Received: 21 May 2018 | Accepted: 27 August 2018

DOI: 10.1111/mec.14854

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Demographic inference in barn swallows using whole-genome data shows signal for bottleneck and subspecies differentiation during the Holocene

Chris C. R. Smith[1] | Samuel M. Flaxman[1] | Elizabeth S. C. Scordato[1,2] | Nolan C. Kane[1] | Amanda K. Hund[1] | Basma M. Sheta[3] | Rebecca J. Safran[1]

[1]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado

[2]Biological Sciences Department, California State Polytechnic University, Pomona, California

[3]Zoology Department, Faculty of Science, Damietta University, Damietta, Egypt

Correspondence
Chris C. R. Smith, BioFrontiers Institute, Boulder, CO.
Email: chriscs@colorado.edu

Funding information
National Science Foundation (NSF) Division of Environmental Biology, Grant/Award Number: 1149942, 1627483; NSF Division of Graduate Education, Grant/Award Number: 1144083; BioFrontiers IT; RMACC Summit Supercomputer; National Science Foundation, Grant/Award Number: ACI-1532235, ACI-1532236; UCB and Colorado State University

Abstract

Accounting for historical demographic features is vital for many types of evolutionary inferences, including the estimation of divergence times between closely related populations. In barn swallow, *Hirundo rustica*, inferring historical population sizes and subspecies divergence times can shed light on the recent co-evolution of this species with humans. Pairwise sequentially Markovian coalescent uncovered population growth beginning on the order of one million years ago—which may reflect the radiation of the broader *Hirundo* genus—and a more recent population decline. Additionally, we used approximate Bayesian computation to evaluate hypotheses about recent timescale barn swallow demography, including population growth due to human commensalism, and a potential founder event associated with the onset of nesting on human structures. We found signal for a bottleneck event approximately 7,700 years ago, near the time that humans began building substantial structures, although there was considerable uncertainty associated with this estimate. Subspecies differentiation and subsequent growth occurred after the bottleneck in the best-supported model, an order of magnitude more recently than previous estimates in this system. We also compared results obtained from whole-genome sequencing versus reduced representation sequencing, finding many similar results despite substantial allelic dropout in the reduced representation data, which may have affected estimates of some parameters. This study presents the first genetic evidence of a potential barn swallow founder effect and subspecies divergence coinciding with the Holocene, which is an important step in analysing the biogeographical history of a well-known human commensal species.

KEYWORDS

approximate Bayesian computation, demographic history, founder effect, PSMC, whole-genome sequencing

1 | INTRODUCTION

One aim of evolutionary biology is to diagnose the mechanisms that give rise to genomic divergence, which is often impacted by drift and demographic history (Whitlock & Lotterhos, 2015). Different components of historical demography, including changes in population size or migration, can leave footprints in the genetic variation among individuals of a population (Kingman, 1982). For example, historical population growth is expected to shift the site frequency spectrum to contain an excess of rare alleles (Tajima,

1989). Such historical changes are sometimes detectable by analysing genetic variation in the present population, if the change was substantial enough and sufficient data are available. Incorporating demographic information may improve the accuracy of population genetic analyses and help address long-standing questions in evolutionary biology (Wolf & Ellegren, 2017). For example, it is crucial to account for demography when estimating the age of divergence between closely related lineages. Although long-term neutral substitution rates are independent of population size, the effect of random drift is magnified in smaller populations, causing allele frequencies to change at a faster rate (Kimura, 1983; Wright, 1931). Therefore, when analysing populations with few or no fixed allelic differences, accounting for historical population sizes and other demographic features may reveal divergence times and evolutionary stories that are quite different than when a stable population size is assumed (Nielsen & Wakeley, 2001).

One of the most well-studied bird species, the barn swallow (*Hirundo rustica*), earned its common name from building its nest in barns, under bridges and on other human structures (Baird, Brewer, & Ridgway, 1874), perhaps because of the availability of water and aerial insect prey (Turner & Rose, 1989). Except for a few documented cases, barn swallows nest exclusively on human-made structures. The ecological relationship between barn swallows and humans is therefore an excellent example of a commensal symbiosis, and it is presumed that *H. rustica* populations have grown in response to the increased number of potential nest sites provided by humans (Møller, 1994). However, previous analyses of barn swallow demographic history are scarce (but see Zink, Pavlova, Rohwer, & Drovetski, 2006) and have not attempted to estimate the timing of population size changes. Although barn swallow ecology is closely intertwined with that of humans, it is uncertain how linked their evolutionary history is with ours.

Within the barn swallow species complex are six subspecies that breed in different regions of the Northern Hemisphere (Figure 1). Despite marked differences in plumage coloration (Turner & Rose, 1989) and differences in song (Wilkins et al., 2018), subspecies have relatively shallow genetic divergence (Safran et al., 2016). Previous studies have analysed the timing of the radiation of this species complex using several mtDNA loci and one nuclear locus (Dor, Safran, Sheldon, Winkler, & Lovette, 2010; Zink et al., 2006). These studies concluded that an early divergence event occurred between two main western and eastern clades on the order of 100,000 years ago (ya) and subsequent differentiation within each clade, meaning that the *H. rustica* subspecies began to differentiate long before substantial human architecture and agriculture began about 10,000 ya (Bogucki, 1999; Diamond, 1997), although within the time frame of some hominid expansions (Armitage et al., 2011).

The above studies were limited to a small number of genetic markers, which increases the potential for error in divergence time estimates. Evolutionary inferences are often unreliable when too few loci are examined (Funk & Omland, 2003), because individual gene trees may be incongruent with other aspects of the history of the population (Maddison, 1997). Incorporating additional loci usually has an averaging effect and conveys information about variation among loci, which can improve the accuracy of phylogenetic and demographic parameter estimates. Reduced representation sequencing approaches (broadly referred to as genotyping by sequencing, GBS) are common in population genetics (Shafer, Gattepaille, Stewart, & Wolf, 2015), because they provide a relatively affordable option for exploring genomewide variation in many individuals and typically generate tens of thousands of SNPs. Whole-genome sequencing (WGS), which generates even more data, is applied increasingly often in studies of nonmodel organisms. In the current study, we reconstruct the demographic history of the barn swallow using two different high-throughput sequencing strategies, WGS and GBS. Thus, in addition to examining the demography of our study system, we are also able to compare parameter estimates derived from the two different genomic data sets.

Advances in sequencing and computational technology, as well as comparative studies within this species (Romano, Costanzo, Rubolini, Saino, & Møller, 2017; Safran et al., 2016; Scordato et al., 2017), justify a contemporary analysis of barn swallow evolutionary history. Here, we first conduct an exploratory analysis of ancient population sizes using the pairwise sequentially Markovian coalescent (PSMC; Li & Durbin, 2011) model. Next, incorporating information gained from PSMC, we simultaneously estimate the timing of subspecies divergence and investigate hypotheses about the recent demography of barn swallows in relation to human civilization. Given what is known about barn swallow ecology, we hypothesized that *H. rustica* population sizes have expanded with the increased availability of potential nest sites on human-made structures. Likewise, it is conceivable that a small number of founding individuals first exploited the niche provided by human structures, representing a founder event. We evaluate these, and other, demographic scenarios using approximate Bayesian computation (ABC).

Due to the intricate phylogeographical history of the *H. rustica* species complex, we analysed only two subspecies, representing the two major clades noted above. In addition to minimizing the number of model parameters, including fewer subspecies has the benefit of retaining more loci with adequate sequencing coverage in each subspecies. *Hirundo rustica savignii* is distributed along the Nile Valley of Egypt and is nonmigratory. *Hirundo rustica erythrogaster* breeds across most of continental North America and migrates seasonally to South America. These subspecies are allopatric representatives of the two main barn swallow clades, and differentiation between them has been accumulating since the time of the oldest divergence event within *H. rustica*. Therefore, dating the age of differentiation between these two subspecies should approach the time when the species complex began to differentiate and thereby enable us to test the hypothesis that divergence and expansion in barn swallows are associated with the onset and expansion of human settlements.

