Received: 3 May 2023 | Revised: 14 August 2023 | Accepted: 30 August 2023

DOI: 10.1111/mec.17130

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Assessing the genetic composition of cotton-top tamarins (*Saguinus oedipus*) before sweeping anthropogenic impact

Linett Rasmussen[1,2] | Claudia Fontsere[1] | Iván D. Soto-Calderón[3] | Rosamira Guillen[4] | Anne Savage[5] | Anders Johannes Hansen[1] | Christina Hvilsom[2] | M. Thomas P. Gilbert[1,6]

[1]Center for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Copenhagen, Denmark

[2]Research and Conservation, Copenhagen Zoo, Frederiksberg, Denmark

[3]Laboratorio de Genética Animal. Grupo Agrociencias, Biodiversidad y Territorio, Instituto de Biología, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellín, Medellín, Colombia

[4]Fundación Proyecto Tití, Barranquilla, Colombia

[5]Proyecto Tití, Inc., Orlando, Florida, USA

[6]University Museum, NTNU, Trondheim, Norway

Correspondence
Linett Rasmussen and Claudia Fontsere, Center for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Øster Farimagsgade 5A, 1352 Copenhagen, Denmark.
Email: linett.rasmussen@sund.ku.dk and claudia.fontsere@sund.ku.dk

Funding information
Alfred Benzon Foundation; Danmarks Grundforskningsfond, Grant/Award Number: 143; Torben & Alice Frimodts Fund

Handling Editor: Brent Emerson

Abstract

During the last century, the critically endangered cotton-top tamarin (*Saguinus oedipus*) has been threatened by multiple anthropogenic factors that drastically affected their habitat and population size. As the genetic impact of these pressures is largely unknown, this study aimed to establish a genetic baseline with the use of temporal sampling to determine the genetic makeup before detrimental anthropogenic impact. Genomes were resequenced from a combination of historical museum samples and modern wild samples at low-medium coverage, to unravel how the cotton-top tamarin population structure and genomic diversity may have changed during this period. Our data suggest two populations can be differentiated, probably separated historically by the mountain ranges of the Paramillo Massif in Colombia. Although this population structure persists in the current populations, modern samples exhibit genomic signals consistent with recent inbreeding, such as long runs of homozygosity and a reduction in genome-wide heterozygosity especially in the greater northeast population. This loss is likely the consequence of the population reduction following the mass exportation of cotton-top tamarins for biomedical research in the 1960s, coupled with the habitat loss this species continues to experience. However, current populations have not experienced an increase in genetic load. We propose that the historical genetic baseline established in this study can be used to provide insight into alteration in the modern population influenced by a drastic reduction in population size as well as providing background information to be used for future conservation decision-making for the species.

KEYWORDS
conservation genomics, cotton-top tamarin, genetic diversity, historical DNA, population genetics, population structure

Linett Rasmussen and Claudia Fontsere contributed equally to this work.

1 | INTRODUCTION

Cotton-top tamarins are a primate species endemic to Colombia (Mast et al., 1993). These tamarins have a dispirited backstory related to their former use in biomedical research facilities. Specifically, in the 1960s, it was discovered that their physiology rendered them a suitable model for studying a variety of human diseases, from Epstein–Barr virus to colitis and colon cancer (Mast et al., 1993). Consequently, this interest led to capturing and mass exportation of cotton-top tamarins, and while the exact numbers of tamarins exported between 1960 and 1973 [when they were declared endangered and legal exportation was banned (Rodríguez et al., 2021)] is uncertain, estimates range from at least 20,000 to as many as 40,000 individuals (Mast et al., 1993; National Research Council, 1976). Although the demographic consequence of this exportation is unknown, it is assumed that it must have led to a decline in their population size. Unfortunately following the ban in 1973, both the illegal trafficking and deforestation continued (Mast et al., 1993; Miller et al., 2004) driving further population declines (Savage et al., 2010, 2016). Thus, cotton-top tamarins are currently listed as critically endangered on the IUCN red list (Rodríguez et al., 2021) with estimates of 7400 individuals left in the wild (Savage et al., 2010).

Very little is known about the genetic diversity (Gyllensten et al., 1994), population structure and evolutionary history of the cotton-top tamarin as the majority of previous studies on the species have focussed on either their relevance for the biomedical industry (Bertone et al., 1998; Wood et al., 1998), their demography, social behaviour and reproduction in long-term field studies (Savage et al., 1996, 1997, 2009, 2021, 2022; Wheaton et al., 2022) or detangling the phylogenetic relationships of the species within the Callitrichidae family (Brcko et al., 2022; Buckner et al., 2015). However, given the history of mass exportation and extensive illegal trafficking of the cotton-top tamarin, together with loss of habitat, it is possible they would have suffered a reduction in genetic variation during the last century, as has been seen in other species that have experienced similar pressures (Sánchez-Barreiro et al., 2021; van der Valk et al., 2019). Therefore, looking into the genomics of past populations could provide insight into the genetic health of the extant population by providing a baseline for comparison.

The analysis of DNA generated from time series of specimens is key to monitoring genetic diversity in species or their populations. While initially studies were constrained to the analysis of single markers, such as mitochondrial DNA (mtDNA), to study population structure of various species including *Alpine ibex*, *Iberian lynx* and multiple fish species (Casas-Marce et al., 2017; Robin et al., 2022; Rodrigues et al., 2018; Star et al., 2017), in recent years, whole genome resequencing has started to be applied, also including ancient DNA (aDNA) and historical DNA (hDNA) in order to characterize how genetic diversity has changed through time (Feng et al., 2019; Pinsky et al., 2021; Sánchez-Barreiro et al., 2021; van der Valk et al., 2019). Importantly, these types of studies can in theory provide baseline information that predates heavy negative anthropogenic impacts, thus establishing the genetic diversity against which levels of genomic erosion can be compared (Díez-del-Molino et al., 2018). Genetic diversity is not only an effect of population size but also a consequence of life history traits (Romiguier et al., 2014) and past events (Díez-del-Molino et al., 2018; Jensen et al., 2022). Since much genetic conservation efforts are focussed around genetic diversity, studies looking into the past to develop a baseline of genetic diversity can provide guidance within conservation and management decisions, that is translocations, and connectivity of current populations (Jensen et al., 2022) as well as evaluating the outcomes of such.

Given the current critically endangered status of cotton-top tamarins (Rodríguez et al., 2021), developing a baseline of genetic diversity of cotton-top tamarins before the mass exportation sparked by biomedical research can provide useful knowledge with respect to conservation efforts. Due to the continuing decline in population sizes and its popularity within the illegal pet trade, more interest is developing in understanding the genetic diversity of the cotton-top tamarin to provide more targeted conservation efforts. Furthermore, due to the high fragmentation of the cotton-top tamarin distribution, knowledge of historical connectivity between populations would assist in decision-making on what populations to prioritize for conservation purposes, where to establish corridors between fragmented forests, as well as indicate potential source populations for translocations and where confiscated cotton-top tamarins from the pet trade could be released. Given these potential uses, we aimed to create a baseline of the structure and genetic diversity of the cotton-top tamarin prior to the heavy pressure that they experienced in the 1960s. We resequenced and then analysed the whole genomes of 26 historical samples drawn from across their historical range and eight modern samples of cotton-top tamarins from the northern and southern range of their current distribution.

2 | MATERIALS AND METHODS

2.1 | Ethical note

This research adhered to the legal requirements of the Colombian government and CITES. The National Authority of Environmental Licenses (ANLA) granted sample collection through Permiso Marco de Recolección #0524 (May 27, 2014) and the Endangered/Endemic Species Permit (8201-2-32,320, 19 December 2019). The Committee on Ethics and Animal Research at the Universidad de Antioquia (3 May 2013) authorized all animal-handling protocols. With regard to the CITES Appendix I status of this species, the following steps were taken to ensure legal transfer of the materials. First, all historical samples were obtained with the permissions of the holding museum and then sent to the Globe Institute, University of Copenhagen under CITES permit (COSE DK014 and DK-2019-0009230-01) for subsequent processing. Second, while the sequencing itself was performed in Denmark, the DNA extraction and library preparation of modern samples prior to sequencing was performed in Colombia.

2.2 | Historical sample collection

We targeted historical specimens that were collected before the 1960s, thus before the heavy exportation of the species for biomedical research. Participating museums subsampled specimens from their collection, ultimately providing the study with 42 historical samples for initial screening (32 resequenced and 26 included in the study, Tables S1 and S2). The majority of samples were skin clippings and a few bone fragments (Table S1). Samples were stored and processed at a facility dedicated to the analysis of ancient DNA at the Globe Institute, University of Copenhagen.

2.3 | Modern sample collection

A total of eight samples were collected from cotton-top tamarins at three sites within their current and historic species distribution in the Colombian departments of Antioquia and Bolívar (Figure 1a, Table S3): four at Santa Catalina (Ceibal, Proyecto Titi), two at San Juan Nepomuceno (Los Titíes de San Juan Forest Reserve, Proyecto Titi) and two at Carepa (Tulenapa Forest, Universidad de Antioquia). Proyecto Titi conducts research at two field sites in the northern region of the species' distribution. Parque Natural Regional Bosque Seco El Ceibal Mono Titi (Perez Barrios, 2013) is a privately owned property in Santa Catalina, Bolívar. Approximately 421 ha of this reserve are categorized as tropical dry forest. Los Titíes de San Juan Forest Reserve is a privately owned protected reserve located in San Juan Nepomuceno, Bolívar, and is approximately 480 ha of tropical forest and land under various restoration regimes. The third site is located in the southernmost region of the species' distribution in Carepa, Antioquia. It is known as the Tulenapa Forest managed by the Universidad de Antioquia, a 150 ha university-protected mosaic of secondary forest mixed with croplands enclaved in an extensive field of banana plantations.

