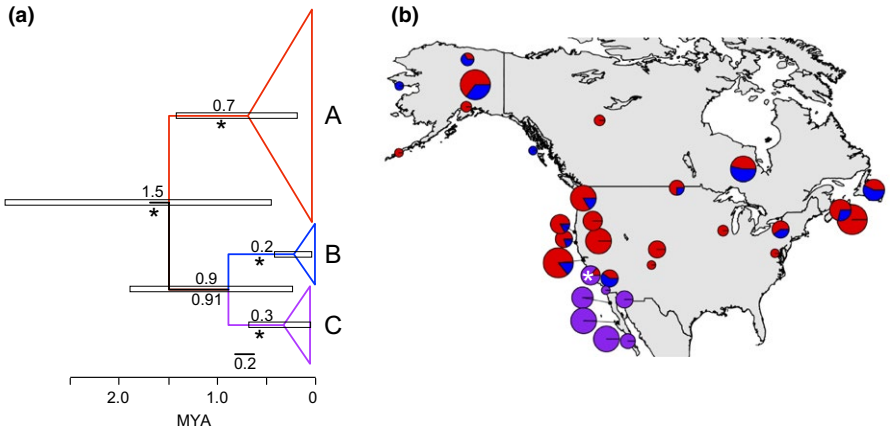

FIGURE 2 Genetic structure within Savannah sparrows based on the mitochondrial locus ND2 across 212 individuals. (a) Bayesian, time-calibrated phylogeny estimated within the program BEAST. The outgroup (Henslow's sparrow) has been removed for visualization. Mean divergence time estimates for each node are denoted above branches with purple error bars representing 95% HPD. Posterior probabilities (PP) for each node are below branches, asterisk signifying PP > 0.95. (b) Mapped distribution of the three ND2 clades recovered by phylogenetic analyses. Red and blue clades correspond to the nominate subspecies group (Figure 1) and are broadly sympatric across the breeding range of the nominate group despite being ~1.3 ma divergent. The purple clade corresponds to salt marsh resident populations in southern California and northwest Mexico. Morro Bay population (subspecies *P. s. alaudinus*) is marked with an asterisk and includes individuals with northwestern Mexico and nominate haplotypes [Colour figure can be viewed at wileyonlinelibrary.com]

Conversely, a lack of population structure in the nuclear genome would either point to the second hypothesis of mtDNA haplotypes arising within a large, panmictic population, or to the third hypothesis of complete admixture of historically isolated lineages. To differentiate the second and third hypotheses, we fit variations on these two demographic models to the site frequency spectrum (SFS) and used population genetic parameter estimates from the best-fit models to parameterize coalescent simulations. The output of these simulations was then used to determine which models could give rise to the observed patterns of mtDNA diversity. This study represents the first attempt to understand the evolution of divergent, but sympatric mtDNA lineages using a genomic data set. Our approach should be broadly applicable and provide deeper insights into the evolution of other unusual locus-specific patterns.

2 | METHODS

2.1 | Sampling

We obtained tissue or blood samples from 191 individual Savannah sparrows, including 15 of the 18 currently recognized subspecies (Dickinson & Christidis, 2014; Wheelwright & Rising, 2008; Supporting Information Table S1). We selected four outgroups based on a recent phylogeny of the family Passerellidae (Klicka et al., 2014): Henslow's sparrow (*Centronyx henslowii*), Le Conte's sparrow (*Ammospiza leconteii*), song sparrow (*Melospiza melodia*) and vesper sparrow (*Pooecetes gramineus*). For all samples, we extracted whole genomic DNA from muscle tissue or whole blood using a Qiagen DNeasy Blood & Tissue extraction kit following the manufacturer's protocols (Valencia, CA). Samples obtained from our fieldwork were collected using methods approved by the University of Illinois, Urbana-Champaign IACUC (protocol #: 13418).

2.2 | mtDNA sequencing and analysis

To more broadly assess the geographic distribution of the two previously identified mtDNA haplogroups, we sequenced the mtDNA gene NADH dehydrogenase subunit 2 (*ND2*) from 102 individuals of the nominate subspecies that were not previously included in any study (see Supporting Information Methods for sequencing details). We also obtained *ND2* sequence data from 110 individuals previously sequenced by Zink et al. (2005) and accessioned in GenBank (AY584869.1–AY584980.1) and *Centronyx henslowii* (AY584982.1) served as an outgroup. Sequences were manually assembled and edited using GENEIOUS version 10.0.6 (Biomatters, Auckland, NZ) and aligned using MUSCLE version 3.7 (Edgar, 2004). During assembly of mtDNA sequences, we did not detect the presence of double peaks, frame shifts or any stop codons in the middle of *ND2* sequences, suggesting that our data are likely free of numts (Bensasson, Zhang, Hartl, & Hewitt, 2001).

We used BEAST version 1.8.4 (Drummond, Suchard, Xie, & Rambaut, 2012) to estimate phylogenetic structure and divergence times within the *ND2* data set. The TN93 model of substitution and partitioning by codon position was identified as the best partitioning scheme, using PARTITIONFINDER2 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2016). We generated input files using BEAUTI version 1.8.4 (Drummond et al., 2012) with a relaxed lognormal clock model (Drummond, Ho, Phillips, & Rambaut, 2006), a mean substitution rate of 0.0105/site/lineage/million years for mtDNA (Weir & Schluter, 2008), and a standard deviation of 0.4 (Walstrom, Klicka, & Spellman, 2011). We ran two independent MCMC chains for 200 million generations each, sampling every 20,000 steps on the CIPRES science portal (Miller, Pfeiffer, & Schwartz, 2010). We ensured priors all had an ESS >200 using TRACER version 1.6.0 (Rambaut, Suchard, Xie, & Drummond, 2014). Trees were merged in LOGCOMBINER version 1.8.4 (Drummond et al., 2012) and 10% of trees were removed in TREEANNOTATOR version 1.8.4 (Drummond et al., 2012).

2.3 | Genotyping-by-sequencing methods

Genotyping-by-sequencing libraries were prepared following Parchman et al. (2012). Briefly, we digested whole genomic DNA with two restriction enzymes (*Eco*RI and *Mse*1), ligated adaptor sequences with unique barcodes for each individual and performed a PCR amplification. We then selected 500–600 bp fragments via gel extraction and purified extracts using a Qiagen Gel Extraction kit (Valencia, CA). Libraries were submitted to the W. M. Keck Sequencing Center at the University of Illinois, Urbana-Champaign, for sequencing on an Illumina HiSeq 2500 platform. Individuals were sequenced on three separate flow-cell lanes, each resulting in over 200 million, 100-nt single-end reads with quality scores greater than 30 and a mean of 1,794,006 reads per individual.

Reads were demultiplexed and barcodes removed using process_radtags in STACKS (Catchen et al., 2011). This resulted in final reads 89 bp in length, which were assembled de novo using the STACKS pipeline (Catchen et al., 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). 89 bp reads that have been aligned and assembled into overlapping "stacks" are referred to as loci for the rest of the paper. STACKS analysis was performed with the following parameter settings: m = 3, M = 3 and n = 3 (see Supporting Information Methods for parameter optimization details). Once within individual reads are assembled into loci, the STACKS pipeline then matches loci across individuals to assemble a catalog of loci that are shared across all individuals (Catchen et al., 2011). Given the computational costs of building a catalog with 195 individuals, we produced a catalog using 73 individuals that included ~30% of individuals from each sampling locality plus the four outgroup species. This design should ensure that we include all common variants in the catalog. Once the catalog was assembled, loci from all 195 individuals were matched against the catalog in sstacks and genotyped at this panel of loci. Finally, a number of output files were generated for downstream analyses using the populations module of STACKS. For all population structure analyses, four output files were generated using a single, randomly sampled SNP per locus (to avoid linkage among SNPs) and the minimum proportion of individuals genotyped at each locus (–r parameter) was set to 0.75, 0.8, 0.9 or 0.95. These different output files are referred to as our 25%, 15%, 10% and 5% missing data sets, respectively. The four categories represent arbitrary levels of missing data selected to determine whether patterns of population structure were robust to different levels of missing data. These four data sets included 13,058 to 200,791 unlinked SNPs (including outgroups) depending on the level of allowed missing data (5%–25%), and without outgroups, the data sets varied from 6,709 to 66,705 unlinked SNPs (Supporting Information Table S2).

2.4 | Population structure analyses

We evaluated range-wide population structure using several complimentary methods. First, we performed principal components analysis on the SNP data set for all 191 Savannah sparrows using the function glPca in the r-package ADEGENET 1.3-1 (Jombart & Ahmed, 2011). To assess the influence of missing data on population structure, we performed this analysis with 5%, 10%, 15% and 25% missing data. We also divided the full data set into northwestern Mexico (n = 56) and nominate (n = 135) groups and performed similar PCAs in Adegenet with these subsets to test for more subtle structure within these groups. Second, we explored genetic structure patterns in FASTSTRUCTURE (Raj, Stephens, & Pritchard, 2014) with different settings for K (1–10) on the full, northwestern Mexico, and nominate data sets for each level of missing data (5%, 10%, 15% and 25%). We also explored how setting a simple or logistic allele frequency prior would impact our results as the logistic prior may perform better in cases with subtle structure (Raj et al., 2014). We used the chooseK.py tool within FASTSTRUCTURE to determine the best-supported number of clusters (K) for each data set. Finally, we performed phylogenetic analyses using RAXML version 8.2.4 (Stamatakis, 2014) on the CIPRES webportal (Miller et al., 2010). Again, these phylogenetic analyses were performed on SNP data sets with 5%, 10%, 15% and 25% missing data. We accounted for biases due to using only variant sites with the Felsenstein correction (Leaché, Banbury, Felsenstein, Oca, & Stamatakis, 2015), which corrects the likelihood for the number of invariant sites present in the data (read length*number of loci − number of SNPs). All analyses were run using an ASC_GTRCAT model with 100 rapid bootstrap replicates to assess topological support.

