Received: 19 April 2021 | Revised: 5 November 2021 | Accepted: 8 November 2021

DOI: 10.1111/mec.16277

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Domestication obscures genomic estimates of population history

Chia-Wei Lu[1] | Cheng-Te Yao[2] | Chih-Ming Hung[1]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[2]Division of Zoology, Endemic Species Research Institute, Nantou, Taiwan

Correspondence
Chih-Ming Hung, Biodiversity Research Center, Academia Sinica, Taipei, Taiwan.
Email: cmhung@gate.sinica.edu.tw

Funding information
Academia Sinica, Grant/Award Number: AS-CDA-108-L05

Abstract

Domesticated species are valuable models to examine phenotypic evolution, and knowledge on domestication history is critical for understanding the trajectories of evolutionary changes. Sequentially Markov Coalescent models are often used to infer domestication history. However, domestication practices may obscure the signal left by population history, affecting demographic inference. Here we assembled the genomes of a recently domesticated species—the society finch—and its parent species—the white-rumped munia—to examine its domestication history. We applied genomic analyses to two society finch breeds and white-rumped munias to test whether domestication of the former resulted from inbreeding or hybridization. The society finch showed longer and more runs of homozygosity and lower genomic heterozygosity than the white-rumped munia, supporting an inbreeding origin in the former. Blocks of white-rumped munia and other ancestry in society finch genomes showed similar genetic distance between the two taxa, inconsistent with the hybridization origin hypothesis. We then applied two Sequentially Markov Coalescent models—PSMC and SMC++—to infer the demographic histories of both. Surprisingly, the two models did not reveal a recent population bottleneck, but instead the PSMC model showed a specious, dramatic population increase in the society finch. Subsequently, we used simulated genomes based on an array of demographic scenarios to demonstrate that recent inbreeding, not hybridization, caused the distorted PSMC population trajectory. Such analyses could have misled our understanding of the domestication process. Our findings stress caution when interpreting the histories of recently domesticated species inferred by PSMC, arguing that these histories require multiple analyses to validate.

KEYWORDS
domestication, inbreeding, PSMC, runs of homozygosity, society finch, white-rumped munia

1 | INTRODUCTION

Domesticating animals and plants is important to human population growth, evolution and cultural development, and domestication has reciprocal impacts on humans and domesticates (Price, 2002; Zeder, 2012). Inferring the history of domestic organisms has also been a major goal in evolutionary biology as it helps determine how phenotypic traits may have evolved in anthropological or historical periods—a relatively short time frame compared with evolution occurring in nature (Zeder, 2018). However, reconstructing domestication history is no trivial task, even for well-known domesticated taxa such as dogs. This is because the analysis involves parsing through a mix of complex processes, such as population bottlenecks, strong

selection, inbreeding and/or introgression (Larson & Burger, 2013). Using genomic data is a promising strategy to examine complex evolutionary processes (Luikart et al., 2003). Nevertheless, it is still challenging to disentangle confounding factors to reveal a domestication history that accurately explains genomic variations.

As genome-sequencing technologies advance, more and more genomes from diverse taxa are being sequenced. Consequently, genomic analyses are widely applied to nonmodel organisms. Genome-based demographic models such as the Pairwise Sequentially Markovian Coalescent (PSMC) model (Li & Durbin, 2011), which requires only singe genomes, are particularly useful and have become highly popular (Beichman et al., 2017). A recent study suggests that PSMC is more robust to fragmented genome assemblies than other genome-based programs in inferring population history (Patton et al., 2019). Many studies have applied PSMC to investigate the population history of extinct or endangered species (Hu et al., 2020; Hung et al., 2014; Saremi et al., 2019), date the divergence time between sister species (Westbuty et al., 2019), and examine interactions between historical climate fluctuations and population changes (Mays et al., 2018; Nadachowska-Brzyska et al., 2015). PSMC has also been used to infer the demographic history of domesticated species and estimate when the domesticated species (or their progenitors) split from their wild relatives (Wang et al., 2021). Despite the popularity and reliability of PSMC for tracing demographic history back to hundreds of thousands of years ago, few studies have explored how recent population changes (e.g., recent population bottlenecks) may affect PSMC inferences. Domesticated organisms arose with human civilization, and their populations have been subject to artificial management for a few hundred or thousand years. Thus, domesticated taxa, especially commercial pets, represent good systems to investigate the power of PSMC in analysing cases with recent population fluctuations.

The society finch (*Lonchura striata domestica*), also known as the Bengalese finch (Figure S1), is a common strain of finch that can be found in pet shops (Weedn, 1993). Since society finches are tame and easy to raise and breed, they are popular among bird fanciers. The society finch is assumed to have been domesticated from the white-rumped munia (*Lonchura striata*; Figure S1) in the past few centuries (Colquitt et al., 2018; Okanoya, 2012, 2015; Svanberg, 2008). It has more variable songs than the white-rumped munia, so scientists have used the society finch to study the evolution of vocal learning in songbirds (Colquitt et al., 2021; Honda & Okanoya, 1999). Its short domestication history makes the society finch a good system to examine the effects of domestication on genomic variations because it is less likely to have experienced multiple domestications that mix hybridization and bottlenecks compared with those with a long-term domestication history (Larson & Burger, 2013).

The origin of the society finch is still controversial. There are two main hypotheses regarding its origin. (1) The inbreeding origin hypothesis: the society finch was a domesticated type of the white-rumped munia that arose simply through inbreeding (Goodwin, 1982; Okanoya, 2012, 2015; Restall, 1996). (2) the hybridization origin hypothesis: the white-rumped munia was crossbred with another finch to generate the society finch (McCarthy, 2006; Oppenborn, 1992). The white-rumped munia is distributed from northcentral India to southeast Asia, including Taiwan and the Malay Peninsula (Goodwin, 1982; del Hoyo & Collar, 2016). One clue to the domestication history of the society finch is its cross-continental trade route: it was first traded from China to Japan, and then from Japan to Europe and North America (Svanberg, 2008). Records show that society finches were bred as different colour variations primarily in Japan before being traded to Europe (Svanberg, 2008; Taka-Tsukasa, 1922). Some assume that the domestication of society finches began in 1763, when a federal lord of Kyushu imported white-rumped munias from Zhejiang (China) to Japan (Okanoya, 2012; Washio, 1996). However, there is little information on how Chinese bird fanciers raised white-rumped munias and whether they domesticated society finches.

Here we aim to use genomic data to clarify whether the society finch was domesticated from the white-rumped munia through inbreeding or crossbreeding with other finches and when the domestication was initiated. Importantly, we also examined whether the recent domestication events bias population history inference by Sequentially Markovian Coalescent models. To obtain reference genomes, we *de novo* assembled the draft genomes for both the society finch and the white-rumped munia. We estimated genomic phylogeny, nucleotide diversity, heterozygosity, runs of homozygosity (ROH) (Ceballos et al., 2018), genetic divergence of genomic blocks with different ancestry and demographic history based on the whole genome sequences of six society finches from two breeds (four pied and two crested white individuals) and four white-rumped munias. Furthermore, we tested whether a recent bottleneck and/or hybridization bias PSMC inferences with both empirical and simulated genome data. The practices help prevent the results of genome-based demographic analyses from being misinterpreted, especially for domestic species. Understanding the origin of the society finch will allow us to gain insight into how inbreeding or hybridization affects behavioural evolution over the course of domestication in the future.

2 | MATERIALS AND METHODS

2.1 | Whole genome sequencing

All society finches used in this study were purchased from bird shops, and white-rumped munias were caught in the wild using mist-nets in Nantou (sample ID: L_UB) or Yunlin (LST037, LST038 and LST040) County, Taiwan. One female pied society finch (L_RO) and one female white-rumped munia (L_UB) were used for *de novo* genome assembly to obtain sequencing reads from Z, W and autosomal chromosomes. We applied the 10× Chromium system for library construction (Weisenfeld et al., 2017), and the prepared libraries were sequenced using the Illumina NovaSeq 6000 and HiSeq 4000 platforms for the society finch and the white-rumped munia, respectively (paired-end; 2 × 150 bp). See the Supporting Methods for details.

2.2 | Whole genome *de novo* assembly

Adapters and low-quality reads were trimmed off using CUTADAPT version 2.3 (Martin, 2011). To avoid possible contamination during DNA extraction, we used KRAKEN2 (Wood et al., 2019) version 2.0.8 to filter out reads that matched our customized database. We used SUPERNOVA version 2.1 (Weisenfeld et al., 2017) to *de novo* assemble the draft genomes. We detected and broke putative misassembled regions using TIGMINT version 1.1.2 (Jackman et al., 2018). We used ARCS version 1.0.5 (Yeo et al., 2018) for scaffolding, GAPCLOSER version 1.12 (Luo et al., 2012) for a gap-closing step and SEQTK version 1.3 (available: github.com/lh3/seqtk) for excluding scaffolds <1000 bp. See Supporting Methods for details.

2.3 | Gene prediction in the assembled genome

We generated training files for gene prediction using BUSCO version 4.0.5 (Seppey et al., 2019; Simão et al., 2015) based on 8,338 Aves genes. To obtain RNA evidence for gene prediction, we extracted whole RNA from the eight tissues of the same individuals used for genome assembly, and conducted RNA sequencing (RNSseq; paired-end; 2 × 150) using the Illumina Hiseq 4000 platform. We mapped trimmed RNA reads with HISAT2 and converted the output file to hint file with AUGUSTUS version 3.3.2. (Stanke et al., 2006). We applied the BUSCO training files and intron hints to the repeat-masked genomes for gene prediction analysis using BRAKER version 2.2.1 (Hoff et al., 2019). See the Supporting Methods for details.

2.4 | Assembled genome quality

We used QUAST version 5.0.1 (Gurevich et al., 2013) to estimate the lengths and numbers of scaffolds, gap lengths and the scaffold N50 of our assembled genomes. The number of undetermined bases was estimated using ASSEMBLY-STAT version 1.0.1 (available online: github.com/sanger-pathogens/assembly-stats). We used BUCSO to infer the completeness of assembled genomes based on 8,338 Aves genes. See the Supporting Methods for details.

