Molecular Biology and Evolution, 2025, **42**, 1–13
https://doi.org/10.1093/molbev/msaf119
Advance access publication 26 May 2025
Discoveries

MBE OXFORD

Accounting for Chimerism in Demographic Inference: Reconstructing the History of Common Marmosets (*Callithrix jacchus*) from High-Quality, Whole-Genome, Population-Level Data

Vivak Soni [1], **Cyril J. Versoza** [1], **Eric J. Vallender** [2,3], **Jeffrey D. Jensen** [1,*,†], **Susanne P. Pfeifer** [1,*,†]

[1]Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA
[2]Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA
[3]Tulane National Primate Research Center, Covington, LA, USA

[†]These authors jointly supervised the project.
*Corresponding authors: E-mails: jeffrey.d.jensen@asu.edu; susanne@spfeiferlab.org.
Associate editor: Kelley Harris

Abstract

As a species of considerable biomedical importance, characterizing the evolutionary genomics of the common marmoset (*Callithrix jacchus*) is of significance across multiple fields of research. However, at least 2 peculiarities of this species potentially preclude commonly utilized population genetic modeling and inference approaches: a high frequency of twin births and hematopoietic chimerism. We here investigate these effects within the context of demographic inference, demonstrating via simulation that neglecting these biological features results in significant misinference of the underlying population history. Based upon this result, we develop a novel approximate Bayesian inference approach accounting for both common twin births and chimeric sampling. In addition, we newly present population genomic data from 15 individuals sequenced to high coverage and utilize gene-level annotations to identify neutrally evolving intergenic regions appropriate for demographic inference. Applying our developed methodology, we estimate a well-fitting population history for this species, which suggests robust ancestral and current population sizes, as well as a size reduction roughly 7,000 years ago likely associated with a shift from arboreal to savanna vegetation in north-eastern Brazil during this period.

Keywords: chimerism, primate, platyrrhine, population history, demography, population genetics

Introduction

Characterized by exudivorous feeding habits and small habitat ranges (~5,000 to 65,000 m^2), the common (or white-tufted-ear) marmoset (*Callithrix jacchus*) is a platyrrhine native to east-central Brazil (Rylands and Faria 1993; Rylands et al. 2009; Garber et al. 2019). Due to its diminutive size (~250 g), early sexual maturity (~15 to 18 months of age), short gestation period (~145 d), and high fecundity (with twin births being the norm), this species has risen in biomedical prominence as a commonly used model for the study of both human neurodevelopmental disorders (e.g. Miller et al. 2016; Philippens and Langermans 2021) and infectious disease dynamics, with the latter partly owing to their seemingly reduced major histocompatibility complex diversity relative to other mammals (Antunes et al. 1998; Wu et al. 2000; Carrion and Patterson 2012).

Together with a high frequency of twin births, *C. jacchus* is also noteworthy for the frequent observation of hematopoietic chimerism—a rare phenomenon among primates. As a consequence, marmoset blood samples have been shown to contain genetic material from both the sampled individual and their twin sibling (Hill 1932; Wislocki 1939; Benirschke et al. 1962; Gengozian et al. 1969; Ross et al. 2007). Importantly, Ross et al. (2007) found that chimerism in marmosets was not only limited to blood samples, but was in fact identified in every sampled tissue examined. However, Sweeney et al. (2012) subsequently suggested the possibility that blood infiltration may give rise to such apparent chimerism across tissues. This result was largely confirmed by del Rosario et al. (2024), who found that chimerism present in liver, kidney, and brain tissues was the result of myeloid and lymphoid cell lineages derived from hematopoietic stem cells and that blood samples contained the greater contribution of sibling nuclei relative to other examined tissue types (see Figure 1 in del Rosario et al. 2024 and the accompanying commentary of Chiou and Snyder-Mackler 2024). Taken together, these results thus do not suggest germline chimerism, but they strongly support chimerism of equal contribution in the sampled blood and chimerism proportional to hematopoietic infiltration in other tissues commonly utilized for sequencing.

Although it may be tempting to try to sidestep the effects of chimerism when performing genetic studies in marmosets by focusing on single births, it is noteworthy that single births are not only rare, but also tend to themselves be chimeric owing to fetal resorption of a dizygotic twin in utero (Jaquish et al. 1996; Windle et al. 1999). It has additionally

Received: February 11, 2025. **Revised:** May 7, 2025. **Accepted:** May 8, 2025

been proposed that one may better utilize samples from tissues potentially less impacted by chimerism (e.g. fingernails, as assessed from lower levels of heterozygosity; Yang et al. 2023); however, the lower (but non-zero) levels of chimerism in non-blood tissue only exacerbate the uncertainty in modeling. Hence, in terms of the sample itself, it appears difficult to reliably avoid the contribution of genetic material from the unsampled twin. Moreover, even if it were possible to sample genetic material from a true single individual in any given generation, the long-term, multi-generation genetic transmission resulting from standard twin births would nonetheless require investigation in order to understand population-level allele frequency dynamics in this species, given this inherent violation of standard modeling assumptions.

The first marmoset genome, assembled from whole-genome shotgun plasmid, fosmid, and BAC end sequences, was published in 2014 (The Marmoset Genome Sequencing and Analysis Consortium 2014). Additional studies have since continued to improve upon this initial assembly (see the review of Vallender 2019), filling in gaps and refining gene annotations, with the most recent reference genome now exhibiting 98.3% completeness (Yang et al. 2021). Population genomic inference has thus far been limited to characterizing general levels of genetic diversity and divergence in the species (Faulkes et al. 2003; Malukiewicz et al. 2014; Yang et al. 2021, 2023; Mao et al. 2024), as well as performing genomic scans, some of which have implicated genes putatively involved in twinning as having experienced long-term positive selection (Harris et al. 2014). Importantly however, no study to date has attempted to model the effects of chimerism on population genetic inference, though one may readily hypothesize that effectively treating sampled chimeras as a single individual (i.e. as though sequenced from a single non-chimeric blood/tissue sample) may well have important implications for observed levels and patterns of genetic variation and thus on downstream evolutionary inference. For example, Mao et al. (2024) found that a non-chimeric closely related platyrrhine (owl monkeys, for which the rate of twinning is also considered to be extremely low; Huck et al. 2014) was characterized by considerably lower divergence to humans compared with marmosets, while Harris et al. (2023) found that marmosets had a generally reduced relative heterozygosity. However, the potential effects of twinning and chimerism themselves on these observations were left unexplored, and it thus remains unclear whether these unusual reproductive dynamics, or, for example, fundamental differences in population size and/or mutation rates, better explain these patterns.

The Importance of Evolutionary Baseline Models

As a focal point of population genetic inference is on quantifying the relative roles of neutral and selective processes in governing levels and patterns of genetic variation, and as chimerism likely plays an additional role in shaping this variation, it is thus necessary to evaluate expectations associated with this inherent "twin-sampling." However, chimerism and twinning aside, disentangling these competing processes is already relatively challenging (for a discussion, see the reviews of Charlesworth and Jensen 2022; Jensen 2023). As but one example, population growth, background selection (BGS; Charlesworth et al. 1993), and recurrent positive selection (Maynard Smith and Haigh 1974) can all result in a similar skew toward rare alleles when examining allele frequency distributions (i.e. the site frequency spectrum [SFS]; Kim 2006; Jensen et al. 2007; Nicolaisen and Desai 2012, 2013; Ewing and Jensen 2014, 2016; Johri et al. 2021; Soni et al. 2023; and see the reviews of Charlesworth and Jensen 2021, 2024). Apart from natural selection and genetic drift, heterogeneity in underlying mutation and recombination rates across the genome can also modify these expectations in significant ways (Dapper and Payseur 2018; Samuk and Noor 2022; Soni et al. 2024a). For these reasons, the case has been made that prior to evaluating genomes for evidence of relatively uncommon evolutionary processes such as positive and balancing selection, one must first construct an evolutionarily appropriate baseline model accounting for these constantly operating evolutionary processes (Bank et al. 2014; Johri et al. 2022b).

Importantly, the appropriate approach to take when accounting for the potentially confounding effects of, for example, selection inference with demographic inference will depend on the specific details of the genome architecture of the species in question. In coding-dense genomes, the large genomic fraction of directly selected sites implies that there may be few genomic regions that are free from the effects of either direct selection or selection at linked sites. As such, the majority of genomic regions are likely shaped by both demography and selection (e.g. Irwin et al. 2016; Sackman et al. 2019; Jensen 2021; Morales-Arce et al. 2022; Terbot et al. 2023a, 2023b; Howell et al. 2023; Soni et al. 2024b). Under this scenario, it is necessary to simultaneously infer population history and selection jointly, and approximate Bayesian computation (ABC) approaches have been developed for this purpose (Johri et al. 2020, 2021, 2023)—though it should be noted that, owing to the large number of underlying parameters involved, these approaches have remained limited to simplified demographic models to date.

By comparison, there are notable advantages to performing evolutionary inference in species with coding-sparse genomes (including marmosets and other primates). Owing to the prevalence of neutral sites at sufficient recombinational distances from functional sites, such that they are unlikely to experience BGS effects (Charlesworth et al. 1993), so-called "2-step" inference approaches become viable (see Soni and Jensen 2025). Here, population history can be inferred from neutral intergenic regions, and numerous well-performing neutral demographic estimators have been developed for such purposes (e.g. Gutenkunst et al. 2009; Excoffier et al. 2013; and see the review of Beichman et al. 2018). Conditional on the population history inferred in this first step, selective processes can then be inferred using functional sites in a second step. Crucially, this baseline model accounting for population history, structure, and gene flow, together with the action of purifying and background selection in and around functional regions, has been found to be important for reducing false-positive rates when scanning for the episodic effects of positive or balancing selection (e.g. Barton 1998; Przeworski 2002; Jensen et al. 2005; Poh et al. 2014; Harris and Jensen 2020; Soni and Jensen 2024; and see Johri et al. 2022a). Such multifaceted analyses of the inter-related evolutionary processes of mutation and crossover/non-crossover rates, with population history and purifying selection, have been executed in a small number of non-chimeric primate species, including in both a haplorrhine (humans; Kong et al. 2002; DeGiorgio et al. 2016; Carlson et al. 2018; Johri et al. 2023; Ragsdale et al. 2023; Soni and Jensen 2025) and a strepsirrhine (aye-ayes; Versoza et al. 2024, 2025a, 2025b; Soni et al. 2024c, 2025a, 2025b; Terbot et al. 2025).

