

UC Merced

UC Merced Previously Published Works

Title

A new lineage of Galapagos giant tortoises identified from museum samples

Permalink

<https://escholarship.org/uc/item/8109t204>

Journal

Heredity, 128(4)

ISSN

0018-067X

Authors

Jensen, Evelyn L
Quinzin, Maud C
Miller, Joshua M
et al.

Publication Date

2022-04-01

DOI

10.1038/s41437-022-00510-8

Peer reviewed

ARTICLE OPEN



A new lineage of Galapagos giant tortoises identified from museum samples

Evelyn L. Jensen^{1,2}✉, Maud C. Quinzin³, Joshua M. Miller⁴, Michael A. Russello⁵, Ryan C. Garrick⁶, Danielle L. Edwards⁷, Scott Glaberman⁸, Ylenia Chiari⁹, Nikos Poulakakis^{10,11}, Washington Tapia^{12,13}, James P. Gibbs¹⁴ and Adalgisa Caccone¹

© The Author(s) 2022

The Galapagos Archipelago is recognized as a natural laboratory for studying evolutionary processes. San Cristóbal was one of the first islands colonized by tortoises, which radiated from there across the archipelago to inhabit 10 islands. Here, we sequenced the mitochondrial control region from six historical giant tortoises from San Cristóbal (five long deceased individuals found in a cave and one found alive during an expedition in 1906) and discovered that the five from the cave are from a clade that is distinct among known Galapagos giant tortoises but closely related to the species from Española and Pinta Islands. The haplotype of the individual collected alive in 1906 is in the same clade as the haplotype in the contemporary population. To search for traces of a second lineage in the contemporary population on San Cristóbal, we closely examined the population by sequencing the mitochondrial control region for 129 individuals and genotyping 70 of these for both 21 microsatellite loci and >12,000 genome-wide single nucleotide polymorphisms [SNPs]. Only a single mitochondrial haplotype was found, with no evidence to suggest substructure based on the nuclear markers. Given the geographic and temporal proximity of the two deeply divergent mitochondrial lineages in the historical samples, they were likely sympatric, raising the possibility that the lineages coexisted. Without the museum samples, this important discovery of an additional lineage of Galapagos giant tortoise would not have been possible, underscoring the value of such collections and providing insights into the early evolution of this iconic radiation.

Heredity; <https://doi.org/10.1038/s41437-022-00510-8>

INTRODUCTION

Remote oceanic islands and archipelagos have piqued the interest of evolutionary biologists for decades as these landscapes offer ideal settings to study colonization and subsequent establishment and diversification of species, largely owing to their isolation and discrete boundaries (Gillespie 2007). Moreover, island systems with well-known geological histories and ages of emergence (e.g., Hawaiian and Galapagos archipelagos) can also add a temporal axis to the study of the emergence of biological diversity (Cowie and Holland 2008; Parent et al. 2008; Shaw and Gillespie 2016).

The Galapagos Archipelago includes 13 major islands larger than 10 km², six smaller islands, and over 200 islets and rocks, located in the Pacific Ocean ~900 km west of the South American continent and straddling the equator (Dirección del Parque Nacional Galápagos 2014). It is volcanic in origin, with islands formed by orogenetic activity along the Nazca plate, such that islands to the west are younger than islands to the east (Geist 1996; Geist et al. 2014). The complex history of the archipelago includes the islands drifting, merging and splitting, with some subsiding into the ocean (Ali and Aitchison 2014). This geological

history impacted phylogeographic patterns of diverse native flora and fauna in some predictable ways (Parent et al. 2008; Grant and Grant 2014; MacLeod et al. 2015; Castañeda-Rico et al. 2019), including the Galapagos giant tortoises. Tortoises arrived on the oldest island and from there colonized the rest of the archipelago by a combination of active dispersal mediated by ocean currents, and vicariance closely matching the merging and splitting of islands (Poulakakis et al. 2012; Poulakakis et al. 2020).

Fourteen species of giant tortoise have been described from the Galapagos Archipelago, with one additional known but undescribed extinct species (Rhodin et al. 2017; Fig. 1). These species are genetically distinct with generally one species per island, except for the tortoises on Isabela and Santa Cruz Islands, where multiple species occur in allopatry (Caccone et al. 1999, 2002; Poulakakis et al. 2015). Three of the recognized species have gone extinct in the last century, and several others are in danger of extinction because of human impacts, including historical overharvesting and habitat degradation (IUCN 2020). These activities have reduced the number of tortoises on the archipelago to 10% of their historical population levels before humans arrived (Tapia et al. 2021).

¹Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. ²School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne, UK. ³MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Department of Biological Sciences, MacEwan University, Edmonton, AB, Canada. ⁵Department of Biology, University of British Columbia, Kelowna, BC, Canada. ⁶Department of Biology, University of Mississippi, Oxford, MS 38677, USA. ⁷Department of Life & Environmental Sciences, University of California, Merced, CA, USA. ⁸Department of Environmental Science and Policy, George Mason University, Fairfax, VA, USA. ⁹Department of Biology, George Mason University, Fairfax, VA, USA. ¹⁰Department of Biology, School of Sciences and Engineering, University of Crete, Irakleio, Greece. ¹¹The Natural History Museum of Crete, School of Sciences and Engineering, University of Crete, Irakleio, Greece. ¹²Galapagos Conservancy, 11150 Fairfax Boulevard #408, Fairfax, VA 22030, USA. ¹³University of Málaga, Campus Teatinos, Apdo. 59, 29080 Málaga, Spain. ¹⁴Department of Environmental and Forest Biology, College of Environmental Science and Forestry, State University of New York, Syracuse, NY, USA. Associate editor: Lounès Chikhi. ✉email: evelyn.jensen@newcastle.ac.uk

Received: 10 November 2020 Revised: 2 February 2022 Accepted: 3 February 2022

Published online: 25 February 2022

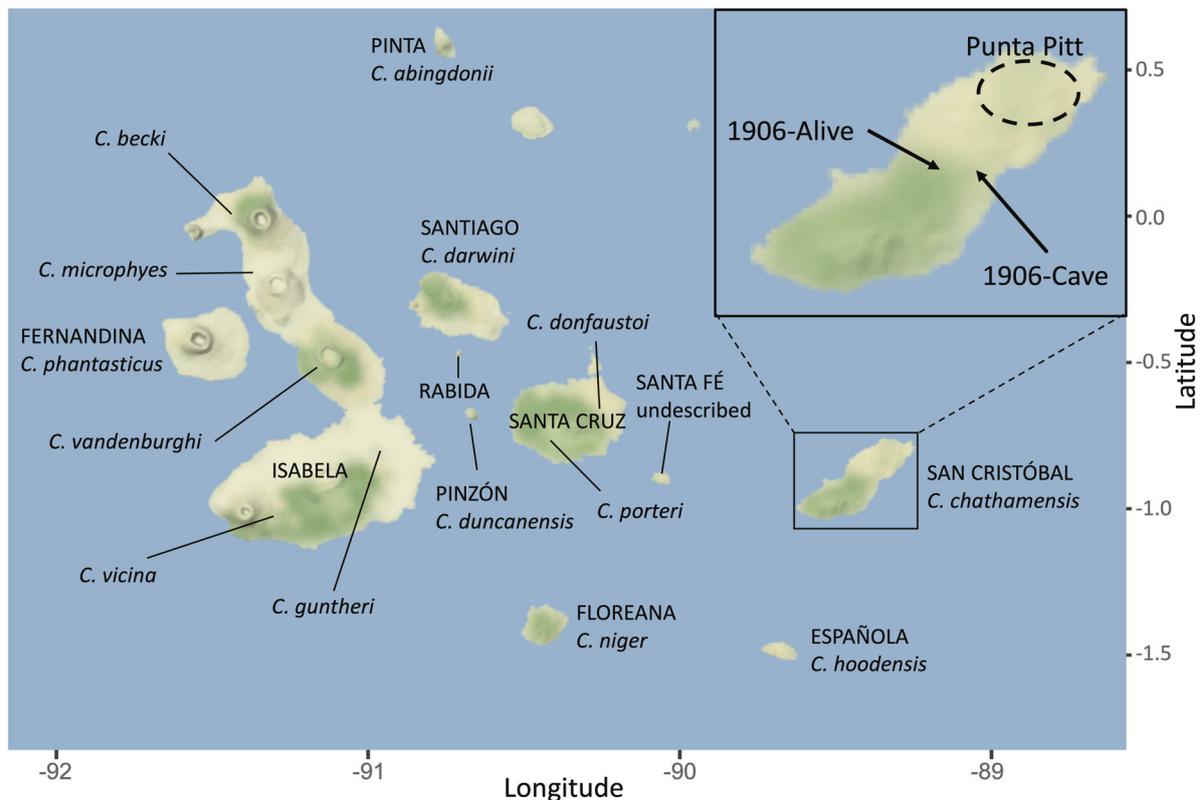


Fig. 1 Map of the Galapagos Archipelago, indicating the locations of each *Chelonoidis* species, with San Cristóbal Island enlarged in the inset map. Island names are in capital letters. The approximate location of the cave where the bones were found in 1906 is marked, as is the approximate location where CAS 8133 was collected alive and the region of Punta Pitt.