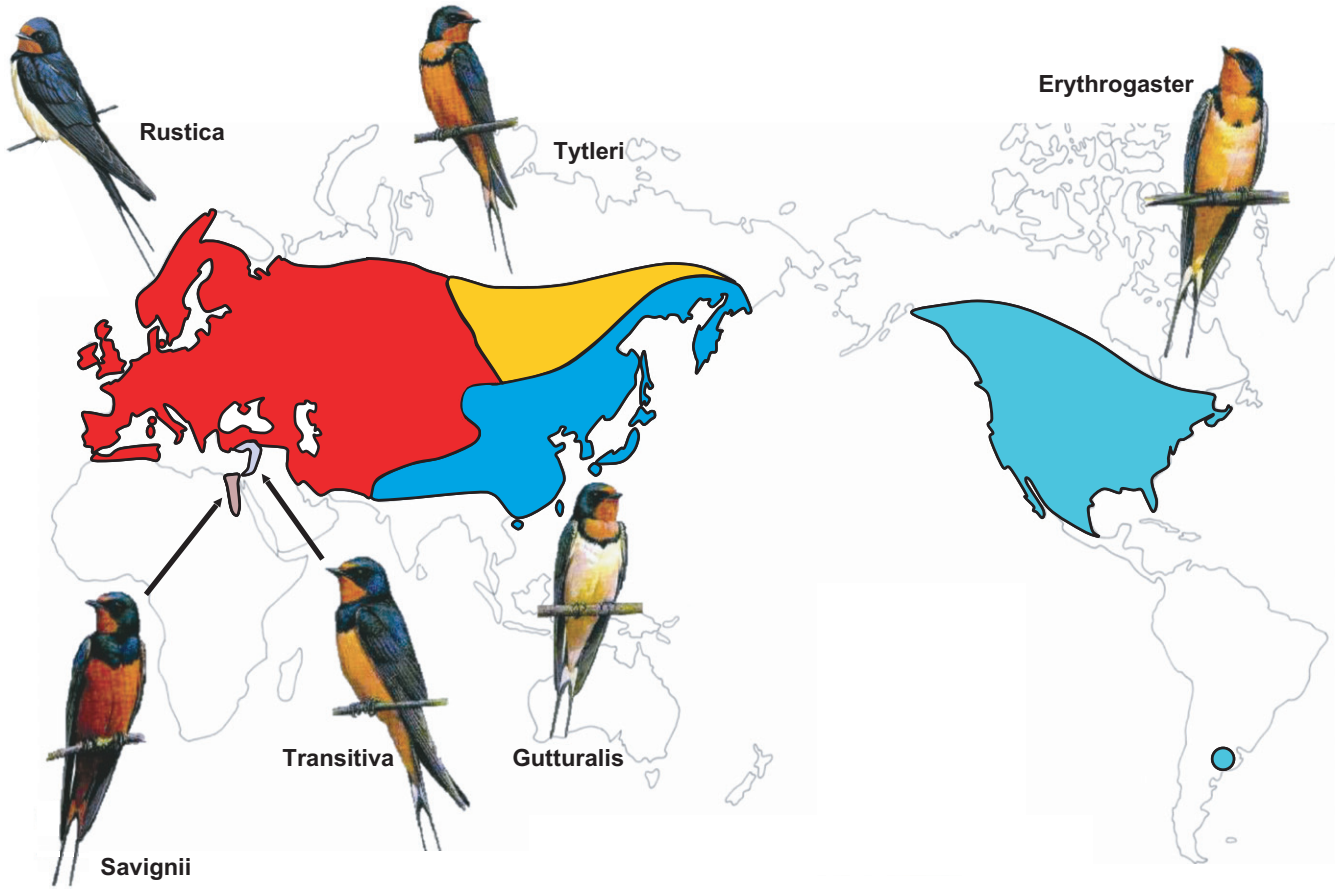

FIGURE 1 Subspecies breeding range map (Scordato et al., 2017) with drawing of a typical male phenotype from each subspecies. Barn swallow drawings courtesy of artist Hilary Burn

2 | MATERIALS AND METHODS

2.1 | Sample collection and DNA extraction

Although we focused our analysis on data from two subspecies, *H. r. savignii* (Egypt) and *H. r. erythrogaster* (Colorado), samples from various parts of the Northern Hemisphere were included in variant calling to improve genotyping accuracy. Sampling was conducted during barn swallow breeding seasons, April–July 2013 in Russia; April–July 2014 in China, Mongolia and Japan; January–June 2015 in Israel; April 2015 in Egypt; May–June 2015 in China and Colorado; and April–May 2016 in Morocco. See Supporting Information Table S1 for a list of sample locations. Birds were caught in mist nets, and approximately 80 μl of blood was collected via brachial venipuncture and stored in Queen's lysis buffer (Seutin, White, & Boag, 1991). We extracted genomic DNA from blood samples using DNeasy Blood and Tissue Kits (Qiagen), following the standard protocol modified to include an overnight digestion. All samples were collected and transported in accordance with permits to RJS from university, state and federal agencies and in accordance with the University of Colorado's Animal Care and Use Policy, IACUC # 1603.01 to RJS. To obtain genetic markers for our analyses, we conducted two different types of sequencing, GBS and WGS.

2.2 | Whole-genome sequencing

Whole-genome sequencing was performed on DNA from 168 individuals chosen to be representative of named subspecies as well as regions of putative hybridization. Eight *H. r. savignii* and eight *H. r. erythrogaster* samples were included. Libraries were prepared at the University of Colorado BioFrontiers sequencing centre using the ILLUMINA NEXTERA XT kit with version 2 dual indexes at 0.5× volume. All individual libraries were combined into a single pool and size-selected from 300 to 700 bp on a 1.5% agarose gel using a PippinPrep. Concentrations were checked by qPCR and TapeStation. The final pool had a concentration of 1.76 ng/μl and average fragment size of approximately 500 bp. Pooled libraries were run on two replicate lanes of an Illumina NovaSeq at Novogene Genome Sequencing Company (Chula Vista, CA, USA), sequencing 150-bp paired-end reads.

2.3 | Genotyping by sequencing

Libraries were prepared following the protocol of Parchman et al. (2012), which is similar to what Elshire et al. (2011) term "genotyping by sequencing." Details of the specific application of this protocol to barn swallows are in Safran et al. (2016) and Scordato

et al. (2017). For brevity, we refer to this sequencing protocol as "GBS." We used the restriction enzymes MseI and EcoRI to digest genomic DNA and ligated a unique 8, 9 or 10-base barcode to the fragment libraries for each individual. We pooled barcoded samples and amplified fragments using standard Illumina primers. Libraries were size-selected (350–400 base pairs) using a PippinPrep quantitative electrophoresis unit (Sage Science). We sequenced 100-bp single-end reads on Illumina HiSeq 2500 and 4000 platforms, with PhiX control to calibrate sequencing error rates and improve cluster discrimination and base calling, at the University of Texas, Austin Genomic Sequencing and Analysis Facility. Raw reads were obtained from the facility without any adapter sequences.

We ran three different libraries on four replicate lanes (12 lanes total) to ensure adequate coverage. Samples from Russia were run in 2013 with 536 samples multiplexed per lane; 512 barn swallow samples from China, Mongolia and Japan, plus 23 samples from another study (535 total samples), were sequenced in multiplex in 2015; and samples from China, Morocco, Egypt and Colorado were run with 546 multiplexed samples in 2016. The 2016 library included 47 samples from the two previous libraries for evaluating potential lane and year effects (see Supporting Information). We ran roughly the same number of samples per lane for each library to ensure consistency in read depth per individual. After combining reads from multiple years for some samples, and excluding two samples with zero reads, the GBS data set totalled 1,545 barn swallow samples, which includes 36 *H. r. savignii* and 26 *H. r. erythrogaster*. All of the following bioinformatics, modelling and analytical steps were applied to both the WGS and GBS data sets, with the exceptions of sample size and locus lengths (see below).

2.4 | Filtering sex chromosomes

This study utilized the barn swallow draft reference genome assembled by Safran et al. (2016). To avoid sex-linked loci, barn swallow scaffolds were required to align to flycatcher or chicken autosomes. Specifically, barn swallow scaffolds were aligned to (a) 27 autosomes (LG16 not available) and Z chromosome from flycatcher (Ellegren et al., 2012); (b) the flycatcher mitochondrial genome; and (c) chromosome W from chicken (International Chicken Genome Sequencing Consortium, 2004) using BLASTn with percentage identity >75 and $e < 10^{-50}$. Barn swallow scaffolds that aligned unambiguously to a single autosome were included in subsequent analyses.