Incorporating Chimerism into an Evolutionary Baseline Model

Given the long-standing and compelling evidence for chimerism in marmosets (see the review of Malukiewicz et al. 2020), together with the high rate of twin births, these features will necessarily be an important component of the evolutionary baseline model for this species in order to fully characterize neutral expectations. Notably, accounting for species-specific biology is not particularly uncommon in baseline model construction. For example, in species characterized by strong reproductive progeny skew (as observed e.g., in both marine and terrestrial broadcast spawners, as well as multiple pathogens), one must similarly account for the impacts of such a Wright–Fisher (WF) model violation on downstream inference, owing to related modifications of expected neutral patterns of variation (Eldon and Wakeley 2006; Matuszewski et al. 2018; Sackman et al. 2019; Sabin et al. 2022; and see the reviews of Tellier and Lemaire 2014; Irwin et al. 2016). Importantly, a neglect of this biological reality in affected organisms has been demonstrated to result in both a mis-inference of population growth and the false detection of widespread positive selection (e.g. Durrett and Schweinsberg 2004; Hallatschek 2018).

In this study, we thus uniquely construct an appropriate population genomic baseline model for *C. jacchus* and firstly examine the effects of twinning and hematopoietic chimerism on equilibrium expectations of levels (e.g. $\hat{\theta}_w$; Watterson 1975) and patterns (e.g. the SFS) of genetic diversity. Secondly, based on these observed deviations, we quantify the extent to which common demographic inference approaches may be biased by the unaccounted for presence of these effects. Thirdly, we develop a novel ABC framework for inferring population history in the presence of both twin births and chimerism, and present a simulation study evaluating the performance of this approach. Finally, we present novel population genomic data from 15 common marmoset individuals sequenced to high coverage and, utilizing gene-level annotations to identify genetic variation in intergenic regions appropriate for demographic inference, we apply our ABC approach to estimate a well-fitting population history for the species. Taken together, these results demonstrate the importance of accounting for these biological peculiarities when performing demographic inference in this species, with the resulting demographic model suggesting a population bottleneck in *C. jacchus* roughly 7,000 years ago, followed by a partial recovery. Paleoclimatic and palynological studies indicate a shift from arboreal to savanna vegetation in north-eastern Brazil during the timeframe of this bottleneck—an event that would be anticipated to impact a species that relies on arboreal locomotion.

Results and Discussion

Whole-Genome, Population-Level Data

In order to infer the population history of common marmosets (*C. jacchus*), we whole-genome sequenced 15 individuals previously housed at the New England Primate Research Center to mean depths of 35× per individual (supplementary table S1, Supplementary Material online). Following best practices in the field (Pfeifer 2017), we mapped individual reads to the current marmoset reference genome and subsequently called, genotyped, and filtered variants using the Genome Analysis Toolkit workflow (van der Auwera and O'Connor 2020).

Table 1 Means and SDs of θ_W (Watterson 1975), Tajima's D (Tajima 1989), and the number of singletons across putatively neutral regions in the empirical data, calculated across 10 kb windows

	θ_W	Tajima's D	No. of singletons
Mean	0.0010	0.2548	0.0007
SD	0.0009	1.1484	0.0013

Additionally, to account for the effects of both direct and background selection—2 evolutionary processes previously shown to bias demographic inference (e.g. Johri et al. 2021; and see the reviews of Charlesworth and Jensen 2021, 2024)—we excluded any sites within, or sufficiently close to, functional regions prior to downstream analyses (see Materials and Methods for further details). The final data set of putatively neutral regions consisted of 1.4 million autosomal, biallelic single nucleotide polymorphisms (SNPs) with a transition–transversion ratio (Ts/Tv) of 2.11 in the accessible genome (supplementary table S2, Supplementary Material online). Table 1 provides the means and standard deviations (SDs) of θ_W (Watterson 1975), Tajima's D (Tajima 1989), and the number of singletons across these putatively neutral regions.

Evaluating the Impact of Chimerism and Twinning on the Performance of Common SFS-Based Estimators of Population History

As marmoset blood samples have been shown to contain genetic material from both the sampled individual and their dizygotic (fraternal) twin (Hill 1932; Wislocki 1939; Benirschke et al. 1962; Gengozian et al. 1969; Ross et al. 2007), we modeled chimerism by simulating a non-WF model in SLiM4 (Haller and Messer 2023) in which monogamous mating pairs produce non-identical twin offspring. The genotypes of these twins were combined post-simulation to create a chimeric individual (Fig. 1; and see Materials and Methods for further details). To evaluate the performance of commonly used SFS-based demographic inference methods, we simulated a single neutrally evolving population under 4 demographic scenarios: (i) population equilibrium (with a single parameter N_{current}, the population size at the time of sampling), as well as 3 instantaneous population size change scenarios (with 3 parameters each: $N_{\text{ancestral}}$, the population size prior to the size change; T, the time of the size change in $N_{\text{ancestral}}$ generations; and N_{current})—namely, (ii) population expansion (population size doubling), (iii) population contraction (population size halving), and (iv) severe population contraction (population size reduced to $0.1N_{\text{ancestral}}$). These scenarios were simulated under both our chimeric model and a standard WF model in order to compare demographic inference power with and without chimerism. We performed demographic inference with the coalescent-based estimator fastsimcoal2 (Excoffier et al. 2013) and the diffusion approximation of $\delta a \delta i$ (Gutenkunst et al. 2009), both of which are commonly used neutral estimators that fit a demographic model to the observed SFS.

Generally, chimerism resulted in mis-inference of demographic parameters across all 4 population histories (see Fig. 2 for the results of the demographic inference of the equilibrium population, and see supplementary figs. S1 to S3, Supplementary Material online for the 3 population size change scenarios). Notably, the variance on parameter

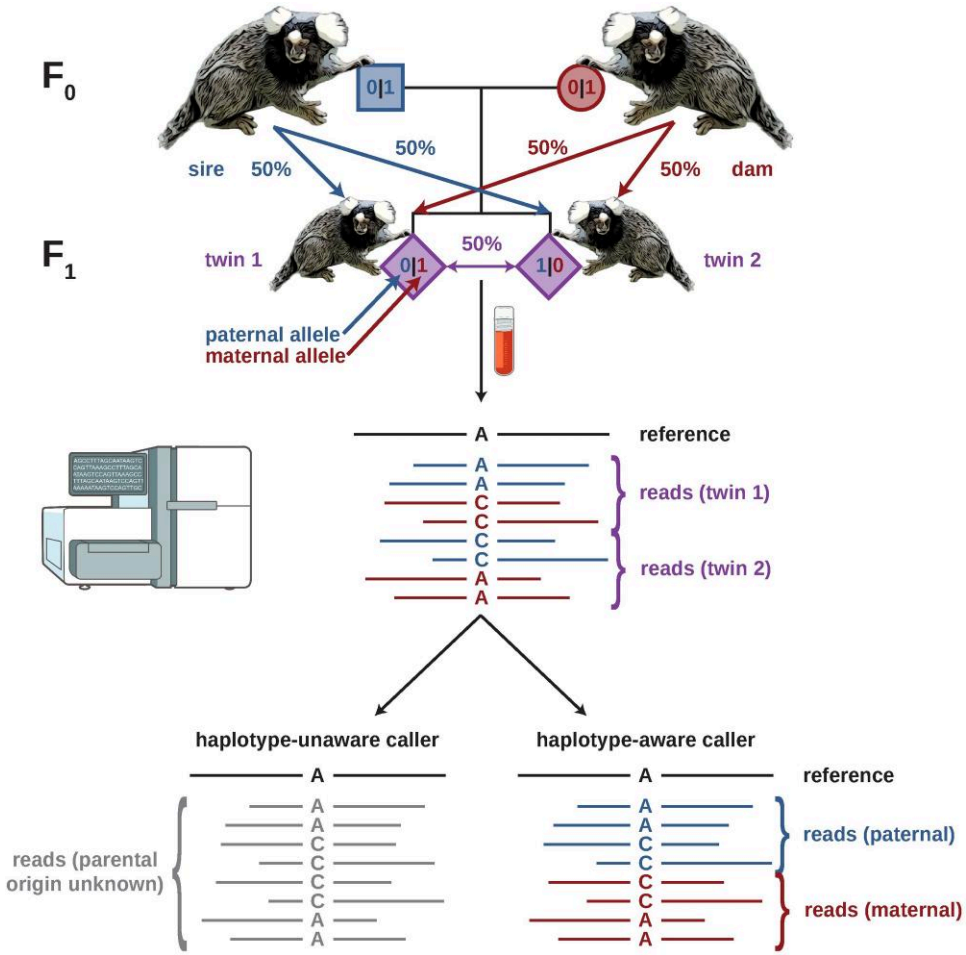

Fig. 1. Modeling of chimerism. Marmoset blood samples contain genetic material from both the sampled individual and their dizygotic (fraternal) twin (at a 1:1 ratio on average). To model chimerism, we thus combined the genotypes from an individual and its non-identical twin in a manner reflecting "haplotype-aware" and "haplotype-unaware" calling strategies, with the former identifying and resolving variants in the context of haplotypes (that is, distinguishing between maternal and paternal alleles) and the latter analyzing each locus either independently or with minimal local context. Marmoset cartoons were adapted from a picture taken by Eric Kilby and shared under the CC BY-SA 2.0 license; the blood vile and sequencer cartoons were taken from NIAID Visual & Medical Arts (07/10/2024). Cryo Blood Vial and Next Gen Sequencer. NIAID NIH BIOART Source: bioart.niaid.nih.gov/bioart/87 and bioart.niaid.nih.gov/bioart/386.

estimates was greatly increased under chimerism relative to the WF population, with the direction of mis-inference depending on the underlying population history. For example, the time of size change, T, was underestimated by both demographic estimators for the expanding population (supplementary fig. S1, Supplementary Material online) but overestimated in the contracting populations (supplementary figs. S2 and S3, Supplementary Material online). This likely owes to the fact that both chimerism and population growth are acting to reduce the correlation between underlying genealogies across the genome, and thus chimerism results in a mis-inference toward more recent growth, whereas chimerism and population decline are acting in opposite directions with regard to these genealogical correlations, and thus chimerism results in a mis-inference toward a more ancient decline. As a general trend, θ_W (Watterson 1975) was reduced under chimerism relative to the WF model. This pattern is likely driven by monogamous non-identical twin births reducing the effective population size, while the increase in Tajima's D (Tajima 1989) relative to the WF model is the result of fewer singletons in the chimeric genome (due to the combining of twin genomes) and thus a skew toward higher frequency alleles in the SFS (see supplementary fig. S4, Supplementary Material online). Combined with the increased variance in inferred parameters, these results suggest that standard neutral demographic estimators are not particularly well suited for inferring the population histories of chimeric species; thus, we have here developed a novel ABC approach for this purpose.