2.5 | SNP and indel calling

To obtain genome resequencing data, we extracted DNA from the blood samples of eight male birds—three pied society finches (LDO253, LDO255 and LDO256), two crested white society finches (LDO004 and LDO005) and three white-rumped munias (LST037, LST038 and LST040). All of the resequenced libraries were sequenced using the Illumina HiSeq X Ten platform. To call the single nucleotide polymorphisms (SNPs) and indels (i.e., variant sites) of the eight male birds plus the two females used for *de novo* assembly, we mapped their trimmed reads to the corresponding reference genomes assembled above using BWA version 0.7.17 (Li & Durbin, 2009). We called variant sites using both Genome Analysis Toolkit (GATK) version 4.1.4 (Van der Auwera et al., 2013) and BCFTOOLS version 1.9 (available: samtools.github.io/bcftools/) and conducted base quality recalibration using GATK. We removed sites with unusual sequencing coverage and nonbiallelic sites. The final data sets were used for all of the downstream analyses unless otherwise noted. To conduct downstream analyses, we excluded sexual chromosomes from the analyses to avoid any problems associated with the haploid Z and W chromosomes in female samples. See the Supporting Methods for details.

2.6 | Phylogenetic relationships

To address the relationships among the finch strains, we mapped the sequencing reads of all 10 individuals together with those of one zebra finch downloaded from NCBI (SRA: SRS5086496, Kinsella et al., 2019) against the white-rumped munia assembled genome to obtain alignable SNPs. The SNP calling procedure was the same as that mentioned in the previous section, and 100,000 SNPs were randomly chosen for phylogenetic analysis. We used RAXML version 8.2.12 (Stamatakis, 2014) to reconstruct the phylogeny based on the ASC_GTRGAMMA model for correcting ascertainment bias and 1,000 bootstrap replicates.

2.7 | Genome-wide genetic heterozygosity and nucleotide diversity

We identified the genotype of each SNP site for all white-rumped munia and society finch genomes using VCFR (Knaus & Grünwald, 2017), and then calculated SNP-based heterozygosity as the frequency of heterozygous genotypes in each taxon (i.e., the white-rumped munia, the pied society finch or the crested white society finch). We estimated average heterozygosity for every 10,000 SNPs to assess statistical differences in genome-wide SNP-based heterozygosity among the taxa. We also estimated nucleotide diversity (π) for each taxon using VCFTOOLS version 0.1.17 (Danecek et al., 2011) with 10-kb nonoverlapping windows combined with customed codes, which corrected problems associated with VCFTOOLS settings when estimating π values (see the Supporting Methods for details). We then statistically compared the π values among the taxa. We applied the Bonferroni correction (Bland & Altman, 1995) for multiple comparisons to evaluate significant differences for the above statistical tests.

2.8 | Runs of homozygosity and estimates of inbreeding time

Genomic homozygosity levels are high in organisms that have experienced population bottlenecks or small population sizes. This is because the haplotypes of different individuals that were inherited

from the same common ancestor have greater chances to form a homozygote in smaller populations. Thus, the number and length of ROH may reflect the demography history of organisms (Ceballos et al., 2018). Here we used PLINK version 1.9 (Purcell et al., 2007) to identify ROH longer than 100 kb (Dong et al., 2021; Li et al., in press) for downstream analyses. We estimated the sum total length of ROH (S_{ROH}), the number of ROH (N_{ROH}) and the mean ROH length (M_{ROH}) for each sequenced genome. To assess the degree of inbreeding, we estimated the inbreeding coefficient, $F_{ROH} = S_{ROH}$ divided by the length of the resequenced genome (Gómez-Sánchez et al., 2018; Keller et al., 2011).

To estimate the time when domestication-associated inbreeding took place, we used the correlation between ROH lengths and the time periods when inbreeding occurred using the equation $g = 100/(2rL)$, where g is the expected time in generations traced back the most recent common ancestor of the paternal and maternal lineages, r is the recombination rate and L is the ROH length (Kardos et al., 2018; Saremi et al., 2019; Thompson, 2013). We used an average recombination rate = 1.5 centimorgans Mb^{-1} estimated from the zebra finch (Backström et al., 2010) for the calculation. Therefore, a long ROH indicated that inbreeding occurred recently, and vice versa. If the society finch domestication occurred through recent inbreeding, this event might leave predominantly long ROH in the genome of this bird. By comparing the distributions of ROH with different lengths between the society finch and white-rumped munia, we might infer the timing of the domestication.

2.9 | Ancestry analyses for testing the hybridization origin hypothesis

We conducted two ancestry analyses to test the hybridization origin hypothesis. First, we mapped the sequencing reads of all 10 individuals against Assembly_WRM to obtain alignable SNPs. We then used PLINK to prune SNPs that were in linkage disequilibrium (LD) based on an approach of 100-SNP sliding windows with 20 SNPs overlapped and $r^2 = .2$. We applied LAMP (Sankararaman et al., 2008) to the LD-pruned data set to infer genomic blocks with white-rumped munia, society finch or mixed (white-rumped munia + society finch) ancestry in society finch genomes as well as white-rumped munia genomes. The blocks with society finch or mixed ancestry probably shared ancestry with the other potential parent species assuming that the hybridization origin hypothesis was true. Thus, this hypothesis would predict that the genetic distance between blocks with white-rumped munia ancestry in society finch genomes and the corresponding regions in white-rumped munia genomes were lower than those of blocks with society finch and mixed ancestry. To test this hypothesis, we estimated D_{xy} values for each ancestry block in society finch genomes against the corresponding regions in white-rumped munia genomes using POPGENOME (Pfeifer et al., 2014). We then compared the D_{xy} values of different types of ancestry blocks. Given that LAMP could only analyse a chromosome/scaffold each time, we only used Scaffold 1 to Scaffold 4 (these four scaffolds made up 19% of Assembly_WRM) from the data set as four separate inputs for the LAMP analyses. We tested a range of mixture proportions (α, from 0.5 to 0.999) with different numbers of generations since admixture (g = 1,000, 200, 150, 100, 20 and 5) for each scaffold (Table S5) and used the results based on the best setting for the D_{xy} analysis (see Results for details).

Second, we examined whether society finch genomes were more likely to show addition of genes from other species than were white-rumped munia genomes, a pattern predicted by the hybridization origin hypothesis. We applied ADMIXTURE (Alexander et al., 2009) to the above LD-pruned data set to infer the ancestry components of each individual for testing against the prediction. We ran the ADMIXTURE analyses from $K = 2$ to $K = 5$ with 100 replicates for each K value, and combined the results using the CLUMPARK server (http://clumpak.tau.ac.il).

2.10 | Demographic history inferences based on individual and multiple genomes

We inferred the historical population size changes in society finches and white-rumped munias based on the 10 sequenced genomes using PSMC. PSMC inferred the change in effective population size (N_e) over time based on data from a single genome. We set $N = 30$, $t = 5$, $r = 5$ and p = "4+30*2+4+6+10" for PSMC analysis (Nadachowska-Brzyska et al., 2015) with 100 bootstrap replicates. The substitution rate was set to 3.44×10^{-9} per site per generation and the generation time to two years according to the PSMC analyses of another passerine, the medium ground finch (*Geospiza fortis*; Nadachowska-Brzyska et al., 2015). To test whether ROH affected PSMC inference, we further masked the ROH regions with SEQTK and ran PSMC analyses again (see Supporting Methods for details).

We also conducted multigenome analysis for demographic history inference using another Sequentially Markovian Coalescent mode, SMC++ (Terhorst et al., 2017). We generated multigenome data sets by mapping the sequencing reads of all individuals against Assembly_WRM and removing scaffolds <10 kb. We separated our samples into three data sets: (i) four white-rumped munia genomes, (ii) four pied society finch genomes and (iii) two crested white society finch genomes. We conducted the SMC++ analyses based on the "estimate" method with composite likelihood for demographic inference. We conducted 20 replicated runs of SMC++ analysis to each of the three data sets with the mutation rate set to 3.44×10^{-9} per site per generation and the generation time to 2 years, the same as those used for the PSMC analyses.

2.11 | Estimating the effect of a recent bottleneck and/or hybridization on PSMC demography inferences based on simulation

To examine the effect of recent inbreeding or hybridization on PSMC inferences, we used the coalescent genome simulator

MACS (Chen et al., 2009) to simulate finch genomes that experienced a bottleneck and/or introgression. Each simulated genome was composed of 30 chromosomes, each 33 Mb long (total length =0.99 Gb). We applied the same mutation rate that we used in PSMC analysis and a recombination rate $=1.5 \times 10^{-8}$ in simulations. For inbreeding scenarios, we used the ROH-unmasked PSMC history of white-rumped munias as a template to simulate one control scenario—(i) no recent bottleneck—and three different recent-bottleneck scenarios—(ii) a short bottleneck occurring 150–200 generations ago, followed by a population size recovery; (iii) a long bottleneck occurring 50–200 generations ago, followed by a population recovery; and (iv) a bottleneck lasting from 200 generations ago to the present (Figure S2). The starting time of the bottleneck was estimated based on the distribution of ROH length (see the "Runs of homozygosity and estimate of inbreeding time section" and Results for details). Each bottleneck event dramatically reduced the population size to 1/1,000th of the original (i.e., N_e from 200,000 to 200). We expected that the longer a simulated bottleneck was, the more severe the inbreeding effect it caused in the population.

We also simulated scenarios in which hybridization between the white-rumped munia and another finch species generated the society finch, followed by either one or no population bottleneck, using MACS (Figure S3). We simulated the white-rumped munia splitting into two lineages 400 years ago (i.e., 200 generations), one of which experienced introgressive hybridization from another finch that diverged from the white-rumped munia 1 million years ago. The introgressed lineage became the society finch and was assumed to have (i) the same N_e as that of the white-rumped munia (N_e =200,000) after the split, (ii) N_e =200,000 after the split immediately followed by a short (50-generation) population bottleneck with N_e shrinking to 200 and then a population size recovery, or (iii) N_e =200 from the split to the present (i.e., a 200-generation bottleneck). We simulated the introgressive hybridization in two ways—(i) 50% of the gene pool in the introgressed lineage was replaced by the other finch's genes in one generation right after the split and (ii) 1% of the gene pool in the former was replaced by the latter's genes per generation for 50 consecutive generations immediately after the split. We then applied the PSMC model to all simulated genomes generated from the above six scenarios (Figure S3) with the same setting as was used in the empirical data. We simulated each of the scenarios descripted in this section 100 times and used all of the 100 replicates for the PSMC analyses.

