the geneticssociety

www.nature.com/hdy

ARTICLE OPEN

Check for updates

Investigating the effects of chimerism on the inference of selection: quantifying genomic targets of purifying, positive, and balancing selection in common marmosets (*Callithrix jacchus*)

Vivak Soni[1], Cyril J. Versoza[1], Susanne P. Pfeifer[1,2]✉ and Jeffrey D. Jensen[1,2]✉

The common marmoset (*Callithrix jacchus*) is of considerable biomedical importance, yet there remains a need to characterize the evolutionary forces shaping empirically observed patterns of genomic variation in the species. However, two uncommon biological traits potentially prevent the use of standard population genetic approaches in this primate: a high frequency of twin births and the prevalence of hematopoietic chimerism. Here we characterize the impact of these biological features on the inference of natural selection, and directly model twinning and chimerism when performing inference of the distribution of fitness effects to characterize general selective dynamics as well as when scanning the genome for loci shaped by the action of episodic positive and balancing selection. Results suggest a generally increased degree of purifying selection relative to human populations, consistent with the larger estimated effective population size of common marmosets. Furthermore, genomic scans based on an appropriate evolutionary baseline model reveal a small number of genes related to immunity, sensory perception, and reproduction to be strong sweep candidates. Notably, two genes in the major histocompatibility complex were found to have strong evidence of being maintained by balancing selection, in agreement with observations in other primate species. Taken together, this work, presenting the first whole-genome characterization of selective dynamics in the common marmoset, thus provides important insights into the landscape of both persistent and episodic selective forces in this species.

Heredity; https://doi.org/10.1038/s41437-025-00804-7

INTRODUCTION

The common marmoset (*Callithrix jacchus*) is a platyrrhine native to Brazil (Rylands and Faria 1993; Rylands et al. 2009; Garber et al. 2019), and a prominent model in biomedical research, with a growing emphasis on aging, gene editing/therapy, neuroscience, and stem cell research in recent years (Antunes et al. 1998; Wu et al. 2000; Carrion et al. 2012; Miller et al. 2016; Philippens and Langermans 2021). Several characteristics of this species are highly beneficial for its usage as an animal model, including its small stature and high reproductive capabilities; however, unlike most non-human primates, *C. jacchus* is also characterized by a high frequency of twin-births and hematopoietic chimerism (meaning that non-germline tissue sampled from a single marmoset contains genetic material both from the individual itself as well as from their twin sibling; Hill 1932; Wislocki 1939; Benirschke et al. 1962; Gengozian et al. 1969; Ross et al. 2007; Sweeney et al. 2012; del Rosario et al. 2024)—two biological peculiarities likely to complicate the application of standard population genetic methodologies when conducting genomic studies in this primate.

Although chimerism was initially assumed to be limited to blood samples, Ross et al. (2007) described the presence of chimerism in numerous tissues. While this was initially interpreted as potentially suggesting germline chimerism as well, Sweeney et al. (2012) proposed that blood infiltration alone could result in the appearance of chimerism across tissue types. This was recently confirmed by del Rosario et al. (2024) in liver, kidney, and brain tissues; though the degrees of chimerism varied, the results continued to suggest that blood samples indeed contain the greatest contribution of sibling nuclei (see Fig. 1 of del Rosario et al. 2024, and the accompanying commentary of Chiou and Snyder-Mackler 2024). Moreover, these results also appeared to confirm that detected chimerism in alternative tissues was proportional to the degree of hematopoietic infiltration (consistent with Sweeney et al. 2012). Taken together, this body of literature does thus not support germline chimerism, but does suggest that sampling the blood of one individual will include the 'well-mixed' chimeric contribution from their sibling. Importantly, this finding also implies that it may not be possible to avoid the contributions of chimerism in genetic analyses of the species (e.g., via the use of alternative tissues; Yang et al. 2023), owing to hematopoietic infiltration. Complicating matters further, circumventing chimeric contributions via the analysis of single births is

[1]Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA. [2]These authors jointly supervised this work: Susanne P. Pfeifer, Jeffrey D. Jensen. Associate editor: Armando Caballero. ✉email: susanne@spfeiferlab.org; jeffrey.d.jensen@asu.edu

Received: 22 July 2025 Revised: 29 September 2025 Accepted: 29 September 2025
Published online: 08 October 2025

likely also compromised, as single births are not only rare in common marmosets (Ward et al. 2014) 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).

Although the first marmoset genome was published more than a decade ago (The Marmoset Genome Sequencing and Analysis Consortium 2014), population genomic inference in the species thus far has primarily been limited to characterizing genome-wide levels of genetic diversity and divergence (Faulkes et al. 2003; Malukiewicz et al. 2014; Yang et al. 2021, 2023; Harris et al. 2023; Mao et al. 2024) as well as conducting genomic scans based on patterns of genetic hitchhiking (Harris et al. 2014). More recently, Soni et al. (2025d) first investigated the effects of twinning and chimerism on neutral demographic inference, demonstrating a potentially serious mis-inference of population history if neglected, and then designed an approximate Bayesian inference approach accounting for these biological peculiarities to describe a well-fitting history of population size change for this species. Notably, however, no study to date has attempted to similarly model or quantify the effects of twinning and chimerism on the population genetic inference of selection, though the results of Soni et al. (2025d) suggest that this may indeed be important given the role of these biological features in shaping levels and patterns of genomic variation. For example, Harris et al. (2023) found that marmosets had a generally reduced relative heterozygosity compared to other primates, whilst Mao et al. (2024) found that owl monkeys—a closely related, non-chimeric platyrrhine species characterized by frequent singleton births—exhibited a considerably lower divergence to humans compared with marmosets. Yet, because chimerism was unexplored in these studies, it is unclear whether these patterns may be explained by the unusual reproductive dynamic of marmosets alone, or whether non-neutral processes need be invoked.

Characterizing general selective dynamics: the distribution of fitness effects

The distribution of fitness effects (DFE) describes the spectrum of selection coefficients associated with new mutations, and tends to be relatively stable over long timescales inasmuch as it largely describes the extent of selective constraint. Characterizing the DFE is crucial in evolutionary genomics in that it summarizes the relative proportion of strongly deleterious, weakly deleterious, and neutral mutations, together with the estimated fraction of beneficial variants, in functional genomic regions (see the reviews of Eyre-Walker and Keightley 2007; Keightley and Eyre-Walker 2010; Bank et al. 2014). Since the majority of fitness-altering mutations are deleterious, their ongoing removal through purifying selection—along with associated background selection effects (Charlesworth et al. 1993)—continuously shapes genomic diversity. Consequently, accurately quantifying these processes is essential for developing reliable evolutionary baseline models in any given species (Comeron 2014, 2017; Ewing and Jensen 2014, 2016; Johri et al. 2022b; Morales-Arce et al. 2022; Howell et al. 2023; Terbot et al. 2023; Soni and Jensen 2025).