2.5 | Trimming and alignment

TRIMMOMATIC (version 0.36) was used to (a) remove bases with quality below 30 from the start and end of each read; and (b) remove reads with average quality below 30 or length below 50 bp. Trimmed reads were aligned to the barn swallow reference using BWA MEM (version 0.7.12).

2.6 | Pairwise sequentially Markovian coalescent

The PSMC program (Li & Durbin, 2011) was used to estimate historical effective population sizes using individuals from North America and Egypt. This method is usually applied to individual diploid samples by inferring historical recombination events, estimating the time to most recent common ancestor (TMRCA) between the alleles at each independent locus and then inferring historical effective population sizes from the distribution of TMRCA values using the theory that population size is inversely proportional to the rate of coalescence. The PSMC does not require explicit demographic hypotheses, or phased data, and is therefore a straightforward way to explore past population size history. However, Li and Durbin point out that the power to analyse very recent, <20 kya, or old, >3 Mya, population sizes is limited with PSMC.

SAMTOOLS (version 1.5) and the BCFTOOLS (version 1.5) consensus callers using a minimum read depth of five were used to call variants in individual samples, and a diploid consensus for each individual was created with vcfutils.pl. Higher minimum read depths resulted in (a) substantial data loss due to the low-to-moderate mean coverage of the WGS data set; and (b) only slight differences in heterozygous genotype calls in the WGS data set (see Supporting Information). The PSMC program and associated scripts were run using the -p flag "4 + 25*2 + 4+6″ (Li & Durbin, 2011), which corresponds to four time intervals spanned by the first population size parameter, two time intervals spanned by the next 25 parameters and the last two parameters spanning four and six intervals. We applied the mutation rate recently estimated for flycatcher using pedigree analysis (Smeds, Qvarnström, & Ellegren, 2016), 2.3×10^{-9}, and assumed 1-year generations (Zink et al., 2006).

2.7 | Sequence processing for ABC

For GBS reads, the restriction sequence was required to be fully intact to include each read, and the restriction sequence was subsequently removed to avoid skewing genetic diversity calculations. Reads from both the WGS and GBS data sets were trimmed and aligned to the barn swallow reference, as above. The SAMtools multiallelic caller was used to call variants in all samples in the WGS data set simultaneously and separately in the GBS samples. All samples were included in variant calling to improve genotyping accuracy, although we focused our analysis only on data from Egypt (*H. r. savignii*) and Colorado (*H. r. erythrogaster*).

Ideally, entire chromosomes would be analysed to maximize the information obtained from whole-genome sequencing. However, in addition to computational constraints, we did not have a genetic map for barn swallows and thus could not appropriately model locus-specific recombination. Therefore, we examined short blocks of sequence, hereafter "loci," to minimize the effects of assuming no recombination within loci (Hung, Drovetski, & Zink, 2017; Lohse, Harrison, & Barton, 2011; Robinson, Bunnefeld, Hearn, Stone, & Hickerson, 2014; Veeramah et al., 2015; Wakeley, King, & Wilton, 2016). Loci were 200–500 bp in the WGS data set, and 50–86 bp in the GBS data set, and were spaced at least 10 kb apart to ensure

independence. As before, a minimum of five reads were required to count an individual sample as covered, and five samples from each of the two subspecies being analysed, *H. r. savignii* and *H. r. erythrogaster*, were required to count a genomic position as sufficiently covered across individuals and populations. Using data from more than five samples theoretically leads to diminishing returns for demographic inference (Robinson et al., 2014). The basic algorithm we used to incorporate the above criteria was as follows: (a) we scanned the alignment files one genomic position at a time; (b) if the position had sufficient coverage across samples, we considered the position as the starting point of a locus for potential inclusion; (c) the locus was extended if each subsequent position was sufficiently covered; (d) if a position was encountered with insufficient coverage, or if the maximum read length was reached, we ended the locus; (e) if the locus had reached the minimum read length, we retained the locus; and (f) we then skipped 10Kb before attempting to start a new locus. Custom scripts used in this study are available on GitHub: https://github.com/c70smith/BarnSwallow2subspeciesAnalysis.

2.8 | Demographic models

Unlike PSMC, which is exploratory and does not rely on explicit hypotheses, ABC evaluates user-defined demographic models. The below scenarios were analysed with ABC to test hypotheses about recent timescale population size changes and the timing of subspecies divergence. However, the initial PSMC analysis indicated a substantial, old population expansion (see Section 3). Therefore, we incorporated a comparable expansion into each of our models for ABC, leaving the magnitude and timing of the expansion uncertain. Specifically, all models included a component of ancient population growth with prior range 100 kya to 2 Mya (all priors are summarized in Table 1). This ancient demographic event inferred by PSMC was included to ensure a sufficient model fit (Gelman, Carlin, Stern, & Rubin, 1995), thereby avoiding erroneous conclusions when evaluating other aspects of demographic history. For example, an ancient expansion that has substantially shaped patterns of genetic variation, if unaccounted for, may lead to a false-positive test for a more recent expansion. Likewise, a model that is unable to simulate data similar to the observed data will most likely produce inaccurate divergence time estimates. More recent population size changes were modelled as described below and summarized in Figure 2.

We first examined a basic null model, Model 1 (Figure 2), with no recent population size changes. Incorporating the expansion inferred by PSMC, the ancestral effective population size prior ranged from 1,000 to 1,000,000 and expanded to achieve a population size between 10,000 and 10,000,000. To address overlapping population size prior ranges, we required expansion or decline events to at least double or halve the population, respectively. Uniform distributions were used for priors spanning less than two orders of magnitude, or conditional event times, and log-uniform priors were used otherwise (Meeker, Hahn, & Escobar, 2017; Ramos & Arregui, 2018; Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010). Model 1 also included a divergence event where the ancestral population splits into two contemporary populations representing the two analysed barn swallow subspecies, *H. r. savignii* and *H. r. erythrogaster*. The divergence time ranged from 100 ya to 1 Mya. As before, we leveraged the mutation rate estimated in flycatcher and used their confidence intervals (1.7×10^{-9} to 3.0×10^{-9} mutations per site per year) as the mutation rate prior range in this study.

TABLE 1 Prior information

Parameter	Abbr.	Prior distribution	Models(s)
Ancestral population size	N_a	log-U(10^3, 10^6)	All
Population 1 size	N_S	log-U(10^4, 10^7)	1,2,3,4,5,6,9
Population 2 size	N_E	log-U(10^4, 10^7)	1,2,3,4,5,6,9
Intermediate population 1 size	N_{i1}	log-U(10^3, 10^7)	2,3
Intermediate population 2 size	N_{i2}	log-U(10^3, 10^7)	3
Expanded population 1 size	N_{e1}	log-U(10^4, 10^7)	4,5,6,7,8
Contracted population 1 size	N_{b1}	log-U(10^3, 10^6)	4,5,6,7,8
Contracted population 2 size	N_{b2}	log-U(10^3, 10^6)	4,6,7,8,9
Expanded population 2 size	N_{e2}	log-U(10^4, 10^7)	6,8
Age of old expansion	T_e	U(10^5, 2×10^6)	All
Age of divergence	T_d	log-U(10^2, 10^6)	All
Age of growth	T_{g1}	log-U(10^2, 10^6)	2,3,4,5,6
Age of population 2 growth	T_{g2}	U(10^2, 10^6)	3,4,6,9
Age of contraction	T_{b1}	log-U(10^3, 10^6)	4,5,6,7,8
Age of population 2 contraction	T_{b2}	U(10^3, 10^6)	6,8
Mutation rate	μ	U(1.7×10^{-9}, 3×10^{-9})	All

"U" is uniform distribution and "log-U" is log uniform. Column four lists the models (Figure 2) that include each parameter. Note that a uniform prior distribution was substituted for event times that were conditionally more recent than the focal parameter, e.g., divergence time in Model 2 used a uniform prior instead of log-uniform.