3 | RESULTS

3.1 | Qualities of whole genome assemblies and genomic variation sites

We obtained a total of 119.5 and 87.5 Gb of raw linked-reads for a pied society finch (L_RO) and white-rumped munia (L_UB), corresponding to around 109× and 80× coverage, respectively. The quality of the assembled genomes for the two species was similar. The scaffold N50 of the assembled society finch genome (Assembly_SF) was 22.4 Mb and 3.25% of the genome contained undetermined bases (Table S1). The genome assembly of the white-rumped munia (Assembly_WRM) had a slightly lower scaffold N50 value and slightly more missing sites (scaffold N50 = 17 Mb and percentage of undetermined bases =5.19%) than Assembly_SF. BUSCO analyses (Table S2) showed similar levels of assembly completeness in Assembly_SF (92.7%) and Assembly_WRM (91.9%). The numbers of genes and transcripts predicted by BRAKER2 were slightly higher in Assembly_SF than in Assembly_WRM (Table S3). When aligned to the zebra finch genome, both Assembly_SF and Assembly_WRM had some scaffolds identified as inversion regions (Figure S4). Nevertheless, only a few segments of zebra finch chromosome W were incorporated into autosomal scaffolds in Assembly_SF and even fewer in Assembly_WRM (Figure S5), suggesting low degrees of misassembly for both genome assemblies.

We obtained 28.3–43.2 Gb of paired-end reads for each of the additional eight individuals (i.e., three pied society finches, two crested white society finches and three white-rumped munias). The mean coverage of each resequenced individual was 21.6×–33.1×. The six society finches had a total of 8,556,621 SNPs and 1,149,067 indels and the four white-rumped munias had 7,847,366 SNPs and 894,825 indels (Table S4).

3.2 | Phylogenetic relationships

The maximum-likelihood phylogenetic tree showed that society finches and white-rumped munias were reciprocally monophyletic (Figure 1a), suggesting that these two taxa diverged with little recent gene flow. In addition, two crested white society finches (LDO004 and LDO005) formed a clade within the society finch monophyletic group (Figure 1a), suggesting the former showed a certain level of genomic differentiation from other society finches. The result also implies that different types of society finches represent diverging lineages that may increase the overall genetic diversity of the society finch, especially if breeders often cross different types.

3.3 | Genome diversity of society finches and white-rumped munias

The genome-wide SNP-based heterozygosity of the white-rumped munia (mean =0.3908 ± 0.0198 [*SD*]) was significantly larger than that of the pied society finch (mean =0.3642 ± 0.0620; *t* test, p = 3.27e−29) and the crested white society finch (mean =0.3550 ± 0.0617; p = 3.24e−43; Figure 1b). The SNP-based heterozygosity of the pied society finch was also significantly larger than that of the crested white society finches (p = 2.42e−10). However, the level of genome-wide nucleotide diversity (π) of the white-rumped munia (mean =0.002860 ± 0.001828) was significantly lower than that of

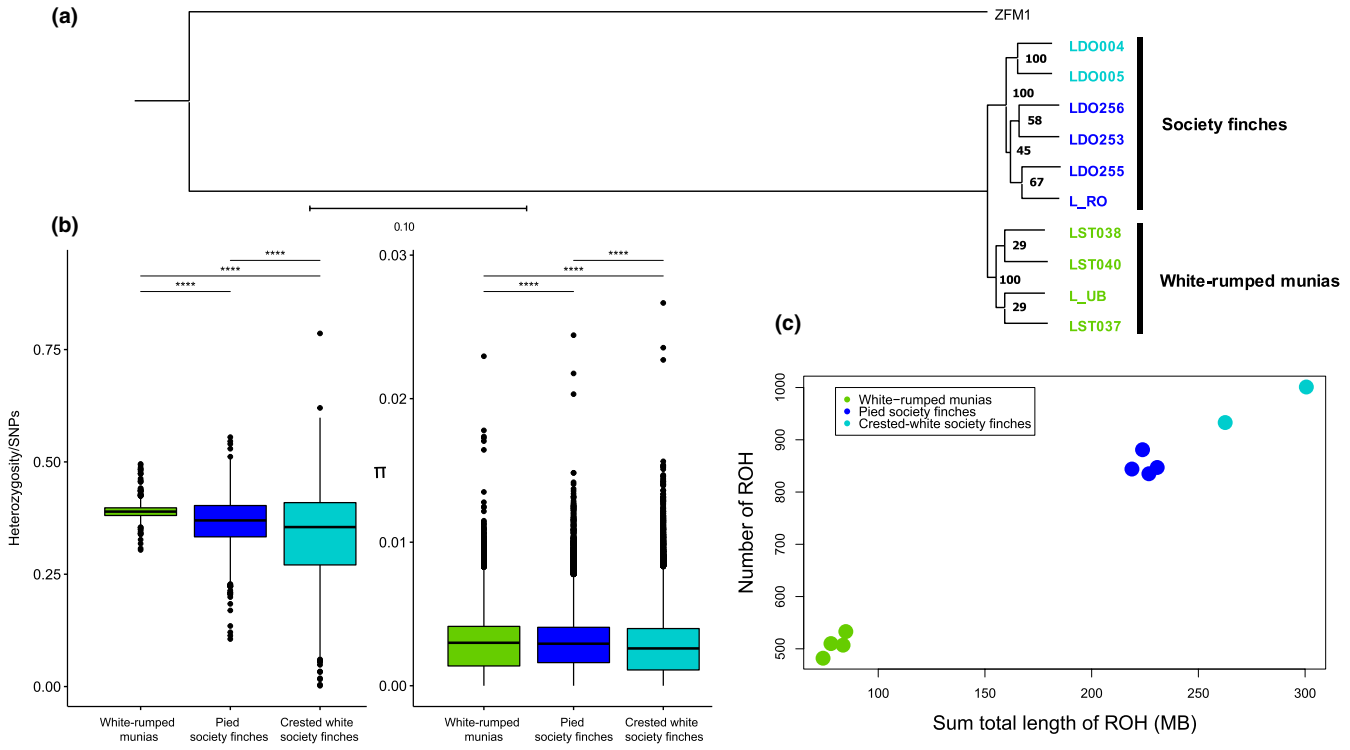

FIGURE 1 (a) Phylogenetic relationship between society finches and white-rumped munias. Support values of each node are based on 1,000 bootstraps. Branch lengths are proportion to numbers of substitutions per site. (b) Genome-wide SNP-based heterozygosity and nucleotide diversity (π) of white-rumped munias, and pied and crested white society finches. Box plots show the minimum (Q1–1.5 interquartile range), 25th percentile (Q1), median, 75th percentile (Q3) and maximum (Q3+1.5 interquartile range) values, and the outliers of the parameters. Horizontal bars indicate significant *t* test results with Bonferroni-adjusted *p*-values <.0001 (****). (c) Comparison between S_{ROH} and N_{ROH} for each sample. LST037, LST038, LST040 and L_UB are white-rumped munias. LDO004 and LDO005 are crested white society finches. LDO253, LDO255, LDO256 and L_RO are pied society finches [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Runs of homozygosity (ROH) variables of each individual

Sample ID	N_{ROH}	S_{ROH} (Mb)	M_{ROH} (kb)	F_{ROH}
LST037	482	74.2	153.8	0.068
LST038	533	84.8	159.1	0.078
LST040	510	77.9	152.6	0.072
L_UB	507	83.5	164.7	0.077
LDO004	933	262.6	281.4	0.240
LDO005	1001	300.6	300.3	0.275
LDO253	881	223.8	254.1	0.205
LDO255	844	218.9	259.3	0.200
LDO256	847	230.7	272.4	0.211
L_RO	835	226.8	271.6	0.207

Note: N_{ROH} indicates the number of ROH, S_{ROH} the sum length of ROH, M_{ROH} the mean length of ROH and F_{ROH} the inbreeding coefficient. LST037, LST038, LST040 and L_UB are white-rumped munias. LDO004 and LDO005 are crested white society finches. LDO253, LDO255, LDO256 and L_RO are pied society finches.

the pied society finch (mean =0.002910 ± 0.001799; $p = 4.65e{-}9$), but significantly higher than that of the crested white society finch (mean =0.002693 ± 0.001991; $p = 8.7e{-}81$; Figure 1b).

The ROH analyses showed stronger signs of inbreeding in society finches than in white-rumped munias. The F_{ROH} values, indicating the level of inbreeding, of society finches (0.200–0.275) were about 3–4 times higher than those of white-rumped munias (0.068–0.078; Table 1). Both the S_{ROH} (218.9–300.6 Mb in society finches, 74.2–84.8 Mb in white-rumped munias; Wilcoxon rank sum test, $p = 9.524e{-}3$) and N_{ROH} values (835–1,001 in society finches, 483–533 in white-rumped munias; Wilcoxon rank sum test, $p = 9.524e{-}3$) were greater in society finches than in white-rumped munias (Figure 1c, Table 1). The M_{ROH} values of society finches (254.1–300.3 kb) were also higher than those of white-rumped munias (152.6–164.7 kb; Wilcoxon rank sum test, $p = 9.524e{-}3$). The two crested white society finches (LDO004 and LDO005) had the two largest values of N_{ROH}, S_{ROH}, M_{ROH} and F_{ROH} among all studied birds (Figure 1c, Table 1), indicating they are subject to even stronger inbreeding than pied society finches.

3.4 | Distribution of ROH length and corresponding time of inbreeding

All white-rumped munia and society finch genomes had similar amounts of ROH with length <166.67 kb (corresponding to

inbreeding occurring >200 generations ago), whereas society finch genomes had more ROH with length >166.67 kb (corresponding to inbreeding occurring <200 generations ago) than white-rumped munia genomes (Figure 2). That is, society finches and white-rumped munias had had similar levels of inbreeding until 151–200 generations ago, when the former started to have higher levels of inbreeding, as indicated by more accumulated long ROH, than the latter. The results suggest that domestication that was expected to dramatically increase the levels of inbreeding of the society finch might have started 151–200 generations ago. Among the society finches, crested white society finches showed more ROH with a length >333.33 kb than other society finches, suggesting that stronger inbreeding, probably resulting from further artificial selection for white plumage colour, occurred in the former starting 101–150 generations ago (Figure 2). Assuming that one generation is 2 years for Passeriformes (Nadachowska-Brzyska et al., 2015), we estimated that domestication of the society finch began around 302–400 years ago, earlier than when Japan started importing society finches. Nevertheless, crested white society finches were strongly selected starting around 202–300 years ago, matching the time periods when society finches with varied plumage forms started being bred in Japan (about 250 years ago).