Inferring the Population History of Common Marmosets Using a Tailored ABC Approach

The first step in demographic inference is to infer the number of populations implied by the empirical data. Based on cross-validation errors (CVE) produced by ADMIXTURE (Alexander et al. 2009), a single population was found to be most likely ($k=1$, CVE = 0.57035; $k=2$, CVE = 0.74668; $k=3$, CVE = 0.97710; $k=4$, CVE = 1.21181; $k=5$, CVE = 1.28517, with k being the number of demes), as may be expected given that our samples were obtained from a single captive colony. Moreover, this result may be viewed as consistent with the current understanding of the natural geographic range of the species being relatively localized, bounded by major river systems, and without obvious fragmentation (e.g. see Figure 3 of Malukiewicz et al. 2020). We therefore inferred parameters for a single population demographic model in

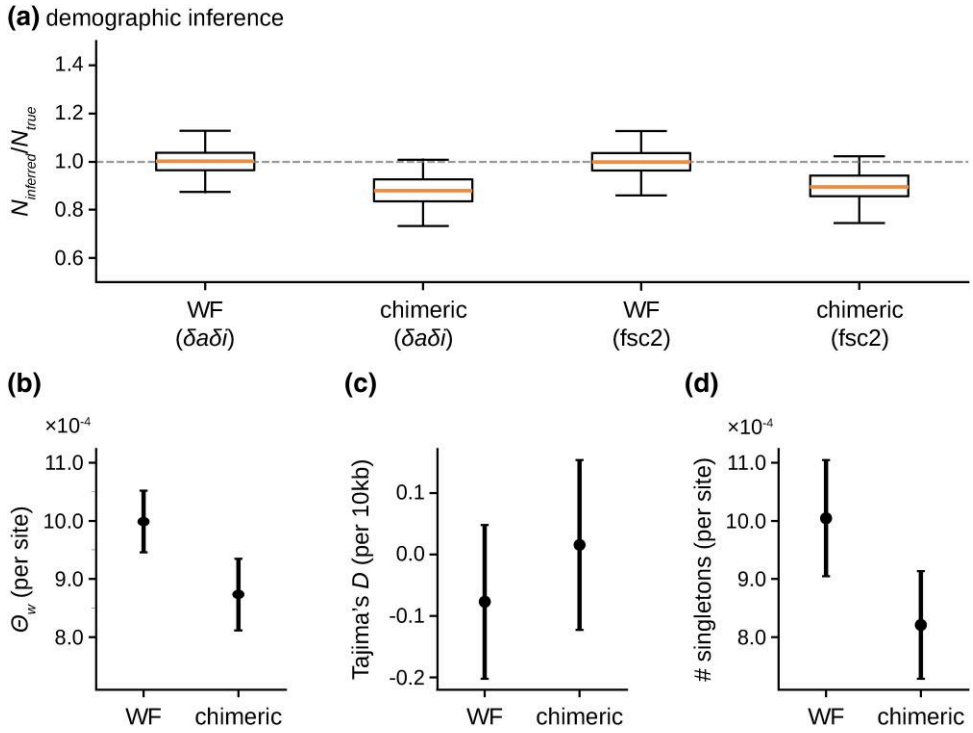

Fig. 2. Demographic inference and summary statistics for simulations of an equilibrium, neutrally evolving population, across 100 simulation replicates, comparing WF and chimeric models. a) Demographic inference results from *δaδi* (Gutenkunst et al. 2009) and fastsimcoal2 (Excoffier et al. 2013). The *y* axis represents the inferred value of the single parameter, the population size (*N*), relative to the true value of this parameter ($N_{inferred}/N_{true}$, with a value of 1 indicating that the inferred and true values are in agreement). The orange line represents the mean inference value across 100 simulation replicates, with boxes representing the 25 and 75 percentiles and whiskers representing minimum and maximum values. b–d) Summary statistics including Watterson's θ (θ_W; Watterson 1975) per site, Tajima's *D* (Tajima 1989) calculated in 10 kb windows, and the number of singletons per site. Points represent the mean values, while confidence intervals represent the variance.

which the population may potentially experience both contraction and/or growth. The model contained 4 parameters: $N_{ancestral}$ (the ancestral diploid population size), N_{change} (the proportionate instantaneous change in population size from $N_{ancestral}$), T_{change} (the time since the instantaneous size change in $N_{ancestral}$ generations), and $N_{current}$ (the population size at the time of sampling, with the size change between the instantaneous size change event and the population size at time of sampling occurring via exponential growth or decline). Parameters were initially drawn from a uniform distribution with ranges: $1{,}000 \leq N_{ancestral} \leq 50{,}000$; $0.01 \leq N_{change} \leq 2$; $0.01 \leq T_{change} \leq 5$; and $1{,}000 \leq N_{current} \leq 50{,}000$, with the upper limit of $N_{ancestral}$ and $N_{current}$ increased to 80,000 following the first round of inference based on 1,000 draws from these priors. Given these parameter ranges, this model also allowed for no population size change (i.e. $N_{change} = 1$ and $N_{current} = N_{ancestral}$), a single step-size change, and exclusive population growth or decline. Simulations for the ABC inference were performed in SLiM4 (Haller and Messer 2023), with 100 replicates for each parameter combination. Although any number of summary statistics might be used for inference with ABC, we found that θ_W, Tajima's *D*, and the number of singletons were the most informative summaries (further details of the ABC inference procedure are provided in the Materials and Methods section).

Figure 3 depicts the posterior distributions for the 4 inferred parameters, the fit of the summary statistics for simulations across 50 ABC inference runs, and the inferred demographic model itself. The simulated demographic model utilizing the point estimates was found to fit the empirical data well, with the mean across simulation replicates capturing the mode of the empirical distribution. The values for these point estimates were $N_{ancestral} = 61{,}198$, $N_{change} = 0.293$, $T_{change} = 0.0287$, and $N_{current} = 33{,}830$, suggesting that the common marmoset underwent a major reduction in population size roughly 3,500 generations ago, before recovering to roughly half of its ancestral size. Notably, these parameter estimates are based on a mean mutation rate of 0.81×10^{-8} per base pair per generation inferred in a closely related platyrrhine, owl monkeys (Thomas et al. 2018). Although a direct mutation rate estimate exists for common marmosets (0.43×10^{-8} per base pair per generation; Yang et al. 2021), this estimate was based on a single trio and the authors did not correct for the effects of chimerism. Thus, while the shape of the inferred demographic model would remain well fitting, the parameter values themselves would vary depending on the mutation rate scaling (e.g. the inferred population sizes would be close to twice as large under the mutation rate estimated by Yang et al. [2021]).

Inferring the Population History of Common Marmosets Using *δaδi*

We ran demographic inference across a range of 1D models with *δaδi* (Gutenkunst et al. 2009) in order to understand the extent of mis-inference when chimerism is unaccounted for (see Materials and Methods). The best-fitting model—based on the Akaike information criterion—was *δaδi*'s growth model, with inferred parameters of $Nu = 0.402$ and $T = 0.270$, where *Nu* is the ratio of ancient to contemporary population size and *T* is the time in the past at which growth began.