Next, we incorporated recent timescale population size changes into the following models (Figure 2). In Model 2, we addressed the hypothesis that *H. rustica* population sizes have grown with the increased availability of potential nest sites on human-made structures by including a component of recent population growth before the divergence event between the two subspecies. The prior range for the timing of this demographic event was kept wide, 100–1,000,000 ya, to simultaneously explore population changes unassociated with human commensalism. The intermediate population size prior ranged from 1,000 to 10,000,000 individuals. Model 3 tested a similar hypothesis, except that the expansion happened after divergence. In Model 4, we tested for evidence of a founder event, where relatively few individuals from the ancestral population first exploited human structures, by modelling a bottleneck, followed by population divergence and subsequent growth. The prior ranges for

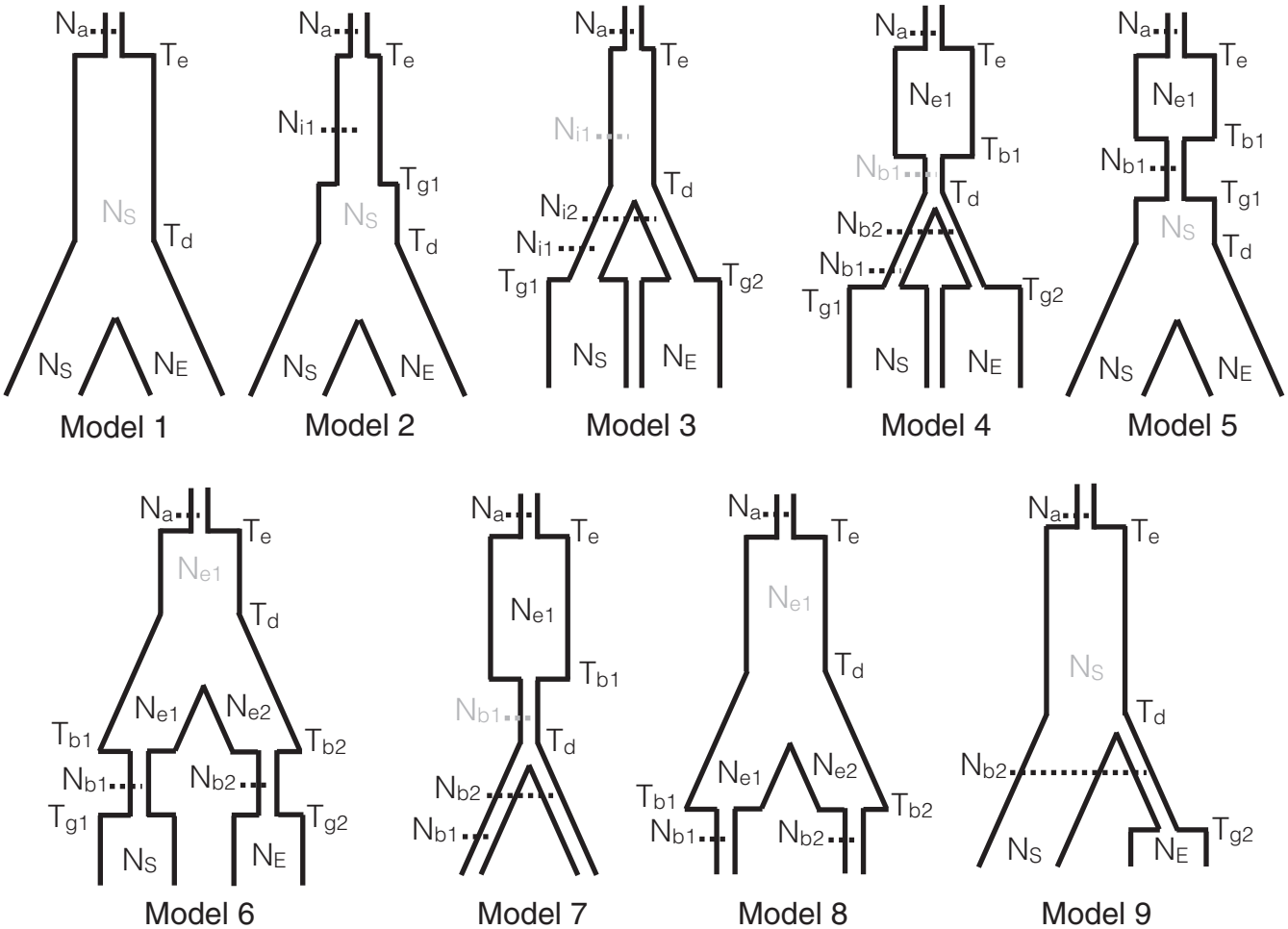

FIGURE 2 Demographic model diagrams, not drawn to scale. Priors are described in Table 1. All models include an ancient population expansion. Model 1 includes a divergence event, without additional population size changes. Model 2 includes a more recent growth event, followed by divergence. In Model 3, growth occurs after divergence. Model 4 includes a bottleneck, followed by divergence and subsequent growth. In Model 5, divergence occurs after recovery. In Model 6, a bottleneck occurs in each population after divergence. Model 7 includes a population size decline without recovery, before divergence. The decline occurs after divergence in Model 8. Model 9 includes a bottleneck in *H. r. erythrogaster* only

the bottleneck time and the contracted population size were 1,000–1,000,000 years ago and 1,000–1,000,000 individuals. Models 5 and 6 were similar to Model 4, except with a different order for events (see Figure 2). Models 7 and 8 represent a hypothesis of population decline without recovery, where the decline occurs before and after divergence, respectively. Last, Model 9 addressed the hypothesis that *H. r. erythrogaster* alone underwent a founder event during the colonization of North America.

We used conditional event time priors to ensure that the events within each simulation followed the specified order. However, applying such conditions unavoidably results in skewed prior distributions for at least some event times, which can affect the resulting biological inferences. To ensure that the parameter of interest can be estimated with as much precision and as little bias as possible, we prevented skewing the focal parameter using one-way conditional priors (Rougemont et al., 2016; Supporting Information Figure S1). Specifically, in Models 1, 3, 6, 8 and 9, the divergence time effective prior remained true log-uniform, while the subsequent event time priors became skewed due to the specified event time conditions; in Model 2, the expansion time prior remained true log-uniform; and in Models 4, 5 and 7, the bottleneck time prior remained true log-uniform. Uniform priors were used for event times conditionally more recent than the focal parameter.

2.9 | Simulations and summary statistics

Neutrally evolving DNA under each demographic model (Figure 2) was simulated using a modified version (see Supporting Information) of the program msABC (Pavlidis, Laurent, & Stephan, 2010), a wrapper of Hudson's ms (Hudson, 2002). We simulated 10^5 data sets for each model and a total of 10^6 data sets under the eventual best-supported model. θ was scaled for each locus to reflect observed locus lengths. Two sequences were simulated to represent each diploid individual in the data set. For each simulated and observed data set, we used msABC to calculate summary statistics including the number of segregating sites (S), π, Watterson's estimator (θ_W), Tajima's D, F_{ST} (as

calculated by Hudson, Boos, & Kaplan, 1992), the proportion of shared variants and the proportion of fixed differences. The average and variance across loci were calculated for each summary statistic, both within individual populations when appropriate and across all individuals, totalling to 30 statistics (Supporting Information Table S2). This set of summary statistics was used for comparing models and estimating demographic parameters. Other statistics calculated by msABC either required out-group DNA or phased data, which we did not have, or were redundant with other statistics (e.g., proportion shared alleles = 1 − proportion private alleles).

2.10 | Model selection and parameter estimation using random forests

Random forests implemented in the R package *abcrf* were used for both model selection and parameter estimation, following the recommendations of Pudlo et al. (2015) and Raynal et al. (2016). Random forests are a supervised machine learning algorithm suited for high-dimension classification or regression problems (Breiman, 2001). The algorithm is named due to training many individual decision trees, or predictive models, on random subsets of the data to achieve various "perspectives," before averaging the predictions of all trees. This method has recently been implemented for ABC and has multiple advantages over preceding ABC techniques, including the avoidance of the ABC rejection step that requires an arbitrary tolerance value to be specified by the user. However, the random forest method does come with its own parameterization, including the number of trees and the number of data sets used by each tree. The default settings were used in this study. The sensitivity to the number of tree predictors was evaluated for each test by visualizing the change in the prediction error obtained with different numbers of trees (Pudlo et al., 2015; Raynal et al., 2016). Event time estimates were interpreted assuming a 1-year generation time (Zink et al., 2006).

2.11 | Pseudo-observed data sets

Ten thousand pseudo observed data sets (PODs) were simulated under the best-supported model using the priors specified above, and the parameter of interest was estimated for each POD using random forests, as described above. The median relative absolute error was calculated by subtracting the point estimate (posterior median) from the true parameter value, taking the absolute value and dividing by the true value for each POD, before finding the median of the resulting values. For visualizing estimates for specific parameter values, a separate group of PODs were simulated with the focal parameter fixed at discrete values ranging between 1,000 and 100,000.