In natural populations, DFE inference approaches utilize polymorphism and/or divergence data to infer either a continuous or discrete distribution of mutational selection coefficients. A commonly used method is the two-step approach of Keightley and Eyre-Walker (2007), in which a demographic model is inferred from synonymous sites in the first step (under the assumption that these sites are evolving neutrally), and then, based on that demographic inference, a DFE is fit to non-synonymous sites (Eyre-Walker and Keightley 2007; Boyko et al. 2008; Eyre-Walker and Keightley 2009; Schneider et al. 2011). More recently, the impact of neglecting background selection effects in such inference was investigated (Johri et al. 2021)—as was the impact of the common neglect of underlying mutation and recombination rate heterogeneity (Soni et al. 2024)—both of which were found to potentially lead to serious mis-inference. Forward-in-time simulation approaches used within an approximate Bayesian setting have thus been developed to jointly infer population history with the DFE, whilst accounting for the effects of selection at linked sites as well as fine-scale mutation and recombination rate heterogeneity (e.g., Johri et al. 2020), utilizing information from the site frequency spectrum (SFS), linkage disequilibrium (LD), as well as divergence data. However, the impact of twinning and chimerism on this inference, of the sort characterizing marmosets, has yet to be explored.

Characterizing the recent, episodic dynamics of positive and balancing selection

Existing methods for detecting the recent fixation of a beneficial mutation by positive selection rely on the expected changes in patterns of variation at linked sites owing to selective sweep dynamics (see the reviews of Stephan 2019; Charlesworth and Jensen 2021). As would be expected, the nature of these hitchhiking effects will depend on, amongst other factors, the strength and age of the beneficial mutation as well as the local recombination environment. In general, this process is associated with a reduction in local nucleotide diversity (Berry et al. 1991) and a skew in the SFS toward both high- and low-frequency derived alleles within the vicinity of the beneficial fixation (Braverman et al. 1995; Simonsen et al. 1995; Fay and Wu 2000). The theoretical expectations under this model of a single, recent selective sweep have been well described, and provide the theoretical basis for the composite likelihood ratio (CLR) test developed by Kim and Stephan (2002), on which multiple subsequent tests of positive selection have been based. Notably however, it has been demonstrated that alternative evolutionary processes (ranging from genetic drift governed by population history to neutral progeny skew) may closely replicate the patterns associated with a selective sweep (e.g., Jensen et al. 2005; Irwin et al. 2016; Charlesworth and Jensen 2022). One commonly used approach for partially addressing this problem was implemented in the SweepFinder software (Nielsen et al. 2005; DeGiorgio et al. 2016), which utilizes a null model derived from the empirically observed SFS as a way of capturing these multi-faceted contributions of alternative processes, and thereby for potentially identifying swept outliers.

Whereas completed selective sweeps reduce variation in their immediate genomic neighborhood, balancing selection maintains genetic variability in populations (see the reviews of Fijarczyk and Babik 2015; Bitarello et al. 2023), potentially over considerable timescales (Lewontin 1987). Based on the genomic signatures left at these varying timescales, Fijarczyk and Babik (2015) characterized different phases of balancing selection, ranging from recent ($<0.4N_e$ generations, where N_e is the effective population size), to intermediate ($0.4–4N_e$ generations), to ancient ($>4N_e$ generations) balancing selection. Notably, the initial trajectory of a newly introduced mutation destined to ultimately experience balancing selection is indistinguishable from that of a partial selective sweep (Soni and Jensen 2024), whereby the newly arisen mutation that has escaped stochastic loss rapidly increases to its equilibrium frequency. These genomic signatures include extended LD due to hitchhiking effects, an excess of intermediate frequency alleles, and a reduction in genetic structure (Schierup et al. 2000; and see the reviews of Crisci et al. 2013; Charlesworth and Jensen 2021). If the selected mutation reaches its equilibrium frequency, it will fluctuate about this frequency if maintained, and the initially generated LD pattern may be broken by subsequent recombination (Wiuf et al. 2004; Charlesworth 2006; Pavlidis et al. 2012). In the case of ancient balancing selection, the targeted allele may continue to segregate as a trans-species polymorphism (Klein et al. 1998; Leffler et al. 2013).

Cheng and DeGiorgio (2020) developed a class of CLR-based methods for the detection of long-term balancing selection,

released under the BalLeRMix software package. This approach involves a mixture model, combining the expectations of the SFS under neutrality and under balancing selection. Based upon this, the expected SFS shape at the putatively selected site and at increasing genomic distances from that site are evaluated. As with SweepFinder2, this class of methods utilizes a null model directly derived from the empirical SFS. Such approaches gain power as the balanced mutation segregates over longer timescales (>25N_e generations in age; Soni and Jensen 2024), as new mutations accumulate on the balanced haplotype.

Notably, however, as with DFE inference, the impact of the inherent twinning and chimerism characterizing marmosets on expected levels and patterns of genomic variation under models of both positive and balancing selection remains unexplored. Yet, these biological phenomena will represent an essential addition to any appropriate baseline model for this species (see Figs. 5 and 6 in Johri et al. 2022a for diagrams summarizing important considerations for constructing such genomic baseline models). Specifically, in addition to the processes of mutation and recombination, the underlying population history of the species, and the pervasive effects of purifying and background selection in and around functional elements, twinning and chimeric sampling represent additional processes necessary to take into consideration when performing selection scans in this species.