3.5 | Ancestry analyses against the hybridization origin hypothesis

The LAMP analysis for ancestry blocks of society finch and white-rumped munia genomes was not consistent with the hybridization origin hypothesis. For the LAMP analysis, we tested different mixture proportions from 0.5:0.5 to 0.9:0.1 and then to 0.991:0.001 with $g = 1{,}000$ (Table S5). We found that the mixture proportion of 0.9:0.1 returned the highest level of white-rumped munia ancestry in the four white-rumped munia genomes, closest to the true pattern (Figure S6). We then tested g values from 5 to 1,000 with the mixture proportions of 0.9:0.1 and 0.6:0.4 and found minor differences among results with different g values except those with $g = 5$ and 20 (Figure S6). We decided to use the results based on a mixture proportion of 0.9:0.1 and $g = 1{,}000$ (Figure 3 for Scaffold 1; Figure S6 for full results) for downstream analyses given that the manual suggested using an overestimated g value if the true value was unknown. The D_{xy} values between blocks with white-rumped munia ancestry in society finch genomes and the corresponding blocks in white-rumped munia genomes were not significantly different from those for blocks with society finch ancestry (t test, $p > .05$) or mixed ancestry ($p > .05$; Figure S7). Thus, the results were inconsistent with the hybridization origin hypothesis.

The ADMIXTURE result at $K = 2$ had the best support (the lowest cross-validation error; Figure S8) and showed two clusters corresponding to the society finch and the white-rumped munia (Figure 4). When $K = 3$, the two taxa showed admixture of the third ancestry with the highest mixture proportions in two crested white society finches; when $K = 4$, all individuals shared the addition of the fourth ancestry, and pied and crested white society finches showed distinct patterns of admixture. The results suggest a certain level of differentiation between the pied and crested white society finches. Importantly, the results also suggest that society finch genomes did not have a higher likelihood of being introgressed by the genes of another species compared with white-rumped munia genomes, inconsistent with the hybridization origin hypothesis.

3.6 | Demographic history inferences based on individual and multiple genomes

Given that the society finch was first domesticated from the white-rumped munia in only the past few hundred years, we expected the two birds to share a population history leading up to this recent,

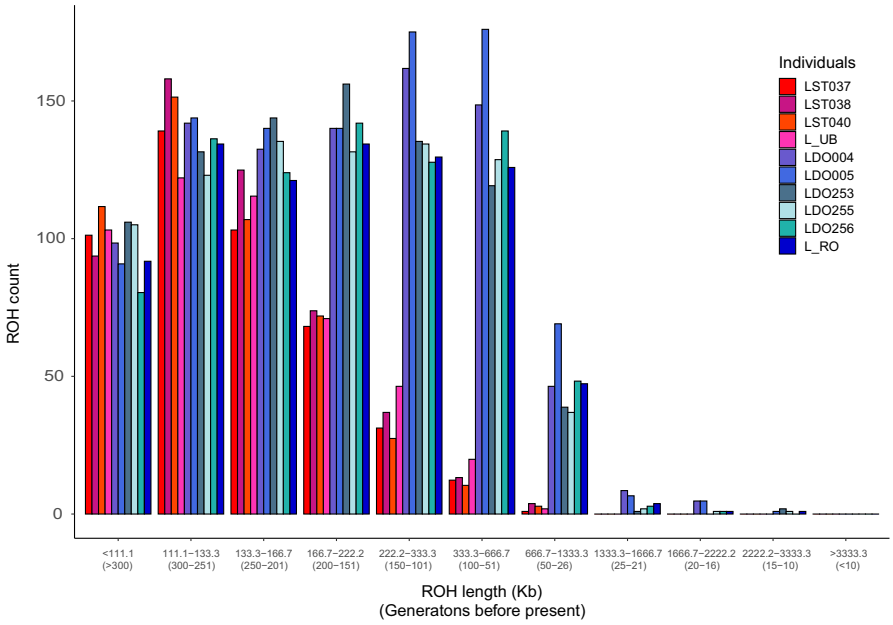

FIGURE 2 Distribution of ROH length and corresponding time of inbreeding (generations before present) for each sample. The expected inbreeding time of a particular length of ROH was estimated using $g = 100/(2rL)$, where g is the expected generations that dates back to when both paternal and maternal lineages shared a common ancestor, r is the recombination rate and L is the length of ROH. LST037, LST038, LST040 and L_UB are white-rumped munias. LDO004 and LDO005 are crested white society finches. LDO253, LDO255, LDO256 and L_RO are pied society finches [Colour figure can be viewed at wileyonlinelibrary.com]

artificial event. Our sampled individual genomes showed various trends of PSMC population history, which were generally divided into two groups corresponding to the society finch and the white-rumped munia (Figure 5; Figures S9 and S10). However, the two groups separated earlier than when they were estimated to first have been domesticated (i.e., a few hundred years ago). The two groups largely shared a stable trend of N_e before 1 million years ago, followed by population expansions of different magnitudes. The white-rumped munia population reached a peak around 200,000 years ago (ka) and then continued to decline until 20 ka, followed by a moderate increase. Surprisingly, the society finch population reached a local peak around 80–200 ka, followed by a decline and then an unexpected spike around 20–40 ka. The N_e of the society finch had been larger than those of white-rumped munia throughout the evolutionary history until the end of the spurious spikes. This N_e spike was observed in all sampled society finches. Interestingly, when ROH regions were masked, the N_e spikes in society finch PSMC plots dramatically diminished, whereas the PSMC trends of white-rumped munias remained largely the same after masking ROH regions, except that the latest population growth became moderately stronger (Figure 6; Figures S11 and S12).

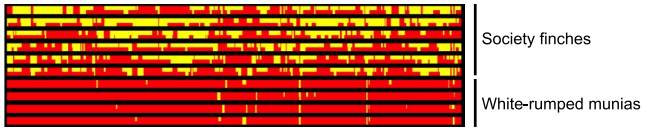

FIGURE 3 Ancestry block analysis for Scaffold 1 using LAMP with $g = 1{,}000$ and a mixture proportion of 0.9:0.1. Each row represents one sample, with six society finches followed by four white-rumped munias. Blocks with society finch ancestry are marked in yellow, blocks with white-rumped munia ancestry are marked in red, and blocks with mixed ancestry are marked in both yellow and red [Colour figure can be viewed at wileyonlinelibrary.com]

The SMC++ analyses based on multiple genomes revealed similar demographic histories among the white-rumped munia, pied society finch and crested white society finch in the past 1 million years (Figure 7). The SMC++ results of the three taxa were somewhat similar to the PSMC results of white-rumped munias. That is, they experienced a population expansion since at least 1 million years ago and reached a peak around 300–400 ka, followed by a decline until 4–10 ka and then a population increase. The moderate inconsistency in the timing of population fluctuations between SMC++ and PSMC was also found in other studies (e.g., Dong et al., 2021). The 20 replicated SMC++ runs showed certain levels of variations, similar to the variation levels in the PSMC bootstrapping results of white-rumped munias (Figure S9). The SMC++ results did not reveal any recent population bottleneck in pied or crested white society finch populations.

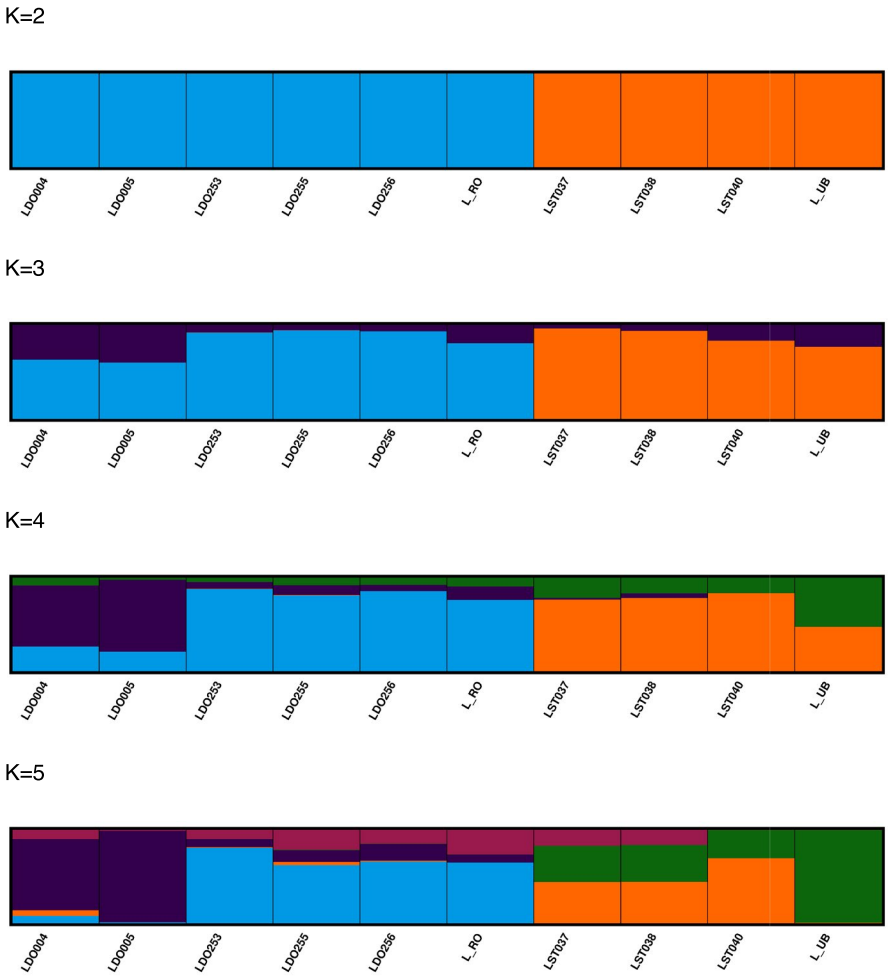

FIGURE 4 Admixture analyses based on the genome-wide SNPs of society finches and white-rumped munias for $K = 2$ to $K = 5$. Cross-validation error values (Figure S8) show that the population structure with $K = 2$ is best supported. LDO004 and LDO005 are crested white society finches. LDO253, LDO255, LDO256 and L_RO are pied society finches. LST037, LST038, LST040 and L_UB are white-rumped munias [Colour figure can be viewed at wileyonlinelibrary.com]