3 | RESULTS

3.1 | Sequence processing

Sequences from both data sets had high quality on average, resulting in little data loss from quality trimming (Supporting Information Figure S2). Samples in the WGS data set had average raw sequencing coverage of 5.8 (Lander & Waterman, 1988) based on the estimated barn swallow genome size from Andrews, Mackenzie, and Gregory (2009), 1.28 Gb. The average aligned read depth excluding zero-depth sites (SAMtools depth) across loci and across samples was 6.5 for the WGS data and much lower, 0.66, for the GBS data. Our locus-filtering procedure produced a total of 79,440 loci from the WGS data and 24,347 loci from the GBS data. The total length of analysed genomic sequence in the final WGS data set represented approximately 2.5% of the genome, which was about 15× that of the GBS data set. Due to our filtering procedure, the proportion of sequence represented by GBS loci that overlapped with the WGS loci was only 6%, and <1% of the retained WGS sites were represented in the GBS loci.

Observed summary statistics appeared mostly consistent between WGS and GBS data sets, although most are not directly comparable due to different locus lengths (Supporting Information Table S2). However, one important discrepancy was the number of fixed alleles between subspecies, which was zero in the WGS loci and four total in the GBS loci. With more samples and fewer loci, fewer fixed differences were expected in the GBS data set. This indicated an overrepresentation of fixed alleles, evidence that allelic dropout is causing heterozygotes to be missed. Comparing genotypes between data sets showed that 31% of confident heterozygotes (≥10 reads) in the WGS analysis were called homozygous in the GBS analysis using a minimum depth of five (see Supporting Information). In the light of this, we repeated our locus selection procedure in the GBS data set requiring a minimum of ten reads at a locus to consider an individual covered at the locus. The new data set contained a reduced 10,914 loci representing about 33× less sequence than the WGS loci and continued to miss 29% of heterozygotes. Due to the apparent shortcomings of the GBS data sets, and much smaller amount of sequence information, we focused on the WGS analysis for empirical inferences in this study and secondarily present the GBS results for comparison.

3.2 | PSMC

The PSMC analysis provided exploratory information about the demographic history of barn swallows (Figure 3a). Population growth older than 100 kya, perhaps older than 1 Mya, was detected in all analysed samples. Note that sudden or rapid population size changes may present as apparent gradual growth using PSMC (Li & Durbin, 2011; Liu & Hansen, 2017). Therefore, in our case, what appears in the PSMC plot as gradual growth concluding ≈150 kya could be caused by older, more rapid growth. Nevertheless, the beginning of this growth, ≈1 Mya, may correspond to the timing of the radiation of the genus, if roughly estimated using the mitochondrial divergence rate 2% per million years (Ho, 2007) and the lower end of mtDNA distances within *Hirundo*, 2% (Dor et al., 2010). The history of population growth indicated by PSMC was consistent with the observed, negative Tajima's *D* of −1.1. Tajima's *D* reflects the shape of the site frequency spectrum, and a strongly negative Tajima's *D*

signifies population growth, assuming neutrally evolving DNA (Tajima, 1989). The observed Tajima's *D* value in this study could not be simulated without an old (>200 kya) expansion (Figure 3b). Therefore, the ancient expansion indicated by PSMC was supported by this classic population genetics statistic.

Following the expansion, the PSMC plot indicated a population decline. Just as a theoretical sudden expansion may cause apparent gradual growth, a more recent, rapid population decline may cause apparent gradual decline. Last, many samples, but not all, showed signal for a very recent expansion after the decline. However, it should be noted that PSMC lacks power for inferring population sizes approaching 20 kya or more recent (Li & Durbin, 2011), because relatively few mutations and recombination events have had time to accumulate in the two analysed sequences. Therefore, the recent timescale population size history of this system requires additional analysis.

Because genomes collected from both subspecies had similar TMRCA distributions, it seems that much of the demographic history of the subspecies is shared. In particular, the PSMC curves from each subspecies overlapped until <100 kya, hinting that divergence may have occurred more recently than previously estimated. Sample 1,607 appeared slightly inconsistent with the other samples, but also had the lowest sequencing coverage. Furthermore, artificial F1 hybrid sequences constructed through "haploidization" were compared between subspecies using PSMC (Cahill, Soares, Green, & Shapiro, 2016), and loci began to coalesce ≈40 kya (Supporting Information Figure S6), representing an upper limit for divergence time.

3.3 | ABC model selection

We next used ABC to test explicit hypotheses about recent population history. Using random forests to choose among all nine demographic scenarios, some of which are qualitatively similar, Model 4 was chosen in the WGS analysis with posterior probability (PP) =

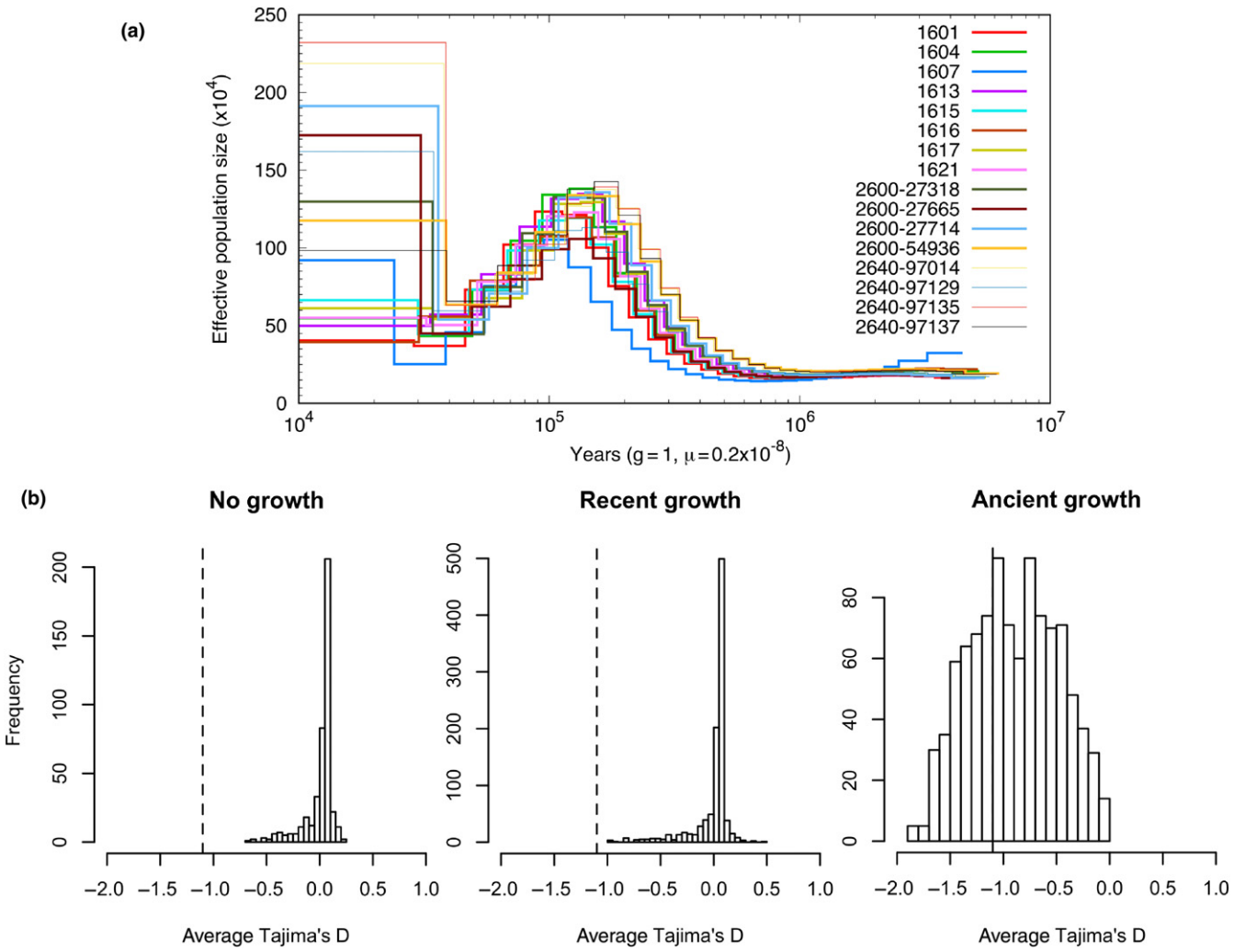

FIGURE 3 (a) PSMC results. The first eight samples in the legend are *H. r. savignii*, and the second eight are *H. r. erythrogaster*. (b) Average Tajima's *D* from 1000 simulations of each of the following variations of Model 1: (left) no growth; (center) population growth more recent than 200 kya; and (right) ancient growth, 100kya to 2mya. The observed Tajima's *D*, -1.1, is indicated by the dashed line