Each modern sample was obtained following the trapping of social groups that had been captured alive in 10- or 12-compartment traps using established methods developed for tamarins (Savage et al., 1993). Tamarins were weighed and sedated with either 12 mg/kg intramuscular Ketamid® (ketamine-midazolam) or with a combination of ketamine HCl (25 mg/kg) and midazolam HCl (0.1 mg/kg). Animals were then given a complete physical examination and ~20–30 abdominal hairs were collected in clean plastic bags or tubes at the Ceibal and San Juan sampling sites, using a clean pair of gloves for each animal. Postcollection hair samples were stored in a location with a dehumidifier until shipped to the University of Antioquia. At the Tulenapa site, between 1.5 and 2.0 mL (≤1% body mass) of blood were also collected in either EDTA or dry MiniCollect® tubes (Monroe) for clinic and genetic analyses. A volume of saline solution equal to the amount of collected blood was administered to replenish the amount of liquid taken (Soto-Calderón et al., 2016). Blood samples were refrigerated and sent to the Universidad de Antioquia within 24 h after collection to be processed. Each individual was returned to the original trap cell; all the group members were collectively released in their home range once they were fully recovered from anaesthesia.

2.4 | Historical sample preparation

To prevent contamination of the historic samples with modern DNA, certain precautions were taken, including reagents of molecular biology grade, blank extraction controls and performing all preamplification steps in a specialized ancient DNA facility that is physically isolated from the location where post-PCR amplified and modern DNA products are manipulated. Standard aDNA approaches for working with historical samples (PPE clothing, double-layered gloves, deep cleaning, facemasks, working inside laminar flow hoods, etc.) were followed. To help decontaminate the skin clippings from exogenous sources of DNA, all samples were pretreated with 5% bleach and then immediately washed with absolute ethanol and water. After washing, more water was added to the samples and incubated at 37°C for an hour to soften the tissue. Water was then replaced with a digestion buffer modified for hides, keratin and cartilage (Campos & Gilbert, 2012), which was left overnight to incubate at 56°C. Bone fragments were crushed into powder and predigested for 1 h at 37°C with a digestion buffer composed of a mixture of proteinase *K*–urea (Ersmark et al., 2016). Predigestion buffer was removed, and fresh digestion buffer was added for overnight digestion. All samples were purified with the Qiagen minElute PCR purification kit (Qiagen) with modifications, including modified PB buffer and double the amount of PB buffer (10×). Libraries were built using the single reaction single-stranded library protocol (Kapp et al., 2021) with modification of adaptors to render the libraries BGISeq compatible (van Grouw et al., 2023). A quantitative PCR was performed on all libraries to identify the number of cycles needed for each sample during library amplification. Libraries were then amplified with the appropriate number of cycles. Samples were then sequenced at BGI Copenhagen with DNBSEQ-G400 using 100 bp single-end chemistry.

2.5 | Modern sample preparation

In the case where the same individual was captured multiple times, we combined all the hair samples before DNA extractions. Approximately 10–40 hairs per cotton-top tamarin were digested with digestion buffer described by (Campos & Gilbert, 2012) with an overnight incubation at 56°C and purified with the MinElute PCR purification kit (Qiagen) per the manufacturer's instructions. Genomic DNA was subsequently extracted and purified with the DNeasy Blood and Tissue kit (Qiagen) per the manufacturer's instructions. Samples were then fragmented as preparation for library build with Covaris LE220-plus ultrasonicator per the manufacturer's instructions with the fragmentation length set to 450 bp. The postfragmentation protocol for the modern samples was the same as the historical samples listed above. Libraries were sequenced at BGI Copenhagen using either DNBSEQ-G400 150 paired-end or 100 paired-end chemistry.

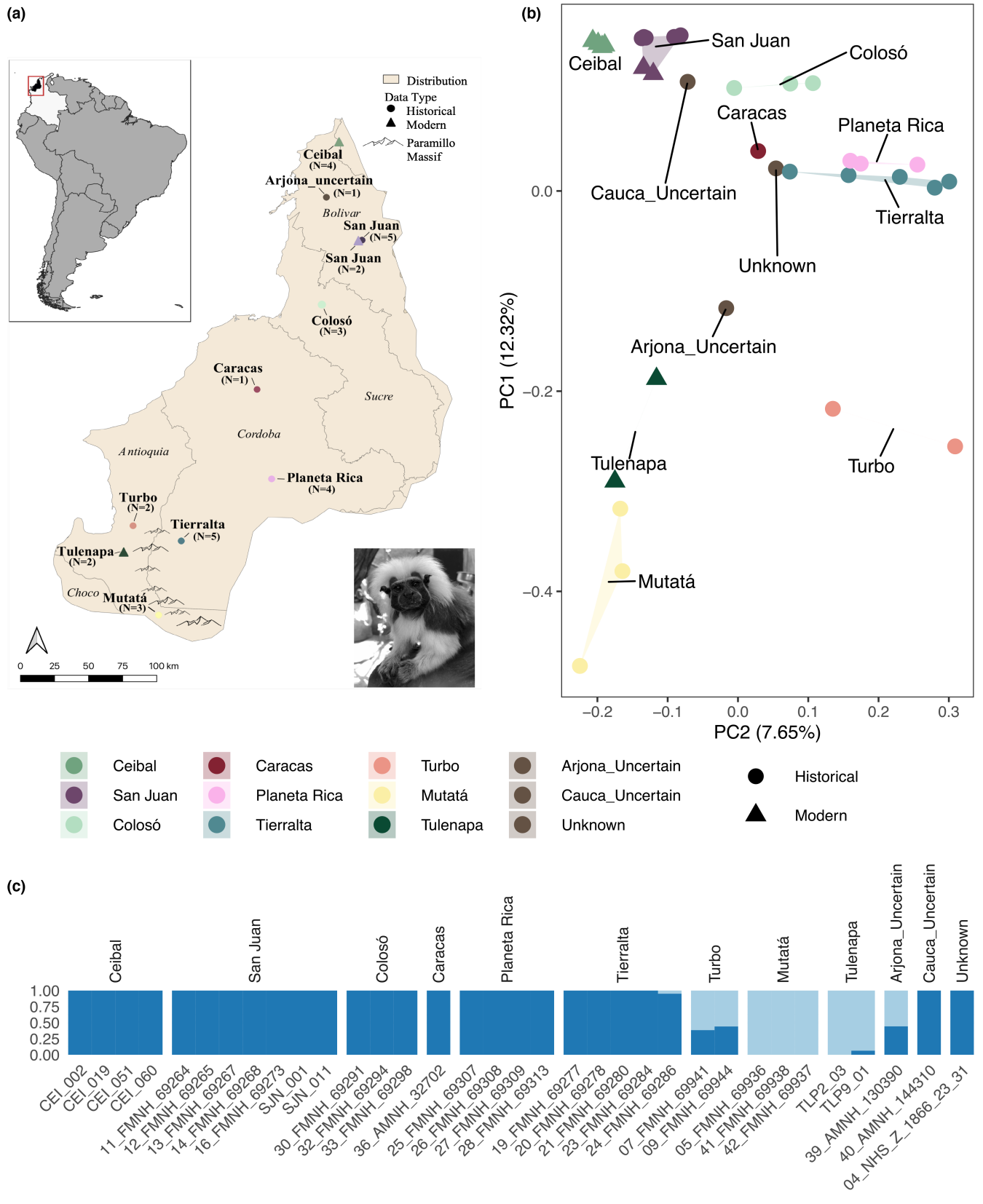

FIGURE 1 Population structure and geographical distribution of cotton-top tamarins. (a) A map representing the historical geographical distribution range of the cotton-top tamarin in Colombia. Sites for this study are colour coded, and the number of samples per site is listed under the site name. The mountain symbols in between Tierralta and, Turbo, Tulanapa and Mutatá represent the Paramillo Massif mountain range. (b) PCA based on 4,522,615 SNPs for the cotton-top tamarin individuals ($N = 34$) using the same colour code as Figure 1a. Historical and modern samples are indicated by circles and triangles, respectively. (c) Admixture analysis at $K = 2$ of all samples ($N = 34$). Each vertical bar represents one individual, and each cluster of vertical bars is representative of a site. In addition, sites have been ordered geographically from north (left) to south (right), with the three uncertain localities (Arjona, Cauca and Unknown) listed furthest to the right.