2.5 | Demographic analysis

A lack of nuclear structure could point to either our second hypothesis of mtDNA clades arising within a large, panmictic population or our third hypothesis of complete admixture of historically isolated lineages. To distinguish between these two latter hypotheses, we took a two-pronged approach. First, we fit a range of demographic models that encompass these two scenarios to the SFS of our SNP data in ∂a∂i version 1.7 (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). Second, we performed a series of coalescent simulations in the program MS (Hudson, 2002) to determine whether parameter estimates from the best-fit models in ∂a∂i could generate observed levels of mtDNA diversity. We repeated these two steps for both the SFS of the nominate population only and the joint SFS of the northwestern Mexico and nominate populations. The latter analysis was performed to account for any gene flow between northwestern Mexico and nominate populations, which if unaccounted for could bias the single-population analyses.

The program ∂a∂i does not handle missing data (Gutenkunst et al., 2009). We excluded all missing data by generating a SFS for our data set based on a restricted number of individuals (northwestern Mexico = 16; nominate = 25) to maximize the number of loci and SNPs that were shared across all individuals (see Supporting Information Methods). The resulting SFS included individuals from across the distribution (i.e., five of six northwestern Mexico subspecies and California, Alaska, Colorado, Washington, New Mexico, North Dakota and Ontario). The number of individuals included should be sufficient to accurately infer a range of demographic scenarios (Robinson, Coffman, Hickerson, & Gutenkunst, 2014). This

41-individual data set was further filtered to remove reads aligning to the z-chromosome, due to its smaller effective population size, and to only include loci with read depth ≥10 that were in Hardy–Weinberg equilibrium (α set to 0.05), to minimize the influence of sequencing errors (see Supporting Information Figure S3). This resulted in a final data set with no missing data of 24,382 SNPs from 8,614 loci. ∂a∂i input files were created using the vcf2dadi function in the R package STACKR version 0.2.6 (Gosselin & Bernatchez, 2016). Analyses were based on a folded SFS given high levels of missing data in outgroups sequenced for phylogenetic analyses, which makes them unsuitable for polarizing the SFS.

2.5.1 | Analysis of the SFS

We fit the following demographic models to the SFS of the nominate group (Supporting Information Figure S4) that would be consistent with our second hypothesis (history of panmixia): a constant population size, exponential growth, a bottleneck and bottleneck followed by exponential growth (Supporting Information Figure S5). We also fit models related to our third hypothesis where a history of isolation was followed by admixture within the nominate group using two different approaches. First, we modelled an initial population split, allowed populations to evolve in isolation, the two populations then merged via admixture, and finally, following admixture one of the nominate populations is removed (the "ghost" population). Second, we divided the SFS into two populations based on their mitochondrial haplotype (A or B; see Figure 2) and fit the SFS to an identical admixture model as above. These two separate approaches to model historical isolation followed by admixture were necessary as mitochondrial and nuclear genomes are unlinked making it difficult to assume that mtDNA divergence will also be reflected in the nuclear genome following secondary contact. Modelling a "ghost" lineage allows for introgression to occur from a second lineage without requiring an explicit assignment of certain individuals to either mtDNA lineage. We next ran a similar series of models in ∂a∂i on the joint SFS of the nominate and northwestern Mexico Savannah sparrows (Supporting Information Figure S6). These models all involve a split between these two populations and model the various demographic scenarios with and without gene flow between populations. As for the single-population analyses, we ran a series of models consistent with a history of panmixia in the nominate population (second hypothesis) or secondary contact and admixture (third hypothesis).

For all models, 10 optimizations were run from different starting parameters using the perturb function in ∂a∂i with max number of iterations set to 10. To ensure that a global optimum for a given model had been reached, we ran one final optimization with 50 iterations using the parameter values estimated from the shorter run with the highest likelihood. We calculated demographic parameter values from the estimated value of theta (4NeμL; L is sequence length) based on a 1-year generation time for Savannah sparrows (Wheelwright & Rising, 2008), the average mutation rate for Passeriformes: 3.3×10^{-9} substitutions/site/year (Zhang et al., 2014), and total sequence length equal to 766,646 bp. We calculated uncertainty for parameter estimates using a nonparametric bootstrapping approach: sampling with replacement over the 8,614 loci, generating frequency spectra from 100 resampled SNP data sets, and using these spectra to calculate parameter uncertainties using the Godambe information matrix (GIM) in ∂a∂i (Coffman, Hsieh, Gravel, & Gutenkunst, 2016). We used an Akaike information theoretical approach to identify the top demographic models. ∂a∂i uses composite likelihoods to estimate model parameters, which assumes independence among SNPs in the SFS. Given the presence of linked sites in our data set (i.e., mean of 2.8 and median 2 SNPs per 89 bp locus), AIC analyses could be associated with increased error and bias towards more complex models and this analysis should be interpreted with caution (Coffman et al., 2016). However, other studies apply AIC methods to composite likelihoods as a first approximation in ranking models (e.g., Meier et al., 2017; Oswald, Overcast, Mauck, Andersen, & Smith, 2017). In cases where the model with the highest likelihood included more parameters than hierarchically nested models with lower likelihood values, we performed a likelihood ratio test with a GIM-based adjustment to account for the use of composite likelihoods (Coffman et al., 2016).

2.5.2 | Linking demographic parameters from nuclear genome to patterns of mtDNA diversity

Assuming that the mitochondrial genome experienced population demographic history similarly to the nuclear genome, we assessed whether the demographic histories estimated using ∂a∂i analyses could result in the observed patterns of mtDNA diversity. To do this, we used point estimates from all single-population ∂a∂i models to parameterize coalescent simulations of mtDNA diversity in the program MS (Hudson, 2002). There was no significant relationship between geographic distance and genetic distance $[F_{st}/(1 - F_{st})]$ in the nominate clade for neither the mtDNA (Mantel statistic r: −0.0745; p-value: 0.84) nor the SNP data set (Mantel statistic r: 0.098; p-value: 0.248). Consequently, we did not set up spatially explicit simulations. We compared estimates of segregating sites (S) and average number of pairwise differences (π) from simulated and observed mtDNA. S is proportional to total tree length and π equals the average lineage length between each individual sequence (Wakely, 2009). This makes these appropriate summary statistics for testing whether the demographic history estimated from nuclear genomic data can produce the observed pattern of divergent, sympatric mtDNA haplogroups. The empirical mtDNA data (see *mtDNA sequencing methods*) were used to calculate nucleotide diversity, π, and S for nominate populations in the program ARLEQUIN 3.5 (Excoffier & Lischer, 2010). A mtDNA substitution rate (μ) of 1.1×10^{-8}/site/generation (Weir & Schluter, 2008) was used for calculation of effective population size (Ne) of *ND2* and theta (Ne*μ*locus length) where generation time was assumed to be 1 year (Wheelwright & Rising, 2008). mtDNA Ne is predicted to be ¼ nuclear Ne; however, selection and other processes can lead to significant deviations from this prediction (Hudson & Turelli, 2003; Lynch, 2007). Rather than assume that mtDNA Ne was equal to 1/4 nuclear Ne, we directly

TABLE 1 The four demographic models and parameters evaluated in the coalescent simulator *ms*

	Ne				Time (Kya)				
Models	Ne ($\times 10^3$)	Ancestor	Bottle	Isolation	Bottle	Bottle Length	Isolation	Admix	*f*
Panmixia Constant Ne	100–2,000 (by 50)	0.1, 0.25, 0.5, 0.75, 1	NA	NA	NA	NA	NA	NA	NA
Panmixia Bottleneck	100–2,000 (by 50)	0.1, 0.25, 0.5, 0.75	0.1, 0.25, 0.5, 0.75	NA	1, 10, 50, 100	10, 50, 100, 200	NA	NA	NA
Admixture Stable Ne	100–2,000 (by 50)	0.1, 0.25, 0.5, 0.75	NA	0.1, 0.25, 0.5, 0.75	NA	NA	10, 50, 100, 200	1, 10, 50, 100	0.5
Admixture Bottleneck	100–2,000 (by 50)	0.1, 0.25, 0.5, 0.75	0.1, 0.25, 0.5	0.75	11, 20, 60, 110	50	70, 110, 160, 260	1, 10, 50, 100	0.5

Note. Ne is effective population size, and Ne for the ancestor, bottleneck population or during isolation prior to admixture is listed as a proportion of contemporary Ne. *f* is the admixture proportion.

estimated a range of Ne for *ND2* based on upper and lower bounds (±1 *SD* of the mean) of nucleotide diversity (0.011 ± 0.006) divided by μ (1.1×10^{-8}/site/generation). This gave bounds on mtDNA Ne from 500,000 to 1,650,000. This estimated range of mtDNA Ne encompasses the predicted estimate that would be calculated by dividing the nuclear Ne estimate from ∂a∂i by four. This suggests that our results are not biased by how we parameterized mtDNA Ne and that the observed mtDNA Ne fits within neutral expectations. We ran a model for each increment of 50,000 between these Ne bounds for a total of 24 different Ne models. For all models, π and *S* were calculated using the sample_stats function within *ms* across 100,000 simulations and the 99th, 95th, 75th, 25th, 5th and 1st percentiles for the distribution of π and *S* were determined in R. We considered a given model to likely explain patterns of mitochondrial diversity if observed values of π and *S* fell between the 99th and 1st percentiles of the simulated distribution.