DFE inference and selection scans in primates

As might be expected, a number of studies have performed DFE inference in humans, beginning with Keightley and Eyre-Walker (2007), who utilized a gene set associated with severe disease or inflammatory response to fit a gamma-distributed DFE. This study estimated a relatively low (~20%) proportion of effectively neutral mutations and a large proportion (~40%) of strongly deleterious mutations. A decade later, both Huber et al. (2017) and Johri et al. (2023) inferred a substantially higher proportion of effectively neutral mutations (~50%) and a smaller proportion of strongly deleterious mutations (~20%)—differences that can likely at least partially be attributed to differences in the underlying gene sets evaluated. In non-human primates, similar inferences have largely been limited to other genera of the great apes (e.g., Castellano et al. 2019; Tataru and Bataillon 2020), with a notable focus on general regulatory regions (e.g., Simkin et al. 2014; Anderson et al. 2020; Kuderna et al. 2024). More recently, Soni et al. (2025c) performed the first DFE inference in an outgroup to the haplorrhine lineage, the aye-aye (a strepsirrhine), utilizing the annotated chromosome-level genome assembly of Versoza and Pfeifer (2024). Drawing on the two-step framework of Soni and Jensen (2025), the authors inferred a DFE by simulating under the demographic model previously inferred for the species from non-coding regions (sufficiently distant from coding regions to avoid background selection effects) by Terbot et al. (2025). This DFE was found to be characterized by a greater proportion of deleterious variants relative to humans, consistent with the larger inferred long-term effective population size of aye-ayes.

In a similar vein to DFE inference, the majority of genomic scans for positive selection in primates have focused on humans, though numerous studies have been conducted in other great apes (e.g., Enard et al. 2010; Locke et al. 2011; Prüfer et al. 2012; Scally et al. 2012; Bataillon et al. 2015; McManus et al. 2015; Cagan et al. 2016; Munch et al. 2016; Nam et al. 2017; Schmidt et al. 2019) as well as in a handful of biomedically-relevant species (e.g., The Rhesus Macaque Genome Sequencing and Analysis Consortium 2007; Pfeifer 2017b). Recently extending this inference to strepsirrhines, Soni et al. (2025a) performed genomic scans for positive and balancing selection in aye-ayes, identifying a number of olfactory-related genes with statistically significant evidence of having experienced long-term balancing selection.

In order to extend this inference to common marmosets as a widely used model system for biomedical research, we have here characterized the effects of twinning and chimerism on general selection inference and have directly modeled this biology when performing DFE inference and genomic scans, all within the context of the recently estimated population history for this species. In this way, we have robustly identified a number of candidate loci implicated in immune, reproductive, and sensory functions that have strong evidence for having experienced recent or ongoing positive and balancing selection effects.

MATERIALS AND METHODS

Animal subjects

Animals previously housed at the New England Primate Research Center 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. All samples were collected during routine veterinary care under approved protocols.

Whole genome, population-level data

We utilized previously collected blood samples from 15 common marmosets to extract DNA using the FlexiGene kit (Qiagen, Valencia, CA) following the standard protocol without any modifications. For each sample, we prepared a PCR-free library that was sequenced to a target coverage of 35× on a DNBseq platform at the Beijing Genomics Institute, resulting in 150 bp paired-end reads. We pre-processed the raw reads following standard quality-control practices (Pfeifer 2017a) using SOAPnuke v.1.5.6, trimming adapters and removing both low-quality and polyX tails (with the parameters: "*-n* 0.01 *-l* 20 *-q* 0.3 *-A* 0.25 *--cutAdaptor -Q* 2 *-G --polyX --minLen* 150"; Chen et al. 2018). We mapped the pre-processed reads with BWA-MEM v.0.7.17 (Li and Durbin 2009) to the common marmoset reference genome (mCalJa1.2.pat.X; GenBank accession number: GCA_011100555.2; Yang et al. 2021). From these read mappings, we identified and marked duplicates using Picard's *MarkDuplicates* embedded within the Genome Analysis Toolkit (GATK) v.4.2.6.1 (Van der Auwera, O'Connor 2020). Following the recommendations of the developers, we generated a gVCF for each sample using the GATK *HaplotypeCaller* (providing the "*--pcr-indel-model* NONE" flag to account for the fact that the reads originated from PCR-free libraries), combined gVCFs across samples with *CombineGVCFs*, and jointly genotyped them using *GenotypeGVCFs*. In all steps, we included both variant and invariant sites (using the "*--emit-ref-confidence* BP_RESOLUTION" and "*--include-non-variant-sites*" flags in the *HaplotypeCaller* and *GenotypeGVCFs*, respectively) genotyped in all individuals ("AN = 30"). To filter out spurious sites, we applied GATK's hard filter criteria (i.e., *FS* > 60.0; *MQ* < 40.0; *MQRankSum* < −12.5; *QD* < 2.0; *QUAL* < 30.0; *ReadPosRankSum* < −8.0; *SOR* > 3.0; with acronyms as defined by GATK) as well as an additional filter criterion based on read coverage ($0.5 \times DP_{ind} \leq DP_{ind} \leq 2.0 \times DP_{ind}$, where DP_{ind} is the average coverage observed in an individual), and removed any sites located within repetitive regions prone to mapping errors with short-read data. Lastly, we limited the dataset to autosomes (chromosomes 1–22) for downstream analyses (Supplementary Table S2).

Calculating exonic divergence

We updated the common marmoset genome included in the 447-way multiple species alignment (Zoonomia Consortium 2020) to the current reference genome available on NCBI (GenBank accession number: GCA_011100555.2; Yang et al. 2021) by first removing the included marmoset genome using the HAL v.2.2 (Hickey et al. 2013) *halRemoveGenome* function and then extracting the neighboring sequences (i.e., the genomes of the closely-related Wied's black-tufted-ear marmoset, *C. kuhlii*, and the ancestral PrimateAnc232) using *hal2fasta*. Afterward, we realigned these genomes with the current marmoset genome assembly in Cactus v.2.9.2 (Armstrong et al. 2020), preserving the original branch lengths. Finally, we reintegrated the updated subalignment using HAL's *halReplaceGenome*.

To estimate fine-scale exonic divergence, we used HAL's *halSummarizeMutations* function to identify fixed differences in exons along the marmoset lineage (i.e., between the genomes of the common marmoset and the ancestral PrimateAnc232). To mitigate the effects of fragmentation present in the *C. kuhlii* genome—which, unlike the common marmoset genome, is currently at the scaffold-level—we kept only those exons found in alignments longer than 10 kb in length. Lastly, we masked any point