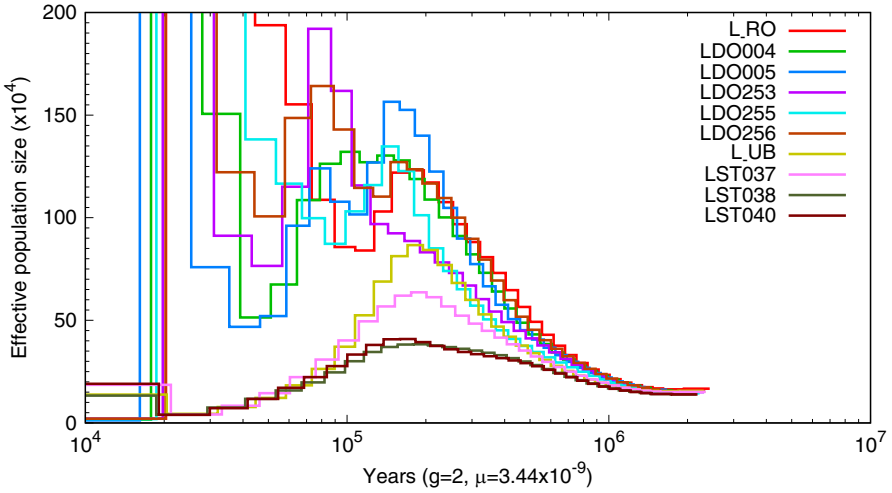

FIGURE 5 PSMC inferences for all samples. We set the generation time (g) as 2 years and the mutation rate (μ) as 3.44×10^{-9} per site per generation. LST037, LST038, LST040 and L_UB are white-rumped munias. LDO004 and LDO005 are crested white society finches. LDO253, LDO255, LDO256 and L_RO are pied society finches [Colour figure can be viewed at wileyonlinelibrary.com]

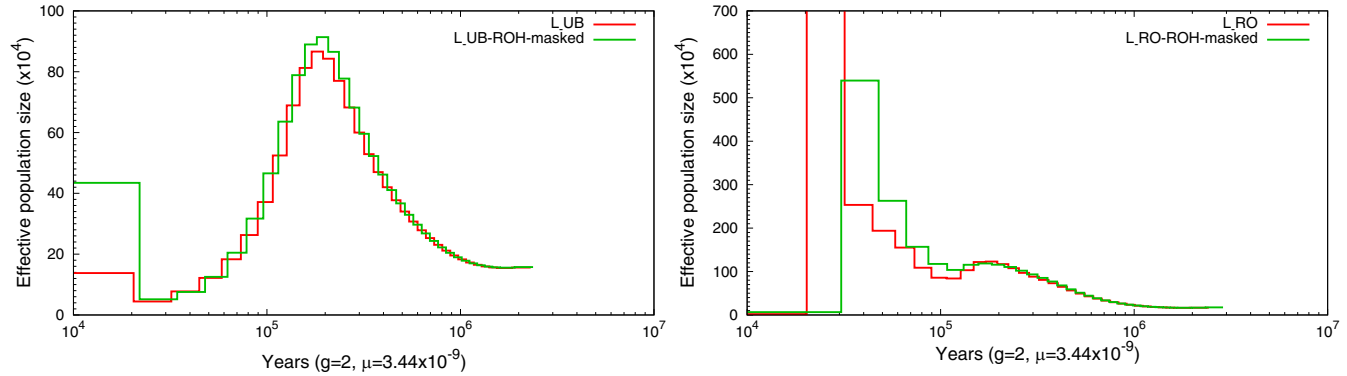

FIGURE 6 Comparisons between ROH-unmasked and ROH-masked PSMC. Red lines indicate ROH-unmasked results, and green lines indicate ROH-masked results. The PSMC plots of ROH-unmasked and ROH-masked genomes are largely similar in the white-rumped munia (L_UB, left panel); however, a spike of effective population size in the society finch (L_RO) is mitigated when ROH are masked (right panel). The PSMC plots of all individuals and full views (showing the whole spike of effective population size) are shown in Figure S8 and S9, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

3.7 | The effect of a recent bottleneck and/or hybridization on inferences of PSMC demography

Our simulations based on the white-rumped munia history, in which a recent population bottleneck was introduced, generated PSMC plots with a population spike occurring 10–50 ka (Figure 8; Figure S13), similar to the empirical PSMC results of society finches (Figure 5). As we increased the span of a bottleneck, which led to a more severe inbreeding effect, the N_e spike became wider and taller. The simulation with a short bottleneck occurring 150–200 generations ago—the inbreeding time inferences based on the distribution of ROH length (Figure 2)—generated an artificial N_e spike in PSMC plots closest to those of society finches in terms of timing and magnitude. The results confirmed that ROH distribution could have accurately reflected the occurrence of a recent bottleneck that separated the society finch history from that of the white-rumped munia. Importantly, our empirical and simulated genomic data demonstrated that a recent bottleneck might mislead PSMC inferences.

We further simulated scenarios of hybridization origin for the society finch with and without a bottleneck. We found that one generation of strong (50%) introgressive hybridization and 50 generations of weak (1%) introgressive hybridization returned similar patterns in PSMC plots, with a slight increase in N_e but a wider time span for a past population growth (Figure S14). However, the hybridization events did not cause an artificial N_e spike in the PSMC plots, which were only revealed in simulated scenarios that included a recent population bottleneck (scenarios 2, 3, 5 and 6 in Figure S14). We also found that a longer population bottleneck led to a larger N_e spike (Figure S14), the same as we found in the simulations with no introgression but a bottleneck. The results indicated that inbreeding played a larger role than hybridization in misleading the PSMC inferences.

4 | DISCUSSION

Understanding how species are domesticated is an important issue in evolutionary biology (Larson & Burger, 2013; Price, 2002; Zeder,

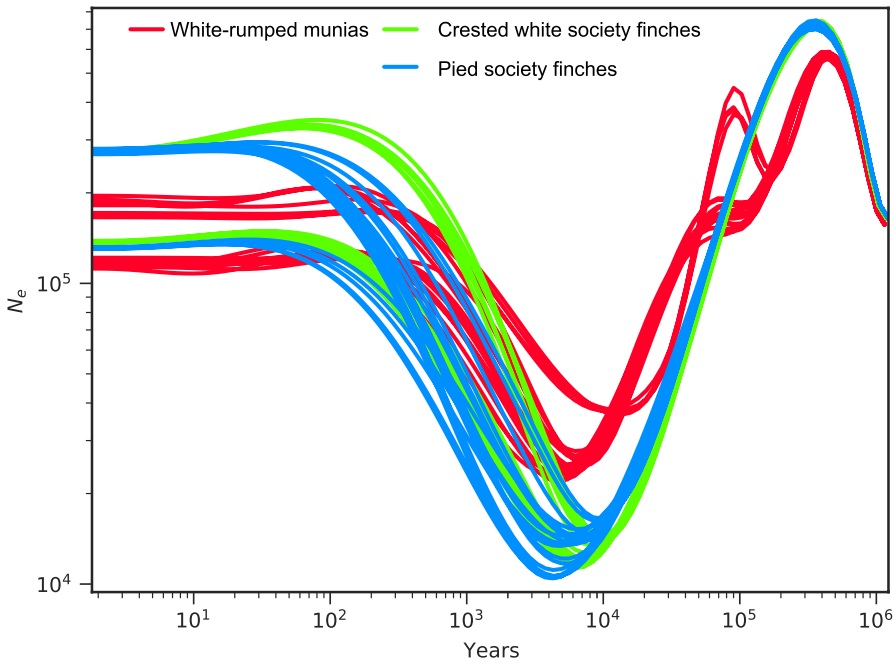

FIGURE 7 SMC++ inferences with 20 replicated runs for each of the white-rumped munia, crested white society finch and pied society finch. We set the generation time (g) as 2 years and the mutation rate (μ) as 3.44×10^{-9} per site per generation. The white-rumped munias comprise LST037, LST038, LST040 and L_UB. The crested white society finches comprise LDO004 and LDO005. The pied society finches comprise LDO253, LDO255, LDO256 and L_RO [Colour figure can be viewed at wileyonlinelibrary.com]