2.6 | The influence of demographic parameter variation on panmixia vs. admixture models

Finally, we assessed whether there are certain areas of parameter space where divergent, sympatric mtDNA haplogroups may be more likely to have arisen under a history of isolation followed by admixture as opposed to within a panmictic population. We performed a series of simulations within *ms* where we modelled: (a) a panmictic population with constant Ne; (b) a panmictic population that experiences a bottleneck; (c) two populations that evolved in isolation then merged via admixture; and (d) an admixture model with a bottleneck occurring during isolation. In each of these models, we varied the ancestral Ne, bottleneck Ne, Ne in isolation and the timing of these events (see Table 1 for parameter values). We ran simulations for each combination of parameter values across a range of Ne from 100,000 to 2,000,000 by increments of 50,000. The number of individuals was set to 163 for all simulations (the same as 163 haploid mitochondrial chromosomes), the number of individuals for which observed mtDNA diversity statistics were calculated in the nominate clade. We ran 780 different parameter combinations for the constant population size model; 9,844 different parameter combinations for both the bottleneck and admixture models; and 7,488 combinations for the admixture with bottleneck model. For all of these different models, we ran 25,000 simulations and calculated π, *S* and their percentiles as above. Based on these simulations, we calculated the 99th percentile Ne, which is the lowest simulated Ne where the observed value of π or *S* falls within the 99th and 1st percentiles of the 25,000 simulations. We plotted variation in this measure to assess the influence of the various time and Ne parameters on mtDNA diversity.

3 | RESULTS

3.1 | mtDNA variation

The time-calibrated mtDNA phylogeny estimated in BEAST showed the nominate subspecies group was represented by two clades, A & B, with clade B sister to the northwestern Mexico clade, C (Figure 2a). All nodes were well supported with high posterior probabilities >0.9. Estimated mean divergence times between clade A and clades B & C was estimated to be 1.5 Ma (95% HPD: 0.5–3.2); and the divergence between clades B and C at 0.9 Ma (95% HPD: 0.3–1.9). We found that clades A & B were broadly sympatric across the breeding range of the nominate subspecies group with the A haplogroup occurring at a higher frequency through much of the range (Figure 2b). The population sampled at Morro Bay, CA, mostly consisted of individuals with the clade C haplotype and one individual with the clade A haplotype.

3.2 | Population structure in the nuclear genome

The first principal component (19.7% of variance) of the full data set revealed three clusters: (a) the nominate Savannah sparrows found across much of North America; (b) populations from northwest Mexico; and (c) an intermediate population from Morro Bay, CA, near the distribution limits of the first two populations. A fourth cluster, corresponding to the subspecies *P. s. sanctorum* restricted to Isla San Benito, separated along the second component axis (1.7% of variance; Figure 3a). Performing PCA on only individuals from

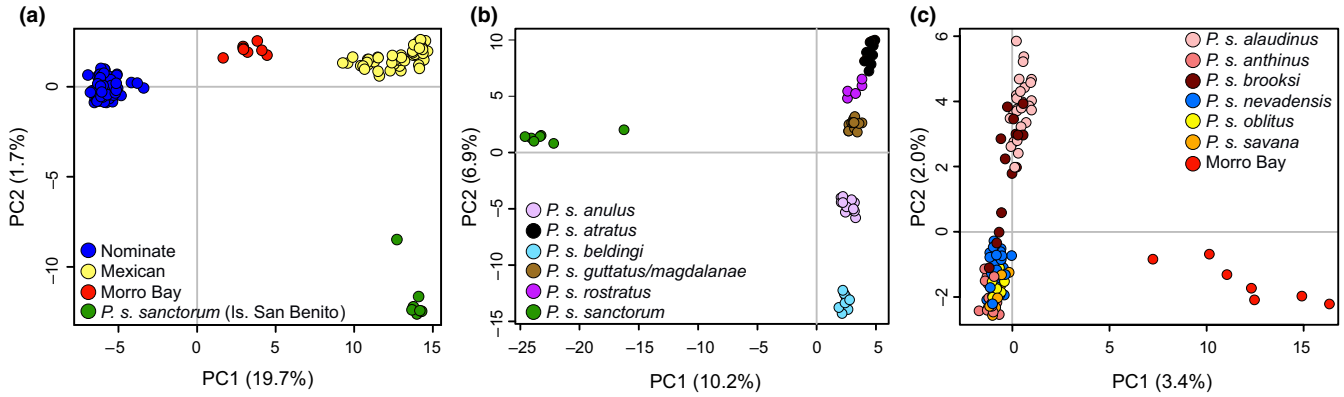

FIGURE 3 Principal components analysis for: (a) full data set (n = 191); (b) northwestern Mexico data set (n = 56); and (c) nominate data set (n = 135). Analysis on data set with 10% missing data is shown [Colour figure can be viewed at wileyonlinelibrary.com]

northwest Mexico also distinguished the Isla San Benito population from others along the first principal component axis (10.2% of variance), while along the second axis (6.9% of variance), four clusters were separated, corresponding to subspecies distributed along the coast of Baja California and Sonora (Figure 3b). Finally, the focused analysis on the nominate group again identified a split between individuals sampled in Morro Bay, CA, from all others along the first axis (3.4% of variance), and while there was some evidence of a split between birds sampled along the Pacific coast in California from those sampled throughout the rest of North America along the second axis (2.0% of variance), individuals of the subspecies *P. s. brooksi* from coastal Washington state were found in both of these clusters (Figure 3c), suggesting that this subtle genetic variation is not strongly geographically structured. Varying the amount of missing data included in these analyses had no effect on overall patterns of population structure (Supporting Information Figure S7).

FASTSTRUCTURE analyses employing the simple prior found K = 2 to maximize the marginal likelihood. Similar to PCA, these clusters reflect the divergence between nominate and northwestern Mexico Savannah sparrows (Figure 4). These analyses provide additional evidence for admixture between the nominate and northwestern Mexico populations at Morro Bay, CA, where individuals are assigned to either group with roughly equal probabilities. Analysis with a logistic prior revealed similar patterns; however, the best-supported model was K = 3 where the third cluster represented the two individuals of *P. s. beldingi* from southern California and some of the genetic variation in Morro Bay birds (Figure 4). Independent analyses of the nominate and northwestern Mexico data sets found K = 1 to be the best-supported model for each, suggesting that the larger genetic divide between these populations was not impeding the identification of more subtle structure within the full data set. Moreover, plotting K = 2–4 for the nominate group did not recover any additional structure (Supporting Information Figure S8).

Phylogenetic analysis in RAXML found *Passerculus sandwichensis* to be a well-supported monophyletic group, bootstrap (BS): 100 (Figure 5). Individuals from northwest Mexico and southern California formed a distinct clade with strong bootstrap support (BS 100); interestingly, individuals from Morro Bay formed an

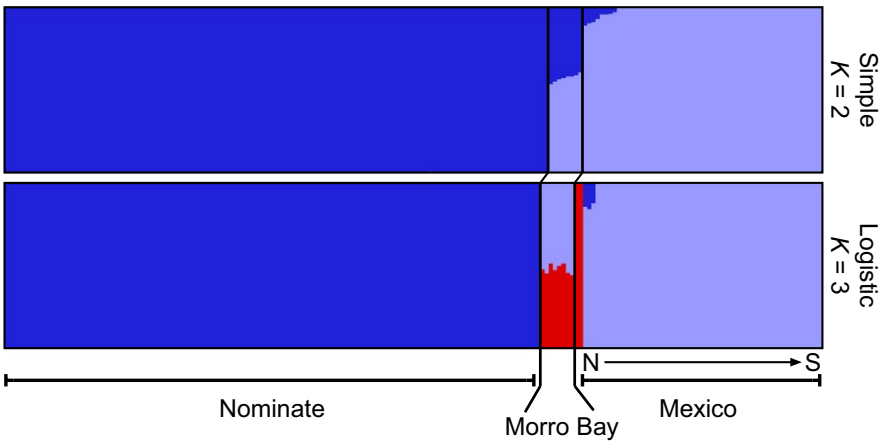

FIGURE 4 Population structure in Savannah sparrows determined in FASTSTRUCTURE. Plots show variation in output when simple prior (upper panel) vs. logistic prior for allele frequencies was set. For analyses using simple prior, K = 2 was found to optimize the marginal likelihood of the clusters, whereas K = 3 was found under a logistic prior. Structure existed between northwestern Mexico and nominate Savannah Sparrows in both analyses. Both also identified an apparent zone of secondary contact at Morro Bay along the California coast. Barplots were generated in the r-package POPHELPER version 1.1.6 (Francis, 2016) [Colour figure can be viewed at wileyonlinelibrary.com]

unresolved polytomy at the basal node of the entire group. Within the northwestern Mexico clade (excluding birds from Morro Bay), patterns of genetic structure were similar to the PCA results. Members of the *P. s. beldingi* subspecies formed a polytomy at the basal node of this group and the subspecies *P. s. anulus*, *P. s. sanctorum*, *P. s. guttatus*, *P. s. magdalanae*, *P. s. atratus* and *P. s. rostratus*