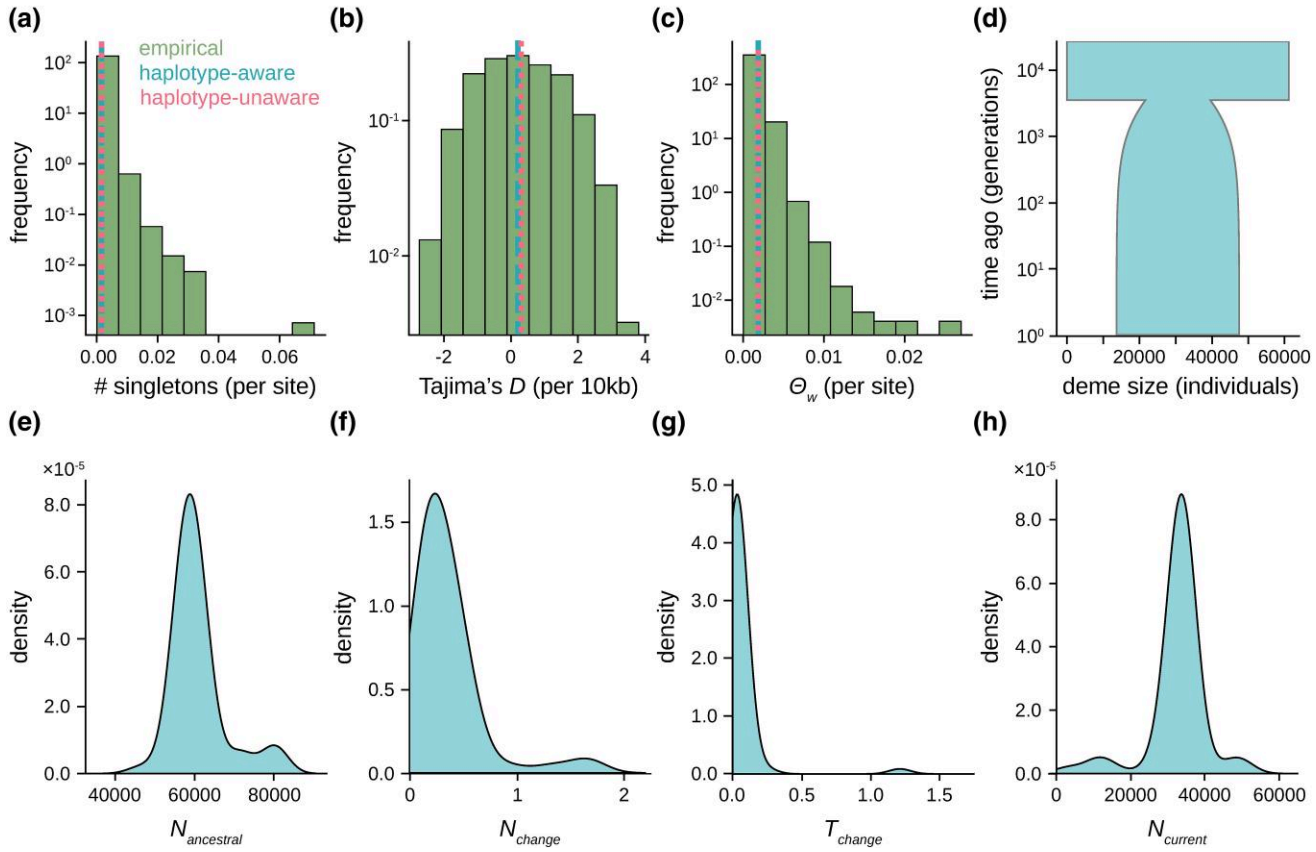

Fig. 3. Results of the ABC demographic inference on the empirical marmoset data. a–c) Distribution of empirical summary statistics (the number of singletons per site, Tajima's *D* [Tajima 1989] calculated in 10 kb windows, and Watterson's θ [θ_W; Watterson 1975] per site, shown in green) and the mean values from simulations of mean parameter values from the posterior distributions using a haplotype-aware modeling framework (blue dashed line) and a haplotype-unaware modeling framework (pink dashed line). d) Schematic of the demographic model with mean parameter values from the posterior distributions. e–h) Posterior distributions for 4 demographic model parameters: $N_{ancestral}$ (the ancestral diploid population size), N_{change} (the proportionate instantaneous change in population size from $N_{ancestral}$), T_{change} (the time since the instantaneous size change, in $N_{ancestral}$ generations), and $N_{current}$ (the population size at time of sampling). Note that the differences in *y* axis scaling owes to the differences in parameter range sizes.

Thus, these parameters represent an ancestral population size of 69,189, with the population undergoing an exponential decline that began 37,397 generations ago, reaching a current day population size of 27,832 diploid individuals (see supplementary fig. S5, Supplementary Material online for a schematic of the best-fitting *δaδi* model and supplementary fig. S6, Supplementary Material online for the fit of the SFS between the inferred model and the empirical data). Although *δaδi* inferred similar ancestral and current day population sizes as our tailored ABC approach, the demographic models are different in shape, with no bottleneck event inferred by *δaδi*. Moreover, the ABC-inferred model includes a recovery post-bottleneck—albeit to a much-reduced size relative to the ancestral population size—whereas *δaδi* inferred a declining population. Crucially, *δaδi* inferred the decline as beginning over 10-fold generations earlier than the timing of our inferred bottleneck event. Our evaluation of neutral demographic estimators (see Evaluating the Impact of Chimerism and Twinning on the Performance of Common SFS-Based Estimators of Population History) demonstrated that the time of size change tends to be inflated by chimerism under models of population contraction (supplementary figs. S2 and S3, Supplementary Material online), which likely explains the substantially older timing of the size change inferred by *δaδi* compared with the ABC model. Indeed, the posterior probability of such an old size change event is extremely low once accounting for chimerism (Fig. 3g). Notably, and as shown in supplementary fig. S6, Supplementary Material online, there is nonetheless an appropriate fit between the SFS inferred by *δaδi* and that observed in the empirical data. These results therefore suggest that neglecting chimerism may result in a potentially mis-inferred but well-fitting demographic model.

Conclusions

Although *C. jacchus* has been designated with the conservation status of least concern by the International Union for Conservation of Nature (Valença-Montenegro et al. 2021)—consistent with the relatively substantial population sizes here inferred—the species is of particular interest due to its frequent use in biomedical research (Vallender 2019) and because of the biologically peculiar phenomenon of chimerism. Assuming generation times in the common marmoset to be around 2 years (Tardif et al. 2003; Park et al. 2016; Schultz-Darken et al. 2016), the demographic model would place the inferred population collapse ~7,000 years ago. This timing corresponds to a period during the Holocene epoch when forest habitats were eroded both by the expansion of human agriculture in north-eastern Brazil and the expansion of savanna vegetation at the expense of arboreal vegetation likely owing to a less humid climate, as evidenced in the

carbon isotope record (Pessenda et al. 2004, 2010). Given that *C. jacchus* rely upon arboreal locomotion (Schiel and Souto 2017), this erosion of arboreal vegetation in that period stands as a potential hypothesis for explaining the population reduction inferred in this study. Moreover, this trend was reversed with a more humid phase that has continued to the modern day, which has been postulated to be associated with arboreal re-expansion (Pessenda et al. 2010). This re-expansion may explain the subsequent population recovery here estimated.

Overall, the well-fitting history inferred here based on a developed framework for modeling both twin births and chimerism via forward-in-time simulation in an ABC framework, together with newly generated, high-quality, whole-genome, population-level genomic data, will prove valuable in future population genomic studies in this species. Moreover, numerous other organisms in addition to callitrichids have been observed to experience varying degrees of chimerism, ranging from tunicates (Casso et al. 2019), cnidarians and sponges (Little 1966; Maier et al. 2012), to particular fungi (Peabody et al. 2000). Within other vertebrates, microchimerism has been reported in cattle (Owen 1945), dogs (Axiak-Bechtel et al. 2013), and horses (Vandeplassche et al. 1970; and see the summary of Kapsetaki et al. 2023). While the biological details of these systems will require direct and individual modeling to account for different levels of chimerism, the framework presented here as well as the justification for incorporating chimerism into evolutionary inference schemes will thus prove valuable across multiple taxa. In addition, the neutral baseline model presented here for *C. jacchus* stands as a necessary prerequisite for further inference of both selective (e.g. the detection of positive and balancing selection) and neutral (e.g. the quantification of population-level recombination maps) processes (Johri et al. 2022a). As our results have demonstrated the important impact of both twinning and chimerism in shaping observed levels and patterns of genomic variation—data summaries on which methods for inferring selective effects, for example, also depend—it is further anticipated that the simulation scheme described here will prove valuable for the construction of future evolutionary genomic analyses.

Materials and Methods

Empirical Data

Samples and Sequencing

Genomic DNA (gDNA) from 15 common marmosets (*C. jacchus*) previously housed at the New England Primate Research Center was whole-genome sequenced to a target coverage of 35× per individual (supplementary table S1, Supplementary Material online). Individuals were genetically determined to be either unrelated or related to no more than the fourth degree (supplementary table S3, Supplementary Material online). All animals were maintained in accordance with the guidelines of the Harvard Medical School Standing Committee on Animals and the Guide for Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council. In brief, blood samples were collected during routine veterinary care under approved protocols. DNA was extracted using the FlexiGene kit (Qiagen, Valencia, CA) following manufacturer protocols. Prior to sequencing, the integrity of each gDNA sample was assessed using Agarose gel electrophoresis and the purity and concentration of samples were quantified using a NanoDrop Spectrophotometer and Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), respectively. Afterward, a PCR-free library was prepared for each sample and paired-end sequenced (2 × 150 bp) on the DNBseq platform at the Beijing Genomics Institute (BGI Group, Shenzhen, China). Sample information and coverage statistics are provided in supplementary table S1, Supplementary Material online.

Read Mapping

Raw sequencing data was preprocessed using SOAPnuke v.1.5.6 (Chen et al. 2018) to remove adaptor sequences, contamination, and low-quality reads (using the following command line options: "*-n* 0.01 *-l* 20 *-q* 0.3 *-A* 0.25 *--cutAdaptor -Q* 2 *-G --polyX --minLen* 150"). Potentially remaining adapter sequences were marked using the Genome Analysis Toolkit (GATK) *MarkIlluminaAdapters* tool v.4.1.8.1 (van der Auwera and O'Connor 2020). Next, reads were mapped to the *C. jacchus* genome assembly of the Vertebrate Genomes Project (consisting of the maternal assembly for all autosomes and chromosome X and the paternal assembly for the Y chromosome; GenBank accession numbers: GCA_011078405.1 and GCA_011100535.1; Yang et al. 2021) using the Burrows–Wheeler Aligner (BWA-MEM) v.0.7.17 (Li and Durbin 2009). To improve alignments, duplicates were marked using GATK *MarkDuplicates* v.4.2.6.1 prior to variant calling.

Variant Calling, Genotyping, and Filtering

For each individual, germline variants were called from high-quality mappings ("*--minimum-mapping-quality* 40") using the GATK *HaplotypeCaller* v.4.2.6.1 in base-pair resolution mode ("*-ERC* BP_RESOLUTION") to obtain calling information at each site of the genome. Thereby, the "*--pcr-indel-model*" was set to "NONE" as a PCR-free library protocol was followed during sequencing. Individual call sets were combined using GATK's *CombineGVCFs* v.4.2.6.1 and jointly genotyped at all sites ("*-all-sites*") using GATK's *GenotypeGVCFs* v.4.2.6.1. Next, the data set was separated into autosomal, biallelic SNPs and monomorphic (i.e. invariant) sites genotyped in all individuals (AN = 30).