In this study, we reconstruct the colonization and evolutionary history of Galapagos giant tortoises (*Chelonoidis* spp.) from San Cristóbal Island (formerly Chatham Island, Fig. 1), recognized as belonging to a single endemic species, *C. chathamensis*. Previous phylogenetic work has shown that the radiation of giant tortoises likely started from an initial colonization of San Cristóbal, the oldest emerged island of the archipelago, with founders sourced from western central South America (Caccone et al. 1999, 2002; Poulakakis et al. 2012, 2020). Accordingly, understanding past and present genetic differentiation of tortoises on San Cristóbal is crucial to understanding the history and pattern of diversification of Galapagos giant tortoises as a whole. San Cristóbal may have been part of a larger landmass 2–3 million years ago (MYA), however during times with higher sea levels its two volcanoes may have been separate islands, that have later re-coalesced into one island (Geist et al. 2014; Karnauskas et al. 2017). This dynamic geological history likely influenced connectivity among tortoises living on San Cristóbal over time. By extension, this might have impacted the colonization of the other islands.

Morphologically, adult Galapagos giant tortoises are characterized as possessing either saddleback or domed carapace shapes. This morphological diversity was recognized by early collectors (Van Denburgh 1914; Fritts 1983; Fritts 1984; Pritchard 1996) and more recently analyzed using geometric morphometrics (Chiari et al. 2009; Chiari 2021). The contemporary population of San Cristóbal tortoises, representing the entire species, *C. chathamensis*, has an intermediate carapace morphology that is highly variable among individuals (Chiari 2021).

On the island today there are ~6700 giant tortoises, having recovered from a bottleneck that reduced the population down to 500–700 individuals in the 1970s. Population recovery was achieved following the elimination of harvesting, removal of invasive species, and a short-lived but successful captive rearing program (Cayot

2008; Tapia et al. 2021). The contemporary population is found predominantly on the arid, northeastern part of the island. However, tortoises formerly resided on the more humid, southwestern and central parts of the island, where historical records suggest they were heavily harvested in the mid 1800s, leading to their extirpation by 1930 (Banning 1933). It was speculated that the southern tortoises represented a distinct species from the northern tortoises on the island (Pritchard 1996). Over time, a small number of tortoises have recolonized the southwestern and central parts of the island. Previous studies on a few samples from the contemporary population that were primarily collected in the northeastern part of the island, found no mitochondrial haplotypic diversity based on assaying a section of the typically highly variable non-coding mitochondrial control region (Caccone et al. 2002; Poulakakis et al. 2012). Previous studies have also not found evidence suggesting more than one genetically distinct taxon based on microsatellite data (Ciofi et al. 2002; Garrick et al. 2015).

In 1906, an expedition on San Cristóbal led by the California Academy of Sciences (CAS) collected a single living tortoise and found the disjointed remains of approximately 17 individuals that had evidently perished after falling into a cave located near the middle of the island (Fig. 1). Currently, pieces of six of these cave specimens are accessioned at the CAS (Van Denburgh 1914). While the single individual found alive had a domed carapace, the two intact carapace remains found in the cave were considered saddleback, although their morphological characteristics fall within the variation observed for intermediate shell shapes (see Chiari 2021). One of these cave specimens (CAS 8127) is the holotype of *C. chathamensis*, the endemic species from San Cristóbal.

Given the dynamic history of tortoises on San Cristóbal Island, and their pivotal role in underpinning the entire tortoise radiation, we analyzed historical samples and the contemporary population in detail. The objectives of this study were to (1) test the

hypothesis that more than one lineage once lived on the island, and (2) describe patterns of genomic variation within the extant population. We collected mitochondrial control region sequence data from six historical samples and 129 samples from the contemporary population and compared these data to available sequences from all other extant and extinct Galapagos giant tortoise species to reconstruct their mitochondrial DNA evolutionary history. We further assessed the contemporary population for evidence of multiple lineages or admixture using two types of nuclear markers: a panel of 21 microsatellite loci and >12 000 single nucleotide polymorphisms (SNPs).

METHODS

Sampling and data collection

Historical samples. Bone fragments were obtained from the California Academy of Sciences collections for six of the seven specimens collected from San Cristóbal Island in 1906: specimen CAS 8133 was collected alive, while the other five (CAS 8127 [long bone], CAS 8128 [long bone], CAS 8129 [long bone], CAS 8130 [skull], CAS 8131 [skull]) were collected from the cave (Fig. 1). One additional specimen collected from the cave in 1906 was not available for this study.

We used a Nikon XT H 225 ST to produce micro-computed tomography (micro-CT) scans of each bone specimen to locate the region with the highest density, and thus the best conditions for DNA preservation. Not all samples had a high-density region (e.g., CAS 8130 and 8131, the skulls), but where possible, we targeted the densest region. In a dedicated ancient DNA facility at Yale University, a Dremel rotary tool with a cutting blade was used to scrape off the surface of the bone, and then cut out a small wedge (~300 mg) for DNA extraction. The pieces of bone were powdered while submerged in liquid nitrogen using a Spex 6770 freezer mill. The bone powder was demineralized in a solution of 0.5 M EDTA pH 8.0, 10% SDS, and Proteinase K, that was incubated at 56 °C overnight in a shaking incubator. The resulting lysate was then mixed with 5x volumes of buffer PB (Qiagen) and centrifuged through a MinElute spin column (Qiagen) to bind the DNA. The MinElute column was washed twice using buffer PE. DNA was eluted from the column using 50 µl of ultra-pure water, warmed to 56 °C. All standard precautions to prevent contamination with contemporary DNA and between historical samples were followed.

We sequenced ~700 bp of the mitochondrial control region in four overlapping fragments using nested PCR and negative controls, as described in Poulakakis et al. (2008) on an ABI 3730 automated sequencer (Applied Biosystems). Multiple amplifications of each fragment were conducted and sequenced in both directions, which along with the overlapping regions of the fragments, gave multiple observations of each position along the sequence. Each position was closely inspected by eye to determine the consensus sequence for each individual in GENEIOUS (version 11.05; Kearse et al. 2012).

We also attempted to genotype the historical individuals at 12 microsatellite loci that have been successfully employed in previous studies of Galapagos giant tortoise museum samples using published protocols (Russello et al. 2007; Poulakakis et al. 2008; Russello et al. 2010), but in this case, they did not yield reliable genotypes.

A fragment of CAS 8128 was sent to the Center for Applied Isotope Studies at the University of Georgia for radiocarbon dating. The sample was mechanically cleaned using a scalpel and wire bristle brush to remove surface contamination and gently reduced to smaller fragments of approximately 3–5 mm in size. The bone fragments were demineralized in cold (4 °C) 1 N HCl for 24 h and rinsed with ultrapure water to neutral. The demineralized bone fragments were treated with 0.1 M NaOH at room temperature and rinsed with ultrapure water to neutral. The samples were rinsed with 1 N HCl to eliminate atmospheric CO₂, rinsed in ultrapure water to pH 4 (slightly acidic), and heated at 80 °C for 8 h. The resulting solution was filtered through glass fiber filters to isolate the total acid insoluble fraction (“collagen”) and freeze-dried.

Approximately 1 mg of collagen was encapsulated in tin, and the elemental concentrations and stable isotope ratios ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) were measured using an elemental analyzer isotope ratio mass spectrometer (EA-IRMS). Values are expressed as $\delta^{13}\text{C}$ with respect to PDB and $\delta^{15}\text{N}$ with respect to AIR, with an error of less than 0.1%.