2.6 | FASTQ trimming, mapping and quality control of DNA sequence data

The 42 historical samples were initially sequenced at a very low depth, aiming at 0.1× coverage although actual output varied considerably due to challenges of pooling aDNA libraries (Table S1), as a form of prescreening to ensure samples contained sufficient levels of endogenous DNA. After prescreening, a total of 32 historical samples (Table S2) and eight modern samples (Table S3) were sequenced for deeper coverage. In addition, NgsRelate v2 (Hanghøj et al., 2019) was used to make sure related individuals were removed from the dataset. The historic dataset was subsequently reduced further to a final set of 26 (Table S4), as two samples were removed due to low coverage, two were identified as related individuals of low coverage, while two others were identified as belonging to another species (Table S2). Reliable information relating to the geographical origin of the historical samples analysed in this study was available for most samples, apart from one sample that did not contain a known locality (04_NHS_Z_1866_23_31). Furthermore, two samples were recorded as collected from locations that fall outside their known natural range or its sampling coordinates derive from the middle of a town, pointing towards a trafficked origin (hence labelled 'uncertain'; Table S4).

The resequenced historical samples and modern samples were mapped with PALEOMIX v1.2.13.2 (Schubert et al., 2014). Adapters were removed with AdapterRemoval v2.2.4 (Schubert et al., 2016). Sequences were aligned to the closely related golden-handed tamarin (*Saguinus midas*) chromosome-level reference genome (ASM2149847v1) and the *Saguinus oedipus* mitochondrial reference genome (NC_021960) (Finstermeier et al., 2013) with the BWA v0.7.16 backtrack algorithm (Li & Durbin, 2009). Furthermore, duplicates were filtered with MarkDuplicates v2.9.1 in Picard (http://broadinstitute.github.io/picard/) and damage of historical samples was assessed with mapDamage v2.0.9 (Jónsson et al., 2013) (Figure S1). Minimum base quality filtering was set to zero for historical samples and 20 for modern samples in order to maximize the amount of reads. More customized filtering was performed in later steps.

2.7 | Variant calling and filtering

We used snpAD v0.3.10 (Prüfer, 2018) to call genotypes in all samples on the autosomal chromosomes. This software was specially developed to account for DNA damage patterns when calling genotypes in ancient and historical specimens. The mapped sequences were transformed from bam-format into snpAD-format files, priors for base composition estimated, and genotypes were called using standard settings. The VCF files were then combined and concatenated with BCFtools (v1.15.1) (Danecek et al., 2021) merge and concat, respectively. We filtered the VCF file with BCFtools filter v1.15.1 (Danecek et al., 2021) to keep only variable sites with minimum read depth of 3, max read depth of 50, genotype quality >30, also excluding indels, multiallelic sites and sites with missingness >0.7. In addition, each sample's allele balance was checked by counting the proportion of reads supporting each allele in heterozygous positions and visualizing the density plots with R v4.2.1 (R Core Team, 2022).

2.8 | Population structure

We used principal component analysis (PCA) and admixture to assess the population structure in the dataset. PCA of all samples (including historical and modern) was done using PLINK v.1.90b (–pca) (Purcell et al., 2007) from the filtered VCF file containing biallelic sites present in at least 70% of the dataset with a minor allele frequency (MAF) of 0.05. Using PLINK v1.9 (Purcell et al., 2007), VCF files were converted to PLINK format (.bed, .bim, & .fam). Admixture was performed 10 times for *K* values of 2 through 10 with ADMIXTURE v1.3.0 (Alexander et al., 2009), cross-validation errors were used to determine best *K*, and likelihoods were used to determine the best run for visualization. In addition, ANGSD v0.921 (–dofasta 2) (Korneliussen et al., 2014) was used to form a consensus of the mtDNA for each sample. All transitions were removed to account for damage patterns of historical samples. mtDNA consensus sequences were used to reconstruct haplotype networks with POPart (Leigh & Bryant, 2015).

2.9 | Population demography

Heterozygosity was calculated with ANGSD v0.940 (Korneliussen et al., 2014) on the autosomes. We first pre-run ANGSD to obtain the list of transversions with the following parameters: -rmTrans 1 -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -C 50 -baq 1 -minMapQ 20 -minQ 20 -setMaxDepth 50 -doCounts 1 -doMajorMinor 1 -GL 2 -doGlf 2 -doMaf 2. Next, we used these sites to obtain the SAF files with the same filtering parameters as before but instead using -sites -doSaf 1 -GL 1 and -noTrans 1 for each sample. Finally, with winsfs v0.7.0 (Rasmussen et al., 2022) we obtained the sfs estimates from which we calculated the heterozygosity values. Heterozygosity estimates were corrected to account for using only transversions by multiplying the values by 3.02 (Ts/Tv + 1; Ts/Tv = 2.02). To account for biases introduced by the uneven coverage of the dataset, we downsampled the highest coverage samples to 5× using samtools view v1.10 (Danecek et al., 2021) and estimated the heterozygosity as previously described. Moreover, the inbreeding coefficient (F_i) of all samples was obtained with NgsRelate v2 (Hanghøj et al., 2019) with the option -F 1 after estimating allele frequencies and obtaining genotype likelihoods with ANGSD v0.939 with the following parameters: -uniqueOnly 1 -remove_bads 1 -noTrans 1 -only_proper_pairs 1 -C 50 -baq 1 -skipTriallelic 1 -gl 2 -minMapQ 30 -nThreads 10 -doGlf 3 -doMajorMinor 1 -doMaf 1 -minMaf 0.00001 -SNP_pval 1e-6. We also estimated the runs of homozygosity (RoH) in the samples with coverage >7× to avoid biases introduced by low coverage and damage patterns as suggested by ROHan (Renaud et al., 2019). We first used ROHan to estimate the damage patterns

using estimateDamage.pl and then obtained RoH in windows of 1 Mb on the autosomal scaffolds using the following parameters: --tstv 2.02 --size 1,000,000 --rohmu 2e-5 --deam5p $file.5p.prof --deam3p $file.3p.prof --auto Chrom_autosomes. The detected RoH per sample were visualized across the chromosome-length scaffolds with R v4.2.1 (R Core Team, 2022) and ggplot2 v3.4.1 package. We then calculated the fraction of the genome in RoH (F_{ROH}) obtained from the output file HMM posterior decoding using *.mid* estimates of heterozygosity, and we binned them by RoH size. We inferred the approximate age of inbreeding based on the RoH length and assuming a constant recombination rate of 1 cM–1 Mbp and a generation time of 6 years (Rodríguez et al., 2021). We used the RoH lengths per individual to time the inbreeding and we used the longest RoH of each sample to determine the time of the potentially most recent inbreeding event. We followed the formula $G = 100/(2^*\text{Mbp})$ (G = generations, Mbp = length of RoHs in Mbp) from (Thompson, 2013) to obtain the age estimates based on RoH length.

Furthermore, ANGSD was used to determine the fixation index (F_{ST}). Due to the limited number of samples per site, F_{ST} was calculated by regions (i.e. greater northeast and southwest) suggested by the population structure analyses.

We used AdmixTools to compute the f3-outgroup statistic and test for shared drift between each sampling site (only with samples with known geographical origin) using qp3Pop (Patterson et al., 2012) on the filtered dataset, using sites with a read depth of 5 and present in at least 80% of the samples, and with *Saguinus midas* as outgroup. We then plotted the pairwise F_3 values with R v4.2.1 (R Core Team, 2022).

We applied EEMS (Petkova et al., 2016) to infer past patterns of effective migration or potential barriers of gene flow between cotton-top tamarin locations. We ran this programme using only the historical samples with known origin (Table S4). We used the filtered VCF file with 5% MAF and containing variable sites present in at least 80% of the samples. We obtained 10 replicate runs of EEMS to ensure consistency with the following parameters nIndiv = 23, nSites = 134,214, nDemes = 1000, diploid = TRUE, numMCMCIter = 2,000,000, numBurnIter = 1,000,000, numThinIter = 9999. Next, we plotted the results in R v4.2.1 (R Core Team, 2022) with the package reemsplots2 (https://github.com/dipetkov/reemsplots2).

Using SNPeff v5.1 (Cingolani et al., 2012), we built a database with the reference genome for the golden-handed tamarin (Shao et al., 2023). SNPeff was then used to calculate the genetic load of the samples using the golden-handed tamarin database with the filtered VCF file not allowing for missingness and only variable positions in the cotton-top tamarins without minor allele frequency filtering and with minimum read depth of 5, maximum read depth of 50 and minimum genotype quality of 30. We counted the total number of derived alleles classified as high impact and moderate impact (total load) per sample and also separated by zygosity: in homozygous state (realized load) and in heterozygous state (masked load). Moderate-impact variants are nondisruptive variants that might change protein effectiveness (i.e. missense variants) and high-impact variants are assumed to have a disruptive impact in the protein, probably causing protein truncation, loss of function (LoF) or triggering nonsense-mediated decay (i.e. stop codons, splice donor variant and splice acceptor and start codon lost). To account for the effect of coverage, we only analysed samples with coverage >5× and we normalized the derived allele count by the derived allele count per sample in synonymous positions. Next, by using the relative counts of genetic load, we estimated the average genetic load per population (grouped by location: greater northeast and southwest, and by sample type: Historical and Modern). Finally, we calculated the total load ratio by comparing the averages of relative counts of genetic load (for high- and moderate-effect variants) per population in different comparisons.