In the absence of a high-quality set of experimentally confirmed variants to train GATK's Variant Quality Score Recalibration algorithm, variant sites were "hard" filtered using GATK's *SelectVariants* and *VariantFiltration* tools v.4.2.6.1 following the developers' recommendations, using only those filter criteria not expected to be impacted by chimerism (i.e. QD < 2.0, SOR > 3.0, FS > 60.0, RMSMapping Quality < 40.0, MappingQualityRankSumTest < −12.5, ReadPosRankSumTest < −8.0; for details, see GATK tutorials—[How to] filter variants by hard filtering). In addition, as repetitive regions are prone to alignment errors and as extremely low or high read coverage is a frequent sign of sequencing and/or assembly issues (see discussion in Pfeifer 2017), SNPs located within repetitive elements as well as those supported by reads with less than half, or more than twice, the average individual autosomal depth of coverage were excluded from further analysis. To obtain information about the number of sites accessible to the study, invariant sites were, when applicable, subjected to the same filter criteria as variant sites. The Ts/Tv—a commonly utilized quality measurement for variants obtained from population genomic data—of the final variant data set was 2.11 (supplementary table S2, Supplementary Material online) which compares well to the

Ts/Tv reported in other primates (e.g., see Wang et al. 2015 and references therein), providing support for the high quality of the final, filtered data set.

Putatively Neutral Regions

To prevent biases in demographic inference due to positive selection, purifying selection, and/or background selection effects, variant and invariant sites were restricted to putatively neutral regions. Specifically, sites within 10 kb of exons (based on 22,355 protein-coding genes annotated in the *C. jacchus* genome; Yang et al. 2021) as well as genomic regions conserved across the primate clade (Kuderna et al. 2024) were masked, resulting in a data set of 1.4 million autosomal, biallelic SNPs with a Ts/Tv of 2.11 in the accessible genome (supplementary table S2, Supplementary Material online). The masking of 10 kb around exons has been shown both analytically and via simulation to be conservative for eliminating likely background selection effects in primate-like genomes, accounting for different DFE shapes, exon lengths, and recombination environments (Johri et al. 2020, 2023). Summary statistics were calculated across 10 kb windows with a 5 kb step size using the Python implementation of libsequence v.1.8.3 (Thornton 2003), with means and SDs of the following statistics: the number of segregating sites (S), the number of singletons, θ_W (Watterson 1975), and Tajima's D (Tajima 1989).

Evaluating the Impact of Chimerism on Demographic Inference

Simulations of Chimerism in a Non-WF Framework

The non-WF framework in SLiM v.4.3 (Haller and Messer 2023) was used to simulate chimerism in dizygotic (fraternal) twins. In each generation, a random pair of individuals was joined as a monogamous breeding pair, with each subsequent reproductive event exclusively generating pairs of non-identical twins. This step thus models the biology of twinning. At the end of each simulation, all individuals were sampled along with their pedigree IDs. For downstream inference, the genomes of chimeric twins were subsampled post-simulation and their genotypes combined in a manner reflecting "haplotype-aware" and "haplotype-unaware" calling strategies, with the former identifying and resolving variants in the context of haplotypes as implemented, for example, in the GATK HaplotypeCaller used in this study (which assembles reads into haplotypes using de Bruijn graphs) and DeepVariant used in several large-scale genomics projects such as the UK Biobank and the 1,000 Genomes Project (which uses a deep-learning model informed by haplotypes; Yun et al. 2021) and the latter analyzing each locus either independently or with minimal local context as implemented, for example, in other popular software such as SAMtools mpileup (Danecek et al. 2021). This choice of chimeric sampling was motived by the fact that in a haplotype-unaware framework, the caller focuses by and large on the allele balance (i.e. the proportion of reads supporting the reference allele and those supporting the alternative allele) at any given site, and hence, many chimeric genotypes would appear as heterozygous (Fig. 1). In contrast, in a more sophisticated haplotype-aware framework as focused upon in this study, callers distinguish between maternal and paternal alleles. Thereby, many variant callers, including GATK, are "designed to be very lenient in order to achieve a high degree of sensitivity"—in other words, GATK is inclined to call a mutation if a mismatch to the reference genome exists in the sequencing reads in order to "minimize the chance of missing real variants" (for additional details, see the GATK Best Practices Workflows—Germline Short Variant Discovery). As such, if a mutation is present in the unsampled twin, it will nonetheless be detected in the sequencing reads and thus interpreted as a variant in the sampled chimeric individual. Practically speaking, this means that sequencing reads from blood samples are being interpreted as containing genetic material from both twins (i.e. sequencing one individual is effectively the same as sequencing 2 individuals and combining their variant calls)—though it should be noted that the proportion of reads needed to support a variant in empirical data will naturally depend on several additional biological and technical factors (such as sequencing depth, per-base confidence scores, mapping qualities, sequence context, e.g. GC-content or proximity to repeats or structural variants, and hematopoietic proportion of sample source) and the statistical details of how a given calling model weighs this evidence to compute genotype probabilities. This step thus models chimeric sampling.

Demographic Model Testing

In order to evaluate the impact of chimerism on demographic inference, simulations were performed under 4 population histories: (i) population equilibrium, (ii) population expansion (instantaneous population size doubling), (iii) population contraction (instantaneous population halving), and (iv) severe population contraction (instantaneous population size reduction to $0.1N$). For each population history, 100 replicates of a 1 Mb region were simulated in a single population of 10,000 diploid individuals, with recombination rate of 1.0×10^{-8} per base pair per generation (Dumont and Payseur 2008) and a mutation rate of 2.5×10^{-8} per base pair per generation (Nachman and Crowell 2000) for testing purposes. After a burn-in of $14N_{ancestral}$ generations (where $N_{ancestral}$ is the initial population size of 10,000), a size change occurred (when applicable) and samples were taken after an additional $0.01N$ generations. Afterward, 10 chimeric and 10 non-chimeric individuals were constructed as described in Simulations of Chimerism in a Non-WF Framework.

Demographic inference was performed separately on chimeric and non-chimeric individuals using 2 commonly used demographic estimators: fastsimcoal2 (version fsc27; Excoffier et al. 2013) and *δaδi* (version 2.0.5; Gutenkunst et al. 2009). In the equilibrium model, a single population size parameter—the current population size ($N_{current}$)—was inferred using fastsimcoal2 whereas in *δaδi*, θ was first estimated from the SFS and $N_{current}$ calculated from θ. In the population size change models, simulated spectra were fitted to the instantaneous size change (growth/decline) model in fastsimcoal2, which fits 3 parameters: the ancestral population size ($N_{ancestral}$), the current population size ($N_{current}$), and the time of change (τ). The parameter search ranges for $N_{ancestral}$ and $N_{current}$ were specified to be uniformly distributed between 10 and 100,000 individuals, while the range for τ was specified to be uniform between 10 and 10,000 generations. For size change inference in *δaδi*, the *two_epoch* model was used to fit 2 parameters: the current population size relative to the ancestral population size (Nu_{opt}) and the time of change relative to the ancestral population size (τ_{opt}), with $N_{current}$ once more calculated from θ. For each simulation replicate, 15 starting values between −2 and 2 were drawn for Nu and 8 starting values between −2 and 2 were drawn for τ,

both evenly distributed in log space, creating a total of 120 different starting parameterizations.

In fastsimcoal2, inference was conducted 100 times per simulation replicate, with 100 optimization cycles per run, and 500,000 coalescent simulations to approximate the expected SFS in each cycle. The best fit was that with the smallest difference between the maximum observed and the maximum estimated likelihood. In *δaδi*, the maximum number of iterations for the optimizer was set to 100 to facilitate convergence. The best fit for each simulation replicate was that with the lowest log-likelihood score.

Inferring the Population History of Common Marmosets

As the initial step in estimating the demographic history of the common marmoset, the extent of population structure was determined by ADMIXTURE v.1.3.0 (Alexander et al. 2009), using a range of k values from 1 to 5, where k is the number of demes. The number of demes and deme assignment of individuals was based on the value of k with the lowest CVE.

Demographic Inference Using ABC on Empirical Data

Following the approach described in Simulations of Chimerism in a Non-WF Framework, 100 replicates of a 1 Mb region were simulated in SLiM v.4.3 (Haller and Messer 2023). Thereby, mutation and recombination rate heterogeneity were modeled by sampling rates from a normal distribution with a mean of 0.81×10^{-8} per base pair per generation (the mutation rate inferred in a closely-related platyrrhine, owl monkeys; Thomas et al. 2018) and 1.0×10^{-8} per base pair per generation (the recombination rate observed in humans; Kong et al. 2002), respectively, and a SD of a quarter of the mean.

For the ABC, parameters were drawn from a uniform distribution with ranges: $1{,}000 \leq N_{\text{ancestral}} \leq 80{,}000$; $0.01 \leq N_{\text{change}} \leq 2$; $0.01 \leq T_{\text{change}} \leq 5$; and $1{,}000 \leq N_{\text{current}} \leq 80{,}000$, with the upper limit of $N_{\text{ancestral}}$ and N_{current} increased to 80,000 following the first round of inference based on 1,000 draws from these priors. A further 100 draws were generated based on the posterior distribution generated using the "neural net" regression method with the default parameters provided by the R package "abc" (Csilléry et al. 2012). A 100-fold cross-validation analysis was performed in order to determine the performance and accuracy of inference for tolerance values of 0.05, 0.08, and 0.1, with a tolerance of 0.08 identified as the most accurate. This value was employed for inference of final parameter values, meaning that 8% of all simulations were accepted by the ABC to estimate the posterior probability of parameter estimates. Inference was performed 50 times, with the mean of the weighted medians of the posterior estimates taken to determine point estimates of the inferred parameters. Finally, the demographic model was simulated in SLiM v.4.3 under these parameter values and plotted against the empirical distribution (Fig. 3).