For accelerator mass spectrometry (AMS) analysis, a 5-mg subsample of collagen was combusted at 575 °C in evacuated and sealed Pyrex tubes in the presence of CuO to produce CO₂. The CO₂ samples were cryogenically purified from the other reaction products and catalytically converted to

graphite using the method of (Vogel et al. 1984). Graphite $^{14}\text{C}/^{13}\text{C}$ ratios were measured using the CAIS 0.5 MeV AMS. Sample ratios were compared to the ratio measured from the Oxalic Acid I standard (NBS SRM 4990).

Contemporary samples. Blood (0.1–1.9 mL) was collected from live individuals during expeditions to San Cristóbal Island in 1999, 2012 and 2016. In 1999, samples were primarily collected from Punta Pitt on the northeastern point of the island (Fig. 1). In 2012, samples were collected from adults at the breeding center on San Cristóbal, some of which were originally found in Punta Pitt, while most were of unknown origin. The sampling in 2016 was conducted as part of a comprehensive census for tortoises on the entire island. In all cases, blood was stored in tubes containing a lysis buffer (Longmire et al. 1997) at ambient temperature in the field and at 4 °C upon arrival in the lab. Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer’s protocol for blood.

The same mitochondrial control region fragment targeted in the historical samples was newly sequenced ($n=72$) or obtained from previous studies ($n=57$) for 129 contemporary samples from San Cristóbal using previously described methods (Caccone et al. 1999).

For a subset of 60 contemporary individuals from the 129 sequenced for the control region, double digest Restriction-site Associated DNA (ddRAD) libraries were created following Peterson et al. (2012), as described in Miller et al. (2018), and sequenced using two lanes of Illumina HiSeq 2000 at the Yale Center for Genome Analysis. Sequences from these new libraries were combined with previously collected data for an additional 10 contemporary San Cristóbal individuals from Miller et al. (2018). These 70 individuals (10 sampled in 1999, 14 in 2006, and 46 in 2016) represent the geographic breadth of sampling locations across San Cristóbal Island. The sequences were processed and aligned to the *C. abingdonii* reference genome (assembly ASM359739v1, Quesada et al. 2019) using the PALEOMIX *bam pipeline* (version 1.2.14, Schubert et al. 2014). Briefly, we used PALEOMIX as a wrapper to efficiently implement read trimming using AdapterRemoval (version 2.3.1, Lindgreen 2012), alignment using BWA *mem* (version 0.7.17, Li and Durbin 2009), and indel realignment using GATK *IndelRealigner* (McKenna et al. 2010). Reads were then filtered to remove any with more than 4 mismatches to the reference using BAMTOOLS (version 2.5.1, Barnett et al. 2011). Variant detection and genotype calling were performed using BCFtools (Li et al. 2009) *mpileup* and *call*, excluding reads with a mapping quality score of less than 30, ignoring indels and outputting only variants. The resulting VCF file was filtered using VCFtools (Danecek et al. 2011) to exclude repetitive regions of the genome, sites with a sequencing read depth greater than one standard deviation above the mean depth (mean = 21.5, SD = 19.0), and sites with >20% missing data. Only biallelic SNP loci with a minor allele count of three were retained. Following these steps, we assessed missingness per individual, and removed six individuals with >50% missing data. Next, we re-filtered the original VCF file with only retained individuals, following the steps above, followed by a filter for Hardy-Weinberg Equilibrium (HWE) using the correction for false discovery rate described by Benjamini and Yekutieli (2001) (adjusted p value = 0.004888) and thinned the loci to retain one SNP per 10,000 bp in order to reduce linkage among SNP loci.

We genotyped these same 70 contemporary individuals from the ddRAD at 21 microsatellite loci, following the procedures described in Quinzin et al. (2019).

To summarize, in the analyses outlined below, a total of 6 historical and 129 contemporary mitochondrial control region sequences were used, and a subset of 70 of the 129 contemporary individuals were also genotyped at the nuclear microsatellites and SNPs (Supplementary Table 1).

Genetic analyses

Phylogenetic analyses—mitochondrial DNA. To investigate phylogenetic relationships among the mitochondrial control region haplotypes found in the contemporary and historical individuals sampled from San Cristóbal, sequences were trimmed in GENEIOUS and aligned using Muscle (Edgar 2004) along with a representative subset of 28 unique haplotypes previously collected from all extinct and extant Galapagos giant tortoise species (Caccone et al. 2002; Beheregaray et al. 2003; Beheregaray et al. 2004; Caccone et al. 2004; Russello et al. 2005; Poulakakis et al. 2008; Chiari et al. 2009; Poulakakis et al. 2012; Edwards et al. 2013). The alignment was verified by eye, giving a final alignment length of 668 bp, and exported as a nexus file and opened in PopArt (<http://popart.otago.ac.nz>) where a statistical parsimony network (Templeton et al. 1992; Clement et al. 2000) was reconstructed.

To place the historical mitochondrial haplotypes within the Galapagos giant tortoise mtDNA-based phylogeny, we used a Bayesian approach implemented in BEAST2 (Bouckaert et al. 2019). The historical mitochondrial sequences from San Cristóbal tortoises were aligned with a more expansive dataset than for the network analysis, consisting of 93 haplotypes previously collected from all extinct and extant Galapagos giant tortoise species and the three outgroup species from continental South America (*Chelonoidis carbonarius*, *C. denticulatus* and *C. chilensis*) (Caccone et al. 2002, 2004), again using Muscle, this time with an alignment length of 718 bp due to gaps caused by alignment to the outgroup sequences. Genbank accession numbers are available in Supplementary Table 2. The BEAST input XML file was created using BEAUTi v. 2.6.2. We used the BModeltest module of BEAST2 to select the best nucleotide substitution model and tested the four combinations of Birth-Death or Yule model for speciation and a relaxed log-normal or strict clock. Four chains of each analysis were run for 100 million Markov chain Monte Carlo (MCMC) replicates, logging the result every 1000 iterations, with a burn-in of 10%. We viewed the log files in TRACER v1.6 (Rambaut et al. 2018) to verify that convergence had been achieved and effective sample size (ESS) values >200 had been obtained for each chain. The tree and log files were combined across the four chains using LogCombiner v. 2.6.2, discarding a burn-in of 10% and maximum clade credibility trees annotated with the mean node heights were generated using TreeAnnotator v. 2.6.2. The combined log files were again viewed in TRACER, and the likelihood of the four models were compared using AICM (Baele et al. 2012).

Population genetic analyses—microsatellites and SNPs. We calculated genetic diversity metrics for the contemporary population's nuclear marker datasets, including observed and expected heterozygosity and the inbreeding coefficient, G_{IS} , using GENODIVE (Meirns and Van Tienderen 2004). Effective population size (N_e) was estimated using the bias-corrected method of linkage disequilibrium (Hill 1981; Waples 2006; Waples and Do 2010), implemented in NeESTIMATOR (Do et al. 2014), excluding alleles with a frequency below 0.05, with confidence intervals determined using the jackknife approach.

We used genotypic clustering analyses, STRUCTURE (Pritchard et al. 2000) and discriminant analysis of principal components (DAPC, Jombart

et al. 2010), to evaluate evidence for substructure within the contemporary population. STRUCTURE was run using the admixture model with correlated allele frequencies, considering values of K (the number of clusters) ranging from one to 10, with 10 iterations per K . For the microsatellite dataset, each run included 1 million MCMC replicates following a burn-in of 100,000, whereas for the SNPs each run included 200 000 MCMC replicates after a burn-in of 100 000. To evaluate the support for each value of K , we plotted the log probability of the data ($\ln \Pr(X|K)$), and calculated the deltaK statistic (Evanno et al. 2005) using STRUCTURE HARVESTER (Earl and vonHoldt 2011). DAPC was implemented using adegenet (Jombart 2008) in R (R Development Core Team 2010). The best value of K was determined using the *find.clusters* function and based on the K with the lowest Bayesian information criterion (BIC) score. The optimal number of principal components to retain was found with the *optim.score* method, which was 1 for both datasets. We also used a principal components analysis to visualize clustering of genotypes, using the R package *Lea* (Frichot and François 2015).