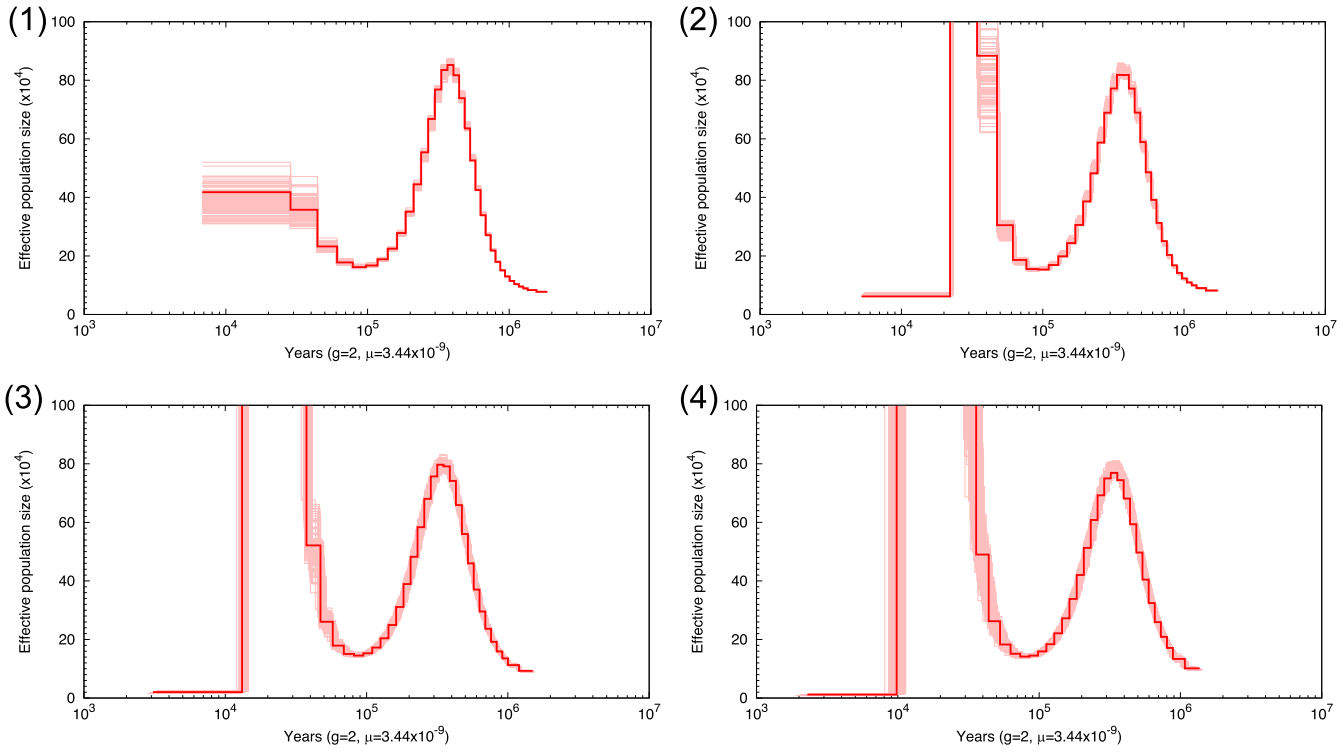

FIGURE 8 PSMC inferences based on 100 simulation replicates for each of the scenarios with different levels of population bottlenecks. Scenario (1) has no recent bottleneck; scenario (2) has a short bottleneck occurring 150–200 generations ago, followed by population size recovery; scenario (3) has a long bottleneck occurring 50–200 generations ago, followed by population recovery; and scenario (4) has a bottleneck lasting from 200 generations ago to the present. Simulation details are shown in Figure S2. The PSMC plots with full views are shown in Figure S10 [Colour figure can be viewed at wileyonlinelibrary.com]

2012). Nevertheless, domestication history is often challenging to infer owing to its complex nature (Larson & Burger, 2013). Society finches have been subject to research on the evolution of singing behaviour (Colquitt et al., 2021; Honda & Okanoya, 1999), and researchers have concluded that the elaborate songs of society finches are a result or byproduct of domestication (Okanoya, 2015). However, no strong evidence has been found involving the evolutionary relationship between the society finch and its presumable ancestor—the white-rumped munia—and thus this relationship remains contentious.

Here, we tested two hypotheses regarding the origin of the society finch—inbreeding vs. hybridization—by assembling the genomes

of it and the white-rumped munia and collecting genome resequencing data from multiple individuals. We used the genomic data to infer their phylogenetic relationships, estimate genetic diversity, ROH distributions and genetic divergence of ancestry blocks across genomes, and reconstruct demographic histories. Importantly, we also used the empirical data and simulated sequences to demonstrate that recent bottlenecks caused by domestication can obscure the signal left by population history so that demographic inferences from genome-based models may be confounded.

4.1 | Inbreeding played a larger role than hybridization in society finch domestication

Our analyses generally favoured the inbreeding origin hypothesis over the hybridization one for the society finch. First, the lower levels of SNP-based heterozygosity in (pied and crested white) society finch genomes than in white-rumped munia genomes are consistent with the inbreeding origin hypothesis. Second, we found that ROH were longer and around twice in number in society finches than in white-rumped munias. In addition, both F_{ROH} and M_{ROH} values were higher in society finches than in white-rumped munias, suggesting an inbreeding origin in the former (Gómez-Sánchez et al., 2018; Keller et al., 2011). Third, the genetic distance between the blocks with white-rumped munia ancestry in society finch genomes and white-rumped munia genomes was not significantly lower than that of blocks with society finch or mixed ancestry. The result is inconsistent with the hybridization origin hypothesis. Fourth, the ADMIXTURE results are incompatible with the prediction of the hybridization origin hypothesis, which predicts that the society finch is more likely to have introgressed from other species than is the white-rumped munia. Finally, we found that inbreeding, rather than hybridization, caused an artificial N_e spike observed in the PSMC models of society finches (see detailed discussion in the previous section). However, the inbreeding history cannot be inferred based merely on their genome-wide nucleotide diversity (π) levels, which were highest in the pied society finch, followed by the white-rumped munia and then the crested white society finch. The results suggest caution in interpreting π-based inference of domestication history.

Given that all society finches showed a stronger signal of inbreeding than white-rumped munias, the higher nucleotide diversity in pied society finches suggests that the society finch might have accumulated mutations faster than the white-rumped munia. The accumulation of new mutations in populations could be a result of relaxed natural selection and/or artificial selection for variable traits (Larson & Burger, 2013). Although domestication can cause inbreeding (Bosse et al., 2017; Lush, 1946), it might also reduce the costs of new mutations associated with predator avoidance, foraging ability and other ecological challenges (Okanoya, 2012), sustaining the genetic variations in the society finch. In addition, humans may intend to select for different colorations or other phenotypes that may add to the genetic variations underlying these traits (Larson & Burger, 2013). It is also possible that pied society finches contain several subdivided inbreeding populations, which may cause inflated estimates of nucleotide diversity for this taxon (Charlesworth, 2003). Finch breeders often cross different types of society finches to generate new breeds in a way that may also increase genomic diversity. However, these breeding practices may not necessarily increase average genetic heterozygosity because each breed may still have a high level of homozygosity, such as that of the crested white society finch (Figure 1c). We conducted extra analyses to support the argument. That is, the π values of a data set mixing the two society finch breeds were significantly higher than those of crested white society finches and nonsignificantly higher than those of pied society finches, whereas the heterozygosity levels of the mixed data set were significantly lower than those of pied society finches although they were significantly higher than those of crested white society finches (Figure S15). Consequently, genomic diversity cannot simply reflect population bottlenecks caused by domestication.

It is still possible that a history of hybridization followed by inbreeding among closely relative offspring caused the higher levels of nucleotide diversity in the pied society finch than the white-rumped munia, although this scenario cannot be fully reflected by the longer and more ROH in the former than the latter. In addition, even though hybridization alone did not cause strong biases in PSMC results, our simulation results also cannot completely reject the possibility of a domestication history that mixed hybridization and inbreeding in the society finch (Figure S14). However, the overall results of the multiple analyses still suggest that inbreeding is more likely to be the main driver of the society finch genomic variation than hybridization. Thus, one should use multiple approaches to fully understand the evolutionary history of species, especially those that may have a complex history mixed with population bottlenecks, growth or hybridizations (Zeder, 2015).

4.2 | The inferred domestication history of the society finch

Here we demonstrate how to use genomic analyses (e.g., the distribution of ROH length) combined with historical documents to infer when and where the domestication of the society finch began and how it has proceeded. Records suggest that the society finch was first shipped from China to Japan for trade, and was then introduced to Europe and North America (Svanberg, 2008). The first record of society finches (i.e., Bengalese finches) in Europe was two specimens with white plumage purchased by the Zoological Society of London in 1860; soon afterward, this type became common in the bird markets of Europe and North America. The name “Bengalese” came from “Bengali,” which might have resulted from an erroneous labelling when the bird was traded from Asia to Europe (Svanberg, 2008). In fact, Japan was the main exporter of society finches to western countries. In 1854, when Japan was forced to end its “closed country” policy and open up to other countries, lots of Japanese products were traded to Europe and North America, including pet birds. Society finches with different colour variations might have been

bred in Japan before being introduced to western countries (Taka-Tsukasa, 1922). The crested white society finches used in this study had more and longer ROH and lower levels of genomic heterozygosity than the pied society finches, probably due to intensive selection for the former's white plumage and crested feathers, which were preferred in Japanese aviculture (Washio, 1996).

Our results support the scenario in which the society finch first arose in China around 400 years ago, experienced another round of inbreeding in Japan, and was introduced to the Western world from Japan one or two centuries later. Before Japan imported the early type of society finches (probably a chocolate or pied type, which resembles the white-rumped munia) from China in the 1760s (Okanoya, 2012; Washio, 1996), it is likely that these birds had already been domesticated in China around 1620–1718 (i.e., 151–200 generations ago, one generation =2 years; Figure 2). We showed that crested (or pure) white society finches emerged around 1720–1818 (i.e., 101–150 generations ago), probably through artificial selection for white plumage in the early form of society finches, which were first imported to Japan in 1763 (Okanoya, 2012; Washio, 1996). In the 1850s, society finches, including the white type, were exported from Japan to Europe and North America.

The society finch has experienced a different domestication process than other well-known domesticated taxa, such as dogs and cattle. In general, there are three pathways for initiating domestication: the commensal, prey and directed pathways (Larson & Burger, 2013; Zeder, 2012). The commensal pathway starts with wild animals being attracted to human settlements or anthropogenic environments, later becoming habituated to humans and establishing a commensal relationship with them (e.g., dogs). By contrast, in the prey pathway, humans first hunt an animal for food and then develop strategies to gradually manage the prey as herds of livestock (e.g., cattle). Finally, in the directed pathway, humans directly capture wild animals (e.g., hamsters) to breed them for particular demand or interest, skipping the habituation and management phases; this pathway is a shorter process, has occurred in more recent history and leads to severer bottlenecks in the domesticates than the commensal or prey pathways (Larson & Burger, 2013; Zeder, 2012). Our results support the view that the society finch arose through the directed pathway, rather than the first two pathways, to become pets or foster parents for other pet birds in the recent past.