3 | RESULTS

In this study, we compiled a genome resequencing dataset of 34 cotton-top tamarins, of which 8 represent modern and 26 historical (collected before 1960) samples (Table S4). The samples span 9 geographic locations (Figure 1a). We sequenced to an average coverage of 5× for the historical samples (ranging from 2 to 14×), and an average coverage of 9× for the modern samples (ranging from 4 to 14×) (Table S4). The historical samples showed deamination patterns but at low levels (<5%) as per their young age (~60 years) (Figure S1). With this dataset, we described 7,467,948 variable positions passing quality filters. Only one cotton-top tamarin sample (CEI_060) exhibited allele imbalance in heterozygous positions (Figure S2), most probably as a result of cross-contamination from another individual from the same site during the combination of samples before extractions (see Section 2). Thus, this sample was removed from further analyses (heterozygosity, inbreeding coefficient and runs of homozygosity).

3.1 | Population structure

We used our dataset to initially explore the population structure of the cotton-top tamarin by constructing a PCA for the 4,522,615 single-nucleotide polymorphisms (SNPs) with 5% MAF. The PCA shows a geographical pattern with a northeast-to-southwest gradient (Figure 1b). This suggests that geography influences the structure of the populations with possibly two clusters, separating the three localities residing in the province Antioquia from the rest.

Next, we performed an admixture analysis at different Ks (Figure S3) with modern and historical samples and found $K = 2$ to be the best fit for the data (Figure S4). Similar to the PCA results, the cotton-top tamarin samples are split into two groups, where the first group contains samples from the sites Ceibal (Santa Catalina), San Juan, Colosó, Caracas (Cereté), Planeta Rica, Tierralta, and two of the uncertain localities (Cauca and unknown) and the second group is comprised of samples from Turbo, Mutatá, Tulenapa (Carepa) and one of the uncertain locality samples (Arjona) (Figure 1c). Henceforth, we refer to the group containing the three localities in

Antioquia and Arjona as the 'southwest' group, and the group containing the remaining samples as the 'greater northeast' group. Further investigations into the structure at larger values of K revealed the emergence of specific components linked to each sampling site. For example, at $K=3$, the first splits included samples from Tierralta and Planeta Rica, from the more northern sites (San Juan and Ceibal) following their geographic distribution, with small instances of admixture in the sites located in the middle such as Caracas and Colosó (Figure S3). Furthermore, samples with uncertain geographical origin show a clear indication of which regions they potentially belong to not only at $K=3$ but also at higher K (Cauca_uncertain is similar to Colosó, Arjona_uncertain is similar to Turbo and the Unknown_uncertain is similar to Caracas; Figure S3), which are also supported by the PCA (Figure 1b). However, the mtDNA haplotypes do not show evidence of the structure suggested by the nuclear DNA (Figure S5). Still, mtDNA haplotypes seem to be more similar within sites rather than between sites.

3.2 | Population demography

We investigated whether the two-population structure has persisted through present times by estimating the weighted F_{ST} between the groups (Table S5). Southwestern historical localities (Mutatá and Turbo) were compared with the modern locality found in the southwest (Tulenapa). This analysis resulted in a weighted F_{ST} of .028, suggesting there is little genetic differentiation between the current and historical southwestern localities. We could see a similar pattern when comparing the greater northeastern historical and modern localities (weighted $F_{ST}=.034$). However, as expected, higher genetic differentiation is present between the southwestern historical localities and the greater northeastern historical localities (weighted $F_{ST}=.056$), confirming the established two-population structure.

Next, historical effective migration between sampling sites was explored with EEMS (Petkova et al., 2016) which suggested possible connectivity between the central and northern areas (Figure 2), similar to the pattern we see in the ADMIXTURE results at $K=2$ and $K=3$ (Figure S3). Moreover, this analysis suggested a barrier of gene flow in the southwest from the rest of the distribution, with less migration than expected under a model of isolation by distance. This strong barrier correlates with the Paramillo Massif mountain range, which is potentially responsible for the two-population structure. The analysis of shared drift with f3-outgroup statistics also supports the two clusters, with the southwest separated from the rest. The greater northeast sampling sites share more drift between them than towards the southwest, highlighting, as well, the barrier of gene flow between Tierralta and the southwestern sites (Figure S6).

We next explored the potential loss of genomic diversity in cotton-top tamarins during the last century due to the population reduction and habitat fragmentation. We estimated genome-wide heterozygosity per sample and compared the values between historical and modern individuals (Table S4). We first checked the effect of coverage on the heterozygosity estimates and found a statistically significant positive correlation (adjusted R-squared: .5783; p-value = 1.704e−07) between both variables (Figure S7). Thus, we restricted the analysis to samples with at least 5× coverage or downsampled to this coverage. We found a significant 1.68× decrease in heterozygosity estimates, from an average of 0.000589 bp^{-1} in the historical samples, to 0.00035 bp^{-1} for the modern samples (Wilcoxon rank-sum test, p-value = .0028; Figure S8). However, when structuring the dataset into the two geographical clusters, we noted

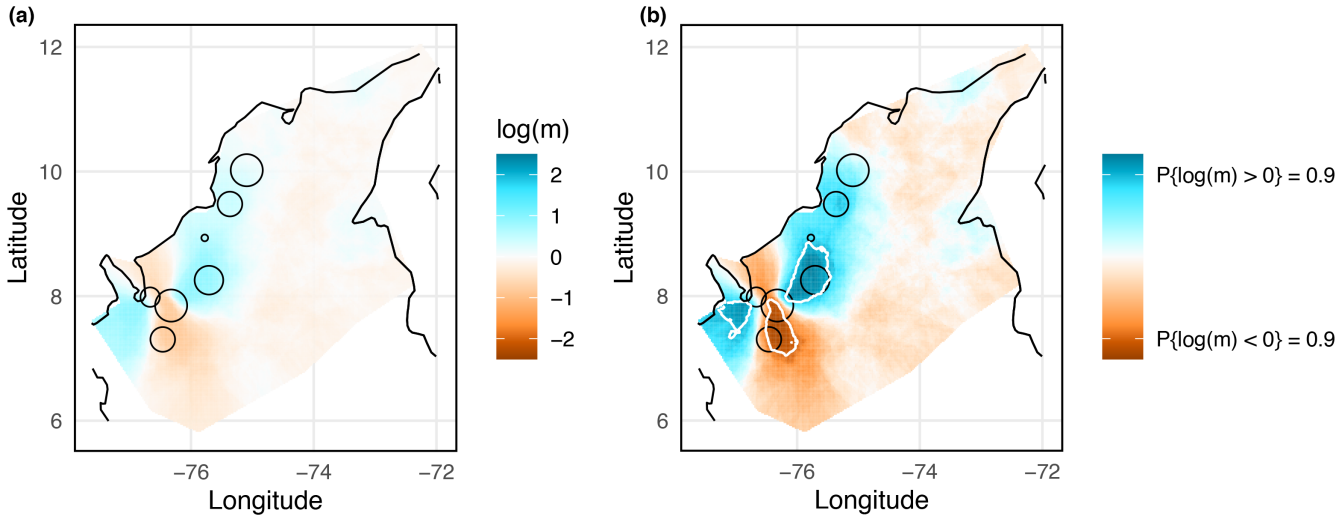

FIGURE 2 Barriers and corridors of gene flow in historical cotton-top tamarin populations. Estimated effective migration surface calculated by EEMS where blue represents higher migration than expected under a model of isolation by distance and brown represents lower migration than expected under isolation by distance. Black circles indicate sampling sites, and their size is relative to the number of samples per site. Only historical samples with known origin were used in this analysis (Table S4). (a) Mean migration rate estimated in log scale. (b) Posterior log mean migration rate with 95% confidence interval highlighted by white outlined polygons.

that the effect was of different magnitude across their range. Specifically, while a significant decrease of 2.1× was seen in the cotton-top tamarins in the greater northeast (average of 0.000581 and 0.000276 bp^{-1}, respectively; Wilcoxon rank-sum test, *p*-value = .016, Figure 3a), the southwestern individuals did not exhibit a significant heterozygosity reduction (1.3× decrease, Wilcoxon rank-sum test, *p*-value = .67), although due to our subsampling only three samples are left in this group (Figure 3a). Heterozygosity was also investigated per location which further confirms a decrease in heterozygosity within modern populations in the greater northeast population, Ceibal and San Juan (Figure S9).