RESULTS

Phylogenetic analyses—mitochondrial DNA

A single mitochondrial control region haplotype was shared by the 106 contemporary individuals. None of the six individuals in the historical sample had this haplotype, but the domed individual sampled alive in 1906 (CAS 8133) had a haplotype that differed by only three substitutions from the contemporary haplotype. The five specimens found in the cave, including the type specimen (CAS 8127), had four novel haplotypes. These haplotypes formed a distinct cluster on the statistical parsimony network that links the haplotypes from *C. hoodensis* and *C. abingdonii* to the main network. All of the cave haplotypes are distinct from the contemporary and CAS 8133 haplotypes, being separated by 15–20 mutational steps (Fig. 2).

The Bayesian Inference phylogenetic analyses for the four parameter combinations each converged across the four chains,

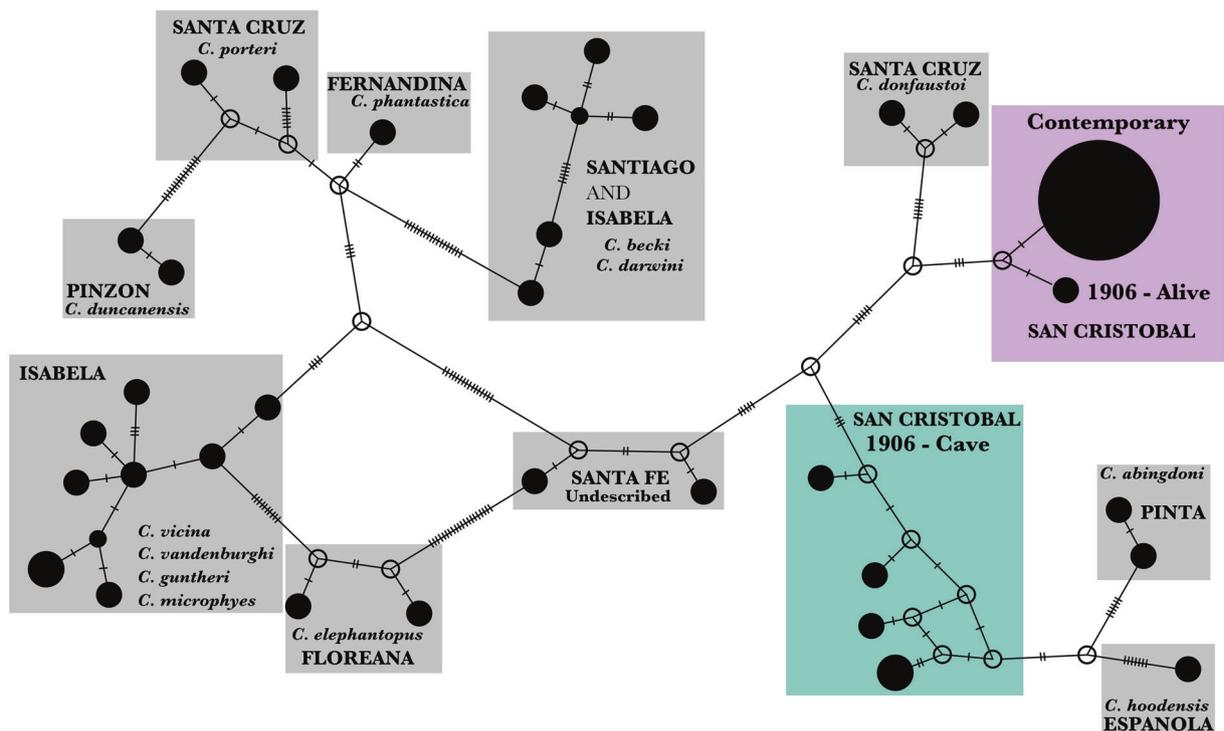


Fig. 2 Statistical parsimony haplotype network of the mitochondrial control region (668 bp) for 129 contemporary individuals from San Cristóbal and six historical specimens collected in 1906, as well as 28 representative haplotypes from the other species of Galapagos giant tortoise. The name of the island where each species occurs is labeled with capital letters, with current taxonomy in italics. Haplotypes are represented as black circles on the network, the size of the circle is proportional to the frequency of the haplotype in the analysis. Open circles represent unsampled, hypothesized haplotypes, and hash marks indicate a single mutational change. Reticulations reflect uncertainty in relationships, or homoplasy.

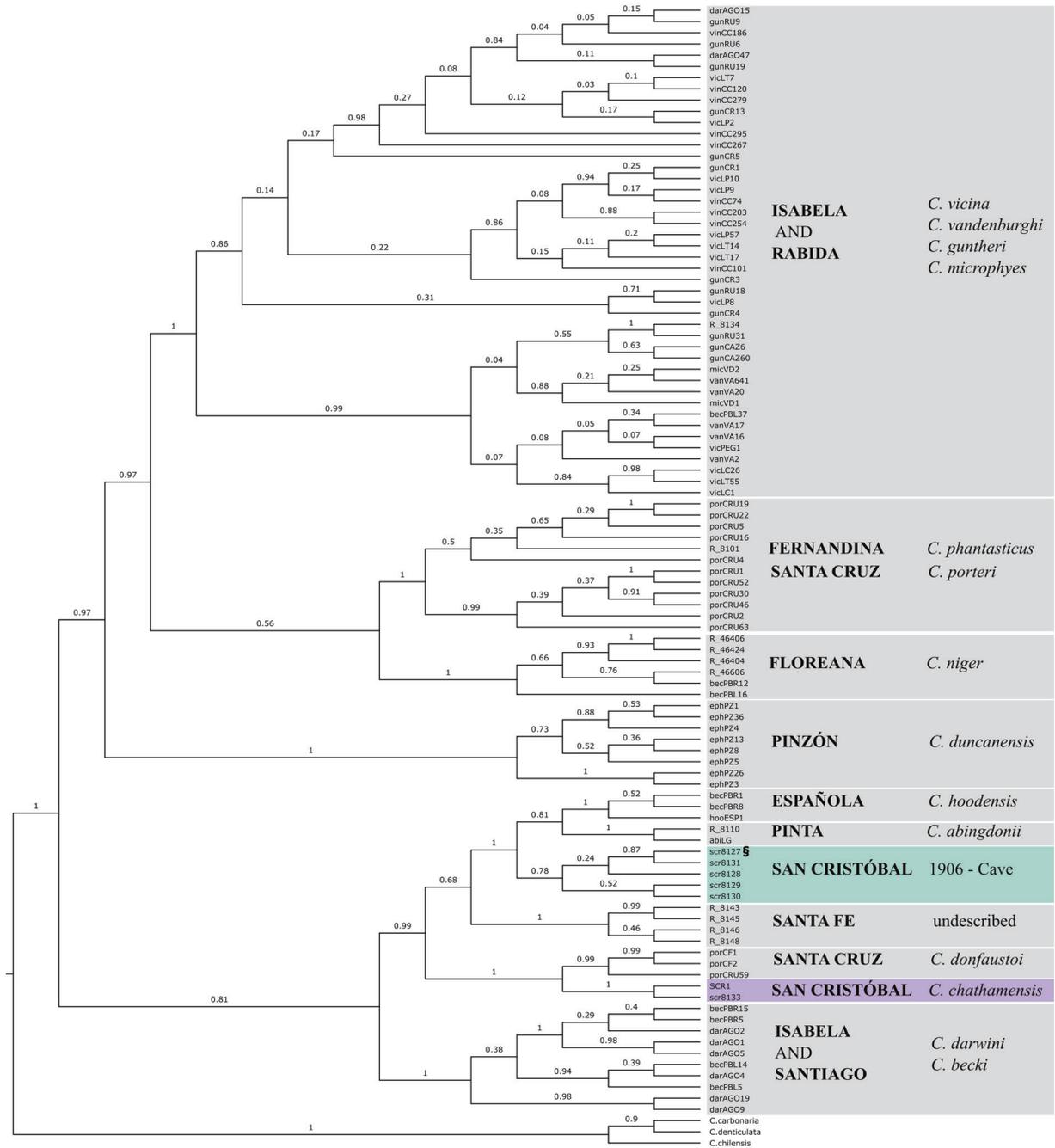


Fig. 3 Bayesian Inference maximum clade credibility cladogram showing relationships among the San Cristóbal historical samples from the cave and collected alive in 1906, and a reference dataset of 93 Galapagos giant tortoise haplotypes and three outgroups based on the mitochondrial control region (alignment length 718 bp), estimated using BEAST with a strict clock and Birth Death tree. The numbers on the branches are the posterior probability support values. § indicates the *C. chathamensis* type specimen. The name of the island where each clade is found is in capitals, with current taxonomy in italics.

with combined ESS values greater than 900. The model comparison indicated that the parameter combination of the strict clock with a Birth Death tree had the lowest AICM value (Supplementary Table 3), and thus the best model fit, so this is the version we present and discuss, although the trees produced by the other models can be found in as Supplementary Fig. 1 A, B and C. The phylogeny (Fig. 3) included 93 representatives from all other species of Galapagos giant tortoise, the contemporary San Cristóbal haplotype, the six historical San Cristóbal individuals, and the three outgroup taxa. The San Cristóbal haplotypes were placed in the same main clade, along with the species *C.*

hoodensis, *C. abingdonii*, *C. donfaustoi* and the undescrbed species from Santa Fe. Within this main clade, the five specimens collected from the cave are in a clade sister to the one containing *C. abingdonii* and *C. hoodensis* (Fig. 3, posterior probability =1.0). As seen in the statistical parsimony network, the individual collected alive in 1906 (CAS 8133) belonged to the same clade as the contemporary San Cristóbal haplotype.