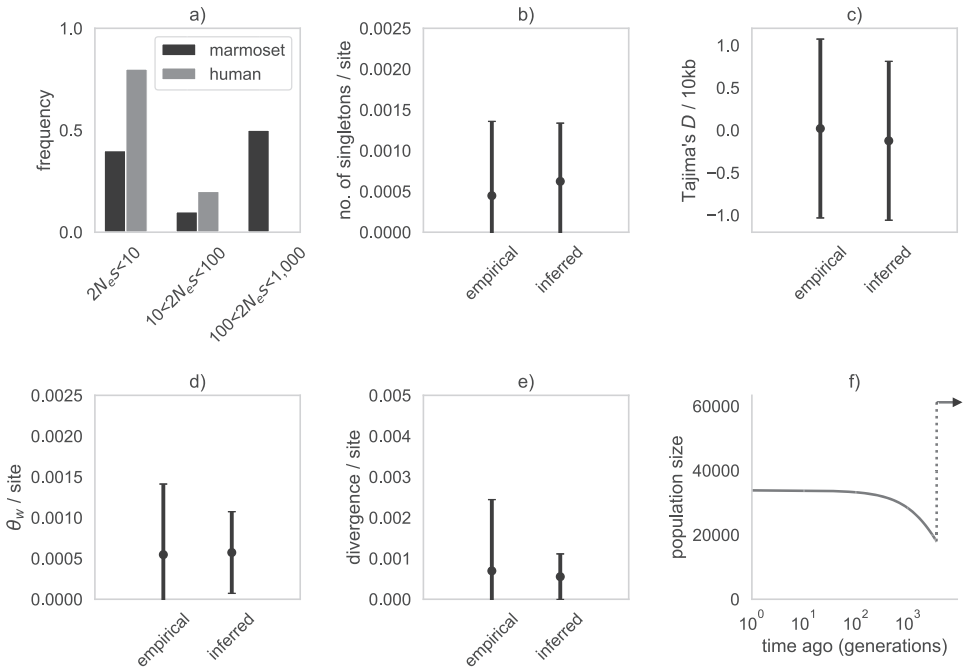

Fig. 1 DFE model-fitting results and summaries. a Comparison between the best-fitting discrete DFE in the common marmoset inferred in this study (shown in black) and the DFE in humans inferred by Johri et al. (2023) (gray). Exonic mutations were drawn from a DFE comprised of three fixed classes: nearly neutral mutations (i.e., $2N_{\text{ancestral}}\ s < 10$), weakly/moderately deleterious mutations ($10 \le 2N_{\text{ancestral}}\ s < 100$), and strongly deleterious mutations ($100 \le 2N_{\text{ancestral}}\ s$). **b**–**e** Comparison of four summary statistics—the number of singletons per site, Tajima's *D* (Tajima 1989) per 10 kb, Watterson's θ_w (Watterson 1975) per site, and exonic divergence per site—between the empirical and simulated data under the inferred DFE. Data points represent the mean value, whilst confidence intervals represent standard deviations. **f** The demographic history previously inferred for this population (re-drawn from Soni et al. 2025d).

mutations known to be polymorphic in the species, and calculated the rate of exonic divergence by dividing the number of substitutions by the number of accessible sites in each exon.

DFE inference

In order to fit a DFE to the coding regions of the common marmoset genome, we used forward-in-time simulations in SLiM v.4.23 (Haller and Messer 2023) to simulate 100 exonic regions, each corresponding to the mean length of exons greater than 1 kb observed in the empirical data (i.e., 3209 bp). More specifically, we performed simulations under the recently published demographic model for the species using the modeling framework of Soni et al. (2025d) to account for both twin-births and chimeric sampling. In brief, monogamous mating pairs produced non-identical twin offspring, and the genotypes of these twins were combined post-simulation to create a chimeric individual (see Fig. 1 of Soni et al. 2025d). The model included a $10N_{ancestral}$ generation burn-in time prior to the start of the demographic model (where $N_{\text{ancestral}}$ is the initial population size of 61,198 individuals). We assumed a branch split time of 0.82 million years (Malukiewicz et al. 2021) and a generation time of 1.5 years (Okano et al. 2012; Han et al. 2022), and randomly drew mutation and recombination rates for each replicate from a normal distribution, such that the mean rates across all 100 simulation replicates were equal to the mean rates inferred in closely related primates (using a mean mutation rate of 0.81×10^{-8} per base pair per generation and a mean recombination rate of 1 cM/Mb, as per Soni et al. 2025d). Following Johri et al. (2020), we drew exonic mutations from a DFE comprised of three fixed classes: nearly neutral mutations ($2N_{\text{ancestral}}\ s < 10$, where $N_{ancestral}$ is the ancestral population size and *s* is the reduction in fitness of the mutant homozygote relative to wild-type), weakly/moderately deleterious mutations ($10 \le 2N_{\text{ancestral}}\ s < 100$), and strongly deleterious mutations ($100 \le 2N_{\text{ancestral}}\ s$).

To infer the DFE, we then performed a grid search by varying the fractions of mutations obtained from each category. Specifically, for each parameter combination, we simulated 100 replicates using the DFE of human functional regions as a starting point (see Johri et al. 2023 for details), and compared the fit of four summary statistics between our empirical and simulated data: Watterson's θ_w (Watterson 1975) per site, Tajima's *D* (Tajima 1989) per 10 kb, the number of singletons per site, and exonic divergence per site. We used pylibseq v.1.8.3 (Thornton 2003) to calculate Watterson's θ_w, Tajima's D, and the number of singletons across 10 kb windows with a 5 kb step size, whereas exonic divergence was calculated as the number of fixations that occurred in our simulated population post-burn-in to enable a direct comparison to the empirically observed number of substitutions along the marmoset branch. Specifically, based upon the fit of a given DFE to the data, a manual grid step was performed to produce an improved fit, until a DFE consistent with the empirical data was identified (as assessed by visual fit of the mean and variance for each of the chosen summary statistics, as illustrated in Fig. 1b–e).

Evaluating the effects of chimerism on genome scans

In order to evaluate the effects of chimerism on genome scans for positive and balancing selection, we simulated 100 replicates of a single marmoset population using SLiM v.4.3 (Haller and Messer 2023). To mimic the genomic architecture of the species, we simulated three functional regions —comprised of nine 130 bp exons separated by 1591 bp introns—isolated by 16,489 bp intergenic regions, for a total region length of 91,161 bp in each replicate. We modeled mutations in intronic and intergenic regions as neutral, and drew exonic mutations from a DFE comprised of four fixed classes with frequencies denoted by f_i: effectively neutral mutations f_0 (i.e., $0 \le 2N_{ancestral}s < 1$), weakly deleterious mutations f_1 ($1 \le 2N_e s < 10$), moderately deleterious mutations f_2 ($10 \le 2N_{ancestral}s < 100$), and strongly deleterious mutations f_3 with ($100 \le 2N_{ancestral}s < 2N_e$), with *s* drawn from a uniform distribution within each bin. We implemented two modeling schemes: an equilibrium population history and the inferred marmoset demographic model of Soni et al. (2025d). For the equilibrium model, we simulated under the twin-birth, chimeric-sampling non-WF framework outlined in Soni et al. (2025d) and a WF model for comparison, both using a $10N_{ancestral}$ burn-in time (where $N_{ancestral} = 10{,}000$). For the marmoset demographic model, we simulated under the twin-birth, chimeric-sampling non-WF framework only, again with a $10N_{ancestral}$ burn-in time (where $N_{ancestral} = 61{,}898$). In all cases, we sampled 15 individuals at the end of each simulation. In each simulation replicate, a single positively selected mutation was introduced. For the selective sweep analysis, we simulated two beneficial population-scaled selection coefficients—$2N_e s = [100, 1000]$—and sampled the population at $\tau = [0.1, 0.2, 0.5]$, where τ is the time since fixation of the beneficial mutation in *N* generations in the