Demographic Inference Using δaδi on Empirical Data

For comparison with the ABC results, demographic inference on the empirical data was also performed using *δaδi* (Gutenkunst et al. 2009). Specifically, inference was performed on 5 1D demographic models, supplied as part of the *δaδi* package: the standard neutral model (SNM), as well as *two_epoch*, *three_epoch*, *growth*, and *bottlegrowth* models (see *δaδi* documentation for further details). Inference was performed 100 times, with 300 maximum iterations per run. Supplementary table S4, Supplementary Material online lists the parameter ranges for each demographic model. The best-fitting model was determined by calculating the Akaike information criterion (Akaike 1974), which weights the likelihood of the model by the number of model parameters.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank the members of the Jensen and Pfeifer Labs for the helpful comments and discussion. Library preparation and sequencing were performed at the Beijing Genomics Institute (BGI Group, Shenzhen, China). Computations were performed on the Sol supercomputer at Arizona State University (Jennewein et al. 2023).

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award numbers R35GM139383 to J.D.J. and R35GM151008 to S.P.P. V.S. was supported by the National Institutes of Health award number R35GM139383 to J.D.J. C.J.V. was supported by the National Science Foundation CAREER Award DEB-2045343 to S.P.P. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Conflict of Interest

None declared.

Data Availability

All sequence data have been deposited under NCBI BioProject PRJNA1215741. Scripts used for this study are available at https://github.com/vivaksoni/marmoset_chimerism_demography.

References

Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974:19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705.

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

Antunes SG, de Groot NG, Brok H, Doxiadis G, Menezes AA, Otting N, Bontrop RE. The common marmoset: a new world primate species with limited MHC class II variability. *Proc Natl Acad Sci U S A*. 1998:95(20):11745–11750. https://doi.org/10.1073/pnas.95.20.11745.

Axiak-Bechtel SM, Kumar SR, Hansen SA, Bryan JN. Y-chromosome DNA is present in the blood of female dogs suggesting the presence of fetal microchimerism. *PLoS One*. 2013:8(7):e68114. https://doi.org/10.1371/journal.pone.0068114.

Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD. Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends Genet*. 2014:30(12):540–546. https://doi.org/10.1016/j.tig.2014.09.010.

Barton NH. The effect of hitch-hiking on neutral genealogies. *Genet Res*. 1998:72(2):123–133. https://doi.org/10.1017/S0016672398003462.
Beichman AC, Huerta-Sanchez E, Lohmueller KE. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu Rev Ecol Evol Syst*. 2018:49(1):433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431.
Benirschke K, Anderson JM, Brownhill LE. Marrow chimerism in marmosets. *Science*. 1962:138(3539):513–515. https://doi.org/10.1126/science.138.3539.513.
Carlson J, Locke AE, Flickinger M, Zawistowski M, Levy S, Myers RM, Boehnke M, Kang HM, Scott LJ, Li JZ, *et al*. Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *Nat Commun*. 2018:9(1):3753. https://doi.org/10.1038/s41467-018-05936-5.
Carrion R, Patterson JL. An animal model that reflects human disease: the common marmoset (*Callithrix jacchus*). *Curr Opin Virol*. 2012:2(3):e62. https://doi.org/10.1016/j.coviro.2012.02.007.
Casso M, Tagliapietra D, Turon X, Pascual M. High fusibility and chimera prevalence in an invasive colonial ascidian. *Sci Rep*. 2019:9(1):15673. https://doi.org/10.1038/s41598-019-51950-y.
Charlesworth B, Jensen JD. Effects of selection at linked sites on patterns of genetic variability. *Annu Rev Ecol Evol Syst*. 2021:52(1):177–197. https://doi.org/10.1146/annurev-ecolsys-010621-044528.
Charlesworth B, Jensen JD. Some complexities in interpreting apparent effects of hitchhiking: a commentary on Gompert. *Mol Ecol*. 2022:31(17):4440–4443. https://doi.org/10.1111/mec.16573.
Charlesworth B, Jensen JD. Population genetics. In: *Encyclopedia of biodiversity*. Elsevier; 2024. p. 467–483. https://linkinghub.elsevier.com/retrieve/pii/B9780128225622000219.
Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993:134(4):1289–1303. https://doi.org/10.1093/genetics/134.4.1289.
Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, *et al*. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience*. 2018:7(1):1–6. https://doi.org/10.1093/gigascience/gix120.
Chiou K, Snyder-Mackler N. Marmosets contain multitudes. *eLife*. 2024:13:e97866. https://doi.org/10.7554/eLife.97866.
Csilléry K, François O, Blum MGB. Abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol*. 2012:3(3):475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x.
Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al*. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021:10(2):giab008. https://doi.org/10.1093/gigascience/giab008.
Dapper AL, Payseur BA. Effects of demographic history on the detection of recombination hotspots from linkage disequilibrium. *Mol Biol Evol*. 2018:35(2):335–353. https://doi.org/10.1093/molbev/msx272.
DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*. 2016:32(12):1895–1897. https://doi.org/10.1093/bioinformatics/btw051.
del Rosario RCH, Krienen FM, Zhang Q, Goldman M, Mello C, Lutservitz A, Ichihara K, Wysoker A, Nemesh J, Feng G, *et al*. Sibling chimerism among microglia in marmosets. *eLife*. 2024:13:RP93640. https://doi.org/10.7554/eLife.93640.
Dumont BL, Payseur BA. Evolution of the genomic rate of recombination in mammals. *Evolution*. 2008:62(2):276–294. https://doi.org/10.1111/j.1558-5646.2007.00278.x.
Durrett R, Schweinsberg J. Approximating selective sweeps. *Theor Popul Biol*. 2004:66(2):129–138. https://doi.org/10.1016/j.tpb.2004.04.002.
Eldon B, Wakeley J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*. 2006:172(4):2621–2633. https://doi.org/10.1534/genetics.105.052175.
Ewing GB, Jensen JD. Distinguishing neutral from deleterious mutations in growing populations. *Front Genet*. 2014:5:684–687. https://doi.org/10.3389/fgene.2014.00007.
Ewing GB, Jensen JD. The consequences of not accounting for background selection in demographic inference. *Mol Ecol*. 2016:25(1):135–141. https://doi.org/10.1111/mec.13390.
Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 2013:9(10):e1003905. https://doi.org/10.1371/journal.pgen.1003905.
Faulkes CG, Arruda MF, Monteiro Da Cruz MA. Matrilineal genetic structure within and among populations of the cooperatively breeding common marmoset, *Callithrix jacchus*. *Mol Ecol*. 2003:12(4):e8. https://doi.org/10.1046/j.1365-294x.2003.01809.x.
Garber PA, Caselli CB, McKenney AC, Abreu F, De la Fuente MF, Araújo A, de Fatima Arruda M, Souto A, Schiel N, Bicca-Marques JC. Trait variation and trait stability in common marmosets (*Callithrix jacchus*) inhabiting ecologically distinct habitats in northeastern Brazil. *Am J Primatol*. 2019:81(7):e23018. https://doi.org/10.1002/ajp.23018.
Gengozian N, Batson JS, Greene CT, Gosslee DG. Hemopoietic chimerism in imported and laboratory-bred marmosets. *Transplantation*. 1969:8(5):633–652. https://doi.org/10.1097/00007890-196911000-00009.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009:5(10):e1000695. https://doi.org/10.1371/journal.pgen.1000695.
Hallatschek O. Selection-like biases emerge in population models with recurrent jackpot events. *Genetics*. 2018:210(3):1053–1073. https://doi.org/10.1534/genetics.118.301516.
Haller BC, Messer PW. SLiM 4: multispecies eco-evolutionary modeling. *Am Nat*. 2023:201(5):E127–E139. https://doi.org/10.1086/723601.
Harris RA, Raveendran M, Warren W, LaDeana HW, Tomlinson C, Graves-Lindsay T, Green RE, Schmidt JK, Colwell JC, Makulec AT, *et al*. Whole genome analysis of SNV and indel polymorphism in common marmosets (*Callithrix jacchus*). *Genes (Basel)*. 2023:14(12):2185. https://doi.org/10.3390/genes14122185.
Harris RA, Tardif SD, Vinar T, Wildman DE, Rutherford JN, Rogers J, Worley KC, Aagaard KM. Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proc Natl Acad Sci U S A*. 2014:111(4):1467–1472. https://doi.org/10.1073/pnas.1316037111.
Harris RB, Jensen JD. Considering genomic scans for selection as coalescent model choice. *Genome Biol Evol*. 2020:12(6):871–877. https://doi.org/10.1093/gbe/evaa093.
Hill JP. II. Croonian lecture.—the developmental history of the primates. *Philos Trans R Soc Lond Ser B Contain Pap Biol Charact*. 1932:221(474–482):45–178. https://doi.org/10.1098/rstb.1932.0002.
Howell AA, Terbot JW, Soni V, Johri P, Jensen JD, Pfeifer SP. Developing an appropriate evolutionary baseline model for the study of human cytomegalovirus. *Genome Biol Evol*. 2023:15(4):1–12. https://doi.org/10.1093/gbe/evad059.
Huck M, Fernandez-Duque E, Babb P, Schurr T. Correlates of genetic monogamy in socially monogamous mammals: insights from Azara's owl monkeys. *Proc Biol Sci*. 2014:281(1782):20140195. https://doi.org/10.1098/rspb.2014.0195.
Irwin KK, Laurent S, Matuszewski S, Vuilleumier S, Ormond L, Shim H, Bank C, Jensen JD. On the importance of skewed offspring distributions and background selection in virus population genetics. *Heredity (Edinb)*. 2016:117(6):393–399. https://doi.org/10.1038/hdy.2016.58.
Jaquish CE, Tardif SD, Toal RL, Carson RL. Patterns of prenatal survival in the common marmoset (*Callithrix jacchus*). *J Med Primatol*. 1996:25(1):57–63. https://doi.org/10.1111/j.1600-0684.1996.tb00194.x.
Jennewein DM, Lee J, Kurtz C, Dizon W, Shaeffer I, Chapman A, Chiquete A, Burks J, Carlson A, Mason N, *et al*. The Sol supercomputer at Arizona State University. In: *Practice and experience in advanced research computing 2023: computing for the common good (PEARC'23)*. New York (NY): Association for Computing Machinery; 2023. p. 296–301.