Population genetic analyses—microsatellites and SNPs
Sequencing of the ddRAD libraries resulted in an average of 19.7 million reads per individual that aligned to the reference genome

Table 1. Diversity measures within the contemporary population on San Cristóbal Island based on 21 microsatellite loci and 12 192 SNPs used to genotype the same 64 individuals.

MARKER	N_A	EFFECTIVE N_A	H_O	H_E	G_{IS}	N_e
SNPS	2	1.41	0.280	0.266	-0.053 ($p = 0.001$)	25.8 (18.2–38.0)
MICROSATELLITES	7.81	3.93	0.665	0.684	0.029 ($p = 0.042$)	34.2 (22.7–55.4)

G_{IS} p-values are based on 999 permutations. In parentheses for N_e are the 95% confidence intervals based on jackknifing.

N_A , mean number of alleles per locus; Effective N_A , the number of alleles that would be expected at a locus based on heterozygosity; H_O , observed heterozygosity; H_E , expected heterozygosity; G_{IS} , inbreeding coefficient; N_e , effective population size.

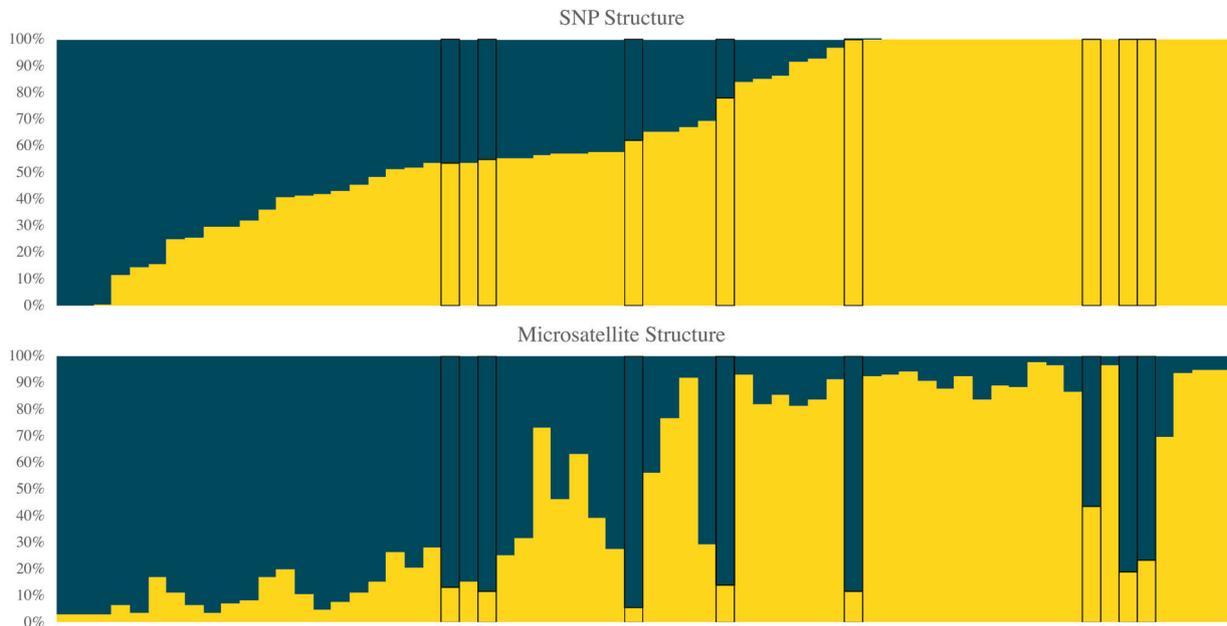


Fig. 4 Barplots depicting $K = 2$ for the STRUCTURE analysis using the SNP (12 192 loci) and microsatellite (21 loci) genotypes for the contemporary San Cristóbal population ($n = 64$). Each bar represents an individual and the proportion of the bar that is each color represents the membership of that individual to the two clusters. The order of individuals is the same in both plots, black boxes around bars highlight individuals with a greater than 0.4 discrepancy in assignment proportions between the analyses.

(range 5.6–87.9 million reads), which were combined with the 10 individuals sequenced previously (Miller et al. 2018). The final filtered ddRAD dataset had 64 individuals retained (six individuals were dropped due to high levels of missing data) that were genotyped at 12,192 loci with a mean depth of 16.5x, and 12% missing data. We omitted these six individuals from the microsatellite analyses in order to have fully overlapping datasets of the same 64 individuals. The microsatellite dataset of 21 loci had less than 1% missing data.

There were conflicting signals of heterozygote excess and deficit based on the type of nuclear marker assayed, with microsatellites showing a heterozygote deficit and SNPs showing heterozygote excess (Table 1). For both datasets, G_{IS} values were significantly different from zero based on 999 permutations, but in opposite directions (Table 1). However, for the microsatellite dataset, when calculating G_{IS} separately for each locus, none were significantly different from zero (Supplementary Table 4). The N_e value estimated from the microsatellites was higher than for the SNPs—34.2 versus 25.8, respectively, but the 95% jackknife confidence intervals were broad and overlapping (Table 1).

The clustering analyses based on microsatellite and SNP genotypic data did not produce strong evidence for substructure within the contemporary San Cristóbal population. For STRUCTURE, $K = 2$ was suggested as most likely when examining the $\ln \Pr(X|K)$ values (i.e. a large increase in $\ln \Pr(X|K)$ between $K = 1$

and $K = 2$, but less substantial increases for larger values of K ; Supplementary Fig. 2A and B) for both the microsatellite and SNP datasets. Although $K = 2, 4$ or 8 were considered likely by the deltaK statistic for the SNP data (Supplementary Fig. 2C), we were unable to estimate deltaK from the microsatellite analysis due to a lack of variance in $\ln \Pr(X|K)$ values between iterations. In some cases, the microsatellite and SNP analyses strongly assigned the same individuals to alternate STRUCTURE clusters (Fig. 4), and in general there is a cline in membership between the two clusters. For DAPC, the BIC scores for $K = 2$ were just slightly lower than for $K = 1$, and density plots for the DAPC showed the two clusters as partially overlapping (Supplementary Fig. 3). The PCA did not reveal any distinct clustering of individuals (Supplementary Fig. 4). When the PCA is color coded according to STRUCTURE $K = 2$ cluster membership, there is a gradient along PC1, however this did not correspond to any identifiable geographic patterns when color coded by sampling region (Supplementary Fig. 4).

Radiocarbon dating

CAS 8128 yielded 28.7% collagen, with an atomic carbon to nitrogen ratio of 3.4. The carbon 14 age was determined to be 410 years old, which corresponds to a calibrated date of 1450–1630 AD (95% probability) using the SHCal13 calibration curve (Hogg et al. 2013), as calculated in OxCal (Bronk Ramsey 2009).