equilibrium model; for the demographic model simulations, we introduced beneficial mutations at times [0.2, 0.5, 1, 2] $N_{ancestral}$ generations prior to the end of the simulation. For the balancing selection analysis, we introduced the balanced mutation at $\tau_b = [10, 50, 75]$, where τ_b is the time since the introduction of the balanced mutation in N generations. We modeled the balanced mutation to experience negative frequency-dependent selection, i.e., in a manner such that the selection coefficient was dependent on the frequency in the population: $S_{bp} = F_{eq} - F_{bp}$, where S_{bp} is the selection coefficient of the balanced mutation, F_{eq} is the equilibrium frequency of the balanced mutation (here set to 0.5), and F_{bp} is the frequency of the balanced mutation in the population. We implemented all simulations such that if the selective sweep failed to fix, or if the balanced mutation was fixed or lost from the population, the simulation would restart at the point of introduction of the selected mutation.

With these simulations on hand, we then performed selective sweep inference using SweepFinder2 v.1.0 (DeGiorgio et al. 2016) to detect signals of positive selection (using the command: "SweepFinder2 *–lu* GridFile FreqFile SpectFile OutFile") and B_{0MAF} (Cheng and DeGiorgio 2020) to detect signals of balancing selection (using the command: "*python3 BalLeRMix+_v1.py -I* FreqFile *--spect* SfsFile *-o* OutFile *–noSub –MAF –rec* 1e-8"), with analyses limited to folded allele frequencies and polymorphic sites only, performing inference at each SNP. Under the demographic model, balancing selection inference was performed in 1 kb windows with a 50 bp step size, to mimic the conditions of the empirical inference performed in this study.

Generating null thresholds for selection inference

In order to generate null thresholds for the inference of selection, we simulated the demographic model of the species (based on the twin-birth, chimeric-sampling, non-WF framework outlined in Soni et al. 2025d), consisting of a relatively recent population bottleneck followed by a recovery via exponential growth to roughly half the ancestral size (see Fig. 1f). In brief, based on this demographic history, we simulated 10 replicates for each of the 22 autosomes in the marmoset genome in SLiM v.4.3 (Haller and Messer 2023), simulating only neutral regions that were at least 10 kb from the nearest coding region, in order to avoid the biasing effects of purifying and background selection. Following Soni et al. (2025d), we assumed a mutation rate of 0.81×10^{-8} per base pair per generation and a recombination rate of 1 cM/Mb. To reduce false positive rates (FPR), we conservatively used the maximum CLR value observed across all null model simulations (i.e., the highest value observed in the absence of positive or balancing selection) as the null threshold for the selection scans with SweepFinder2 v.1.0 (DeGiorgio et al. 2016) and B_{0MAF} (Cheng and DeGiorgio 2020) on the empirical data.

Inferring recent positive and balancing selection in the marmoset genome

With the null thresholds, we applied the inference schema discussed above with SweepFinder2 and B_{0MAF} on the empirical autosomal data. We performed sweep inference at each SNP and balancing selection inference across 1 kb windows, with a step size of 50 bp. We manually curated our candidate loci by identifying genes under the significant likelihood surface, and evaluated the obtained dataset via the NCBI database (Sayers et al. 2022) and Expression Atlas (Madeira et al. 2022) in order to identify function and expression patterns in different primate species. Additionally, we used the Database for Annotation, Visualization, and Integrated Discovery (DAVID; Ma et al. 2023) to perform a gene ontology analysis (The Gene Ontology Consortium et al. 2023).

Identifying synonymous and non-synonymous mutations

Extensive genome fragmentation can impact comparative genomic analyses. As many primate genome assemblies present in the 447-way multiple species alignment are currently still at the scaffold (rather than chromosome) level, we limited our analyses to a subset of high-quality genomes representing different branches along the primate clade, with humans (hg38; Schneider et al. 2017) as a representative for the haplorrhines, common marmosets (mCalJa1.2.pat.X; Yang et al. 2021) as a representative for the platyrrhines, and aye-ayes (DMad_hybrid; Versoza and Pfeifer 2024) as a representative for strepsirrhines. Additionally, we further curated the alignments to only include exonic regions of the genes identified by SweepFinder2 ($n = 35$) using the genome annotation of the common marmoset assembly (Yang et al. 2021) and the *hal2maf* function implemented in Cactus v.2.9.3 (Armstrong et al. 2020). Using a custom script, we identified fixed differences between marmosets and both humans and aye-ayes (i.e., positions at which both humans and aye-ayes carry the same nucleotide but marmosets carry a different nucleotide), removing any sites known to segregate in the species. We then utilized SnpEff v.5.2.1 (Cingolani et al. 2012) to identify synonymous (i.e., *SYNONYMOUS_CODING*, *SYNONYMOUS_START*, and *SYNONYMOUS_STOP*) and non-synonymous (i.e., *NON_SYNONYMOUS_CODING*, *NON_SYNONYMOUS_START*, and *NON_SYNONYMOUS_STOP*) mutations.

Identifying highly divergent genes

We calculated mean divergence per gene, and performed gene functional analysis using DAVID (Ma et al. 2023) on the subset of genes that overlapped between sweep candidates, and those with a mean divergence value greater than the 75th, 95th, and 99th percentiles of neutral divergence.

RESULTS AND DISCUSSION

Utilizing patterns of exonic divergence to infer the DFE in common marmosets

To quantify fine-scale exonic divergence, we first updated the common marmoset genome included in the 447-way mammalian multiple species alignment (Zoonomia Consortium 2020) to the current reference genome available on NCBI (Yang et al. 2021), before retrieving substitutions along the marmoset branch. Examining the dataset, we found that only a small number of exons exhibited rates of exonic fixation higher than the maximum neutral divergence observed in non-coding regions (Soni et al. 2025b) in 1 Mb windows, whereas no exons exceeded the maximum neutral divergence observed in 1 kb windows (Supplementary Fig. S1)—an anticipated observation given the pervasiveness of purifying selection in functional regions (Charlesworth et al. 1993).