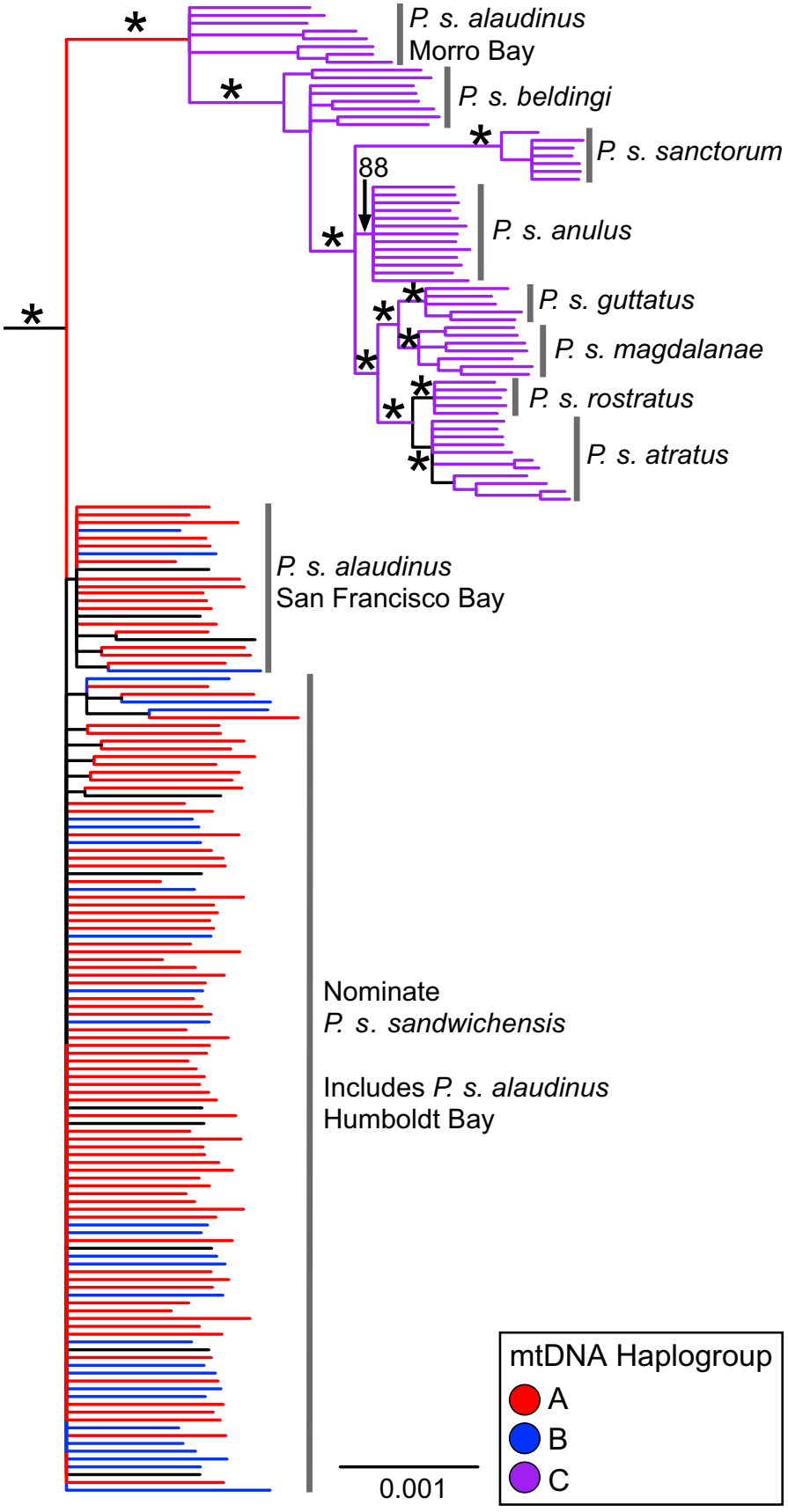

FIGURE 5 Maximum-likelihood phylogeny of *Passerculus sandwichensis* generated in RAXML 8.2.4. The phylogeny is rooted by four closely related sparrow taxa (*Ammospiza leconteii*, *Centronyx henslowii*, *Melospiza melodia* and *Pooecetes gramineus*) that have been removed for visualization purposes. Phylogeny constructed with 25% missing data set and using the Felsenstein correction for ascertainment bias. Asterisks denote nodes with >95 bootstrap support. Nodes with bootstrap support <50 have been collapsed. Outgroups not shown. Individuals on the phylogeny are coloured by mtDNA haplogroup (Figure 2). Black branches signify individuals for which we lack mtDNA [Colour figure can be viewed at wileyonlinelibrary.com]

each formed well-supported monophyletic clades with the *P. s. anulus* clade receiving the lowest support (BS: 88; Figure 5). Again, we found little evidence for structure in the nominate group, corroborating results from FASTSTRUCTURE and PCA of the full data set (Figure 3a). In sum, the bulk of our population structure results does not support our first hypothesis that mitochondrial introgression has exceeded nuclear introgression upon contact of historically isolated lineages.

3.3 | Demographic analyses

3.3.1 | Nominate population SFS analysis

A single panmictic population that experienced a bottleneck was found to be the best-supported model (Table 2). Four other models also had high AIC scores, including a bottleneck-population growth model, and three models involving admixture from a "ghost" population. Adjusted likelihood ratio tests confirm that the bottleneck model was the best-fit model (likelihood: −44.3) compared to the bottleneck-population growth model (likelihood: −46.0; adjusted *D* statistic: 137.43; *p*-value: 0) and the admixture with growth model (likelihood: −44.2; adjusted *D* statistic: 0.17; *p*-value: 0.34). In contrast, simulations based on parameter estimates from single-population ∂a∂i analyses (Supporting Information Figures S9–S10) indicate that a constant Ne was the only history that could result in observed levels of mtDNA diversity. Indeed, for a constant Ne, the observed levels of π (11.42) fell within the 99th and 95th percentiles for the entire range of potential mtDNA Ne (500,000 to 1,650,000) tested in the simulations (Supporting Information Figure S9) and for segregating sites (*S*), the observed *S* (76) fell within the 99th percentile at Ne greater than 650,000 and the 95th percentile above an Ne of 800,000 (Supporting Information Figure S10). Despite conflicting results between the best-supported model from ∂a∂i (bottleneck) and the *ms* simulations (constant Ne), both are consistent with our second hypothesis that divergent mtDNA clades within the nominate population arose within a panmictic population. See Supporting Information Table S3 for the parameter estimates from each of the models fit to the nominate SFS.

3.3.2 | Analysis of northwestern Mexico and nominate joint SFS

By far the best-fit model to the joint SFS (ΔAIC for next best model: 23.96) was an isolation-with-migration (IM) model where low levels of gene flow continued following divergence between the northwestern Mexico and nominate clades with the nominate clade maintaining a constant population size and the northwestern Mexico population experiencing a bottleneck (Table 3; Figure 6). In general, IM models showed much greater likelihoods, and all top models involved a bottleneck or exponential growth within the northwestern Mexico population (Table 3). Parameter estimates from the best-fit model suggest a population split occurring ~480,000 generations ago (95% CI: 240,829–2,066,377), following this split, the nominate population maintained a large Ne of ~1,950,000 (95% CI: 758,589–3,684,813). The northwestern Mexico population experienced a bottleneck roughly ~380,000 generations ago (95% CI: 240,829–541,737) during which the Ne was reduced to ~7,400 (95% CI: 0–22,132). This bottleneck ended ~100,000 generations ago (95% CI: 59,429–147,642) and the current northwestern Mexico Ne was estimated to be ~150,000 (95% CI: 51,276–303,320). Migration rates between the two populations were roughly symmetric though slightly greater from the northwestern Mexico into the nominate population (Table 4). Simulations of the best-fit two-population model (Figure 6) also matched observed π (13.48) and *S* (86) for both populations (Figure 7). The observed levels of π fell within the 99th percentile at Ne greater than 1,050,000 and the 95th percentile above

TABLE 2 Comparisons of results from analysis of nominate population SFS in ∂a∂i (22,066 SNPs; Supporting Information Figure S4)

Models	LL	*K*	AIC	ΔAIC	Weights
Bottleneck	−44.3	4	96.5	0	0.608
Bottleneck–exponential growth	−46.0	3	98.0	1.5	0.290
Admix "ghost," constant	−44.4	6	100.9	4.3	0.069
Admix "ghost," exponential growth	−44.2	7	102.5	5.9	0.031
Admix "ghost," bottleneck	−44.3	10	108.5	12.0	0.001
Exponential growth	−141.6	2	287.3	190.8	0
Admix mtDNA, exponential growth	−183.7	5	377.5	281.0	0
Admix mtDNA, constant	−183.9	5	377.9	281.3	0
Admix mtDNA, bottleneck	−182.0	8	379.9	283.3	0
Constant	−6,501	1	13,004	12,908	0

Note. Models in bold are consistent with the second hypothesis that divergent mtDNA clades arose within a panmictic population, and nonbolded models are consistent with the third hypothesis of historical isolation followed by admixture. Values included are likelihoods (LL), number of parameters (*K*), AIC, ΔAIC and Akaike weights (the relative likelihood of each model).

TABLE 3 Comparisons of results from analysis of joint SFS of the northwestern Mexico and nominate population in ∂a∂i

Models	LL	*K*	AIC	ΔAIC	Weights
IM, constant size Nom, bottle Mex	−473.08	9	964.16	0	1.000
IM, bottle both	−481.06	13	988.12	23.96	0.000
IM, Admix "ghost," bottle both	−503.88	16	1,039.76	75.6	0.000
IM, Admix "ghost," constant size Nom., bottle Mex	−600.3	12	1,224.6	260.44	0.000
IM, constant size Nom, exp grow Mex	−610.97	6	1,233.94	269.78	0.000
IM, exp grow both	−611.92	7	1,237.84	273.68	0.000
IM Admix "ghost," exp grow Nom, bottle Mex	−668.01	14	1,364.02	399.86	0.000
IM, exp grow Nom, constant size Mex	−1,074.29	6	2,160.58	1,196.42	0.000
IM, Admix "ghost," constant size both	−1,083.26	8	2,182.52	1,218.36	0.000
IM, constant size both	−1,086.71	5	2,183.42	1,219.26	0.000
IM, bottle Nom, constant size Mex	−1,085.14	9	2,188.28	1,224.12	0.000
IM, Admix mtDNA, constant size both	−1,651.71	8	3,319.42	2,355.26	0.000
IM Admix mtDNA, constant size Nom, bottle Mex	−1,806.19	12	3,636.38	2,672.22	0.000
No mig., constant size Nom, bottle Mex	−2,142.44	7	4,298.88	3,334.72	0.000
No mig., bottle both	−2,154	11	4,330	3,365.84	0.000
No mig., exp grow both	−2,199.18	5	4,408.36	3,444.2	0.000
No mig., constant size Nom, exp grow Mex	−2,211.58	4	4,431.16	3,467	0.000
No mig., exp grow Nom, constant size Mex	−2,322.55	4	4,653.1	3,688.94	0.000
No mig., bottle Nom, constant size Mex	−2,319.7	7	4,653.4	3,689.24	0.000
IM, Admix mtDNA, bottle both	−2,331.49	16	4,694.98	3,730.82	0.000
No mig., constant size both	−2,383.22	3	4,772.44	3,808.28	0.000
Mig., Admix mtDNA, exp grow Nom, bottle Mex	−3,023.11	14	6,074.22	5,110.06	0.000

Note. Models in bold are consistent with the second hypothesis that divergent mtDNA clades arose within a panmictic population, and nonbolded models are consistent with the third hypothesis of historical isolation followed by admixture. Values included are likelihoods (LL), number of parameters (*K*), AIC, ΔAIC and Akaike weights (the relative likelihood of each model).