Next, we explored inbreeding by estimating the inbreeding coefficient (F_i) of all samples (Table S4) with NgsRelate (Hanghøj et al., 2019) and the fraction of the genome in RoH (F_{ROH}) in the subset of samples with the highest coverage (>7×, *N* = 9, Table S6) using ROHan (Renaud et al., 2019). In line with the decreases in heterozygosity seen in the genomes of the modern samples, we also noted that although they exhibit higher inbreeding coefficients than

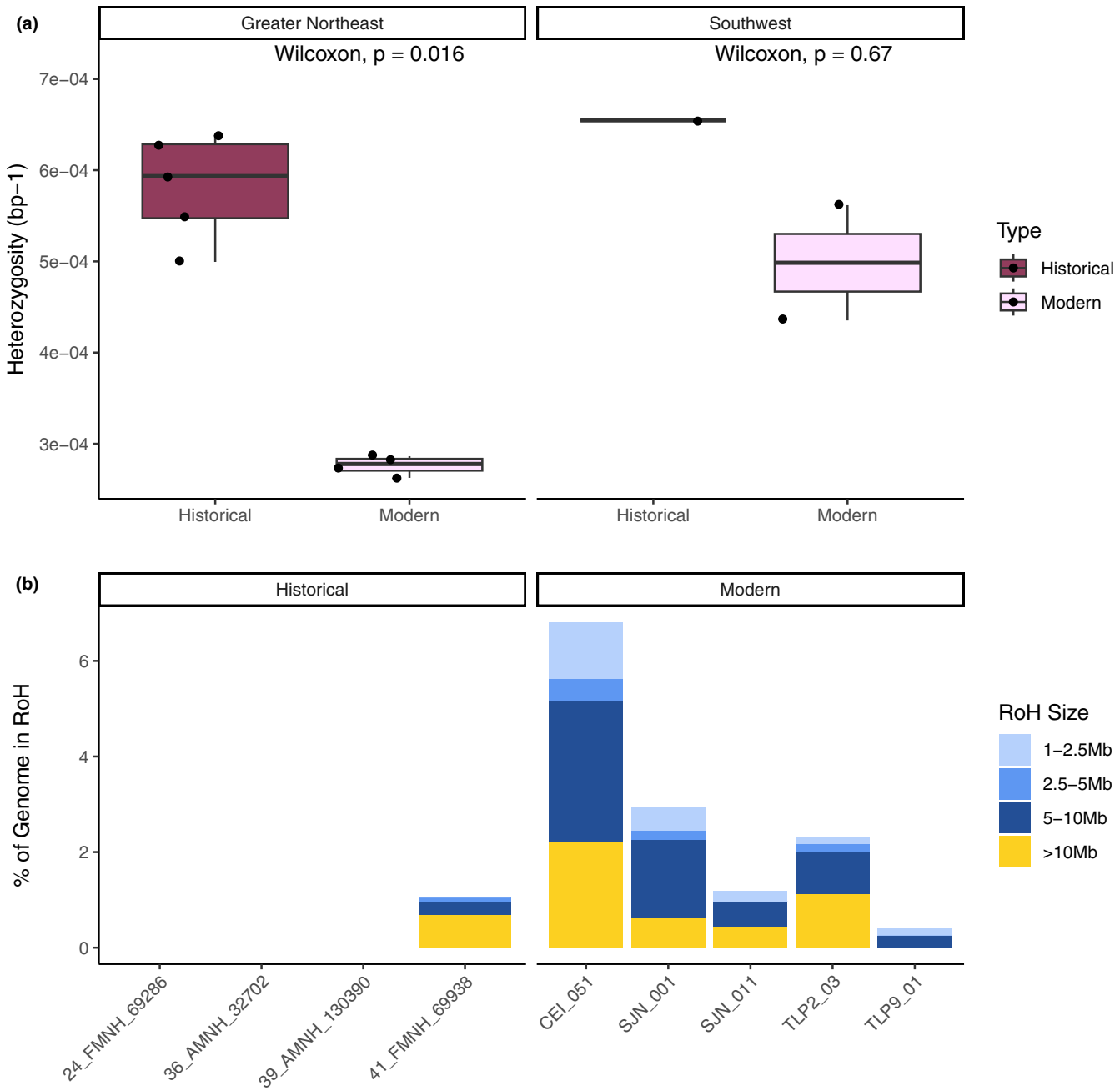

FIGURE 3 Genetic diversity and inbreeding in cotton-top tamarin individuals. (a) Genome-wide heterozygosity calculated with ANGSD for samples at 5× coverage. Statistical significance was calculated with Wilcoxon rank-sum test. Regarding the box plots: the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends no further than 1.5 × IQR (interquartile range), and the lower whisker extends to the smallest value at most 1.5 × IQR of the hinge. The middle line represents the median. (b) The fraction of the genome in RoH for samples above 7× coverage estimated with ROHan. Colour is representative of the RoH size in megabases (Mb).

historical samples, (Figure S10) they do not suffer from extreme levels of inbreeding. Similarly, modern samples have accumulated RoH of >1 Mb (Figures 3b and S11), something that is almost absent from the genomes of the historical samples, with only one sample from Mutatá exhibiting some degree of inbreeding with both F_i and F_{ROH} (41_FMNH_69938). In particular, for modern samples, on average only 2.73% of the genome is in RoH (ranging from 1.19% to 6.80%), with an average RoH length of 5.97 Mb. This would be compatible with a scenario of inbreeding happening approximately ~12 generations ago or ~73 years ago, based on average RoH length (Thompson, 2013). However, this inbreeding continued to much more recent times, up to ~2–3 generations ago in some individuals (~15 years ago) based on the coalescence time of longest fragments (Table S6).

Finally, we estimated the genetic load of samples with at least 5× coverage ($N = 16$, Table S7) and compared it between historical and modern samples as well as between geographical origin (greater northeast and southwest) for both high-impact and moderate-impact variants. We also disaggregated the genetic load into masked (in heterozygous state) and realized load (in homozygous state). Derived allele counts were normalized by the derived allele synonymous counts (Figure 4a). When restricting the analysis on a subset of samples with at least 7× coverage ($N = 9$), we found similar results that varied only slightly for the greater northeast historical population (Figure S12). Although the historical data show that both the greater northeast and the southwest population had accumulated similar levels of genetic load, the amount was slightly lower in the greater northeast population (Figure 4b), a relation that did not persist to the present day. In modern samples, we detected a relative reduction of deleterious variants (especially high-impact variants) compared with the historical population, possibly due to selection being more efficient in purging highly deleterious variants in homozygosis (Figure 4b). Although the total load observed in modern samples from the greater northeast populations is not lower than in their historical counterpart, we find that they have transformed the masked load into realized load, as expected due to their higher inbreeding (Figure 4a). When comparing modern greater northeast with southwest populations, the former has more total genetic load than the southwest population (Figure 4b), pointing to the modern southwest population being more efficient at purging deleterious variants. However, given the limited and nonsignificant diversity loss in the southwest population, this result should be taken with caution. We cannot rule out potential biases on calling high-impact variants in historical southwest population samples due to their lower genotype quality (Figure S13; Pinsky et al., 2021). Also, the limited sample size and the possibility we are not sampling continuous genetic diversity limit our interpretation of these results.

4 | DISCUSSION

Historical samples of cotton-top tamarins were collected across their natural distribution and sequenced to determine the historical population structure and demography. This knowledge can form a baseline to understand the consequences of anthropogenic effects, such as exportations and habitat fragmentation on the species genetic diversity. For that, we also collected and sequenced a small number of samples representing the current populations at three different locations. Although the historical dataset contains a relatively even distribution across their natural range, there are still areas that are lacking representation, such as in the southern portion of the Department of Sucre. Furthermore, due to the low number of modern samples, assumptions made during the comparison of the historical and modern data should be seen as initial insights, rather than firm knowledge, as they do not necessarily belong to the same continuous population (as fine-scale structure appears with higher Ks, see Figure S3). Thus, to confirm these results, more modern samples from the entire distribution should be collected.

4.1 | Structure and demography of historical cotton-top tamarin populations

The analyses performed in this study suggest that the historical cotton-top tamarins can be grouped into two populations: one in the greater northeast and one in the southwest. This pattern seems to be present in the modern population as well, although more samples from the central region are needed to confirm this result. This grouping of two populations is supported not only by the PCA and admixture analyses (Figure 1) but also with the F_{ST} and f3-outgroup analyses (Table S5 and Figure S6). In addition, the results of the EEMS analysis clearly show that there has been a geographical barrier separating the southwest region from the greater northeast (Figure 2). This barrier area suggested by EEMS overlaps with a rather mountainous area (Paramillo Massif mountain range) that could be acting as a geographical barrier to gene flow.

Moreover, EEMS suggests that the northeastern population has a higher-than-expected effective migration rate throughout the region, even suggesting a corridor of gene flow between Caracas, Colosó, Planeta Rica and Tierralta. Although the PCA and ADMIXTURE analyses support a structure of two populations, the PCA shows a north-to-south gradient in the greater northeast and $K = 3$ in admixture shows admixture in Caracas and Colosó, suggesting the possibility of two subpopulations (northern and central population) in the greater northeast. Furthermore, the structure from the PCA and the admixture results provides us with insight useful for determining the likely origins of samples that were provided with uncertain localities. For example, the 'Arjona_uncertain' specimen was obtained in the far north; however, its genetics reveals that it probably originated from somewhere near Turbo (Antioquia) in the southwest region (Figure S3), thus was likely trafficked and on its way to a Colombian port such as Cartagena (Connelly & Peyronnin, 2021), or airport such as Barranquilla (National Research Council, 1976). According to the results, the other two samples with uncertain localities (Cauca_uncertain and Unknown) likely originated in the greater northeast: Cauca_uncertain possibly between San Juan Nepomuceno (Bolívar)