The DFE is expected to remain relatively stable over long evolutionary timescales, and divergence is therefore an informative summary statistic when performing inference of long-term patterns of selection. We utilized forward-in-time simulations in SLiM (Haller and Messer 2023) under the recently published demographic model for the population (re-drawn in Fig. 1f), and using the modeling framework of Soni et al. (2025d) to account for both twin-births and chimeric sampling (see "Materials and methods"), to fit the observed fine-scale patterns of exonic divergence as well as summaries of both genetic variation and the SFS (the number of singletons, Watterson's θ_w (Watterson 1975), and Tajima's D (Tajima 1989)) with a DFE shape consisting of neutral, weakly/moderately deleterious, and strongly deleterious mutational classes. We evaluated this DFE under a divergence time of 0.82 million years (Malukiewicz et al. 2021) between the common marmoset and the closely related Wied's black-tufted-ear marmoset (*C. kuhlii*), assuming a generation time of 1.5 years (Okano et al. 2012; Han et al. 2022), by comparing the data summaries between simulated and empirical exonic data in order to quantify a well-fitting DFE. As depicted in Fig. 1, a DFE of new mutations characterized by a large proportion of strongly deleterious and nearly neutral variants, and a small proportion of weakly/moderately deleterious mutations, fit the summary statistics observed in common marmosets well. A recent estimate of the DFE from human populations by Johri et al. (2023) has been included for comparison—as it relied upon the same inference approach used here—which was characterized by a higher density of neutral variants and a considerably lower density of strongly deleterious variants relative to the common marmoset. These patterns are consistent with the much larger long-term effective population size in common marmosets (Soni et al. 2025d), resulting in an increased efficacy of purifying selection (as the strength of selection experienced by an individual mutation is the product of the effective population size, N_e, and the selection coefficient, s).

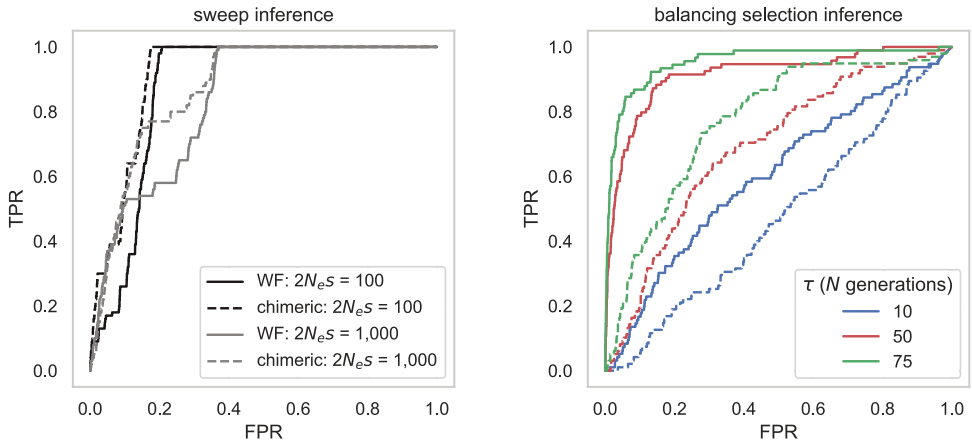

Fig. 2 ROC plots based on 100 simulated replicates of a constant population size model for selective sweep (left) and balancing selection (right) inference using the SweepFinder2 and the B_{0MAF} method, respectively. Solid lines represent standard Wright-Fisher (WF) simulations, whilst dashed lines represent the twin-birth, chimeric-sampling non-WF framework outlined in Soni et al. (2025d), with the false-positive rate (FPR) and the true-positive rate (TPR) provided on the *x*-axis and *y*-axis, respectively. For selective sweep inference, power analyses were conducted across two selection regimes—population-scaled strengths of selection of $2N_eS = 100$ and 1000—with each simulation terminating at the point of fixation of the beneficial mutation. For balancing selection inference, the simulation was terminated at three values of τ (the time since the introduction of the balanced mutation): 10*N*, 50*N*, and 75*N* generations, where $N = 10{,}000$ (for details of simulation and inference schema, see "Materials and methods").

Evaluating the effects of twinning and chimerism on the inference of recent positive and balancing selection

Prior to performing genomic scans for selection, simulations were utilized to assess whether twinning and chimerism might impact the power and false positive rates when employing SFS-based methods for the inference of selective sweeps and balancing selection. Specifically, we performed forward-in-time simulations under an equilibrium population history for both the twin-birth, chimeric-sampling framework outlined in Soni et al. (2025d) and a standard Wright-Fisher (WF) model for comparison, in order to assess the statistical performance of the genomic scan approaches implemented in SweepFinder2 (DeGiorgio et al. 2016) and B_{0MAF} (Cheng and DeGiorgio 2020). Figure 2 provides receiver operating characteristic (ROC) plots across 100 simulated replicates, for two different population-scaled strengths of selection ($2N_eS = 100$ and 1000 for sweep inference), and three different times since the introduction of the mutation experiencing negative frequency-dependent selection for balancing selection inference. Given previous results demonstrating that SFS-based methods have little power to detect balanced variants segregating for $<25\,N$ generations (Soni and Jensen 2024), we evaluated introduction times of $\tau_b = 25\,N$, $50\,N$ and $75\,N$ generations prior to sampling, where the population size $N = 10{,}000$. For selective sweep simulations, only replicates in which the beneficial mutation fixed were considered; similarly, for balancing selection simulations, only replicates in which the balanced allele was segregating at the time of sampling were considered. Though twinning and chimerism did not appear to greatly affect sweep inference power, the power to detect balancing selection was notably reduced (Fig. 2). More specifically, though the effect of twinning individually can be viewed as an increase in relatedness in the population and a reduction in effective population size, the effect of chimeric sampling has been shown to skew the SFS toward intermediate frequency alleles (see Supplementary Figs. S4 and S6 in Soni et al. 2025d). Thus, taken together, this likely interferes with the detection of balancing selection owing to the similar signal being produced, whilst the expected selective sweep pattern remains rather distinctive in comparison.