Jensen JD. Studying population genetic processes in viruses: from drug-resistance evolution to patient infection dynamics. In: Bamford DH, Zuckerman M, editors. *Encyclopedia of virology*. Elsevier; 2021. p. 227–232. https://doi.org/10.1016/B978-0-12-814515-9.00113-2.

Jensen JD. Population genetic concerns related to the interpretation of empirical outliers and the neglect of common evolutionary processes. *Heredity (Edinb)*. 2023:130(3):109–110. https://doi.org/10.1038/s41437-022-00575-5.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*. 2005:170(3):1401–1410. https://doi.org/10.1534/genetics.104.038224.

Jensen JD, Thornton KR, Bustamante CD, Aquadro CF. On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. *Genetics*. 2007:176(4):2371–2379. https://doi.org/10.1534/genetics.106.069450.

Johri P, Aquadro CF, Beaumont M, Charlesworth B, Excoffier L, Eyre-Walker A, Keightley PD, Lynch M, McVean G, Payseur BA, *et al.* Recommendations for improving statistical inference in population genomics. *PLoS Biol*. 2022a:20(5):e3001669. https://doi.org/10.1371/journal.pbio.3001669.

Johri P, Charlesworth B, Jensen JD. Toward an evolutionarily appropriate null model: jointly inferring demography and purifying selection. *Genetics*. 2020:215(1):173–192. https://doi.org/10.1534/genetics.119.303002.

Johri P, Eyre-Walker A, Gutenkunst RN, Lohmueller KE, Jensen JD. On the prospect of achieving accurate joint estimation of selection with population history. *Genome Biol Evol*. 2022b:14(7):evac088. https://doi.org/10.1093/gbe/evac088.

Johri P, Pfeifer SP, Jensen JD. Developing an evolutionary baseline model for humans: jointly inferring purifying selection with population history. *Mol Biol Evol*. 2023:40(5):1–14. https://doi.org/10.1093/molbev/msad100.

Johri P, Riall K, Becher H, Excoffier L, Charlesworth B, Jensen JD. The impact of purifying and background selection on the inference of population history: problems and prospects. *Mol Biol Evol*. 2021:38(7):2986–3003. https://doi.org/10.1093/molbev/msab050.

Kapsetaki SE, Fortunato A, Compton Z, Rupp SM, Nour Z, Riggs-Davis S, Stephenson D, Duke EG, Boddy AM, Harrison TM, *et al.* Is chimerism associated with cancer across the tree of life? *PLoS One*. 2023:18(6):e0287901. https://doi.org/10.1371/journal.pone.0287901.

Kim Y. Allele frequency distribution under recurrent selective sweeps. *Genetics*. 2006:172(3):1967–1978. https://doi.org/10.1534/genetics.105.048447.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, *et al.* A high-resolution recombination map of the human genome. *Nat Genet*. 2002:31(3):241–247. https://doi.org/10.1038/ng917.

Kuderna LFK, Ulirsch JC, Rashid S, Ameen M, Sundaram L, Hickey G, Cox AJ, Gao H, Kumar A, Aguet F, *et al.* Identification of constrained sequence elements across 239 primate genomes. *Nature*. 2024:625(7996):735–742. https://doi.org/10.1038/s41586-023-06798-8.

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

Little FJ. Biospecific chimerid sponges from the antarctic. *Nature*. 1966:211(5047):436–438. https://doi.org/10.1038/211436a0.

Maier E, Buckenmaier A, Tollrian R, Nürnberger B. Intracolonial genetic variation in the scleractinian coral *Seriatopora hystrix*. *Coral Reefs*. 2012:31(2):505–517. https://doi.org/10.1007/s00338-011-0857-9.

Malukiewicz J, Boere V, de Oliveira MAB, D'arc M, Ferreira JVA, French J, Housman G, de Souza CI, Jerusalinsky L, R de Melo F, *et al.* An introduction to the *Callithrix* genus and overview of recent advances in marmoset research. *ILAR J*. 2020:61(2–3):110–138. https://doi.org/10.1093/ilar/ilab027.

Malukiewicz J, Boere V, Fuzessy LF, Grativol AD, French JA, de Oliveira e Silva I, Pereira LC, Ruiz-Miranda CR, Valença YM, Stone AC. Hybridization effects and genetic diversity of the common and black-tufted marmoset (*Callithrix jacchus* and *Callithrix penicillata*) mitochondrial control region. *Am J Phys Anthropol*. 2014:155(4):522e36. https://doi.org/10.1002/ajpa.22605.

Mao Y, Harvey WT, Porubsky D, Munson KM, Hoekzema K, Lewis AP, Audano PA, Rozanski A, Yang X, Zhang S, *et al.* Structurally divergent and recurrently mutated regions of primate genomes. *Cell*. 2024:187(6):1547–1562.e13. https://doi.org/10.1016/j.cell.2024.01.052.

Matuszewski S, Hildebrandt ME, Achaz G, Jensen JD. Coalescent processes with skewed offspring distributions and nonequilibrium demography. *Genetics*. 2018:208(1):323–338. https://doi.org/10.1534/genetics.117.300499.

Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res*. 1974:23(1):23–35. https://doi.org/10.1017/S0016672300014634.

Miller CT, Freiwald WA, Leopold DA, Mitchell JF, Silva AC, Wang X. Marmosets: a neuroscientific model of human social behavior. *Neuron*. 2016:90(2):219–233. https://doi.org/10.1016/j.neuron.2016.03.018.

Morales-Arce AY, Johri P, Jensen JD. Inferring the distribution of fitness effects in patient-sampled and experimental virus populations: two case studies. *Heredity (Edinb)*. 2022:128(2):79–87. https://doi.org/10.1038/s41437-021-00493-y.

Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics*. 2000:156(1):297–304. https://doi.org/10.1093/genetics/156.1.297.

Nicolaisen LE, Desai MM. Distortions in genealogies due to purifying selection. *Mol Biol Evol*. 2012:29(11):3589–3600. https://doi.org/10.1093/molbev/mss170.

Nicolaisen LE, Desai MM. Distortions in genealogies due to purifying selection and recombination. *Genetics*. 2013:195(1):221–230. https://doi.org/10.1534/genetics.113.152983.

Owen RD. Immunogenetic consequences of vascular anastomoses between bovine twins. *Science*. 1945:102(2651):400–401. https://doi.org/10.1126/science.102.2651.400.

Park JE, Zhang XF, Choi SH, Okahara J, Sasaki E, Silva AC. Generation of transgenic marmosets expressing genetically encoded calcium indicators. *Sci Rep*. 2016:6(1):34931. https://doi.org/10.1038/srep34931.

Peabody RB, Peabody DC, Sicard KM. A genetic mosaic in the fruiting stage of *Armillaria gallica*. *Fungal Genet Biol*. 2000:29(2):72–80. https://doi.org/10.1006/fgbi.2000.1187.

Pessenda LCR, de Souza Ribeiro A, Marques Gouveia SE, Aravena R, Boulet R, Albertino Bendassolli J. Vegetation dynamics during the late Pleistocene in the Barreirinhas region, Maranhão State, northeastern Brazil, based on carbon isotopes in soil organic matter. *Quat Res*. 2004:62(2):183–193. https://doi.org/10.1016/j.yqres.2004.06.003.

Pessenda LCR, Gouveia SEM, Ribeiro ADS, De Oliveira PE, Aravena R. Late Pleistocene and Holocene vegetation changes in northeastern Brazil determined from carbon isotopes and charcoal records in soils. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2010:297(3-4):597–608. https://doi.org/10.1016/j.palaeo.2010.09.008.

Pfeifer SP. From next-generation resequencing reads to a high-quality variant data set. *Heredity (Edinb)*. 2017:118(2):111–124. https://doi.org/10.1038/hdy.2016.102.

Philippens I, Langermans JAM. Preclinical marmoset model for targeting chronic inflammation as a strategy to prevent Alzheimer's disease. *Vaccines (Basel)*. 2021:9(4):388. https://doi.org/10.3390/vaccines9040388.

Poh Y-P, Domingues VS, Hoekstra HE, Jensen JD. On the prospect of identifying adaptive loci in recently bottlenecked populations. *PLoS One*. 2014:9(11):e110579. https://doi.org/10.1371/journal.pone.0110579.

Przeworski M. The signature of positive selection at randomly chosen loci. *Genetics*. 2002:160(3):1179–1189. https://doi.org/10.1093/genetics/160.3.1179.

Ragsdale AP, Weaver TD, Atkinson EG, Hoal EG, Möller M, Henn BM, Gravel S. A weakly structured stem for human origins in Africa. *Nature*. 2023:617(7962):755–763. https://doi.org/10.1038/s41586-023-06055-y.

Ross CN, French JA, Ortí G. Germ-line chimerism and paternal care in marmosets (*Callithrix kuhlii*). *Proc Natl Acad Sci U S A*. 2007:104(15):6278–6282. https://doi.org/10.1073/pnas.0607426104.

Rylands AB, Coimbra-Filho AF, Mittermeier RA. The systematics and distributions of the marmosets (*Callithrix*, *Callibella*, *Cebuella*, and *Mico*) and *Callimico* (*Callimico*) (Callitrichidae, Primates). In: Ford SM, Porter LM, Davis LC, editors. *The smallest anthropoids: the Marmoset/Callimico radiation*. New York (NY): Springer; 2009. p. 25–61.