1,550,000; and the observed levels of *S* fell into the 99th and 95th percentiles at population sizes of 1,100,000 and 1,250,000, respectively. We also ran simulations based on the upper and lower bounds for parameter estimates. Observed levels of π and *S* again fell within the 99th and 95th percentiles for a similar range of mtDNA effective population sizes for simulations based on parameter estimates from both the upper and lower bounds (Supporting Information Figures S11–S12). Results from demographic analysis of the joint SFS and *ms* simulations strongly support a demographic history where divergent mtDNA clades within the nominate clade arose within a panmictic population of Savannah sparrows further supporting our second hypothesis.

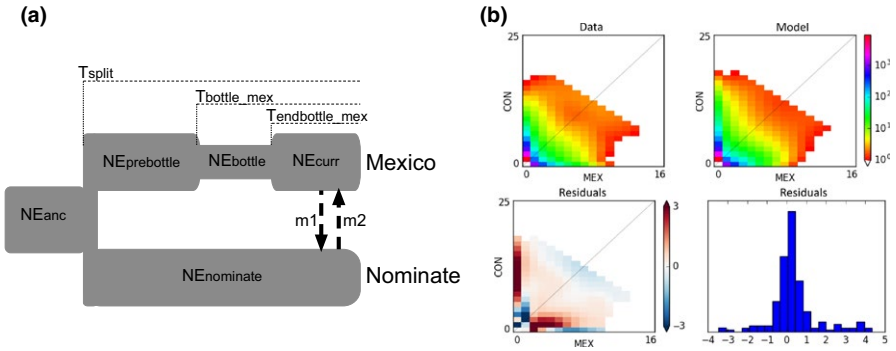

FIGURE 6 (a) Schematic of best-supported model, where northwestern Mexico and nominate populations diverged at Tsplit followed by a bottleneck in the northwestern Mexico population size from Tbottle_mex to Tendbottle_mex, whereas the nominate population maintained a constant population size following the split. Labelled parameters correspond to point estimates and uncertainties listed in Table 3. (b) Comparison of observed site frequency spectrum to simulated spectra given parameter values from model in Figure 6a. Lower panels represent residuals between fit of model and data [Colour figure can be viewed at wileyonlinelibrary.com]

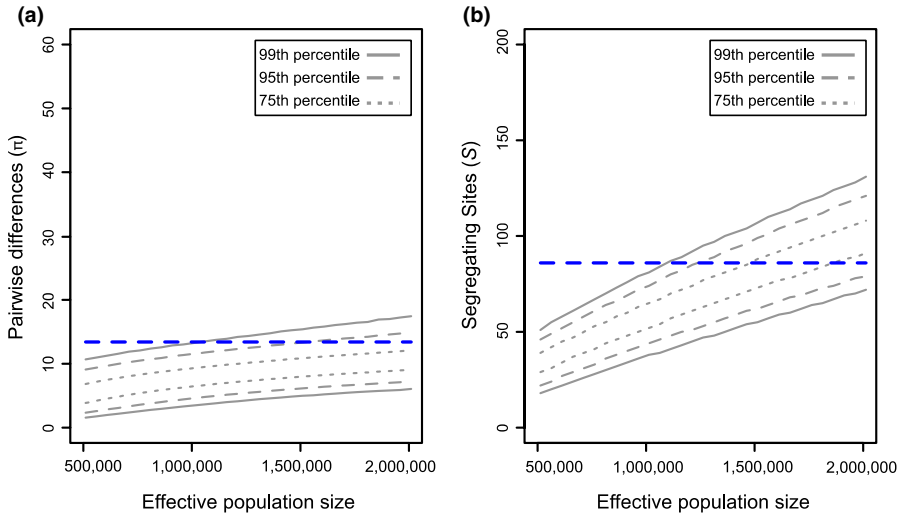

FIGURE 7 Results from *ms* simulations where parameter estimates from the best-supported demographic model from ∂a∂i analyses (Figure 6) were used to assess whether this demographic model could result in observed mtDNA diversity within Savannah sparrows. Simulations are based on 204 individuals across a range of effective population sizes relevant to empirical estimates of mtDNA Ne with the range of Ne based on variance in nucleotide diversity. For average pairwise differences (a) and number of segregating sites (b) solid, dashed and dotted grey lines represent 99%, 95% and 75% quantiles, respectively, of these summary statistics calculated from across 100,000 simulations for each effective population size. Blue line represents empirical estimate of mtDNA average pairwise differences (13.48) and number of segregating sites (86) [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | The influence of panmixia vs. admixture histories on mtDNA diversity

The simulations described above allowed us to identify the most likely demographic history for Savannah sparrows based on the nuclear data set. We complimented these analyses with a series of additional simulations to determine: (a) the range of demographic histories that could produce divergent but sympatric haplotype clades and (b) whether this evolutionary outcome was more likely under a history of isolation followed by secondary contact compared to a history of lineage sorting within a panmictic population. Population size proved to be the key parameter contributing to patterns of mtDNA diversity across all simulations. In the absence of a bottleneck, both model classes (panmixia and admixture) required a population size of over 350,000 to generate observed levels of mtDNA diversity with little differences existing between the two sets of models (Supporting Information Figures S13, S15). Fluctuations in population size due to bottlenecks or increases from a small ancestral population proved to have the greatest influence on mtDNA diversity across all simulated models. Lower population sizes in the ancestral population or during bottlenecks required a much greater contemporary effective population size to generate observed levels of mtDNA diversity (Supporting Information Figures S13–S16). Variation in parameters related to the duration of bottlenecks or isolation and time since admixture had comparatively little influence on the effective population size required to generate observed patterns

(Supporting Information Figures S14–S16). In cases where time parameters did influence 99th percentile Ne, it was usually dependent on population size parameters. For example, longer bottlenecks led to decreases in the 99th percentile Ne in instances where the bottleneck was less severe (i.e., 25%–50% reduction in Ne; Supporting Information Figures S14, S16). In summary, these simulations show that there is little reason to assume a priori that divergent, sympatric mtDNA haplogroups are more likely to arise when historically isolated populations come into secondary contact than within a single large and panmictic population. Rather, we show that if populations remain relatively large and constant through time, both scenarios can produce the occurrence of divergent but sympatric haplotype clades across a wide range of parameter space.

4 | DISCUSSION

Unusual mtDNA patterns, such as the presence of divergent but sympatric mtDNA lineages, have often been interpreted as evidence for specific demographic scenarios (e.g., Avise, 2000). However, the same pattern of mtDNA variation can often be produced by multiple demographic histories. Revisiting these unusual mtDNA patterns with genomic data sets is needed to test whether common interpretations of these phylogeographic patterns are justified, and to better understand the processes that give rise to irregular patterns of mtDNA variation. To this end, we generated a large, nuclear SNP data set for the Savannah sparrow to resolve the demographic history underlying the pattern of two divergent (~2%) mtDNA lineages that are sympatric across much of the breeding distribution of the species (Figure 2; Zink et al., 2005). First, we showed that this mtDNA pattern exists despite little geographic structure within the nuclear genome of the nominate clade (Figures 3–5). Second, based on the best-fit models from ∂a∂i analyses (Tables 2 and 3; Figure 6) and corresponding simulations (Figure 7) we conclude that the two divergent mtDNA lineages within the nominate clade arose in a large, panmictic population. Although this pattern of divergent and sympatric mtDNA lineages in the absence of nuclear structure has frequently been interpreted as evidence for admixture of historically isolated lineages (Avise, 2000; Hogner et al., 2012; Quinn et al., 1992), our results clearly show that sympatry of divergent mtDNA clades cannot be interpreted as evidence for a history of secondary contact and admixture in the absence of other corroborating data.

4.1 | Patterns of genetic structure

Across all population structure analyses (Figures 3–5), the greatest divergence was found between the nominate group and birds restricted to tidal marshes in northwest Mexico and southern California. Savannah sparrows exhibit a surprising lack of structure within the nominate clade given that samples spanned most of the breeding distribution from Alaska to Newfoundland and south to California, New Mexico and Virginia (Figures 3–5). PCA of the nominate group did suggest the existence of some subtle population structure along the Pacific coast of California (Figure 3c). We consider this potential structure to be weak given that (a) this structure only explains two per cent of the genetic variance found within the nominate clade (including birds from Morro bay); (b) individuals sampled from coastal Washington span both clusters; and (c) this cluster was not recovered in any other analyses of population structure. This weak structure could reflect multiple scenarios, including some degree of isolation along the California coast, gene flow with northwest Mexico populations along the Pacific coast and/or local adaptation to tidal marsh environments. Regardless of the interpretation, this subtle structure does not appear to have influenced downstream demographic analyses or the conclusion that sympatric, divergent mtDNA lineages arose within a large, contiguous population. In contrast, we did identify clear population structure in PCA (Figure 3b) and phylogenetic (Figure 5) analyses within the northwestern Mexico clade that largely corresponds to recognized subspecies (van Rossem, 1947). Structure among these different subspecies appears to be maintained despite high postbreeding dispersal to other salt marshes and coastal areas in southern California and northwest Mexico (Garrett, 2008; van Rossem, 1947). Finally, given our increased sampling along the Pacific coast relative to previous phylogeographic studies, we identified a zone of secondary contact between the two Savannah sparrow lineages at Morro Bay, CA (Figures 2–4), that had been previously suggested by intermediate morphology and plumage of birds collected at this location (e.g., Grinnell & Miller, 1944). Further sampling and analyses will be required to characterize the precise history and extent of introgression among Savannah sparrow populations along the Pacific coast.