DISCUSSION

A new mitochondrial lineage of tortoises on San Cristóbal Island

Based on evidence from the mitochondrial control region, we have discovered a new lineage within the Galapagos giant tortoise radiation. All five of the skeletal remains collected in 1906 from the cave on San Cristóbal Island have haplotypes that are part of a clade that is distinct from the haplotype found in the contemporary population on the same island (Fig. 3). We evaluated mitochondrial control region sequences from 129 contemporary individuals collected across the island (~1.5% of the current estimated population on the island, Tapia et al. 2021), but found the same single *C. chathamensis* haplotype previously identified, suggesting that the “cave” lineage has gone extinct. Herein, we refer to this as the “extinct” San Cristóbal lineage. The sister clade to the extinct lineage is the one including the Pinta and Española species (*C. abingdonii* and *C. hoodensis*). Using the clade ages from Poulakakis et al. (2020), this would place the divergence between the extant and extinct lineages from San Cristóbal at around 0.72 MYA. These data (Fig. 3) clearly show that the extinct and contemporary lineages on San Cristóbal are not each other’s closest relatives based on their mitochondrial DNA, leaving their origin and relationship to the other Galapagos tortoise species unclear (Poulakakis et al. 2012, Poulakakis et al. 2020). At this point we can only formulate alternative hypotheses, some more likely than others, based on the known geology of the archipelago, the history of the group, and the data available. One possibility we can clearly dismiss is that the two mitochondrial DNA lineages derive from tortoises that colonized Galapagos from the continent at different times, as all the species in the Galapagos tortoise radiation are included in a single clade, separated from the continental species of *Chelonoidis*. A more likely scenario, following from our reconstructed phylogeny (Fig. 3), is that when tortoises initially dispersed from San Cristóbal to colonize the islands of Santiago, Santa Cruz, and Pinzón, there was only a single lineage on the island. This lineage may have subsequently split within San Cristóbal (perhaps due to high sea levels causing the island to be divided into two), with the ancestors of the extinct lineage colonizing Española, while the ancestors of the contemporary lineage colonized Santa Cruz, giving rise to *C. donfaustoi*. Alternatively, it could be that one of the two mitochondrial DNA lineages on San Cristóbal is derived from a colonization event from nearby islands, thus representing double colonization from two different sources, at different times (i.e., one continental, and subsequently, one intra-archipelago). However, we think that this is relatively unlikely, as the re-invasion would have occurred from islands to the West, such as Santa Cruz, while all studies so far support an East to West colonization pattern across the archipelago rather than West to East (Beheregaray et al. 2004; Poulakakis et al. 2012, 2020). A more likely scenario is that there was potentially a proto-island that included the two modern islands of Española and San Cristóbal when the first tortoises arrived from the continent (Poulakakis et al. 2012, 2020). On this island, tortoises might have differentiated into two different groups. Once the islands split, there could have been a recolonization event from Española to San Cristóbal, leading to the coexistence of two different taxa with different mitochondrial DNA lineages, with the one derived from the Española recolonization eventually going extinct, leaving as its only trace the mitochondrial DNA sequences retrieved from the cave bone remains. To evaluate the likelihood of these two possibilities we would need additional sequence data to provide more resolution as well as historical specimens from Española Island.

Examining the contemporary population

It is unfortunate that we were unable to obtain microsatellite genotypes from the historical specimens, because assessing

nuclear divergence will be critical to understanding the degree to which the extinct and contemporary taxa are distinct. Given the geographic proximity of the collection sites for the cave and live specimens in 1906 (Fig. 1), and the radiocarbon date of CAS 8128 being between 1450 and 1630 AD, it seems likely that the lineages were sympatric at least in recent history. Such a situation raises the possibility that the lineages come from two different gene pools that had fused, resulting in a single population with mitochondrial haplotypes from both ancestral lineages, that became fixed (or very nearly fixed) for the single haplotype following the subsequent bottleneck in the 1900s. Lineage fusion is documented to be occurring in the Galapagos giant tortoise species *C. becki*, where secondary contact between descendants of two separate colonizations of Volcano Wolf, both from Santiago Island, is resulting in introgressive hybridization (Garrick et al. 2014). The intermediate carapace morphology of the contemporary San Cristóbal population could be explained by this type of scenario, as elsewhere in the Galapagos, intermediate shell morphology is likely due to secondary contact among species with different carapace morphology (Fritts 1984; Russello et al. 2007; Poulakakis et al. 2008; Chiari 2021).

To evaluate evidence of possible admixture or substructure in the contemporary population that might be the legacy of past lineage fusion, we examined nuclear genetic markers in the contemporary population. Using both 21 microsatellites and >12 000 SNPs from ddRAD sequencing, we found some support for $K = 2$. This result, if reliable, would suggest the existence of two nuclear lineages on the island. However, evaluating the relative support for $K = 1$ and $K = 2$ can be difficult and it is important to interpret the results in light of the total evidence for genetic subdivision (Janes et al. 2017; Cullingham et al. 2020). In this case, the SNPs and microsatellites assigned some individuals with high membership to opposite clusters (Fig. 4), suggesting that there is not true signal to support the assignment. There is also no identifiable biologically meaningful pattern to the $K = 2$ clusters based on geography, illustrated by the PCA color coded by sampling region (Supplementary Fig. 4), where the single cloud of points has a random scattering of each color. Taken together, these results do not constitute strong evidence of population substructure that would support the continued existence of two nuclear lineages on the island.

The documented bottleneck down to an estimated 500–700 individuals on San Cristóbal in the 1970s has evidently led to the fixation of the contemporary population to a single mitochondrial control region haplotype based on our sequencing of over 125 individuals. Studies that have compared the levels of nuclear diversity among the Galapagos giant tortoise species have found San Cristóbal to have moderate diversity (Garrick et al. 2015; Miller et al. 2018). Compared to our findings here, Garrick et al. (2015) found higher expected heterozygosity (0.73 vs 0.68) and effective population size (68, 95% CI 32–4592 vs 34, 95% CI 23–55, Table 1). However, Garrick et al. (2015) results were based on 12 microsatellites genotyped for 19 individuals sampled in 1999 from Punta Pitt, whereas ours included 10 of those same individuals plus 54 tortoises sampled from a broader area. It is not possible to directly compare the SNP diversity measures calculated here with those from Miller et al. (2018) because that study identified variants across all species of Galapagos giant tortoise, thus including many loci that are not variable within the San Cristóbal population. The difference in G_{IS} calculated between the SNPs and microsatellites in this study is probably due to locus AGG68, which has only two alleles (Supplementary Table 4), the minor with a frequency of just 5%. When this locus is excluded, the G_{IS} from microsatellites does not significantly differ from zero (data not shown), and the G_{IS} value for the SNPs, although significant, is only slightly negative and close to zero.

The estimates of N_e for the current population on San Cristóbal are 26 and 34 for the SNPs and microsatellites, respectively.

Although N_e estimates based on the linkage disequilibrium method have been shown to be fraught, especially when the assumptions of an ideal population are violated (Waples and Yokota 2007; Waples 2010), we think that the N_e estimates here are reliable because they are based on a sample size ($n = 64$) that is larger than the estimated N_e for both markers, a condition which provides some extra confidence in the estimate (England et al. 2006). This is also supported by findings from previous SNP analyses in the Pinzón species of giant tortoise (Jensen et al. 2018a). In that study it was found that N_e estimates based on sample sizes above 60 individuals were relatively consistent. The comparison of SNP-based N_e estimates in these two species offers some additional insights into the demographic history of the San Cristóbal tortoises. The N_e estimate for the contemporary Pinzón population ($N_e = 59$) (Jensen et al. 2018a) is larger than that for San Cristóbal ($N_e = 24$), despite the former species having experienced a much more severe 20th century bottleneck to just 250 individuals. If we use the N_e and bottleneck census size for the Pinzón and San Cristóbal species to calculate $N_e:N_c$ ratios, we obtain dramatically different values (0.24 and 0.03, respectively), providing additional support for San Cristóbal having experienced an extremely pronounced bottleneck that led to the fixation of a single mitochondrial haplotype.

CONCLUSIONS AND FUTURE DIRECTIONS

Understanding the evolutionary history of giant tortoises on San Cristóbal Island may hold further clues to reconstructing the broader tortoise radiation across the Galapagos Archipelago. Due to the highly degraded nature of the bones found in the cave, we were unable to collect genotypic data from microsatellites in this study, despite having used micro-CT scans to identify the region of bone that should yield the best-preserved DNA. Future efforts using shotgun sequencing or capture approaches may be able to yield a full mitochondrial genome (e.g., Jensen et al. 2018b), reduced representation of nuclear SNP loci (e.g., Gauthier et al. 2020), or a low-coverage nuclear genome (e.g., Yao et al. 2020), which could provide critical information on the level of divergence between the extinct lineage and other recognized taxa, and shed light on the possibility of lineage fusion on San Cristóbal. Based on the data in hand, we can confidently conclude that there was a previously unrecognized mitochondrial lineage, but without nuclear genetic data, it is unclear whether this lineage persisted as a distinct entity until its extinction in the 20th century.