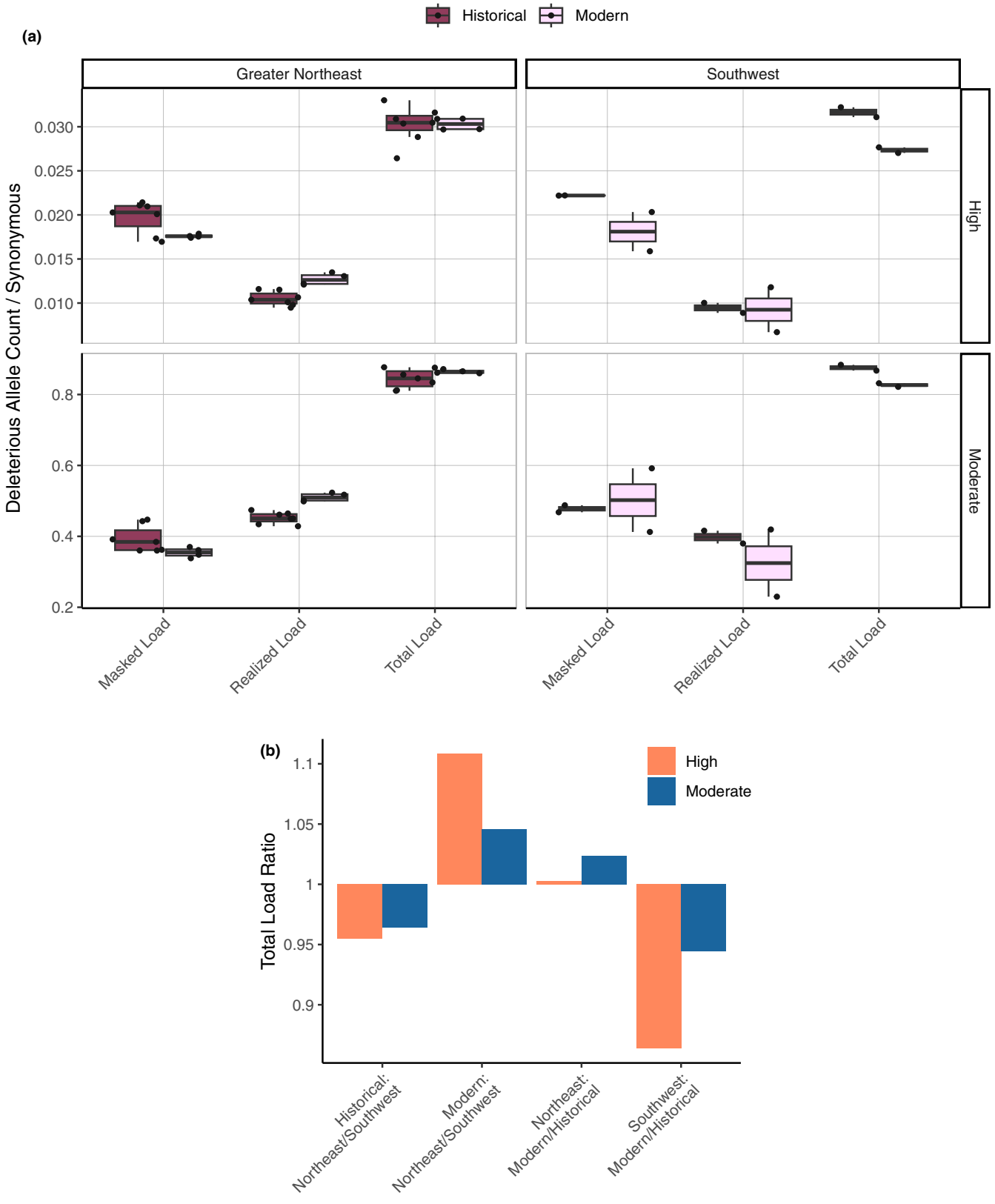

FIGURE 4 Genetic load for samples over 5× coverage for the greater northeast and southwest regions in modern and historical samples. (a) The relative count of genetic load is calculated as the ratio between derived allele counts in high- or moderate-impact mutations by the derived allele count in synonymous mutations. Masked load represents the count in heterozygous state, realized load represents the count in homozygous state, and total load represents the global count of derived alleles in deleterious positions. Regarding the box plots: the lower and upper hinges correspond to the first and third quartiles, the upper whisker extends no further than 1.5×IQR (interquartile range), and the lower whisker extends to the smallest value at most 1.5×IQR of the hinge. The middle line represents the median. (b) The total load ratio is estimated by comparing the averages of relative count of genetic load in different populations (by sample type and by geographical origin).

and Colosó (Sucre), and Unknown around Caracas sampling site (Cereté, Córdoba) (Figure S3).

4.2 | Loss of genetic diversity and inbreeding in present-day populations

Valuable insights can be gained by comparing historical and modern samples. Although we have both historical and modern samples, only one sampling site (San Juan) has samples from both periods. Thus, fine-scale structure (only present at higher Ks, Figure S3) might impact the temporal comparison. In this study, we found a significant decrease in heterozygosity for the greater northeast population (Figure 3a) when comparing the historical with the modern samples. This reduction in heterozygosity is similar to that seen in Grauer's gorillas (van der Valk et al., 2019) and in the white rhinoceros pre- versus post-bottleneck (Sánchez-Barreiro et al., 2021). It was notable that reduced and non-significant decrease in heterozygosity was observed when analysing the southwest samples on their own. Although on the one hand, we hypothesize that this might simply derive from the need for more samples from this region. On the other hand, we cannot rule out that the mountainous topography of the southwestern region may simply make it more challenging for traffickers and loggers to reach some of the areas that they inhabit, thus conferring a degree of protection on the cotton-top tamarins that live there.

We also observed that the modern samples showed a slight increase in inbreeding (Figure S10), which is also supported by the increase in RoH in the modern relative to historic samples (Figure 3b). The estimated RoH length (with on average >5 Mbp) reflects historic inbreeding, occurring around 70 years ago (~1950) and at least as recently as ~15 years ago based on our timing inferences. This indicates that inbreeding emerged as a consequence of the massive population decline due to exportations of cotton-top tamarins and continued after its banning, possibly due to continuous habitat degradation and fragmentation. Interestingly, and despite this, when comparing the genetic load between modern and historical populations, our data preliminarily suggest that the species has managed to avoid overall increases of harmful mutations in their genomes, as seen in other species facing similar circumstances (Grossen et al., 2020; Robinson et al., 2018; van der Valk et al., 2021). Although both the sequencing of more samples to a higher depth of coverage and sampling of more locations representing the southwest population would be necessary to reach a more conclusive result, we speculate that if accurate, this finding might derive from their complex social structure and high rates of emigration/immigration among social groups (Savage et al., 1996). The greater northeast population seems to have undergone a decrease in the accumulation of high-impact variants through time. This suggests that while strongly potentially deleterious mutations were purged, this has not been the case for less deleterious mutations (moderate-effect variants), and that there has been conversion from masked load into realized load. The southwest population also experienced a decrease in genetic load, even higher than the greater northeast, without conversion of masked load into realized load. This may indicate that the southwest populations are able to overall maintain higher genetic diversity as a result of experiencing less pressure than that faced by the greater northeast populations.

4.3 | Conservation insights

The knowledge provided by the historical populations can be leveraged to assist conservation of the species. After exploring the historical population structure of the cotton-top tamarin, we have identified two largely isolated populations that should be considered in future conservation actions for the species. Currently, the majority of conservation efforts are focussed on the northern region of the distribution. Our results point to the importance of also establishing conservation efforts in the southwestern region. Furthermore, the information gained from the genomic insights could be used for planning the establishment of forest corridors to connect fragmented habitats. For example, this study suggests that corridors should be planted to reconnect the northern and central parts of the greater northeast, as well as connecting forest in highly fragmented habitats.

As a consequence of the illegal pet trade, many cotton-top tamarins are confiscated without knowing where they were taken from in the wild. We found a correlation between geography and genetics and assigned origins to the individuals in our dataset with unknown or uncertain origin. Our dataset can therefore assist in the development of a reference panel to provide information about the geographical origin of illegally trafficked individuals. Due to the strong geographical signal in the analysis, the historical samples could be used in addition to modern samples to supply locality information. Future research could develop this dataset further by designing arrays to identify specific geolocations of confiscated individuals, similar to that done in the chimpanzee (Fontsere et al., 2022; Frandsen et al., 2020).

In conclusion, this study allowed for the development of a genetic baseline for cotton-top tamarins through the exploration of the historical museum samples. We have provided insight into the historical structure and demography before detrimental anthropogenic activity. This knowledge can be used to understand the level of genetic diversity that is in the modern populations and help guide conservation efforts.

AUTHOR CONTRIBUTIONS

All authors contributed to the design of the study. The modern sample collection for the northern sites was performed under the management and guidance of Anne Savage (AS) and Rosamira Guillen (RG), whereas the southern site was managed by Iván D. Soto-Calderón (IDSC). Collection of museum samples was led by Linett Rasmussen (LR) with assistance from Christina Hvilsom (CH) and M. Thomas P. Gilbert (MTPG). All laboratory work was carried out by LR and IDSC. Claudia Fontsere (CF) and LR performed data analysis on the dataset. LR and CF drafted the manuscript,

while all other authors participated in maturing and editing the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank the Field Museum (USA), the Swedish Royal Museum of Natural History Museum, the Natural History Museum of Oslo, the National Museum of Scotland and the American Museum of Natural History for providing samples for this study. Also, a special thank you to Ania Brown for sampling at the American Museum of Natural History. Furthermore, we would like to thank the many people from Proyecto Titi and Universidad de Antioquia that participated in the collection of modern samples. We thank CARDIQUE for providing permission to collect samples at the Proyecto Titi field sites. The authors would like to thank Long Zhou and Yang Zhou for providing us with the annotated reference genome files needed for the genetic load analysis and Katerina Guschanski, Morten Limborg and Thomas Bataillon for critical feedback during the preparation of the final manuscript. We thank Bastiaan Star and an anonymous reviewer for their constructive feedback during revision. We thank the Alfred Benzon Foundation, Torben & Alice Frimodts Fund and DNRF Award 143 for financial support. We thank Sarah Mak for guidance and assistance in the laboratory and Emily Cavill and Beth Shapiro for technical guidance in implementing the SCR method.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All raw data have been uploaded to the European Nucleotide Archive (ENA; Accession no.: PRJEB59492), and VCF files have been uploaded to ERDA server from the University of Copenhagen and can be downloaded using the following link: https://sid.erda.dk/sharelink/G35uxa3W8N. The metadata for these samples can be found in the Data S1. Furthermore, code and input files for data analyses performed in this study are accessible through GitHub (https://github.com/claudefa/CottonTop_Tamarins).