4.3 | Recent inbreeding misleads the PSMC model

The amount of published genome-wide data has been increasing in the past decade, facilitating research on evolutionary biology and conservation. Taking advantage of these increases in available genome sequences, the PSMC model has become a powerful and widely used tool for inferring demographic history for various taxa (e.g., Mammalia, Aves, Testudines, Osteichthyes and Arthropoda; Li & Durbin, 2011; Hung et al., 2014; Nadachowska-Brzyska et al., 2015; Chen et al., 2016; Barth et al., 2017; Westbuty et al., 2019; Quesada et al., 2019; Shingate et al., 2020), with inferences mostly or only reliable for the distant past (Dong et al., 2021; Patton et al., 2019). One study showed that fragmentation in genome assemblies has little effect on PSMC analysis (Patton et al., 2019). In contrast, another study suggested that genome-based models, including PSMC, have difficulty reconstructing complex demographic histories (Beichman et al., 2017). Studies also showed that linked selection and population structure could bias Sequentially Markovian Coalescent model inferences (Johri et al., 2021; Mazet et al., 2016). In addition, PSMC might misestimate splitting times between domestical species and their wild relatives because this model does not take into account the possible effect of admixture (Wang et al., 2021). However, few studies have examined whether recent fluctuations in population size (<1,000 years ago) may influence PSMC inferences. The society finches were domesticated around 400 years ago, making them a good system to examine how PSMC performs in species with recent bottlenecks or other demographic changes.

In this study, the PSMC results of society finches showed a spurious spike in N_e from 20,000 to 40,000 years ago; this spike largely plummeted when the ROH regions of genomes were masked. Interestingly, the PSMC plots of white-rumped munias did not show an N_e spike and masking ROH regions had little impact on their PSMC results. Nadachowska-Brzyska et al. (2015) suggested that PSMC results were not affected by the removal of ROH regions from several avian genomes that, however, did not show an N_e spike in the first place. In contrast, our simulations that assumed a severe, recent bottleneck returned PSMC plots with a large, artificial N_e spike well predating this bottleneck. This indicates that recent demographic events such as inbreeding may affect PSMC inferences; thus, these inferences should be interpreted with caution, especially when a spike in N_e occurs. Although we cannot completely rule out the possibility that other factors such as linked selection and population structure might also create an artificial increase of N_e in the PSMC inferences of society finches (Johri et al., 2021; Mazet et al., 2016), our results still support that a recent bottleneck can cause a N_e spike in PSMC inferences.

The spurious spike could be misleading because it may be erroneously interpreted as a signal of population expansion or hybridization. Saremi et al. (2019) found a similar spike in the PSMC plot of one puma and argued that it was the result of hybridization between two divergent lineages. However, their argument was based on a hypothesis in which some hybrid individuals were released into the gene pool of the studied population around 50 years ago and analyses of just one single genome. Here we used coalescence-based simulations with an array of demographic scenarios to show that a recent population bottleneck, rather than hybridization, was responsible for the spurious N_e spike. Such a spike was observed in all six society finch genomes. Thus, this pattern is likely to be found in species that were domesticated through the directed pathway with intensive artificial selection on wild species followed by commercial breeding (Larson & Burger, 2013; Zeder, 2012).

Another recently domesticated species, the budgerigar (*Melopsittacus undulatus*), which was imported from Australia to Europe in 1840 and has been bred as a pet ever since (Daniell & Murray, 1986), also showed a population spike in its PSMC analysis (Nadachowska-Brzyska et al., 2015). The PSMC inference for the budgerigar—an N_e spike occurring 50,000–70,000 years ago that was

about twice the size of the second highest N_e in its history—is similar to those of society finches (Figure S16; we regenerated the budgerigar PSMC plot following the method of Nadachowska-Brzyska et al., 2015 with some modifications; see Supporting Methods for details). The spikes in the N_e of society finches were relatively higher than that of the budgerigar (about 2–30 times higher than the second highest N_e in the society finch history; Figure S10), probably because the budgerigar was domesticated more recently. This observation corroborates our finding that recent domestication may bias demographic inferences with the PSMC model.

In this study, we also used a demographic model based on multiple genomes—SMC++—for which the inbreeding history did not cause any artificial N_e spike. However, this multigenome model also did not reveal a recent population bottleneck in the society finch, as in the PSMC model. Thus, inferring recent population size changes is arduous even using demographic models based on either single or multiple genomes. Our findings suggest that if populations experienced particularly recent bottlenecks due to domestication, then their inferred history may show erroneous estimates in the direction, magnitude and timing of changes in N_e and mislead our understanding of domestication processes. Thus, we have to rethink the domestication history of many taxa, especially that of recently domesticated ones inferred from the PSMC model.

5 | CONCLUSIONS

In this study, we examined the evolutionary relationship between the society finch and white-rumped munia by constructing their draft genomes with long scaffolds. We combined both empirical and simulated data to provide genetic evidence supporting that humans domesticated the white-rumped munia probably through inbreeding to generate the society finch around 400 years ago. We also found that such domestication-associated inbreeding may not always reduce the nucleotide diversity in domesticates, probably because domestication leads to relaxed natural selection or selective breeding for diverse strains. However, we still cannot completely reject the possibility that there was a history of hybridization followed by inbreeding among closely related offspring in domestication of the society finch. Genomic data from other Estrildid finches of East Asia with which the white-rumped munia might be hybridized to generate the society finch will be useful to further assess this possibility. Importantly, our simulations reveal that PSMC inferences are vulnerable to recent demographic events such as bottlenecks, even though this model is designed to depict population size changes in a more distant past. In addition, the whole genome data of the society finch and the white-rumped munia will be valuable resources for understanding genetic mechanisms governing the morphological and behavioural changes that separate a domesticate from its wild ancestor.

ACKNOWLEDGMENTS

We thank Yi-Ying Chen for taking care of the experimental animals. We are grateful to Ming-Yuan Tsai and Sue-Jen Huang for their important contributions to fieldwork. We also thank Royal Finch breeder Bangladesh for authorizing the use of a photo of crested white society finches. We thank Noah Last of Third Draft Editing for his English language editing. This project was supported by the Career Development Award of Academia Sinica (AS-CDA-108-L05).

AUTHOR CONTRIBUTIONS

C.M.H. and C.W.L. designed the research. C.W.L. and C.T.Y. performed the research. C.W.L. analysed the data. C.W.L. and C.M.H. wrote the paper. All authors commented on the paper.

DATA AVAILABILITY STATEMENT

The genome assemblies and raw sequencing reads are available in GenBank (PRJNA779480, SAMN23039802–SAMN23039811). The codes used for all analyses in this study including genome assembly, SNP calling, phylogeny, SNP-based heterozygosity, nucleotide diversity, ROH, ADMIXTURE and LAMP estimates, MACS simulation, PSMC and SMC++ are available in the Zenodo repository (https://doi.org/10.5281/zenodo.5672294).

ORCID

Chih-Ming Hung https://orcid.org/0000-0002-4785-6370

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Backström, N., Fostmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., Webster, M. T., Öst, T., Schneider, M., Kempenaers, B., & Ellegren, H. (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research*, *20*, 485–495. https://doi.org/10.1101/gr.101410.109

Barth, J. M. I., Damerau, M., Matschiner, M., Jentoft, S., & Hanel, R. (2017). Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna populations. *Genome Biology and Evolution*, *9*(4), 1084–1098. https://doi.org/10.1093/gbe/evx067

Beichman, A. C., Phung, T. N., & Lohmueller, K. E. (2017). Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3 - Genes Genomes Genetic*, *7*, 3605–3620. https://doi.org/10.1534/g3.117.300259

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, *310*, 170. https://doi.org/10.1136/bmj.310.6973.170

Bosse, M., Megens, H.-J., Derks, M. F. L., de Cara, Á. M. R., & Groenen, M. A. M. (2017). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary Applications*, *12*, 6–17. https://doi.org/10.1111/eva.12691

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics*, *19*, 220–234. https://doi.org/10.1038/nrg.2017.109

Charlesworth, D. (2003). Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1434), 1051–1070. https://doi.org/10.1098/rstb.2003.1296

Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., Liu, S., Lu, H., Tain, S., Li, R., & Shi, W. (2016). Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan n. spp. Molecular Biology*

and Evolution, *33*(5), 1337–1348. https://doi.org/10.1093/molbev/msw017

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, *19*, 135–142. https://doi.org/10.1101/gr.083634.108

Colquitt, B. M., Merullo, D. P., Konopka, G., Roberts, T. F., & Brainard, M. S. (2021). Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits. *Science*, *371*(6530), eabd9704. https://doi.org/10.1126/science.abd9704

Colquitt, B. M., Mets, D., & Brainard, M. S. (2018). Draft genome assembly of the Bengalese finch, *Lonchura straita domestica*, a model for motor skill variability and learning. *GigaScinence*, *7*, 1–6. https://doi.org/10.1093/gigascience/giy008

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. & 1000 Genomes Project Analysis Group (2011). The Variant Call Format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daniell, A., & Murray, N. D. (1986). Effects of inbreeding in the budgerigar *Melopsittacus undulates* (Aves: Psittacidae). *Zoo Biology*, *5*, 233–239. https://doi.org/10.1002/zoo.1430050216

del Hoyo, J., & Collar, N. J. (2016). *HBW and BirdLife International Illustrated Checklist of the Birds of the World. Volume 2: Passerines*. Lynx Edicions and BirdLife International, Barcelona, Spain and Cambridge, UK.