Next, we constructed a baseline model consisting of the estimated demographic history for this population (characterized by a population size reduction followed by recovery), twin-births and chimeric sampling (Soni et al. 2025d), the DFE in coding regions capturing purifying and background selection effects (as estimated in this study), as well as mutation and recombination rate heterogeneity by randomly drawing both rates from normal distributions for each 1 kb window of our simulated data (for full simulation details, see "Materials and methods"). For individual selective sweeps, we assessed population-scaled strengths of selection of $2N_eS = 100$ and 1000 at four different times since the introduction of the beneficial mutation ($\tau = 0.2$, 0.5, 1, and 2, scaled in *N* generations). As shown in Fig. 3, the power to detect selective sweeps is naturally dependent on the strength of selection acting on the beneficial mutation, the time since the introduction of the beneficial mutation, and the window size evaluated. The marmoset demographic model consists of a nearly 70% reduction in population size followed by a partial recovery to roughly half of its ancestral size. Notably, although population bottlenecks can replicate patterns of variation associated with selective sweeps (e.g., Barton 1998; Poh et al. 2014; Harris and Jensen 2020), this reduction in marmosets is not sufficiently severe to strongly reduce power or increase FPRs, and thus considerable power remains for sweep detection in our study. The pattern of slightly improved performance for weaker selection coefficients observed in Fig. 3 owes to the fact that the time since introduction rather than the time since fixation is here modeled; that is, stronger selection coefficients are characterized by older fixation times owing to their rapid sojourn time, thus resulting in a subtle loss in power. As illustrated in Fig. 4, the detection of balancing selection is more sensitive to these factors given the ability of both twinning with chimerism, as well as moderate population bottlenecks, to generate an excess of intermediate frequency alleles. Furthermore, and consistent with previous work (Soni and Jensen 2024), the power to detect balancing selection depends strongly on the time since the introduction of the balanced mutation.

Based upon these results, and in order to robustly identify candidate loci experiencing recent positive and/or balancing selection within the context of this marmoset biology and population history, we performed genome-wide selection scans using the CLR methods SweepFinder2 (DeGiorgio et al. 2016) to detect signals of positive selection (performing sweep inference at each SNP individually) and B_{0MAF} (Cheng and DeGiorgio 2020) to detect signals of balancing selection (performing sweep inference

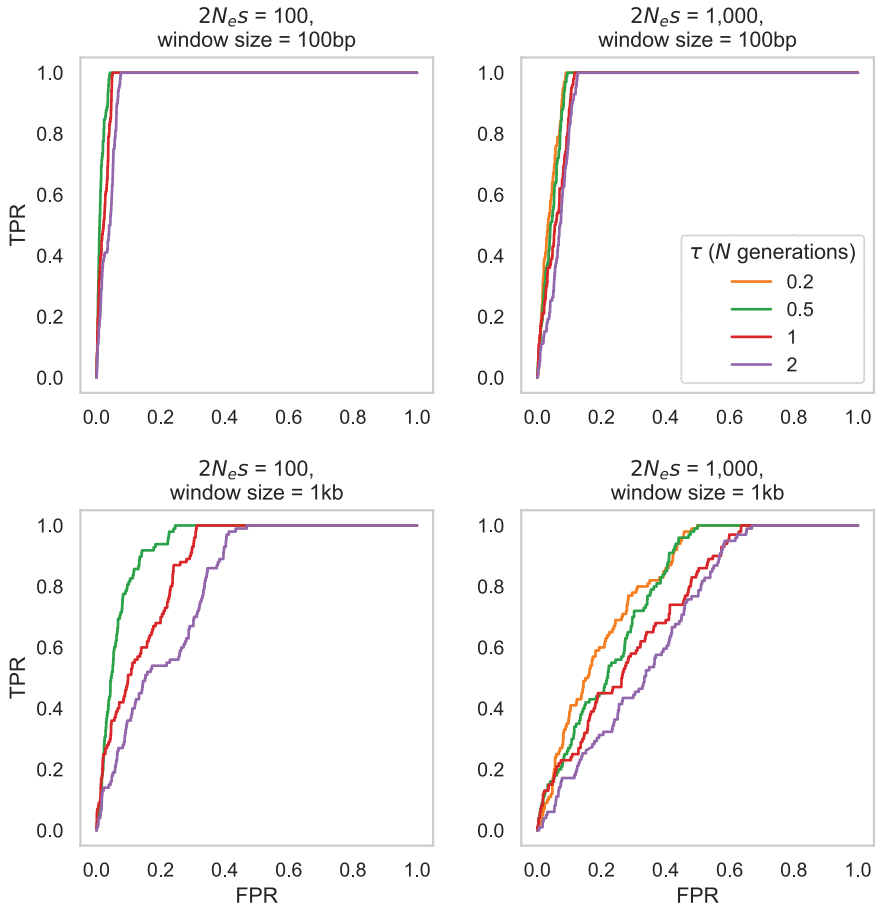

Fig. 3 ROC plots based on 100 simulated replicates of the common marmoset demographic model inferred by Soni et al. (2025d) for selective sweep inference using SweepFinder2, with mutation and recombination rates drawn from a normal distribution such that the mean rate per replicate is equal to the fixed rate (see "Materials and methods" for details). The false-positive rate (FPR) and the true-positive rate (TPR) are provided on the *x*-axis and *y*-axis of the ROC plots, respectively. Power analyses were conducted across two selection regimes—population-scaled strengths of selection of $2N_eS = 100$ and 1000—four times of introduction of the beneficial mutation ($\tau = 0.2N$, $0.5N$, $1N$, $2N$; where $N = 61{,}898$), and two window sizes (100 bp and 1 kb). Note that no ROC could be plotted for simulation schemata in which none of the beneficial mutations reached fixation at the time of sampling.

in 1 kb windows with a 50 bp step size). Although outlier approaches—i.e., approaches that assume that genes in the chosen tail of the distribution (often 5% or 1%) are likely sweep candidates, regardless of the underlying evolutionary model (Harris et al. 2018, and see the discussions in Howell et al. 2023; Jensen 2023; Johri et al. 2023; Terbot et al. 2023)—are commonly used to identify candidate regions experiencing positive selection, such approaches have been shown to be associated with high FPRs (Teshima et al. 2006; Thornton and Jensen 2007; Jensen et al. 2008; Jensen 2023; Soni et al. 2023). We therefore instead constructed an evolutionarily appropriate baseline model accounting for commonly operating evolutionary processes, as recommended by Johri et al. (2022a, b). Based on the maximum CLR values observed under neutrality across simulation replicates of the inferred marmoset population history with twinning and chimerism, we set the null thresholds for positive and balancing selection inference as 82.16 for SweepFinder2 and 87.57 for B_{0MAF}, respectively (see "Materials and methods" for details), and identified candidate regions as any empirical CLR values in excess of these thresholds. For orientation, a naive outlier approach would have resulted in 174,035 (at 5%) or 34,807 (at 1%) detections for sweep scans, while our baseline model approach represented 0.3% of loci being sweep candidates. For balancing selection scans, 3481 (at 5%) and 696 (at 1%) candidate windows would have been detected, while our baseline approach represented 0.02% of windows.