BENEFIT-SHARING STATEMENT

For this study, an international collaboration was formed to ensure the participation of scientists from Colombia, where the cotton-top tamarins reside. Furthermore, we incorporated scientific institutions, such as Proyecto Titi and the Copenhagen Zoo, that could apply findings to the species conservation efforts. As a result, the authors of this manuscript built a group committed to the conservation of the cotton-top tamarins and finding ways to apply results and develop relevant studies for the future.

ORCID

Linett Rasmussen https://orcid.org/0000-0002-8050-5732
Claudia Fontsere https://orcid.org/0000-0003-2233-6026
Iván D. Soto-Calderón https://orcid.org/0000-0002-6311-8378
Rosamira Guillen https://orcid.org/0000-0001-9501-1834
Anne Savage https://orcid.org/0000-0002-4738-8490
Christina Hvilsom https://orcid.org/0000-0001-7870-6888
M. Thomas P. Gilbert https://orcid.org/0000-0002-5805-7195

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Bertone, E. R., Giovannucci, E. L., King, J., Petto, A. J., & Johnson, L. D. (1998). Family history as a risk factor for ulcerative colitis-associated colon cancer in cotton-top tamarin. *Gastroenterology*, *114*(4), 669–674. https://doi.org/10.1016/S0016-5085(98)70580-3

Brcko, I. C., Carneiro, J., Ruiz-García, M., Boubli, J. P., Silva-Júnior, J. S. E., Farias, I., Hrbek, T., Schneider, H., & Sampaio, I. (2022). Phylogenetics and an updated taxonomic status of the tamarins (Callitrichinae, Cebidae). *Molecular Phylogenetics and Evolution*, *173*, 107504. https://doi.org/10.1016/J.YMPEV.2022.107504

Buckner, J. C., Lynch Alfaro, J. W., Rylands, A. B., & Alfaro, M. E. (2015). Biogeography of the marmosets and tamarins (Callitrichidae). *Molecular Phylogenetics and Evolution*, *82*(Pt B), 413–425. https://doi.org/10.1016/J.YMPEV.2014.04.031

Campos, P. F., & Gilbert, T. M. P. (2012). DNA extraction from keratin and chitin. *Methods in Molecular Biology*, *840*, 43–49. https://doi.org/10.1007/978-1-61779-516-9_6/COVER

Casas-Marce, M., Marmesat, E., Soriano, L., Martínez-Cruz, B., Lucena-Perez, M., Nocete, F., Rodríguez-Hidalgo, A., Canals, A., Nadal, J., Detry, C., Bernáldez-Sánchez, E., Fernández-Rodríguez, C., Pérez-Ripoll, M., Stiller, M., Hofreiter, M., Rodríguez, A., Revilla, E., Delibes, M., & Godoy, J. A. (2017). Spatiotemporal dynamics of genetic variation in the *Iberian lynx* along its path to extinction reconstructed with ancient DNA. *Molecular Biology and Evolution*, *34*(11), 2893–2907. https://doi.org/10.1093/MOLBEV/MSX222

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, *6*(2), 80–92. https://doi.org/10.4161/FLY.19695

Connelly, B., & Peyronnin, H. (2021). Taking off. Wildlife trafficking in the Latin America and Caribbean Region. https://www.traffic.org/site/assets/files/14510/routes_taking_off_report.pdf

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying temporal genomic erosion in endangered species. *Trends in Ecology & Evolution*, *33*(3), 176–185. https://doi.org/10.1016/J.TREE.2017.12.002

Ersmark, E., Klütsch, C. F. C., Chan, Y. L., Sinding, M. H. S., Fain, S. R., Illarionova, N. A., Oskarsson, M., Uhlén, M., Zhang, Y. P., Dalén, L., & Savolainen, P. (2016). From the past to the present: Wolf phylogeography and demographic history based on the mitochondrial control region. *Frontiers in Ecology and Evolution*, *4*, 134. https://doi.org/10.3389/FEVO.2016.00134/BIBTEX

Feng, S., Fang, Q., Barnett, R., Li, C., Han, S., Kuhlwilm, M., Zhou, L., Pan, H., Deng, Y., Chen, G., Gamauf, A., Woog, F., Prys-Jones, R., Marques-Bonet, T., Gilbert, M. T. P., & Zhang, G. (2019). The genomic footprints of the fall and recovery of the crested ibis. *Current Biology*, *29*(2), 340–349.e7. https://doi.org/10.1016/J.CUB.2018.12.008

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., & Roos, C. (2013). A mitogenomic phylogeny of living

primates. *PLoS One*, *8*(7), e69504. https://doi.org/10.1371/JOURNAL.PONE.0069504

Fontsere, C., Kuhlwilm, M., Morcillo-Suarez, C., Alvarez-Estape, M., Lester, J. D., Gratton, P., Schmidt, J. M., Dieguez, P., Aebischer, T., Álvarez-Varona, P., Agbor, A., Angedakin, S., Assumang, A. K., Ayimisin, E. A., Bailey, E., Barubiyo, D., Bessone, M., Carretero-Alonso, A., Chancellor, R., … Marques-Bonet, T. (2022). Population dynamics and genetic connectivity in recent chimpanzee history. *Cell Genomics*, *2*(6), 100133. https://doi.org/10.1016/j.xgen.2022.100133

Frandsen, P., Fontsere, C., Nielsen, S. V., Hanghøj, K., Castejon-Fernandez, N., Lizano, E., Hughes, D., Hernandez-Rodriguez, J., Korneliussen, T. S., Carlsen, F., Siegismund, H. R., Mailund, T., Marques-Bonet, T., & Hvilsom, C. (2020). Targeted conservation genetics of the endangered chimpanzee. *Heredity*, *125*(1–2), 15–27. https://doi.org/10.1038/s41437-020-0313-0

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in *Alpine ibex*. *Nature Communications*, *11*(1), 1001. https://doi.org/10.1038/s41467-020-14803-1

Gyllensten, U., Bergström, T., Josefsson, A., Sundvall, M., Savage, A., Blumer, E. S., Humberto Giraldo, L., Soto, L. H., & Watkins, D. I. (1994). The cotton-top tamarin revisited: MHC class I polymorphism of wild tamarins, and polymorphism and allelic diversity of the class II DQA1, DQB1, and DRB loci. *Immunogenetics*, *40*(3), 167–176. https://doi.org/10.1007/BF00167076

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience*, *8*(5), giz034. https://doi.org/10.1093/gigascience/giz034

Jensen, E. L., Díez-del-Molino, D., Gilbert, M. T. P., Bertola, L. D., Borges, F., Cubric-Curik, V., de Navascués, M., Frandsen, P., Heuertz, M., Hvilsom, C., Jiménez-Mena, B., Miettinen, A., Moest, M., Pečnerová, P., Barnes, I., & Vernesi, C. (2022). Ancient and historical DNA in conservation policy. *Trends in Ecology & Evolution*, *37*(5), 420–429. https://doi.org/10.1016/j.tree.2021.12.010

Jónsson, H., Ginolhac, A. A., Schubert, M., Johnson, P. L. F. F., Orlando, L., Jonsson, H., Ginolhac, A. A., Schubert, M., Johnson, P. L. F. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *The Journal of Heredity*, *112*(3), 241–249. https://doi.org/10.1093/JHERED/ESAB012

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Leigh, J. W., & Bryant, D. (2015). Popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Mast, R. B., Rodriguez, J. V., & Mittermeier, R. A. (1993). The Colombian cotton-top tamarin in the wild. In N. K. Clapp (Ed.), *A primate model for the study of colitis and colonic carcinoma the cotton-top tamarin* (Saguinus oedipus) (pp. 3–43). CRC Press. https://doi.org/10.1201/9781351069410-1

Miller, L., Savage, A., & Giraldo, H. (2004). Quantifying remaining forested habitat within the historic distribution of the cotton-top tamarin (*Saguinus oedipus*) in Colombia: Implications for long-term conservation. *American Journal of Primatology*, *64*(4), 451–457. https://doi.org/10.1002/AJP.20091

National Research Council. (1976). Neotropical primates: Field studies and conservation: Proceedings of a symposium on the distribution and abundance of neotropical primates. In *Neotropical Primates*. National Academies Press. https://doi.org/10.17226/18666

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Perez Barrios, L. E. (2013). Por medio de la cual se declara El Parque Natural Regional Bosque Seco El Mono Titi, ubicado en el municipio de Santa Catalina, Departamento de Bolivar y se adoptan otras determinaciones (Report No. 0004). Corporacion Autonoma Regional Del Canal Del Dique, Cardique. https://vlex.com.co/vid/seco-ceiba-l-mono-catalina-determinaciones-489894990