0.60 (model classification votes in Supporting Information Table S3) and prior error rate (PER), or misclassification rate = 0.36. Model 4 represents a bottleneck scenario—population size decline followed by recovery—where the recovery occurs in each sister population after subspecies divergence. This model was designed to test the hypothesis that barn swallows experienced a founder effect during the transition to nesting on human-made structures. Model 4 was able to replicate each of the observed summary statistics (Supporting Information Figure S3). For each model choice and parameter estimation step using random forests, we found that the default number of trees, 500, was sufficient to stabilize prediction error, with larger numbers of trees giving diminishing returns on statistical precision (Supporting Information Figure S4).

Because several demographic models were qualitatively very similar, we next examined particular subsets of models. When the models other than Model 4 that include population decline—Models 5, 6, 7 and 8—were excluded, Model 4 was more easily distinguished from the models without a decline (PP = 0.86; PER = 0.28). Population decline was therefore supported by both PSMC and our ABC analysis. To evaluate the evidence of a postbottleneck recovery, we compared only Models 4 and 7, which are identical except that Model 7 does not include a postbottleneck expansion, and Model 4 was supported (PP = 0.67; PER = 0.12). However, the signal for postbottleneck recovery was perhaps not as strong as that of the initial population decline. To evaluate how important the relative order of the divergence and recovery events was, we compared Models 4 and 5, which are similar except that recovery comes before divergence in Model 5, and Model 4 was supported (PP = 0.71; PER = 0.11). In summary, these results show evidence for a population size decline before divergence, with a postbottleneck expansion in each population. Note that the posterior probabilities and error rates associated with these tests indicated uncertainty in model selection, and therefore, the other models could not be completely refuted, although they received less support than Model 4.

The more stringently filtered GBS data set also supported Model 4 (PP = 0.55). Using the less stringently filtered GBS data set, Model 7 was chosen (PP = 0.52; PER = 0.36); however, when Models 4 and 7 were compared alone, Model 4 was supported (PP = 0.84; PER = 0.37), which highlights that (a) statistical power using the GBS data set may be lacking for model selection or (b) that the classification algorithm depends on the number or combination of models included.

3.4 | ABC parameter estimation

Parameters were estimated for the best-supported demographic scenario, Model 4, which included a population bottleneck. The timing of the bottleneck is of primary interest, because if the decline was as recent as the beginning of human architecture, it would support the founder effect hypothesis. The age of the decline was estimated to be 7,700 ya (the median of the posterior is reported here and subsequently) with 95% credible interval (CI) 1,100 to 163,000 ya in the WGS analysis, indicating that older bottleneck times are possible (Figure 4). Estimating the focal parameter for pseudo-observed data sets simulated under Model 4 gave reasonably accurate point estimates (Supporting Information Figure S5); specifically, the median relative absolute error for bottleneck time was 0.33, indicating that half of the estimates were within 33% of the true value. When the pseudo-observed bottleneck age was fixed at 25,000 ya or 50,000 ya, the parameter was estimated to be as small as 7,700 ya in only 16% and 7% of PODs, respectively. The following were point estimates for other Model 4 parameters, although the CIs were wide for all parameters (Supporting Information Table S4), and signal for some parameters may not be as strong as signal for the focal parameter. The bottleneck decreased the estimated ancestral population size of 7×10^6 down to 4.6×10^4. Subspecies divergence was estimated to have occurred 3,550 ya, followed by expansion 1,450 and 1,700 ya in *H. r. savignii* and *H. r. erythrogaster*, respectively, expanding the effective population sizes to an estimated 1.7×10^6 and 1.1×10^6.

Using the ten-minimum read depth GBS data set, we estimated the timing of the bottleneck in Model 4, which gave 47,000 ya (CI 1,200–596,000 ya). Using the five-minimum read depth GBS data set gave 80,000 ya (CI 1,200–780,000 ya) for the same parameter. Both the model selection and parameter estimation results perhaps indicate a trend of decreasing power with GBS, especially when insufficient filtering steps are used.

4 | DISCUSSION

4.1 | Barn swallow demographic history

Natural historians report that barn swallows as a species once nested in caves and on cliffs and only relatively recently have experienced an opportunistic ecological transition to nesting in barns and

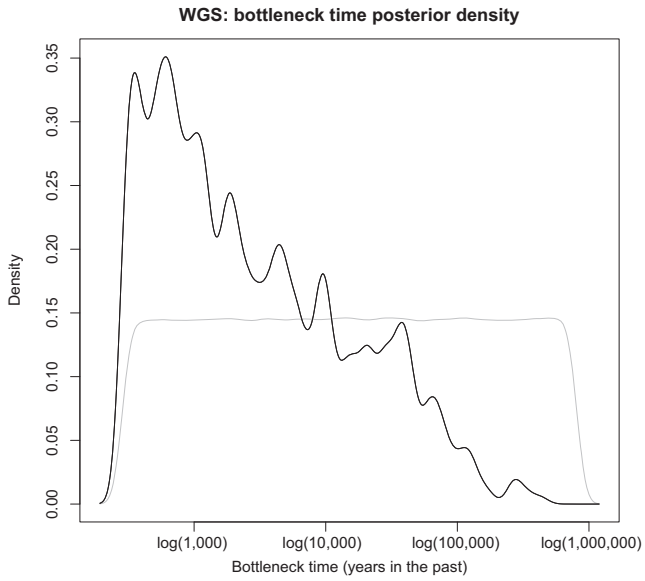

FIGURE 4 Posterior histogram for bottleneck time in Model 4, generated using the *abcrf* package. The x-axis is log-scaled. The black and grey lines are the posterior and prior densities, respectively

on bridges and other human structures (Turner, 2010). Evidence to support this hypothesis includes (a) anecdotal sightings of barn swallow on cliffs in the western USA during the nineteenth century (Baird et al., 1874); (b) anecdotal reports of an extant population nesting in caves on the Channel Islands in California; and (c) genetic analyses that estimated major barn swallow divergence events on the order of 10^5 years ago (Dor et al., 2010; Zink et al., 2006). However, the latter studies had access to the limited data. Our study leverages whole-genome data to uncover multiple population size changes within *H. rustica*. We demonstrate evidence for a population bottleneck on the order of 10^4 years ago, rather than 10^5 years ago, and even more recent divergence between subspecies.

We arrived at that conclusion as follows. The initial PSMC analysis uncovered population growth on the order of 1 million ya. Next, we used ABC to evaluate explicit hypotheses about more recent timescale demography, while accounting for the ancient growth inferred from PSMC. In the ABC analysis, a scenario with a population bottleneck was supported, and the bottleneck was estimated to have occurred 7,700 ya, although with upper CI limit >100,000 ya. The wide CIs for bottleneck time are likely attributable to the wide prior ranges and complex models used, but could also be due to a lack of strong signal. Because we included only two subspecies, representing the oldest split in the species complex, additional analyses are required to estimate fine-scale divergence times and other parameters among the other *H. rustica* subspecies.