Dong, F., Kuo, H.-C., Chen, G.-L., Wu, F., Shan, P.-F., Wang, J., Chen, D., Lei, F.-M., Hung, C.-M., Liu, Y., & Yang, X.-J. (2021). Population genomic, climatic and anthropogenic evidence suggest the role of human forces in endangerment of green peafowl (*Pavo muticus*). *Proceeding of the Royal Society B: Biological Sciences*, *288*, 20210073. https://doi.org/10.1098/rspb.2021.0073

Gómez-Sánchez, D., Olalde, I., Sastre, N., Enseñat, C., Carrasco, R., Marques-Bonet, T., Lalueza-Fox, C., Leonard, J. A., Vilà, C., & Ramírez, O. (2018). On the path to extinction: Inbreeding and admixture in a declining grey wolf population. *Molecular Ecology*, *27*, 3599–3612. https://doi.org/10.1111/mec.14824

Goodwin, D. (1982). *Estrildid finches of the world. British Museum (Natural History)*. Oxford University Press, UK.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods in Molecular Biology*, *1962*, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

Honda, E., & Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the white-backed munia (*Lonchura striata*) and its domesticated strain, the Bengalese finch (*Lonchura striata var. domestica*). *Zoological Science*, *16*, 319–326. https://doi.org/10.2108/zsj.16.319

Hu, Y., Thapa, A., Fan, H., Ma, T., Wu, Q., Ma, S., Zhang, D., Wang, B., Li, M., Yan, L., & Wei, F. (2020). Genomic evidence for two phylogenetic species and long-term population bottlenecks in red pandas. *Science Advances*, *6*, eaax5751. https://doi.org/10.1126/sciadv.aax5751

Hung, C.-M., Shaner, P.-J.- L., Zink, R. M., Liu, W.-C., Chu, T.-C., Huang, W.-S., & Li, S.-H. (2014). Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proceedings of the National Academy of Sciences*, *1111*(29), 10636–10641. https://doi.org/10.1073/pnas.1401526111

Jackman, S. D., Coombe, L., Chu, J., Warren, R. L., Vandervalk, B. P., Yeo, S., Xue, Z., Mohamadi, H., Bohlmann, J., Jones, S. J. M., & Birol, I. (2018). Tigmint: Correcting assembly errors using linked reads from large molecules. *BMC Bioinformatics*, *19*, 393. https://doi.org/10.1186/s12859-018-2425-6

Johri, P., Riall, K., Becher, H., Excoffier, L., Charlesworth, B., & Jensen, D. J. (2021). The impact of purifying and background selection on the inference of population history: Problems and prospects. *Molecular Biology and Evolution*, *38*(7), 2986–3033. https://doi.org/10.1093/molbev/msab050

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C., & Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution*, *2*(1), 124–131. https://doi.org/10.1038/s41559-017-0375-4

Keller, M. C., Visscher, P. M., & Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, *189*(1), 237–249. https://doi.org/10.1534/genetics.111.130922

Kinsella, C. M., Ruiz-Ruano, F. J., Dion-Côté, A.-M., Charles, A. J., Gossmann, T. I., Cabrero, J., Kappei, D., Hemmings, N., Simons, M. J. P., Camacho, J. P. M., Forstmeier, W., & Suh, A. (2019). Programmed DNA elimination of germline development genes in songbirds. *Nature Communication*, *10*, 5468. https://doi.org/10.1038/s41467-019-13427-4

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Larson, G., & Burger, J. (2013). A population genetics view of animal domestication. *Trends in Genetics*, *29*(4), 197–205. https://doi.org/10.1016/j.tig.2013.01.003

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*, 493–496. https://doi.org/10.1093/bioinformatics/btp324

Li, S. H., Liu, Y., Yeh, C. F., Fu, Y., Yeung, C. K., Lee, C. C., Chiu, C. C., Kuo, T. H., Chan, F. T., Chen, Y. C., Ko, W. Y. & Yao, C. T. (in press) Not out of the woods yet: signatures of the prolonged negative genetic consequences of a population bottleneck in a rapidly re-expanding wader, the black-faced spoonbill Platalea minor. *Molecular Ecology*. https://doi.org/10.1111/mec.16260

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Review Genetics*, *4*, 981–994. https://doi.org/10.1038/nrg1226

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, Y., Han, C., Cheng, D. W., Yiu, S.-M., … Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*, 18. https://doi.org/10.1186/2047-217X-1-18

Lush, J. L. (1946). Chance as a cause of changes in gene frequency within pure breeds of livestock. *American Naturalist*, *80*(792), 318–342. https://doi.org/10.1086/281448

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

Mays, H. L. Jr, Hung, C. M., Shaner, P. J., Denvir, J., Justice, M., Yang, S. F., Roth, T. L., Oehler, D. A., Fan, J., Rekulapally, S., & Primerano, D. A. (2018). Genomic analysis of demographic history and ecological niche modeling in the endangered Sumatran rhinoceros *Dicerorhinus sumatrensis*. *Current Biology*, *28*(1), 70–76. https://doi.org/10.1016/j.cub.2017.11.021

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, *116*, 362–371. https://doi.org/10.1038/hdy.2015.104

McCarthy, E. M. (2006). Handbook of Avian Hybrids of the World. *Oxford*, 285–286.

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal dynamics of avian populations during Pleistocene

revealed by whole-genome sequences. *Current Biology*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Okanoya, K. (2012). Behavioural factors governing song complexity in Bengalese finches. *International Journal of Comparative Psychology*, *25*, 44–59.

Okanoya, K. (2015). Evolution of song complexity in Bengalese finches could mirror the emergence of human language. *Journal of Ornithology*, *156*, 65–72. https://doi.org/10.1007/s10336-015-1283-5

Oppenborn, G. (1992). *Japaniche Mövchen*. Herkunft, Zucht, Haltung.

Patton, A. H., Margres, M. J., Stahlke, A. R., Hendricks, S., Lewallen, K., Hamede, R. K., Ruiz-Aravena, M., Ryder, O., McCallum, H. I., Jones, M. E., Hohenlohe, P. A., & Storfer, A. (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian devils. *Molecular Biology and Evolution*, *36*(12), 2906–2921. https://doi.org/10.1093/molbev/msz191

Pfeifer, B., Wittelsbuerger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*, 1929–1936. https://doi.org/10.1093/molbev/msu136

Price, E. O. (2002). *Animal domestication and behavior*. CABI.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Quesada, V., Freitas-Rodríguez, S., Miller, J., Pérez-Silva, J., Jiang, Z.-F., Tapia, W., Santiago-Fernández, O., Campos-Iglesias, D., Kuderna, L. F. K., Quinzin, M., Álvarez, M. G., Carrero, D., Beheregaray, L., Gibbs, J. P., Chiari, Y., Glaberman, S., Ciofi, C., Araujo-Voces, M., Mayoral, P., … López-Otín, C. (2019). Giant tortoise genomes provide insights into longevity and age-related disease. *Nature Ecology and Evolution*, *3*, 87–95. https://doi.org/10.1038/s41559-018-0733-x

Restall, R. (1996). *Munias and Mannikins*. Pica Press, 81, 83.

Sankararaman, S., Sridhar, S., Kimmel, G., & Halperin, E. (2008). Estimating local ancestry in admixed populations. *American Journal of Human Genetics*, *8*(2), 290–303. https://doi.org/10.1016/j.ajhg.2007.09.022

Saremi, F. N., Supple, M. A., Byrne, A., Cahill, J. A., Coutinho, L. L., Dalén, L., Figueiró, H. V., Jason, W. E., Milne, H. J., O'Brien, S. J., O'Connell, B., Onorato, D. P., Riley, S. P. D., Sikich, J. A., Stahler, D. R., Villela, P. M. S., Vollmers, C., Wayne, R. K., Eizirik, E., … Shapiro, B. (2019). Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nature Communications*, *10*, 4769. https://doi.org/10.1038/s41467-019-12741-1

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, *1962*, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Shingate, P., Ravi, V., Prasad, A., Tay, B.-H., Garg, K. M., Chattopadhyay, B., Yap, L.-M., Rheindt, F. E., & Venkatesh, B. (2020). Chromosome-level assembly of the horseshoe crab genome provides insights into genome evolution. *Nature Communications*, *11*, 2322. https://doi.org/10.1038/s41467-020-16180-1

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

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

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, *7*, 62. https://doi.org/10.1186/1471-2105-7-62

Svanberg, I. (2008). Towards a cultural history of the Bengalese finch (*Lonchura domestica*). *Der Zoollogische Garten*, *77*, 334–344. https://doi.org/10.1016/j.zoolgart.2008.05.003

Taka-Tsukasa, N. (1922). Aviculture in Japan. *The Avicultural Magazine*, *13*, 15–22.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*, 303–309. https://doi.org/10.1038/ng.3748

Thompson, E. A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*, *194*(2), 301–326. https://doi.org/10.1534/genetics.112.148825

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 1–33. https://doi.org/10.1002/0471250953.bi1110s43

Wang, M. S., Zhang, J. J., Guo, X., Li, M., Meyer, R., Ashari, H., Zheng, Z.-Q., Wang, S., Peng, M.-S., Jiang, Y. U., Thakur, M., Suwannapoom, C., Esmailizadeh, A., Hirimuthugoda, N. Y., Zein, M. S. A., Kusza, S., Kharrati-Koopaee, H., Zeng, L., Wang, Y.-M., … Zhang, Y.-P. (2021). Large-scale genomic analysis reveals the genetic cost of chicken domestication. *BMC Biology*, *19*, 1–16. https://doi.org/10.1186/s12915-021-01052-x

Washio, K. (1996). *Enigma of Bengalese finches (in Japanese)*. Kindai Bungei Sha.

Weedn, L. (1993). The society finch: A bounty for beginners. *AFA Watchbird*, *20*(5), 29–30.

Weisenfeld, N. I., Kumar, V., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, *27*, 757–767. https://doi.org/10.1101/gr.214874.116

Westbuty, M. V., Petersen, B., Garde, E., Heide-Jøgensen, M. P., & Lorenzen, E. D. (2019). Narwhal genome reveals long-term low genetic diversity despite current large abundance size, *iScience*, *15*, 592–599. https://doi.org/10.1016/j.isci.2019.03.023

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken2. *Genome Biology*, *20*, 257. https://doi.org/10.1186/s13059-019-1891-0

Yeo, S., Coombe, L., Warren, R. L., Chu, J., & Birol, I. (2018). ARCS: scaffolding genome drafts with linked reads. *Bioinformatics*, *34*(5), 725–731. https://doi.org/10.1093/bioinformatics/btx675

Zeder, M. A. (2012). The domestication of animals. *Journal of Anthropological Research*, *68*(2), 161–190. https://doi.org/10.3998/jar.0521004.0068.201

Zeder, M. A. (2015). Core questions in domestication research. *Proceedings of the National Academy of Sciences*, *112*(11), 3191–3198. https://doi.org/10.1073/pnas.1501711112

Zeder, M. A. (2018). Why evolutionary biology needs anthropology: Evaluating core assumptions of the extended evolutionary synthesis. *Evolutionary Anthropology*, *27*, 267–284. https://doi.org/10.1002/evan.21747

SUPPORTING INFORMATION

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

How to cite this article: Lu, C.-W., Yao, C.-T., & Hung, C.-M. (2022). Domestication obscures genomic estimates of population history. *Molecular Ecology*, 31, 752–766. https://doi.org/10.1111/mec.16277