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., Gondek-Wyrozemska, A. T., Baalsrud, H. T., Brieuc, M. S. O., Kjesbu, O. S., Godiksen, J. A., Barth, J. M. I., Matschiner, M., Stenseth, N. C., Jakobsen, K. S., Jentoft, S., & Star, B. (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(15), e2025453118. https://doi.org/10.1073/PNAS.2025453118/-/DCSUPPLEMENTAL

Prüfer, K. (2018). SNPAD: An ancient DNA genotype caller. *Bioinformatics*, *34*(24), 4165–4171. https://doi.org/10.1093/bioinformatics/bty507

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 tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Rasmussen, M. S., Garcia-Erill, G., Korneliussen, T. S., Wiuf, C., & Albrechtsen, A. (2022). Estimation of site frequency spectra from low-coverage sequencing data using stochastic EM reduces overfitting, runtime, and memory usage. *Genetics*, *222*(4), iyac148. https://doi.org/10.1093/genetics/iyac148

Renaud, G., Hanghøj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics*, *212*(3), 587–614. https://doi.org/10.1534/GENETICS.119.302057

Robin, M., Ferrari, G., Akgül, G., Münger, X., von Seth, J., Schuenemann, V. J., Dalén, L., & Grossen, C. (2022). Ancient mitochondrial and modern whole genomes unravel massive genetic diversity loss during near extinction of *Alpine ibex*. *Molecular Ecology*, *31*(13), 3548–3565. https://doi.org/10.1111/MEC.16503

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in Island foxes. *Current Biology*, *28*(21), 3487–3494.e4. https://doi.org/10.1016/J.CUB.2018.08.066/ATTACHMENT/8480CF63-3010-4FFA-94DB-E732DEDEDD08/MMC1.PDF

Rodrigues, A. T., McKechnie, I., & Yang, D. Y. (2018). Ancient DNA analysis of indigenous rockfish use on the Pacific coast: Implications for marine conservation areas and fisheries management. *PLoS One*, *13*(2), e0192716. https://doi.org/10.1371/JOURNAL.PONE.0192716

Rodríguez, V., Link, A., Guzman-Caro, D., Defler, T. R., Palacios, E., Stevenson, P. R., & Mittermeier, R. A. (2021). *Saguinus oedipus* (amended version of 2020 assessment). The IUCN Red List of Threatened Species 2021. https://doi.org/10.2305/IUCN.UK.2021-1.RLTS.T19823A192551067.en

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. A. T., Weinert, L. A., Belkhir, K., Bierne, N., … Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, *515*(7526), 261–263. https://doi.org/10.1038/nature13685

Sánchez-Barreiro, F., Gopalakrishnan, S., Ramos-Madrigal, J., Westbury, M. V., De Manuel, M., Margaryan, A., Ciucani, M. M., Vieira, F. G., Patramanis, Y., Kalthoff, D. C., Timmons, Z., Sicheritz-Pontén, T., Dalén, L., Ryder, O. A., Zhang, G., Marquès-Bonet, T., Moodley, Y., & Gilbert, M. T. P. (2021). Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (*Ceratotherium simum*). *Molecular Ecology*, *30*(23), 6355–6369. https://doi.org/10.1111/MEC.16043

Savage, A., Giraldo, L. H., Sblumer, E., Soto, L. H., Burger, W., & Snowdon, C. T. (1993). Field techniques for monitoring cotton-top tamarins (*Saguinus oedipus oedipus*) in Colombia. *American Journal of Primatology*, *31*(3), 189–196. https://doi.org/10.1002/AJP.1350310304

Savage, A., Giraldo, L. H., Soto, L. H., & Snowdon, C. T. (1996). Demography, group composition, and dispersal in wild cotton-top tamarin (*Saguinus oedipus*) groups. *American Journal of Primatology*, *38*, 8–100. https://doi.org/10.1002/(SICI)1098-2345(1996)38:1

Savage, A., Shideler, S. E., Soto, L. H., Causado, J., Giraldo, L. H., Lasley, B. L., & Snowdon, C. T. (1997). Reproductive events of wild cotton-top tamarins (*Saguinus oedipus*) in Colombia. *American Journal of Primatology*, *43*, 329–337. https://doi.org/10.1002/(SICI)1098-2345(1997)43:4

Savage, A., Snowdon, C. T., Soto, L., Castro, J., Medina, F., Emeris, G., Garcia, L. C., & Guillen, R. (2022). The impacts of seasonal variation and climate on food utilization in a population of critically endangered cotton-top tamarins (*Saguinus oedipus*) in Colombia: A 22-year longitudinal study. *American Journal of Primatology*, *84*(8), e23415. https://doi.org/10.1002/AJP.23415

Savage, A., Snowdon, C. T., Soto, L., Medina, F., Emeris, G., & Guillen, R. (2021). Factors influencing the survival of wild cotton-top tamarin (*Saguinus oedipus*) infants. *American Journal of Primatology*, *83*(7), e23262. https://doi.org/10.1002/AJP.23262

Savage, A., Soto, L., Medina, F., Emeris, G., & Soltis, J. (2009). Litter size and infant survivorship in wild groups of cotton-top tamarins (*Saguinus oedipus*) in Colombia. *American Journal of Primatology*, *71*(8), 707–711. https://doi.org/10.1002/AJP.20696

Savage, A., Thomas, L., Feilen, K. L., Kidney, D., Soto, L. H., Pearson, M., Medina, F. S., Emeris, G., & Guillen, R. R. (2016). An assessment of the population of cotton-top tamarins (*Saguinus oedipus*) and their habitat in Colombia. *PLoS One*, *11*(12), e0168324. https://doi.org/10.1371/JOURNAL.PONE.0168324

Savage, A., Thomas, L., Leighty, K. A., Soto, L. H., & Medina, F. S. (2010). Novel survey method finds dramatic decline of wild cotton-top tamarin population. *Nature Communications*, *1*(3), 30. https://doi.org/10.1038/ncomms1030

Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, *9*(5), 1056–1082. https://doi.org/10.1038/nprot.2014.063

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 88. https://doi.org/10.1186/s13104-016-1900-2

Shao, Y., Zhou, L., Li, F., Zhao, L., Zhang, B., Shao, F., Chen, J., Chen, C., Bi, X., Zhuang, X., Zhu, H., Hu, J., Sun, Z., Li, X., Wang, D., Rivas-González, I., Wang, S., Wang, Y., Chen, W., … Wu, D. (2023). Phylogenomic analyses provide insights into primate evolution. *Science*, *380*(6648), 913–924. https://doi.org/10.1126/science.abn6919

Soto-Calderón, I. D., Acevedo-Garcés, Y. A., Álvarez-Cardona, J., Hernández-Castro, C., & García-Montoya, G. M. (2016). Physiological and parasitological implications of living in a city: The case of the white-footed tamarin (*Saguinus leucopus*). *American Journal of Primatology*, *78*(12), 1272–1281. https://doi.org/10.1002/AJP.22581

Star, B., Boessenkool, S., Gondek, A. T., Nikulina, E. A., Hufthammer, A. K., Pampoulie, C., Knutsen, H., André, C., Nistelberger, H. M., Dierking, J., Petereit, C., Heinrich, D., Jakobsen, K. S., Stenseth, N. C., Jentoft, S., & Barrett, J. H. (2017). Ancient DNA reveals the Arctic origin of Viking age cod from Haithabu, Germany. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(34), 9152–9157. https://doi.org/10.1073/PNAS.1710186114

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 Valk, T., de Manuel, M., Marques-Bonet, T., & Guschanski, K. (2021). Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. *BioRxiv*, *696831*. https://doi.org/10.1101/696831

van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology*, *29*(1), 165–170.e6. https://doi.org/10.1016/j.cub.2018.11.055

van Grouw, H., Hernández-Alonso, G., Cavill, E., & Gilbert, M. T. P. (2023). The founding feathers: The true ancestry of the domestic barbary dove. *Bulletin of the British Ornithologists Club*, *143*(2), 153–171.

Wheaton, C. J., Feilen, K. L., Soto, L. H., Medina, F., Emeris, G., Guillen, R., & Savage, A. (2022). Seasonality of reproduction in wild cotton-top tamarins (*Saguinus oedipus*) in Colombia. *American Journal of Primatology*, *84*(2), e23359. https://doi.org/10.1002/AJP.23359

Wood, J. D., Peck, O. C., Tefend, K. S., Rodriguez-M, M. A., Rodriguez-M, J. V., Hernández-C, J. I., Stonerook, M. J., & Sharma, H. M. (1998). Colitis and colon cancer in cotton-top tamarins (*Saguinus oedipus oedipus*) living wild in their natural habitat. *Digestive Diseases and Sciences*, *43*(7), 1443–1453. https://doi.org/10.1023/A:1018842210330

SUPPORTING INFORMATION

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

How to cite this article: Rasmussen, L., Fontsere, C., Soto-Calderón, I. D., Guillen, R., Savage, A., Hansen, A. J., Hvilsom, C., & Gilbert, M. T. P. (2023). Assessing the genetic composition of cotton-top tamarins (*Saguinus oedipus*) before sweeping anthropogenic impact. *Molecular Ecology*, *32*, 5514–5527. https://doi.org/10.1111/mec.17130