4.2 | Divergent mtDNA haplogroups likely arose within a panmictic nominate clade

A lack of nuclear structure within the nominate clade of Savannah sparrows could indicate that the sympatric mtDNA clades arose within a large, panmictic population or that there was a history of isolation and complete admixture between historically isolated lineages. The best-fit models across all ∂a∂i analyses were only consistent with the former that sympatric, divergent mtDNA clades arose within a single panmictic population. Although we found no support for a hypothesis of secondary contact and admixture (Tables 2 and 3), we were unable to fully resolve the exact demographic history experienced by a panmictic nominate group. Analysis of the nominate group SFS identified a suite of models with a bottleneck or population expansion that were well supported (Table 2). However, in our *ms* simulations based on parameter estimates for each of the models compared within ∂a∂i, only a model of constant population size was able to generate observed levels of mtDNA π and S (Supporting Information Figures S9–S10). It is possible that these conflicts reflect the greater demographic information contained within the SFS vs. the summary statistics π and S, but in our view, a more likely explanation for the discordant results is that individuals from the northwestern Mexico clade were not included in the analysis of the nominate SFS. Migration from unsampled populations can lead to high support

TABLE 4 Parameter estimates from best model (Figure 6) with 95% CI calculated based on GIM uncertainties

Parameter	ML estimate	95% CI low	95% CI high
Theta (4NeμL)	2,646.28	1,771.31	3,520.69
$NE_{ancestor}$	261,497	175,035	347,904
$NE_{nominate}$	1,951,471	758,589	3,684,813
$NE_{prebottle\_mex}$	14,878,428	0	111,940,918
NE_{bottle_mex}	7,408	0	22,132
NE_{curr_mex}	152,296	51,276	303,320
T_{split}	481,386	240,829	2,066,377
T_{bottle_mex}	383,490	240,829	541,737
$T_{endbottle\_mex}$	99,879	59,429	147,642
m1 (mex → nom)	6.31E−06	5.83E−06	6.56E−06
m2 (nom → mex)	1.37E−06	8.75E−07	1.63E−06

for a bottleneck model (e.g., Nielsen & Beaumont, 2009). When the northwestern Mexico population is included, analysis of the joint SFS strongly supported a model with migration between the nominate and northwestern Mexico clades with the nominate population maintaining a constant population size following the initial divergence (Table 4; Figure 6). Coalescent simulations performed on parameter estimates from this best-fit model reproduced observed levels of mtDNA variation (Figure 7). Our simulations also showed that this model was robust to parameter estimate error with simulations parameterized on upper and lower confidence intervals also reproducing observed mtDNA diversity (Supporting Information Figure S11–S12). Additional simulations (see below) indicate that stable, large population sizes are a key factor in the generation of observed mtDNA diversity, further supporting our inference that nominate Savannah sparrows maintained a large and constant population size following divergence from northwestern Mexico birds.

An alternative demographic scenario that we did not explicitly consider is suggested by mtDNA structure (Figure 2), which shows a sister relationship between nominate clade B and northwestern Mexico clade C. This suggests the possibility of historical introgression from northwestern Mexico into nominate populations across North America followed by isolation and divergence of the B and C haplogroups. Unidirectional introgression from northwestern Mexico to nominate populations would require the spread of mtDNA haplotypes from a population of low Ne into a larger population (see Table 4). This runs counter to theoretical and empirical inferences that have shown asymmetric introgression tends to occur in the direction from larger to smaller populations (e.g., Currat et al., 2008; Sarver et al., 2017). Alternatively, selection for the northwestern Mexico mtDNA haplotypes could result in its introgression and sweep through the nominate population (e.g., Sloan et al., 2016). In this case, we would expect the favoured mtDNA haplotype to either completely displace the less advantageous haplotype or displace it in parts of the species range where it is favourable. The latter scenario would lead to geographic structure in mtDNA haplogroups and not the observed sympatry (Figure 2). Second, a selective sweep would also reduce genetic diversity in the mtDNA genome (e.g., Toews, Mandic, Richards, & Irwin, 2014), which contrasts with the high mtDNA genetic diversity observed in Savannah sparrows. Balancing selection could maintain both haplotypes, but McDonald–Kreitman tests on the mtDNA genes ND2 and ND3 from the nominate group support a history of neutral evolution in Savannah sparrow mtDNA (Zink et al., 2005). Given these considerations, we favour the hypothesis that the two sympatric mtDNA lineages within nominate Savannah sparrows most likely arose within a panmictic population as suggested by our best-fit model. Under this scenario, a colonization or vicariance event led to the isolation of the northwestern Mexico population after the emergence of two sympatric mtDNA lineages within the nominate population. At this time point, the northwest Mexico population was either comprised solely of individuals possessing the mtDNA lineage ancestral to clades B and C or individuals derived from both of the nominate lineages with the A lineage subsequently lost due to stochastic processes. Regardless, divergence between clades B and C would have occurred following isolation of the northwest Mexican population with clade C emerging in an isolated northwest Mexico population and the lineage evolving into clade B persisting in sympatry with clade A in the nominate population.

Our results suggest that the nominate clade of Savannah sparrows maintained a large and constant population size throughout much of the Pleistocene (divergence time between clades: 0.24–2 mya; Table 4). This contrasts with many other North American taxa studied that show evidence of population fragmentation (e.g., Shafer, Cullingham, Côté, & Coltman, 2010; Weir & Schluter, 2004) and/or population contraction followed by rapid expansion (e.g., Ball, Freeman, James, Bermingham, & Avise, 1988; Milá, Girman, Kimura, & Smith, 2000) in response to Pleistocene glacial cycles. Resistance to population contraction and fragmentation during the Pleistocene in Savannah sparrows is consistent with several hypotheses that link ecological and life history attributes to patterns of population genetic differentiation. First, certain habitat associations can influence patterns of genetic structure with species that occupy more open and/or ephemeral habitats exhibiting less genetic structure than forest/closed habitat species (e.g., Burney & Brumfield, 2009; Drovetski, Zink, Ericson, & Fadeev, 2010; Harvey, Aleixo, Ribas, & Brumfield, 2017). One hypothesis explaining this pattern is that open country taxa exhibit increased dispersal propensity to exploit ephemeral or patchy resources (e.g., Burney & Brumfield, 2009). Natal and adult dispersal patterns of Savannah sparrows support this hypothesis. Although site fidelity is highly variable among Savannah sparrow populations (Wheelwright & Rising, 2008), the lowest levels of annual return rates reported for both adults and nestlings banded at a breeding site (<6%) were found in grassland habitats of the northern great plains (Jones, Dieni, Green, & Gouse, 2007). Low return rates were not unique to Savannah sparrows in this region and the authors argued that high interannual variation in resource availability selects for an opportunistic and flexible breeding life history (Jones et al., 2007). Second, the specialist–generalist variation hypothesis

posits that more specialized species will exhibit greater population structure, lower levels of gene flow and lower genetic diversity relative to generalist species (Li, Jovelin, Yoshiga, Tanaka, & Cutter, 2014). Although this hypothesis has received mixed support (e.g., Hung, Drovetski, & Zink, 2017; Matthee, Engelbrecht, & Matthee, 2018; Titus & Daly, 2017), a generalist ecology likely contributes to the lack of population genetic structure and large effective population sizes observed within nominate Savannah sparrows. Savannah sparrows occupy a diversity of open habitats ranging from tundra to salt marshes to agricultural fields (Wheelwright & Rising, 2008). Palynological, fossil and other climatic evidence suggests that contiguous tundra, grassland and steppe habitats existed across the southern edge of advancing glaciers and further south in the United States during the last glacial maxima (Dyke, 2005; Williams, Shuman, & Bartlein, 2009; Williams, Shuman, Webb, Bartlein, & Leduc, 2004). Successful exploitation of these habitats by generalist Savannah sparrows may have enabled them to maintain population connectivity when other more specialized species re-treated to isolated refugia.

4.3 | Sympatry of divergent mtDNA lineages can arise under a range of demographic scenarios

This study provides important empirical support for predictions that divergent but sympatric mtDNA lineages could emerge through coalescent stochasticity in a large, stable population. Although the bulk of our evidence supports this hypothesis, the parameter space in which similar patterns could arise under different demographic scenarios has been unexplored. To this end, we performed additional simulations to test if a history of isolation followed by recent admixture might be more likely to produce the observed mtDNA patterns under different demographic conditions. Based on these simulations, we found that either history could produce the observed patterns of mtDNA diversity present in Savannah sparrows over a broad range of population genetic parameter values (Supporting Information Figures S13, S15). However, a major result stemming from this analysis was that the patterns of mtDNA diversity in Savannah sparrows could only arise in large populations (mtDNA Ne > 350,000). Parameters related to historical fluctuations in Ne also played a critical role in shaping patterns of genetic diversity with small ancestral Ne or bottlenecks significantly eroding mtDNA diversity. Even in a model of recent contact and admixture, populations would need to maintain a large Ne throughout their history to produce the observed patterns of mtDNA diversity. The importance of large population size in the emergence of divergent, sympatric mtDNA lineages is supported by theory (e.g., Edwards & Beerli, 2000; Slatkin & Hudson, 1991) and empirical evidence. For example, passenger pigeons (*Ectopistes migratorius*)—legendary for their enormous flocks—exhibit divergent, sympatric mtDNA lineages and mtDNA Ne was recently estimated to be 13 million (Murray et al., 2017). Hogner et al. (2012) also report relatively high nucleotide diversity within the European redstart (*Phoenicurus phoenicurus*), a species with sympatric mtDNA lineages that are ~5% divergent.