Based on these results, we put forth a new hypothesis regarding the natural history of the *H. rustica* species complex. The earliest evidence of substantial human architecture is 12,000–15,000 years old, earlier than farming in some regions of the Middle East and Europe (Barker, 2009; Bogucki, 1999; Iakovleva & Djindjian, 2005; Mithen, 2004; Potts, 2012). Agriculture is thought to have originated in the Middle East as early as 12,000–13,000 ya (Hillman & Davies, 1990; Hillman, Hedges, Moore, Colledge, & Pettitt, 2001; Salamini, Özkan, Brandolini, Schäfer-Pregl, & Martin, 2002; Snir et al., 2015; Zeder, 2008). Early human settlements would have had some characteristics of the modern-day rural environments that barn swallows prefer, particularly the structures they use for nesting with nearby open fields and edge habitats that provide good foraging. In our analysis, the statistical evidence for a bottleneck in the ancestral population 7,700 years ago approaches the timing of the first human structures, which would support the hypothesis of a founder effect associated with the origin of human commensalism. In this scenario, a relatively small number of founding individuals may have taken advantage of the ecological niche created by human structures and agriculture and subsequently diverged from the ancestral swallow population, before expanding from their founding population size to occupy their current, Holarctic range (Turner & Rose, 1989). This scenario is arguably the more parsimonious explanation for the evolution and geographical expansion of barn swallows, compared with a scenario where *H. rustica* differentiated into subspecies and colonized most of the Northern Hemisphere before experiencing selection strong enough to cause convergent evolution of subspecies on all continents to nest on human structures.

Other biogeographical hypotheses are possible. For example, our estimates for bottleneck time coincide somewhat with the receding of glaciers in Beringia, before the covering of the Bering land bridge by the sea (Goebel, Waters, & O'rourke, 2008). It is possible that the subspecies *H. r. erythrogaster* experienced a population bottleneck during the colonization of North America during this time. However, we evaluated a model that included a bottleneck only in *H. r. erythrogaster* and instead found support for a bottleneck in the ancestral population which shaped genetic variation in both analysed subspecies. Furthermore, a land bridge is likely not required for swallows to disperse between Siberia and the Aleutian Islands, where barn swallows nest currently (Gibson, 1981). The decline in the past 100 ky indicated by PSMC may correspond to population reduction during the most recent ice age, while warming temperatures at the beginning of the Holocene ≈12,000 ya may have stimulated population growth in barn swallows, irrespective of human activity. However, our ABC analysis demonstrated signal for a population decline more recent than the previous ice age.

4.2 | Comparison of sequencing strategies

Multiple high-throughput sequencing technologies are now available to researchers for sequencing thousands of genetic markers simultaneously. GBS and other variations of reduced representation sequencing using restriction enzymes are common in population genetics (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013). However, WGS is used increasingly often in studies of nonmodel organisms (Ellegren, 2014). The primary differences between the two strategies are that WGS explores a greater proportion of the genome, while GBS produces data for a larger number of samples per locus. Additionally, WGS gives more even sampling at both alleles at heterozygous loci and is less subject to null alleles and other biases, thus giving more reliable estimates of allele frequencies. For example, 29% of confident heterozygous sites in our WGS data appeared homozygous using GBS with ≥10 read depth.

In our demographic inference application using ABC, similar model selection results were obtained using both data sets, although parameter estimates were different. The degree of similarity in model selection is perhaps surprising, considering that (a) the WGS data set explored much more of the genome; and (b) the proportion of overlapping loci between the two sequencing strategies is small due to our locus selection procedure. However, we did notice a difference in parameter estimation using GBS, which is likely due to allelic dropout. The WGS analysis indicated a bottleneck time near to the start of human agriculture and architecture, while the GBS data sets gave estimates that were much older. We give more weight to the WGS results, here, and infer that the GBS data are disadvantaged, because 15–33× more sequence is covered in the WGS data set and because heterozygotes were genotyped more accurately. Importantly, these results are dependent on the demography of the biological system, the specific models being analysed and the data processing steps used. In particular, if greater depth of sequencing is used with a reduced representation approach, it may

alleviate allelic dropout issues while retaining a sufficient number of loci for parameter estimation.

4.3 | Methodological considerations

The results from this study illustrate the utility of the complementary PSMC and ABC analyses applied to whole-genome data for demographic inference. PSMC is a straightforward method for exploring historical effective population sizes, while ABC provides a framework for evaluating explicit demographic hypotheses. The main improvements over previous research on barn swallows were including more sequence data and accounting for recent population size history when attempting to estimate subspecies divergence time. A fundamental problem this study engaged is attempting to infer very recent population size changes, for example <20 kya. Such recent events may be difficult to detect, because too few mutations and recombination events have had time to accumulate to reflect the demographic change.

Studies like ours would potentially benefit from more sophisticated summary statistics. For example, more confident phasing methods or longer sequencing reads would allow the calculation of haplotype diversity and linkage-based summary statistics. Likewise, sequencing DNA from an out-group would allow the calculation of summary statistics that require the identification of ancestral and derived alleles. In our case, however, estimating the TMRCA between barn swallows and an extant out-group would provide only an upper limit for the timing of the radiation of the barn swallow subspecies. It should also be noted that this study utilized the flycatcher mutation rate which is expected to differ slightly from that of barn swallow, and did not analyse potential population structure within subspecies. Other factors that may improve demographic analyses with WGS data include a genetic map and computationally efficient methods for simulating very long sequences with recombination. A promising alternative technique that requires phased input is the multiple sequential Markovian coalescent method (Schiffels & Durbin, 2014).

In summary, this study analysed the demographic history of two barn swallow subspecies and found evidence of growth in the ancestral population and a more recent bottleneck. The timing of the bottleneck approached the timing of the earliest human architecture, suggesting that the barn swallow ancestor experienced a founder effect during the transition to human commensalism before subspecies divergence. Furthermore, our analysis compared WGS and GBS sequencing technologies and showed that the two analyses gave similar tendencies for model selection, although perhaps differed in fine-scale parameter estimation.

ACKNOWLEDGEMENTS

We are grateful to the following people for assistance in the field: Matthew Wilkins, Georgy Semenov, Alexander Rubtsov, Gennady Bachurin, Nikolai Markov, Olga Zayetseva, Elena Shnayder and Yulia Sheina (Russia); Liu Yu, Caroline Glidden, Rachel Lock and Wei Liang (China); Sundev Gomboobaatar, Unurjargal Enkhbat, Bayanmunkh Dashnyam and Davaadorj Enkhbayar (Mongolia); Kazuo Koyama, Wataru Kitamura, Takashi Tanioka and Yuta Inaguma (Japan); Mamdouh Ahmed (Egypt); Sheela Turbek, Saad Hanane, Najib Magri, Said Hajib and Imad Cherkaoui (Morocco); and Joanna Hubbard and Yoni Vortman (Israel). This study was supported by funding from the National Science Foundation (NSF) Division of Environmental Biology awards 1149942 and 1627483, to RJS and SMF, respectively, and NSF Division of Graduate Education 1144083 award to CCRS. This work utilized both the BioFrontiers Computing Core at the University of Colorado at Boulder (UCB) supported by BioFrontiers IT and the RMACC Summit Supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), UCB and Colorado State University.

DATA ACCESSIBILITY STATEMENT

DNA sequences are associated with BioProject PRJNA323498 in the Sequence Read Archive.

AUTHOR CONTRIBUTIONS

E.S.C.S., N.C.K., A.K.H., B.M.S. and R.J.S. produced the data. C.C.R.S., S.M.F., E.S.C.S. and R.J.S. analysed the data. C.C.R.S., S.M.F., E.S.C.S., N.C.K., A.K.H. and R.J.S. wrote the manuscript.

ORCID

Chris C. R. Smith http://orcid.org/0000-0002-6470-3413
Elizabeth S. C. Scordato http://orcid.org/0000-0003-0224-8280

REFERENCES

Andrews, C. B., Mackenzie, S. A., & Gregory, T. R. (2009). Genome size and wing parameters in passerine birds. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1654), 55–61. https://doi.org/10.1098/rspb.2008.1012

Armitage, S. J., Jasim, S. A., Marks, A. E., Parker, A. G., Usik, V. I., & Uerpmann, H. P. (2011). The southern route "out of Africa": Evidence for an early expansion of modern humans into Arabia. *Science*, *331*(6016), 453–456. https://doi.org/10.1126/science.1199113

Baird, S. F., Brewer, T. M., & Ridgway, R. (1874). *A history of North American birds* (Vol. 2). Boston, MA: Little, Brown, and Company. https://doi.org/10.5962/bhl.title.49274

Barker, G. (2009). *The agricultural revolution in prehistory: Why did foragers become farmers?* Oxford, UK: Oxford University Press on Demand.

Bogucki, P. (1999). *The origins of human society*. Oxford, UK: Blackwell Publishers.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cahill, J. A., Soares, A. E., Green, R. E., & Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1699), 20150138. https://doi.org/10.1098/rstb.2015.0138

Diamond, J. (1997). *Guns, germs, and steel: The fates of human societies*. New York, NY: W.W. Norton & Company.