The recognition of an additional lineage of giant tortoise on San Cristóbal Island may have implications for the taxonomy of *C. chathamensis*. The holotype for *C. chathamensis* is one of the cave specimens (CAS 8127), which has a very distinct mitochondrial haplotype from the contemporary population, which goes by that name. If future evidence based on nuclear genetic markers confirms that the extinct lineage warrants species status, a new name and type specimen would be required to represent the lineage still living on San Cristóbal today.

This study also provides important lessons that go beyond their relevance for this particular group of organisms, highlighting the importance of including a diachronic perspective to reconstruct the evolutionary history of a group. We are aware that this is not always possible, as our inference of the evolutionary past is routinely inferred from contemporary samples only. This study shows what could have been missed if samples collected at two different time points were not included. Without the ability to analyze genetic data from bone samples found in the cave together with samples from the contemporary population on the same island, the new mitochondrial lineage would remain unknown. This in turn has opened doors to new research directions into the taxonomy, systematics, and evolutionary history of this iconic radiation.

DATA AVAILABILITY

The microsatellite Structure file and SNP vcf file are available on Dryad (doi:10.5061/dryad.xwdbvr1fp). The demultiplexed ddRAD fastq files are available on the NCBI SRA as BioProject PRJNA802698. The newly generated mitochondrial control region sequences are available from Genbank (Accession MT899437—MT899442).

REFERENCES

- Ali JR, Aitchison JC (2014) Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J Biogeogr* 41:1227–1241
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evolution* 29:2157–2167
- Banning GH 1933. Hancock expedition to the Galapagos Islands, 1933. *Bulletin of the Zoological Society of San Diego* 10.
- Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT (2011) BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692
- Beheregaray LB, Ciofi C, Caccone A, Gibbs JP, Powell JR (2003) Genetic divergence, phylogeography and conservation units of giant tortoises from Santa Cruz and Pinzón, Galápagos Islands. *Conserv Genet* 4:31–46
- Beheregaray LB, Gibbs JP, Havill N, Fritts TH, Powell JR, Caccone A (2004) Giant tortoises are not so slow: rapid diversification and biogeographic consensus in the Galápagos. *Proc Natl Acad Sci USA* 101:6514–6519
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Poppinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu C-H, Xie D, Zhang C, Stadler T, Drummond AJ (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biol* 15:e1006650
- Bronk Ramsey C (2009) Bayesian analysis of radiocarbon dates. *Radiocarbon* 51:337–360
- Caccone A, Gentile G, Burns CE, Sezzi E, Bergman W, Ruelle M, Saltonstall K, Powell JR (2004) Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. *Mol Phylogenetics Evol* 31:794–798
- Caccone A, Gentile G, Gibbs JP, Fritts TH, Snell HL, Betts J, Powell JR (2002) Phylogeography and history of giant Galápagos tortoises. *Evolution* 56:2052–2066
- Caccone A, Gibbs JP, Ketmaier V, Suatoni E, Powell JR (1999) Origin and evolutionary relationships of giant Galápagos tortoises. *Proc Natl Acad Sci USA* 96:13223–13228
- Castañeda-Rico S, Johnson SA, Clement SA (2019) Insights into the evolutionary and demographic history of the extant endemic rodents of the Galápagos Islands. *Therya* 10:213–228
- Cayot L (2008) The restoration of giant tortoise and land iguana populations in Galápagos. *Galápagos Res* 65:39–43
- Chiari Y (2021). Chapter 8 - Morphology. In Gibbs JP, Cayot LJ, Aguilera WT eds. *Galapagos giant tortoises*. Academic Press, p. 139–156
- Chiari Y, Hyseni C, Fritts TH, Glaberman S, Marquez C, Gibbs JP, Claude J, Caccone A (2009) Morphometrics parallel genetics in a newly discovered and endangered taxon of Galápagos tortoise. *PLoS One* 4:e6272
- Ciofi C, Milinkovitch MC, Gibbs JP, Caccone A, Powell JR (2002) Microsatellite analysis of genetic divergence among populations of giant Galápagos tortoises. *Mol Ecol* 11:2265–2283
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659
- Cowie RH, Holland BS (2008) Molecular biogeography and diversification of the endemic terrestrial fauna of the Hawaiian Islands. *Philos Trans R Soc B: Biol Sci* 363:3363–3376
- Cullingham CI, Miller JM, Peery RM, Dupuis JR, Malenfant RM, Gorrell JC, Janes JK (2020) Confidently identifying the correct K value using the ΔK method: when does $K = 2$? *Mol Ecol* 29:862–869
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, G. 1000 Genomes Project Analysis (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Dirección del Parque Nacional Galápagos (2014) Plan de Manejo de las Áreas Protegidas de Galápagos para el Buen Vivir. Puerto Ayora, Galápagos, Ecuador
- Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Mol Ecol Resour* 14:209–214

- Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Edwards DL, Benavides E, Garrick RC, Gibbs JP, Russello MA, Dion KB, Hyseni C, Flanagan JP, Tapia W, Caccone A (2013) The genetic legacy of Lonesome George survives: giant tortoises with Pinta Island ancestry identified in Galápagos. *Biol Conserv* 157:225–228
- England PR, Cornuet J-M, Berthier P, Tallmon DA, Luikart G (2006) Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conserv Genet* 7:303
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods Ecol Evolution* 6:925–929
- Fritts TH (1983) Morphometrics of Galápagos tortoises: evolutionary implications. In: Bowman R, Berson M, Leviton A eds *Patterns of Evolution in Galapagos Organisms*. Pacific Division of the American Association for the Advancement of Science, San Francisco, CA
- Fritts TH (1984) Evolutionary divergence of giant tortoises in Galápagos. *Biol J Linn Soc* 19:165–176
- Garrick RC, Benavides E, Russello MA, Hyseni C, Edwards DL, Gibbs JP, Tapia W, Ciofi C, Caccone A (2014) Lineage fusion in Galapagos giant tortoises. *Mol Ecol* 23:5276–5290
- Garrick RC, Kajdacs B, Russello MA, Benavides E, Hyseni C, Gibbs JP, Tapia W, Caccone A (2015) Naturally rare versus newly rare: demographic inferences on two timescales inform conservation of Galápagos giant tortoises. *Ecol Evolution* 5:676–694
- Gauthier J, Pajkovic M, Neuenschwander S, Kaila L, Schmid S, Orlando L, Alvarez N (2020) Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Mol Ecol Resour* 20:1191–1205
- Geist D (1996) On the emergence and submergence of the Galápagos Islands. *Noticias de Galápagos* 56:5–8
- Geist DJ, Snell H, Snell H, Goddard C, Kurz MD (2014) A paleogeographic model of the Galápagos islands and biogeographical and evolutionary implications. In: Harpp KS, Mittelstaedt E, d'Ozouville N, Graham DW eds *The Galápagos: a natural laboratory for the earth science*. John Wiley & Sons, Inc, New Jersey, p 145–166
- Gillespie RG (2007) Oceanic islands: models of diversity. In: Levin SA ed. *Encyclopedia of Biodiversity*. Elsevier, Oxford, UK
- Grant PR, Grant BR (2014) 40 Years of Evolution: Darwin's Finches on Daphne Major Island. Princeton University Press, <https://doi.org/10.2307/j.ctt5hhncc>
- Hill WG (1981) Estimation of effective population-size from data on linkage disequilibrium. *Genetical Res* 38:209–216
- Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, Guilderson TP, Heaton TJ, Palmer JG, Reimer PJ, Reimer RW, Turney CSM, Zimmerman SRH (2013) SHCal13 Southern Hemisphere Calibration, 0–50,000 Years cal BP. *Radiocarbon* 55:1889–1903
- IUCN. 2020. The IUCN Red List of Threatened Species. Version 2020-2. <https://www.iucnredlist.org>
- Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL (2017) The $K = 2$ conundrum. *Mol Ecol* 26:3594–3602
- Jensen EL, Edwards DL, Garrick RC, Miller JM, Gibbs JP, Cayot LJ, Tapia W, Caccone A, Russello MA (2018a) Population genomics through time provides insights into the consequences of decline and rapid demographic recovery through head-starting in a Galapagos giant tortoise. *Evolut Appl* 11:1811–1821
- Jensen EL, Miller JM, Edwards DL, Garrick RC, Tapia W, Caccone A, Russello MA (2018b) Temporal Mitogenomics of the Galapagos Giant Tortoise from Pinzon Reveals Potential Biases in Population Genetic Inference. *J Heredity* 109:631–640
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet* 11:94
- Karnauskas KB, Mittelstaedt E, Murtugudde R (2017) Paleoceanography of the eastern equatorial Pacific over the past 4 million years and the geologic origins of modern Galápagos upwelling. *Earth Planet Sci Lett* 460:22–28
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, S. Genome Project Data Processing (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
- Lindgreen S (2012) AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* 5:337
- Longmire J, Maltbie M, Baker R 1997. Use of 'lysis buffer' in DNA isolation and its implication for museum collections. *Occasional Papers the Museum Texas Tech University*:1–3.
- MacLeod A, Rodriguez A, Vences M, Orozco-terWengel P, Garcia C, Trillmich F, Gentile G, Caccone A, Quezada G, Steinfartz S (2015) Hybridization masks speciation in the evolutionary history of the Galapagos marine iguana. *Proc Biol Sci* 282:20150425. <https://doi.org/10.1098/rspb.2015.0425>
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* 4:792–794
- Miller JM, Quinzin MC, Edwards DL, Eaton DAR, Jensen EL, Russello MA, Gibbs JP, Tapia W, Rueda D, Caccone A (2018) Genome-wide assessment of diversity and divergence among extant Galapagos giant tortoise species. *J Heredity* 109:611–619
- Parent CE, Caccone A, Petren K (2008) Colonization and diversification of Galapagos terrestrial fauna: a phylogenetic and biogeographical synthesis. *Philos Trans R Soc Lond B Biol Sci* 363:3347–3361
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7:e37135
- Poulakakis N, Edwards DL, Chiari Y, Garrick RC, Russello MA, Benavides E, Watkins-Colwell GJ, Glaberman S, Tapia W, Gibbs JP, Cayot LJ, Caccone A (2015) Description of a new Galápagos giant tortoise species (*Chelonoidis*; Testudines: Testudinidae) from Cerro Fatal on Santa Cruz Island. *PLoS One* 10:e0138779
- Poulakakis N, Glaberman S, Russello M, Beheregaray LB, Ciofi C, Powell JR, Caccone A (2008) Historical DNA analysis reveals living descendants of an extinct species of Galápagos tortoise. *Proc Natl Acad Sci USA* 105:15464–15469
- Poulakakis N, Miller JM, Jensen EL, Beheregaray LB, Russello MA, Glaberman S, Boore J, Caccone A (2020) Colonization history of Galapagos giant tortoises: insights from mitogenomes support the progression rule. *J Zool Syst Evolut Res* 58:1262–1275
- Poulakakis N, Russello M, Geist D, Caccone A (2012) Unravelling the peculiarities of island life: vicariance, dispersal and the diversification of the extinct and extant giant Galápagos tortoises. *Mol Ecol* 21:160–173
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pritchard PCH (1996) The Galápagos tortoises—nomenclatural and survival status. *Chelonian Res Monogr* 1:1–85
- Quesada V, Freitas-Rodríguez S, Miller J, Pérez-Silva JG, Jiang Z-F, Tapia W, Santiago-Fernández O, Campos-Iglesias D, Kuderna LFK, Quinzin M, Álvarez MG, Carrero D, Beheregaray LB, Gibbs JP, Chiari Y, Glaberman S, Ciofi C, Araujo-Voces M, Mayoral P, Arango JR, Tamargo-Gómez I, Roiz-Valle D, Pascual-Torner M, Evans BR, Edwards DL, Garrick RC, Russello MA, Poulakakis N, Gaughran SJ, Rueda DO, Bretones G, Marqués-Bonet T, White KP, Caccone A, López-Otin C (2019) Giant tortoise genomes provide insights into longevity and age-related disease. *Nat Ecol Evolution* 3:87–95
- Quinzin MC, Sandoval-Castillo J, Miller JM, Beheregaray LB, Russello MA, Hunter EA, Gibbs JP, Tapia W, Villalva F, Caccone A (2019) Genetically informed captive breeding of hybrids of an extinct species of Galapagos tortoise. *Conserv Biol* 33:1404–1414
- R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 67:901–904
- Rhodin AGJ, Iverson JB, Bour R, Fritz U, Georges A, Shaffer HB, van Dijk PP. 2017. Turtles of the world: annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status. Pp 1–292 in Rhodin AGJ, Iverson JB, van Dijk PP, Saumure RA, Buhlmann KA, Pritchard PCH, Mittermeier RA, eds. *Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group*. Chelonian Research Monographs.
- Russello MA, Beheregaray LB, Gibbs JP, Fritts T, Havill N, Powell JR, Caccone A (2007) Lonesome George is not alone among Galápagos tortoises. *Curr Biol* 17: R317–318
- Russello MA, Glaberman S, Gibbs JP, Marquez C, Powell JR, Caccone A (2005) A cryptic taxon of Galápagos tortoise in conservation peril. *Biol Lett* 1:287–290

- Russello MA, Poulakakis N, Gibbs JP, Tapia W, Benavides E, Powell JR, Caccone A (2010) DNA from the past informs *ex situ* conservation for the future: an “extinct” species of Galápagos tortoise identified in captivity. *PLoS One* 5:e8683
- Schubert M, Ermini L, Der Sarkissian C, Jonsson H, Ginolhac A, Schaefer R, Martin MD, Fernandez 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. *Nat Protoc* 9:1056–1082
- Shaw KL, Gillespie RG (2016) Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. *Proc Natl Acad Sci* 113:7986–7993
- Tapia AW, Sevilla C, Malaga J, Gibbs JP (2021) Chapter 20 - Tortoise populations after 60 years of conservation. In Gibbs JP, Cayot LJ, Aguilera WT, eds. *Galapagos Giant Tortoises*. Academic Press, p. 401–432
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation *Genet* 132:619–633
- Van Denburgh J (1914) Expedition of the California Academy of Sciences to the Galápagos Islands 1906–1906. *Proc Calif Acad Sci* 2:203–374
- Vogel JS, Southon JR, Nelson DE, Brown TA (1984) Performance of catalytically condensed carbon for use in accelerator mass spectrometry. *Nucl Instrum Methods* 223:289–293
- Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv Genet* 7:167–184
- Waples RS (2010) Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Mol Ecol Resour* 10:785–796
- Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol Appl* 3:244–262
- Waples RS, Yokota M (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics* 175:219–233
- Yao L, Witt K, Li H, Rice J, Salinas NR, Martin RD, Huerta-Sánchez E, Malhi RS (2020) Population genetics of wild *Macaca fascicularis* with low-coverage shotgun sequencing of museum specimens. *Am J Phys Anthropol* 173:21–33

ACKNOWLEDGEMENTS

We thank the California Academy of Sciences for providing access to their specimens, C. Mariani for help with genotyping microsatellites, and C. Hadden for help with carbon dating. We thank the Galápagos National Park Directorate for their collaboration. We thank three anonymous reviewers for their helpful comments and suggestions. This work was supported by a number of agencies that funded either the field expeditions to gather the samples or the genetic analyses. They include the National Geographic Society, The Eppley Foundation, the Galapagos Conservancy, and the Mohamed Bin Zayed Species Conservation Fund, Turtle Conservancy, and Swiss Friends of Galapagos.

AUTHOR CONTRIBUTIONS

EJ: Formal analysis, investigation, data curation, writing- original draft, writing—review and editing, visualization. MQ: Conceptualization, formal analysis, investigation, data

curation, writing—review and editing. JM: Conceptualization, writing—review and editing. MR: Sample and data curation, conceptualization, writing—review and editing. RG: Writing—review and editing. DE: Resources, data curation, writing—review and editing. SG: Sample and data curation, writing—review and editing. YC: Resources, writing—review and editing. NP: Writing—review and editing. WT: Resources, writing—review and editing. JG: Resources, writing—review and editing. AC: Conceptualization, methodology, resources, writing—original draft, writing—review and editing, supervision, project administration, funding acquisition.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Samples collected were collected and transported under CITES permits 12US209142, 15US209142/9, Galápagos Park Permit PC-75–16, Framework Contract for Access to Genetic Resources MAE-DNB-CM-2016-0060, and in accordance with Yale Institutional Animal Care and Use Committee (IACUC) permit number 2016–10825.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-022-00510-8>.

Correspondence and requests for materials should be addressed to Evelyn L. Jensen.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022