These empirical and simulation results have broad implications for the role of effective population size in phylogeographic inference. First, widespread theoretical and empirical work has elucidated the myriad issues associated with reliance on mtDNA as a marker for phylogeographic inference (e.g., Ballard & Whitlock, 2004; Edwards & Beerli, 2000; Edwards & Bensch, 2009; Good et al., 2015; Irwin, 2002), yet the foundational importance of mtDNA in phylogeography means that it continues to play an outsized role in reconstructing species history. Given this history of the field, unusual mtDNA patterns have often been submitted as compelling evidence for rare historical events (e.g., historical isolation followed by admixture); however, our simulations show that in species with large effective population sizes, conventional interpretations of mtDNA patterns may be particularly misleading. In these cases, it will be impossible to tease apart different demographic hypotheses in the absence of corroborating data (e.g., Kearns et al., 2018). Second, recent population separation and high migration rate are common explanations for the absence of population differentiation in species; however, our results along with other bird species (e.g., Hung et al., 2017) and many marine fish species (e.g., Cano, Shikano, Kuparinen, & Merilä, 2008) indicate that effective population size may also be an important population genetic parameter influencing patterns of population differentiation at neutral loci. Indeed, simulations performed by Cano et al. (2008) showed that even in the absence of migration, large populations will not exhibit differentiation after several thousand generations. This suggests that population bottlenecks may be key to the fixation of neutral alleles in isolation and the emergence of population genetic structure during isolation in refugia. Further comparative work will be required, but resolution of the role population size plays in population differentiation will provide a more nuanced understanding of the various ways certain geological processes (e.g., glaciation) could influence patterns of population divergence. Finally, despite a lack of neutral genetic variation within nominate Savannah sparrows, our sampling of this group included eight different subspecies. Large Ne may provide an explanation for the phenotypic–genotypic discordance observed in this and other species (e.g., Harris, Alström, Ödeen, & Leaché, 2018; Mason & Taylor, 2015). Although in large populations, neutral alleles will persist for long periods without drifting to fixation, alleles under selection will be rapidly driven to fixation (e.g., Charlesworth, 2009). Adaptation to spatially varying ecological pressures in different Savannah sparrow populations could have driven favourable alleles underlying phenotypic differences to fixation, even as neutral regions of the genome remain undifferentiated.

5 | CONCLUSIONS

Using a large genomic SNP data set, we revisited a previous phylogeography of the Savannah sparrow where two mtDNA haplogroups were found to be sympatric across much of the species' distribution (Zink et al., 2005). Our analyses revealed that these

two mitochondrial clades were not associated with significant population structure in the nuclear genome. Instead, a series of demographic analyses suggest that the divergent mtDNA lineages arose within a single large and panmictic population. These results contrast with the most frequent interpretation of this pattern, which is often assumed to reflect a history of isolation, secondary contact and admixture. This contributes to the growing body of phylogeographic analyses based on genomic data sets that have shown previous interpretations based on mtDNA data sets to be inaccurate or misleading (e.g., Emerson et al., 2010; Good et al., 2015). In the case of Savannah sparrows, our estimated history of constant, large population sizes led to starkly different inferences into the role habitat associations and effective population size may play in explaining phylogeographic patterns than previous interpretations based on mtDNA alone. Revisiting previous mtDNA-based phylogeographic studies with genomic data sets will thus be critical for accurately inferring species history and illuminating how species' biology shapes phylogeographic patterns. Our analytical approach should be broadly applicable to address this in any species with broad geographic sampling.

ACKNOWLEDGEMENTS

We thank the following museums and individuals for providing tissue samples for this study: J. Cracraft, P. Sweet, & T. Trombone (AMNH); K. Bostwick (CUMV); D. Dittmann & F. Sheldon (LSUMZ); C. Witt & A. Johnson (MSB); C. Cicero & R. Bowie (MVZ); J. Rising, R. Zink & M. Westberg (ROM & BELL); P. Unitt & K. Burns (SDNHM & SDSU Museum of Biodiversity); K. Winker (UAM); C. Gebhard, & H. James (USNM); S. Birks (UWBM); K. Zyskowski (YPM). Adolfo Navarro-Siqüenza was instrumental in providing necessary permits and assistance with field logistics for work in Mexico. We also thank J. Bates, S. Hackett, B. Marks, D. Brinkmeier, J. Maley, J. McCormack and A. Gordillo Martínez for their assistance with permits and field logistics. S. M. Robles Bello, A. Hernández Cardona, D. Levey and D. Senner provided excellent field assistance in collecting samples. J. Jones and A. Hernandez helped with sequencing and laboratory work. Personnel at CDFW, IDNR and USFWS assisted with access to federal and state lands. Finally, we thank N. Sly, M. Stager, C. Wolf, R. Schweizer, J. Velotta, N. Senner, E. Beckman, the AEGG discussion group at the University of Montana, Irby Lovette, Bryan Carstens, Brian Smith and two anonymous reviewers for useful comments on the manuscript. Funding for field work and sequencing was provided by AMNH Frank M. Chapman Memorial Fund, SSE Rosemary Grant Award, SICB GIAR, Sigma-XI GIAR, UIUC Animal Biology departmental grants, Illinois Ornithological Society, Systematics Research Fund, AOU research award, Center for Latin America & Caribbean Studies Tinker Fellowship, and start-up funds to ZAC from UM and UIUC.

AUTHOR CONTRIBUTIONS

P.M.B. and Z.A.C. designed the study. P.M.B. performed all sequencing and data analysis. P.M.B. and Z.A.C. wrote the paper.

DATA ACCESSIBILITY

mtDNA sequence data have been submitted to GenBank under Accession nos MK425508—MK425598. Short-read Illumina data have been submitted to the NCBI sequence read archive (SRA accession: PRJNA521441). All code for ∂a∂i analyses and *ms* simulations can be found on github: https://github.com/phbenham/SAVSphylogeography_code.

ORCID

Phred M. Benham https://orcid.org/0000-0003-3276-6177

REFERENCES

Andersson, M. (1999). Hybridization and skua phylogeny. *Proceedings of the Royal Society B: Biological Sciences*, *266*, 1579–1585. https://doi.org/10.1098/rspb.1999.0818

Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Cambridge, MA: Harvard University Press.

Bachtrog, D., Thornton, K., Clark, A., & Andolfatto, P. (2006). Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution*, *60*, 292–302. https://doi.org/10.1111/j.0014-3820.2006.tb01107.x

Ball, R. M., Freeman, S., James, F. C., Bermingham, E., & Avise, J. C. (1988). Phylogeographic population structure of Red-winged Blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *85*, 1558–1562. https://doi.org/10.1073/pnas.85.5.1558

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x

Beckman, E. J., & Witt, C. C. (2015). Phylogeny and biogeography of the New World siskins and goldfinches: Rapid, recent diversification in the Central Andes. *Molecular Phylogenetics and Evolution*, *87*, 28–45. https://doi.org/10.1016/j.ympev.2015.03.005

Bensasson, D., Zhang, D. X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology and Evolution*, *16*, 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6

Block, N. L., Goodman, S. M., Hackett, S. J., Bates, J. M., & Raherilalao, M. J. (2015). Potential merger of ancient lineages in a passerine bird discovered based on evidence from host-specific ectoparasites. *Ecology and Evolution*, *5*, 3743–3755. https://doi.org/10.1002/ece3.1639

Burney, C. W., & Brumfield, R. T. (2009). Ecology predicts levels of genetic differentiation in neotropical birds. *The American Naturalist*, *174*, 358–368. https://doi.org/10.1086/603613

Cano, J. M., Shikano, T., Kuparinen, A., & Merilä, J. (2008). Genetic differentiation, effective population size and gene flow in marine fishes: Implications for stock management. *Journal of Integrated Science and Technology*, *10*, 1–10.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D. J. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*, 171–182.

Catchen, J. M., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation.

Nature Reviews Genetics, *10*, 195–205. https://doi.org/10.1038/nrg2526

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, *33*, 591–593. https://doi.org/10.1093/molbev/msv255

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, *62*, 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Dickinson, E. C., & L. Christidis (Eds.). (2014). *The Howard & Moore complete checklist of the birds of the world*, vol. *2*, 4th ed. East Sussex, UK: Passerines. Aves Press.

Drovetski, S. V., Zink, R. M., Ericson, P. G. P., & Fadeev, I. V. (2010). A multilocus study of pine grosbeak phylogeography supports the pattern of greater intercontinental divergence in Holarctic boreal forest birds than in birds inhabiting other high-latitude habitats. *Journal of Biogeography*, *37*, 696–706. https://doi.org/10.1111/j.1365-2699.2009.02234.x

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*, 699–710. https://doi.org/10.1371/journal.pbio.0040088

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973. https://doi.org/10.1093/molbev/mss075

Dyke, A. S. (2005). Late quaternary vegetation history of northern North America based on pollen, macrofossil and faunal remains. *Geographie Physique Et Quaternaire*, *59*, 211–262. https://doi.org/10.7202/014755ar

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113.

Edwards, S. V., & Beerli, P. (2000). Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, *54*, 1839–1854. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x

Edwards, S. V., & Bensch, S. (2009). Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, *18*, 2930–2933. https://doi.org/10.1111/j.1365-294X.2009.04270.x

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 16196–16200. https://doi.org/10.1073/pnas.1006538107

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Francis, R. M. (2016). pophelper: An r package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*, 27–32.

Garrett, K. L. (2008). Large-billed savannah sparrow (Passerculus sandwichensis rostratus). In W. D. Shuford, & T. Gardali (Eds.), *California bird species of special concern: a ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California*. *Studies of Western Birds*, vol. *1*, pp. 388–392. Sacramento, CA: Western Field Ornithologists, Camarillo, California, and California Department of Fish and Game.