Dor, R., Safran, R. J., Sheldon, F. H., Winkler, D. W., & Lovette, I. J. (2010). Phylogeny of the genus Hirundo and the Barn Swallow

subspecies complex. *Molecular Phylogenetics and Evolution*, *56*(1), 409–418. https://doi.org/10.1016/j.ympev.2010.02.008

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., … Uebbing, S. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*(7426), 756. https://doi.org/10.1038/nature11584

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 397–423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Gibson, D. D. (1981). Migrant birds at Shemya Island, Aleutian Islands, Alaska. *Condor*, *83*, 65–77. https://doi.org/10.2307/1367606

Goebel, T., Waters, M. R., & O'rourke, D. H. (2008). The late Pleistocene dispersal of modern humans in the Americas. *Science*, *319*(5869), 1497–1502. https://doi.org/10.1126/science.1153569

Hillman, G. C., & Davies, M. S. (1990). Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. *Journal of World Prehistory*, *4*(2), 157–222. https://doi.org/10.1007/BF00974763

Hillman, G., Hedges, R., Moore, A., Colledge, S., & Pettitt, P. (2001). New evidence of Lateglacial cereal cultivation at Abu Hureyra on the Euphrates. *The Holocene*, *11*(4), 383–393. https://doi.org/10.1191/095968301678302823

Ho, S. Y. (2007). Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology*, *38*(4), 409–414. https://doi.org/10.1111/j.0908-8857.2007.04168.x

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338. https://doi.org/10.1093/bioinformatics/18.2.337

Hudson, R. R., Boos, D. D., & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, *9*(1), 138–151.

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*(14), 3775–3784. https://doi.org/10.1111/mec.14152

Iakovleva, L., & Djindjian, F. (2005). New data on Mammoth bone settlements of Eastern Europe in the light of the new excavations of the Gontsy site (Ukraine). *Quaternary International*, *126*, 195–207. https://doi.org/10.1016/j.quaint.2004.04.023

International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511623486

Kingman, J. F. (1982). On the genealogy of large populations. *Journal of Applied Probability*, *19*(A), 27–43. https://doi.org/10.2307/3213548

Lander, E. S., & Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics*, *2*(3), 231–239. https://doi.org/10.1016/0888-7543(88)90007-9

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493. https://doi.org/10.1038/nature10231

Liu, S., & Hansen, M. M. (2017). PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular Ecology Resources*, *17*(4), 631–641. https://doi.org/10.1111/1755-0998.12606

Lohse, K., Harrison, R., & Barton, N. H. (2011). A general method for calculating likelihoods under the coalescent process. *Genetics*, *189*, 977–987. https://doi.org/10.1534/genetics.111.129569

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/sysbio/46.3.523

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). *Statistical intervals: A guide for practitioners and researchers* (Vol. *541*). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781118594841

Mithen, S. (2004). *After the ice: A global human history, 20,000–5000 BC*. Cambridge, MA: Harvard University.

Møller, A. P. (1994). *Sexual selection and the barn swallow*. Oxford: Oxford University Press.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–2847. https://doi.org/10.1111/mec.12350

Nielsen, R., & Wakeley, J. (2001). Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics*, *158*(2), 885–896.

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*(12), 2991–3005. https://doi.org/10.1111/j.1365-294X.2012.05513.x

Pavlidis, P., Laurent, S., & Stephan, W. (2010). msABC: A modification of Hudson's ms to facilitate multi-locus ABC analysis. *Molecular Ecology Resources*, *10*(4), 723–727. https://doi.org/10.1111/j.1755-0998.2010.02832.x

Potts, D. T. (Ed.) (2012). *A companion to the archaeology of the ancient Near East*. Chichester, UK: John Wiley & Sons.

Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2015). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866.

Ramos, A. A., & Arregui, I. (Eds.) (2018). *Bayesian astrophysics* (Vol. *26*). Cambridge, UK: Cambridge University Press.

Raynal, L., Marin, J. M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2016). ABC random forests for Bayesian parameter inference. *arXiv preprint arXiv:1605.05537*.

Robinson, J. D., Bunnefeld, L., Hearn, J., Stone, G. N., & Hickerson, M. J. (2014). ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology*, *23*(18), 4458–4471. https://doi.org/10.1111/mec.12881

Romano, A., Costanzo, A., Rubolini, D., Saino, N., & Møller, A. P. (2017). Geographical and seasonal variation in the intensity of sexual selection in the barn swallow *Hirundo rustica*: A meta-analysis. *Biological Reviews*, *92*(3), 1582–1600. https://doi.org/10.1111/brv.12297

Rougemont, Q., Roux, C., Neuenschwander, S., Goudet, J., Launey, S., & Evanno, G. (2016). Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. *PeerJ*, *4*, e1910. https://doi.org/10.7717/peerj.1910

Safran, R. J., Scordato, E. S. C., Wilkins, M. R., Hubbard, J. K., Jenkins, B. R., Albrecht, T., … Nosil, P. (2016). Genome-wide differentiation in closely related populations: The roles of selection and geographic isolation. *Molecular Ecology*, *25*(16), 3865–3883. https://doi.org/10.1111/mec.13740

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., & Martin, W. (2002). Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetics*, *3*(6), 429. https://doi.org/10.1038/nrg817

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919. https://doi.org/10.1038/ng.3015

Scordato, E. S., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology*, *26*, 5676–5691. https://doi.org/10.1111/mec.14276

Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology*, *69*(1), 82–90. https://doi.org/10.1139/z91-013

Shafer, A., Gattepaille, L. M., Stewart, R. E., & Wolf, J. B. (2015). Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: In silico evaluation of power, biases and proof of concept in Atlantic walrus. *Molecular Ecology*, *24*(2), 328–345. https://doi.org/10.1111/mec.13034

Smeds, L., Qvarnström, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research*, *26*(9), 1211–1218. https://doi.org/10.1101/gr.204669.116

Snir, A., Nadel, D., Groman-Yaroslavski, I., Melamed, Y., Sternberg, M., Bar-Yosef, O., & Weiss, E. (2015). The origin of cultivation and proto-weeds, long before Neolithic farming. *PLoS One*, *10*(7), e0131422. https://doi.org/10.1371/journal.pone.0131422

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, *123*(3), 597–601.

Turner, A. (2010). *The barn swallow*. London, UK: Bloomsbury Publishing.

Turner, A., & Rose, C. (1989). *Swallows and martins: an identification guide and handbook* (No. 598.2 TUR).

Veeramah, K. R., Woerner, A. E., Johnstone, L., Gut, I., Gut, M., Marques-Bonet, T., … Hammer, M. F. (2015). Examining phylogenetic relationships among gibbon genera using whole genome sequence data using an approximate Bayesian computation approach. *Genetics*, *200*(1), 295–308. https://doi.org/10.1534/genetics.115.174425

Wakeley, J., King, L., & Wilton, P. R. (2016). Effects of the population pedigree on genetic signatures of historical demographic events. *Proceedings of the National Academy of Sciences*, *113*(29), 7994–8001. https://doi.org/10.1073/pnas.1601080113

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 116. https://doi.org/10.1186/1471-2105-11-116

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *The American Naturalist*, *186*(S1), S24–S36. https://doi.org/10.1086/682949

Wilkins, M. R., Scordato, E. S. C., Semenov, G., Karaardıç, H., Shizuka, D., Rubtsov, A., … Safran, R. J. (2018). Patterns of Barn Swallow (*Hirundo rustica*) song divergence at continental and global scales. *Biological Journal of the Linnaean Society*, *21*, 10.

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87. https://doi.org/10.1038/nrg.2016.133

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the national Academy of Sciences*, *105*(33), 11597–11604. https://doi.org/10.1073/pnas.0801317105

Zink, R. M., Pavlova, A., Rohwer, S., & Drovetski, S. V. (2006). Barn swallows before barns: Population histories and intercontinental colonization. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1591), 1245–1251. https://doi.org/10.1098/rspb.2005.3414

SUPPORTING INFORMATION

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

How to cite this article: Smith CCR, Flaxman SM, Scordato ESC, et al. Demographic inference in barn swallows using whole-genome data shows signal for bottleneck and subspecies differentiation during the Holocene. *Mol Ecol*. 2018;00:1–13. https://doi.org/10.1111/mec.14854