Rylands AB, Faria DS. Habitats, feeding ecology, and home range size in the genus *Callithrix*. In: Rylands AB, editors. *Marmosets and tamarins: systematics, behaviour, and ecology*. Oxford: Oxford University Press; 1993. p. 262–272.

Sabin S, Morales-Arce AY, Pfeifer SP, Jensen JD. The impact of frequently neglected model violations on bacterial recombination rate estimation: a case study in *Mycobacterium canettii* and *Mycobacterium tuberculosis*. *G3 (Bethesda)*. 2022:12(5):jkac055. https://doi.org/10.1093/g3journal/jkac055.

Sackman AM, Harris RB, Jensen JD. Inferring demography and selection in organisms characterized by skewed offspring distributions. *Genetics*. 2019:211(3):1019–1028. https://doi.org/10.1534/genetics.118.301684.

Samuk K, Noor MAF. Gene flow biases population genetic inference of recombination rate. *G3 (Bethesda)*. 2022:12(11):jkac236. https://doi.org/10.1093/g3journal/jkac236.

Schiel N, Souto A. The common marmoset: an overview of its natural history, ecology and behavior. *Dev Neurobiol*. 2017:77(3):244–262. https://doi.org/10.1002/dneu.22458.

Schultz-Darken N, Braun KM, Emborg ME. Neurobehavioral development of common marmoset monkeys. *Dev Psychobiol*. 2016:58(2):141–158. https://doi.org/10.1002/dev.21360.

Soni V, Jensen JD. Temporal challenges in detecting balancing selection from population genomic data. *G3 (Bethesda)*. 2024:4:1–14. https://doi.org/10.1093/g3journal/jkae069.

Soni V, Jensen JD. Inferring demographic and selective histories from population genomic data using a two-step approach in species with coding-sparse genomes: an application to human data. *G3 (Bethesda)*. 2025:15(4):jkaf019. https://doi.org/10.1093/g3journal/jkaf019.

Soni V, Johri P, Jensen JD. Evaluating power to detect recurrent selective sweeps under increasingly realistic evolutionary null models. *Evolution*. 2023:77(10):2113–2127. https://doi.org/10.1093/evolut/qpad120.

Soni V, Pfeifer SP, Jensen JD. The effects of mutation and recombination rate heterogeneity on the inference of demography and the distribution of fitness effects. *Genome Biol Evol*. 2024a:16(2):evae004. https://doi.org/10.1093/gbe/evae004.

Soni V, Terbot JW, Jensen JD. Population genetic considerations regarding the interpretation of within-patient SARS-CoV-2 polymorphism data. *Nat Commun*. 2024b:15(1):3240. https://doi.org/10.1038/s41467-024-46261-4.

Soni V, Terbot JW II, Versoza CJ, Pfeifer SP, Jensen JD. A whole-genome scan for evidence of positive and balancing selection in aye-ayes (*Daubentonia madagascariensis*) utilizing a well-fit evolutionary baseline model. *G3 (Bethesda)*. 2025a:jkaf078. https://doi.org/10.1093/g3journal/jkaf078.

Soni V, Versoza C, Pfeifer SP, Jensen JD. Estimating the distribution of fitness effects in aye–ayes (*Daubentonia madagascariensis*), accounting for population history as well as mutation and recombination rate heterogeneity. *Am J Primatol*. 2025b: In press. https://doi.org/10.1101/2025.01.02.631144v1.

Soni V, Versoza C, Terbot J, Jensen JD, Pfeifer SP. Inferring fine-scale mutation and recombination rate maps in aye-ayes (*Daubentonia madagascariensis*). bioRxiv 2024.12.28.630620. https://doi.org/10.1101/2024.12.28.630620, 28 December 2024c, preprint: not peer reviewed.

Sweeney CG, Curran E, Westmoreland SV, Mansfield KG, Vallender EJ. Quantitative molecular assessment of chimerism across tissues in marmosets and tamarins. *BMC Genomics*. 2012:13(1):98. https://doi.org/10.1186/1471-2164-13-98.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989:123(3):585–595. https://doi.org/10.1093/genetics/123.3.585.

Tardif SD, Smucny DA, Abbott DH, Mansfield K, Schultz-Darken N, Yamamoto ME. Reproduction in captive common marmosets (*Callithrix jacchus*). *Comp Med*. 2003:53:364–368.

Tellier A, Lemaire C. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *Mol Ecol*. 2014:23(11):2637–2652. https://doi.org/10.1111/mec.12755.

Terbot JW, Cooper BS, Good JM, Jensen JD. A simulation framework for modeling the within-patient evolutionary dynamics of SARS-CoV-2. *Genome Biol Evol*. 2023a:15(11):1–10. https://doi.org/10.1093/gbe/evad204.

Terbot JW, Johri P, Liphardt SW, Soni V, Pfeifer SP, Cooper BS, Good JM, Jensen JD. Developing an appropriate evolutionary baseline model for the study of SARS-CoV-2 patient samples. *PLoS Pathog*. 2023b:19(4):e1011265. https://doi.org/10.1371/journal.ppat.1011265.

Terbot JW, Soni V, Versoza CJ, Pfeifer SP, Jensen JD. Inferring the demographic history of aye-ayes (*Daubentonia madagascariensis*) from high-quality, whole-genome, population-level data. *Genome Biol Evol*. 2025:17(1):evae281. https://doi.org/10.1093/gbe/evae281.

The Marmoset Genome Sequencing and Analysis Consortium. The common marmoset genome provides insight into primate biology and evolution. *Nat Genet*. 2014:46:850e7. https://doi.org/10.1038/ng.3042.

Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DST, Murali SC, Williams LE, Doddapaneni H, Muzny DM, *et al.* Reproductive longevity predicts mutation rates in primates. *Curr Biol*. 2018:28(19):3193–3197.e5. https://doi.org/10.1016/j.cub.2018.08.050.

Thornton K. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics*. 2003:19(17):2325–2327. https://doi.org/10.1093/bioinformatics/btg316.

Valença-Montenegro MM, Bezerra BM, Ruiz-Miranda CR, Pereira DG, Miranda JMD, Bicca-Marques JC, Oliveira L, da Cruz MAOM, Valle RR, Mittermeier RA. "*Callithrix jacchus*". IUCN Red List of Threatened Species. 2021. https://doi.org/10.2305%2FIUCN.UK.2021-1.RLTS.T41518A191705043.en.

Vallender EJ. The genome of the common marmoset. In: *The common marmoset in captivity and biomedical research*. Elsevier; 2019. p. 313–333. https://doi.org/10.1016/B978-0-12-811829-0.00019-4.

Vandeplassche M, Podliachouk L, Beaud R. Chimerism in horses. In: XIth European conference on animal blood groups and biochemical polymorphism. Dordrecht: Springer; 1970. p. 459–462. https://doi.org/10.1007/978-94-017-5226-8_75.

van der Auwera GA, O'Connor BD. *Genomics in the cloud: using Docker, GATK, and WDL in terra*. Sebastopol: O'Reilly Media; 2020.

Versoza CJ, Ehmke EE, Jensen JD, Pfeifer SP. Characterizing the rates and patterns of *de novo* germline mutations in the aye-aye (*Daubentonia madagascariensis*). *Mol Biol Evol*. 2025a:42(3):msaf034. https://doi.org/10.1093/molbev/msaf034.

Versoza CJ, Jensen JD, Pfeifer SP. The landscape of structural variation in aye-ayes (*Daubentonia madagascariensis*). bioRxiv 2024.11.08.622672. https://doi.org/10.1101/2024.11.08.622672, 11 November 2024, preprint: not peer reviewed.

Versoza CJ, Lloret-Villas A, Jensen JD, Pfeifer SP. A pedigree-based map of crossovers and non-crossovers in aye-ayes (*Daubentonia madagascariensis*). *Genome Biol Evol*. 2025b:17(5):evaf072. https://doi.org/10.1093/gbe/evaf072.

Wang J, Raskin L, Samuels DC, Shyr Y, Guo Y. Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics*. 2015:31(3):318–323. https://doi.org/10.1093/bioinformatics/btu668.

Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Pop Biol*. 1975:7(2):256–276. https://doi.org/10.1016/0040-5809(75)90020-9.

Windle CP, Baker HF, Ridley RM, Oerke AK, Martin RD. Unrearable litters and prenatal reduction of litter size in the common marmoset (*Callithrix jacchus*). *J Med Primatol*. 1999:28(2):73e83. https://doi.org/10.1111/j.1600-0684.1999.tb00254.x.

Wislocki GB. Observations on twinning in marmosets. *Am J Anat*. 1939:64(3):445–483. https://doi.org/10.1002/aja.1000640305.

Wu MS, Tani K, Sugiyama H, Hibino H, Izawa K, Tanabe T, Nakazaki Y, Ishii H, Ohashi J, Hohjoh H, *et al*. MHC (major histocompatibility complex)-DRB genes and polymorphisms in common marmoset. *J Mol Evol*. 2000:51(3):e22. https://doi.org/10.1007/s002390010083.

Yang C, Zhou Y, Marcus S, Formenti G, Bergeron LA, Song Z, Bi X, Bergman J, Rousselle MMC, Zhou C, *et al*. Evolutionary and biomedical insights from a marmoset diploid genome assembly. *Nature*. 2021:594(7862):227–233. https://doi.org/10.1038/s41586-021-03535-x.

Yang X, Mao Y, Wang XK, Ma DN, Xu Z, Gong N, Henning B, Zhang X, He G, Shi YY, *et al*. Population genetics of marmosets in Asian primate research centers and loci associated with epileptic risk revealed by whole-genome sequencing. *Zool Res*. 2023:44(5):837–847. https://doi.org/10.24272/j.issn.2095-8137.2022.514.

Yun T, Li H, Chang PC, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics*. 2021:36(24):5582–5589. https://doi.org/10.1093/bioinformatics/btaa1081.