Giska, I., Sechi, P., & Babik, W. (2015). Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. *BMC Evolutionary Biology*, *15*, 217. https://doi.org/10.1186/s12862-015-0488-9

Good, J. M., Vanderpool, D., Keeble, S., & Bi, K. (2015). Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, *69*, 1961–1972. https://doi.org/10.1111/evo.12712

Gosselin, T., & Bernatchez, L. (2016). stackr: GBS/RAD data exploration, manipulation and visualization using R. R package version 0.4.6. https://github.com/thierrygosselin/stackr. doi: https://doi.org/10.5281/zenodo.154432

Grinnell, J., & Miller, A. H. (1944). The distribution of the birds of California. *Pacific Coast Avifauna*, *27*, 1–608.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Harris, R. B., Alström, P., Ödeen, A., & Leaché, A. D. (2018). Discordance between genomic divergence and phenotypic variation in a rapidly evolving avian genus (Motacilla). *Molecular Phylogenetics and Evolution*, *120*, 183–195. https://doi.org/10.1016/j.ympev.2017.11.020

Harvey, M. G., Aleixo, A., Ribas, C. C., & Brumfield, R. T. (2017). Habitat association predicts genetic diversity and population divergence in Amazonian birds. *The American Naturalist*, *190*, 631–648. https://doi.org/10.1086/693856

Hoelzer, G. A., Dittus, W. P., Ashley, M. V., & Melnick, D. J. (1994). The local distribution of highly divergent mitochondrial DNA haplotypes in toque macaques *Macaca sinica* at Polonnaruwa, Sri Lanka. *Molecular Ecology*, *3*, 451–458. https://doi.org/10.1111/j.1365-294X.1994.tb00123.x

Hogner, S., Laskemoen, T., Lifjeld, J. T., Porkert, J., Kleven, O., Albayrak, T., … Johnsen, A. (2012). Deep sympatric mitochondrial divergence without reproductive isolation in the common redstart *Phoenicurus phoenicurus*. *Ecology and Evolution*, *2*, 2974–2988.

Hudson, R. (2002). Ms a program for generating samples under neutral models. *Bioinformatics*, *18*, 337–338.

Hudson, R. R., & Turelli, M. (2003). Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, *57*, 182–190.

Hung, C.-M., Drovetski, S. V., & Zink, R. M. (2017). The roles of ecology, behaviour and effective population size in the evolution of a community. *Molecular Ecology*, *26*, 3775–3784. https://doi.org/10.1111/mec.14152

Irwin, D. E. (2002). Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, *56*, 2383–2394. https://doi.org/10.1111/j.0014-3820.2002.tb00164.x

Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, S. L., Dieni, J. S., Green, M. T., & Gouse, P. J. (2007). Annual return rates of breeding grassland songbirds. *The Wilson Journal of Ornithology*, *119*, 89–94. https://doi.org/10.1676/05-158.1

Kazancioğlu, E., & Arnqvist, G. (2014). The maintenance of mitochondrial genetic variation by negative frequency-dependent selection. *Ecology Letters*, *17*, 22–27. https://doi.org/10.1111/ele.12195

Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., … Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, *9*(1), 906. https://doi.org/10.1038/s41467-018-03294-w

Klicka, J., Barker, K. F., Burns, K. J., Lanyon, S. M., Lovette, I. J., Chaves, J. A., & Bryson, R. W. (2014). A comprehensive multilocus assessment of sparrow (Aves: Passerellidae) relationships. *Molecular Phylogenetics and Evolution*, *77*, 177–182. https://doi.org/10.1016/j.ympev.2014.04.025

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*, 772–773. https://doi.org/10.1093/molbev/msw260

Leaché, A. D., Banbury, B. L., Felsenstein, J., De Oca, A. N. M., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies.

Systematic Biology, *64*, 1032–1047. https://doi.org/10.1093/sysbio/syv053

Li, S., Jovelin, R., Yoshiga, T., Tanaka, R., & Cutter, A. D. (2014). Specialist versus generalist life histories and nucleotide diversity in Caenorhabditis nematodes. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20132858.

Lynch, M. (2007). *The origins of genome architecture*. Sunderland, MA: Sinauer Associates, Inc. Publishers.

Mason, N. A., & Taylor, S. A. (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, *24*, 3009–3025. https://doi.org/10.1111/mec.13140

Matthee, C. A., Engelbrecht, A., & Matthee, S. (2018). Comparative phylogeography of parasitic Laelaps mites contribute new insights into the specialist-generalist variation hypothesis (SGVH). *BMC Evolutionary Biology*, *18*, 1–11. https://doi.org/10.1186/s12862-018-1245-7

Meier, J. I., Sousa, V. C., Marques, D. A., Selz, O. M., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Demographic modeling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular Ecology*, *26*, 123–141.

Milá, B., Girman, D. J., Kimura, M., & Smith, T. B. (2000). Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. *Proceedings of the Royal Society B: Biological Sciences*, *267*, 1033–1040.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). *Creating the CIPRES science gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE 2010)* (pp. 45–52). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Mims, M. C., Hulsey, C. D., Fitzpatrick, B. M., & Streelman, J. T. (2010). Geography disentangles introgression from ancestral polymorphism in Lake Malawi cichlids. *Molecular Ecology*, *19*, 940–951. https://doi.org/10.1111/j.1365-294X.2010.04529.x

Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J., … Shapiro, B. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, *954*, 1–14. https://doi.org/10.1126/science.aao0960

Nielsen, R., & Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Molecular Ecology*, *18*, 1034–1047. https://doi.org/10.1111/j.1365-294X.2008.04059.x

Oswald, J. A., Overcast, I., Mauck, W. M., Andersen, M. J., & Smith, B. T. (2017). Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. *Molecular Ecology*, *26*, 1386–1400. https://doi.org/10.1111/mec.14013

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*, 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Quinn, T. W. (1992). The genetic legacy of Mother Goose—phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Molecular Ecology*, *1*, 105–117.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589. https://doi.org/10.1534/genetics.114.164350

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Rising, J. (2017). Belding's Sparrow (*Passerculus guttatus*). In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions. (Retrieved from http://www.hbw.com/node/61919 on 5 April 2017).

Robinson, J. D., Coffman, A. J., Hickerson, M. J., & Gutenkunst, R. N. (2014). Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evolutionary Biology*, *14*, 254. https://doi.org/10.1186/s12862-014-0254-4

Sarver, B. A. J., Demboski, J. R., Good, J. M., Forshee, N., Hunter, S. S., & Sullivan, J. (2017). Comparative phylogenomic assessment of mitochondrial introgression among several species of chipmunks (*Tamias*). *Genome Biology and Evolution*, *9*, 7–19.

Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*, 4589–4621. https://doi.org/10.1111/j.1365-294X.2010.04828.x

Slatkin, M., & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, *129*, 555–562.

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2016). The on-again-off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*, 2212–2236. https://doi.org/10.1111/mec.13959

Spottiswoode, C. N., Stryjewski, K. F., Quader, S., Colebrook-Robjent, J. F. R., & Sorenson, M. D. (2011). Ancient host specificity within a single species of brood parasitic bird. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 17738–17742. https://doi.org/10.1073/pnas.1109630108

Stamatakis, A. (2014). RaxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Titus, B. M., & Daly, M. (2017). Specialist and generalist symbionts show counterintuitive levels of genetic diversity and discordant demographic histories along the Florida Reef Tract. *Coral Reefs*, *36*, 339–354. https://doi.org/10.1007/s00338-016-1515-z

Toews, D. P. L., Mandic, M., Richards, J. G., & Irwin, D. E. (2014). Migration, mitochondria, and the yellow-rumped warbler. *Evolution*, *68*, 241–255. https://doi.org/10.1111/evo.12260

van Rossem, J. (1947). A synopsis of the savannah sparrows of northwestern Mexico. *Condor*, *49*, 97–107. https://doi.org/10.2307/1364357

Wakely, J. (2009). *Coalescent theory: An introduction*. Greenwood Village, Colorado: Roberts & Company Publishers.

Walstrom, V. W., Klicka, J., & Spellman, G. M. (2011). Speciation in the white-breasted nuthatch (*Sitta carolinensis*): A multilocus perspective. *Molecular Ecology*, *21*, 907–920.

Weckstein, J., Zink, R. M., Blackwell-Rago, R. C., & Nelson, D. A. (2001). Anomalous variation in mitochondrial genomes of white-crowned (*Zonotrichia leucophrys*) and golden-crowned (*Z. atricapilla*) sparrows: Pseudogenes, hybridization, or incomplete lineage sorting? *The Auk*, *118*, 231–236.

Weir, J. T., & Schluter, D. (2004). Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society B: Biological Sciences*, *271*, 1881–1887. https://doi.org/10.1098/rspb.2004.2803

Weir, J. T., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*, *17*, 2321–2328. https://doi.org/10.1111/j.1365-294X.2008.03742.x

Wheelwright, N. T., & Rising, J. D. (2008). Savannah Sparrow (*Passerculus sandwichensis*). In A. Poole (Ed.), *Birds of North American Online*. Ithaca, NY: Cornell Lab of Ornithology.

Williams, J. W., Shuman, B. N., & Bartlein, P. J. (2009). Rapid responses of the prairie-forest ecotone to early Holocene aridity in mid-continental North America. *Global and Planetary Change*, *66*, 195–207. https://doi.org/10.1016/j.gloplacha.2008.10.012

Williams, J. W., Shuman, B. N., Webb, T., Bartlein, P. J., & Leduc, P. L. (2004). Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecological Monographs*, *74*, 309–334. https://doi.org/10.1890/02-4045

Xiao, J. H., Wang, N. X., Murphy, R. W., Cook, J., Jia, L. Y., & Huang, D. W. (2012). Wolbachia infection and dramatic intraspecific mitochondrial DNA divergence in a fig wasp. *Evolution*, *66*, 1907–1916. https://doi.org/10.1111/j.1558-5646.2011.01561.x

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., … Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution

and adaptation. *Science*, *346*, 1311–1320. https://doi.org/10.1126/science.1251385

Zink, R. M., Rising, J. D., Mockford, S., Horn, A. G., Wright, J. M., Leonard, M., & Westberg, M. C. (2005). Mitochondrial DNA variation, species limits, and rapid evolution of plumage coloration and size in the Savannah sparrow. *The Condor*, *107*, 21–28. https://doi.org/10.1650/7550

How to cite this article: Benham PM, Cheviron ZA. Divergent mitochondrial lineages arose within a large, panmictic population of the Savannah sparrow (*Passerculus sandwichensis*). *Mol Ecol*. 2019;28:1765–1783. https://doi.org/10.1111/mec.15049

SUPPORTING INFORMATION

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