Signatures of positive selection in the common marmoset genome

Utilizing the thresholds from the evolutionary baseline model, we identified 10,599 loci with statistically significant evidence of having experienced selective sweep effects, mapping to 216 genes within the marmoset genome (see Fig. 5 for the genome-wide scan results for SweepFinder2, Supplementary Figs. S2–S23 for scan results along individual chromosomes, and Supplementary File S1 for candidate gene regions with CLR values greater than the null threshold). Performing a gene functional analysis of these 216 candidates using the DAVID (Ma et al. 2023), we identified several enriched categories related to cellular components, sequence features, and gene regulation (Supplementary Table S1). Additionally, we compared the identified highly divergent genes on the *C. jacchus* branch in the multi-species alignment with these sweep candidates (see Supplementary Figs. S2–S23 for per-chromosome exonic divergence overlayed with the genomic scan results). These analyses highlighted a number of genes and categories of interest:

(1) Immune-related genes are among the most rapidly evolving across vertebrates, due to the evolutionary arms race in response to pathogen exposure (e.g., George et al. 2011; Rausell and Telenti 2014; Soni et al. 2025c). Three tumor-related genes were identified as candidates for recent positive selection in the peaks of the likelihood surface.

Specifically, loci in the gene *EPHB1* had the highest CLR value in our genome scans—a gene that plays a critical role in immune cell development and function, and that has been implicated in several nervous system diseases, cardiovascular diseases, and cancers (Xie et al. 2024). The tumor suppressor *RPH3AL* was both a candidate for recent positive selection and exhibited an unusually high rate of fixations. The *BRAF* gene regulating cell growth is a proto-oncogene in which strongly activating mutations have been argued to experience selective sweep dynamics in humans (Gopal et al. 2019). Additionally, *TSPAN2* and *TSBP1* exhibited a rate of divergence higher than the 99th percentile neutral divergence (as estimated by Soni et al. 2025b). *TSPAN2* plays a role in cell motility and has been implicated in both nervous system development and cancer progression (Otsubo et al. 2014, and see the review of Yaseen et al. 2017). The most highly diverged gene identified in our dataset, *TSBP1*, is located in the major histocompatibility complex (MHC). Hoh et al. (2020) identified a genomic region encompassing *TSBP1* as undergoing positive selection in native human populations in North Borneo, hypothesizing that the *Plasmodium* parasite endemic to the region was the driver of local adaptation. Given that the common marmoset is one of numerous New World monkeys that are targets of infection by *P. brasilianum* (Alvarenga et al. 2017), this parasite may similarly be driving these dynamics.

(2) The candidate gene *CDC14B* is also noteworthy in that its retrogene *CDC14B2* originated by retroduplication in the hominoid ancestor ~18–25 million years ago (Marques et al. 2005), with evidence of experiencing positive selection in African apes ~7–12 million years ago (Rosso et al. 2008). In mice, *CDC14B2*-deficient cells have been shown to accumulate more endogenous DNA damage than wild-type cells, consistent with premature aging (Wei et al. 2011)—a result of particular interest given that marmosets remain as a widely-used model organism for the study of neurodegeneration and aging (Perez-Cruz and Rodriguez-Callejas 2023).

(3) Given that many of the unique aspects of marmoset biology revolve around reproduction, identifying genes involved in ovulation and gestation is of considerable biomedical interest. Three genes linked to reproductive function were observed to fall in the high-divergence gene set. *SPACA7*, for example, plays a vital role in spermatogenesis (Aisha and Yenugu 2023), and has been argued to be experiencing positive selection across the primate clade (van der Lee et al. 2017); additionally, there is evidence that *ZP2*, involved in

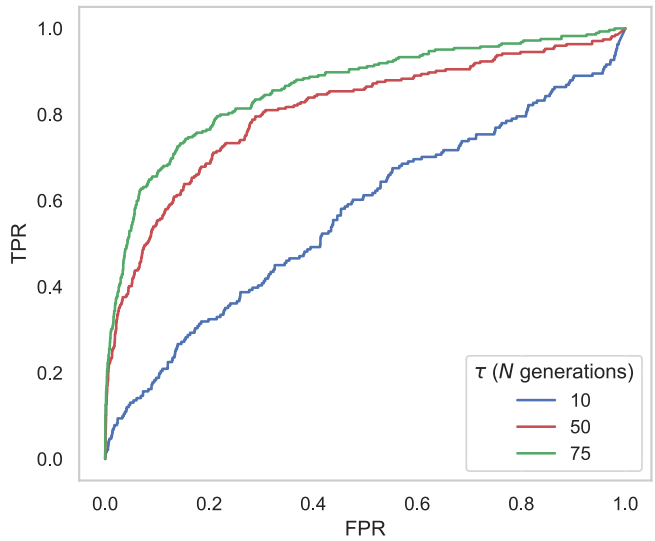

Fig. 4 ROC plots based on 100 simulated replicates of the common marmoset demographic model inferred by Soni et al. (2025d) for balancing selection inference using B_{0MAF}, with mutation and recombination rates drawn from a uniform distribution such that the mean rate per replicate is equal to the fixed rate (see "Materials and methods" for details). The false-positive rate (FPR) and the true-positive rate (TPR) are provided on the *x*-axis and *y*-axis of the ROC plots, respectively. Power analyses were conducted across three times of introduction of the balanced mutation ($\tau = 10N$, $50N$, and $75N$ generations; where $N = 61{,}898$).

Fig. 5 Genome-wide selective sweep scan results using SweepFinder2 (shown in blue) and empirical exonic divergence (red). The *x*-axis shows the position along each autosome (chromosomes 1–22), the left *y*-axis shows the composite likelihood ratio (CLR) value of the sweep statistic at each SNP, and the right *y-axis* shows the mean gene divergence at each SNP. The horizontal blue dashed line represents the null threshold for sweep detection, and the horizontal red dashed line represents the 75th percentile neutral divergence.

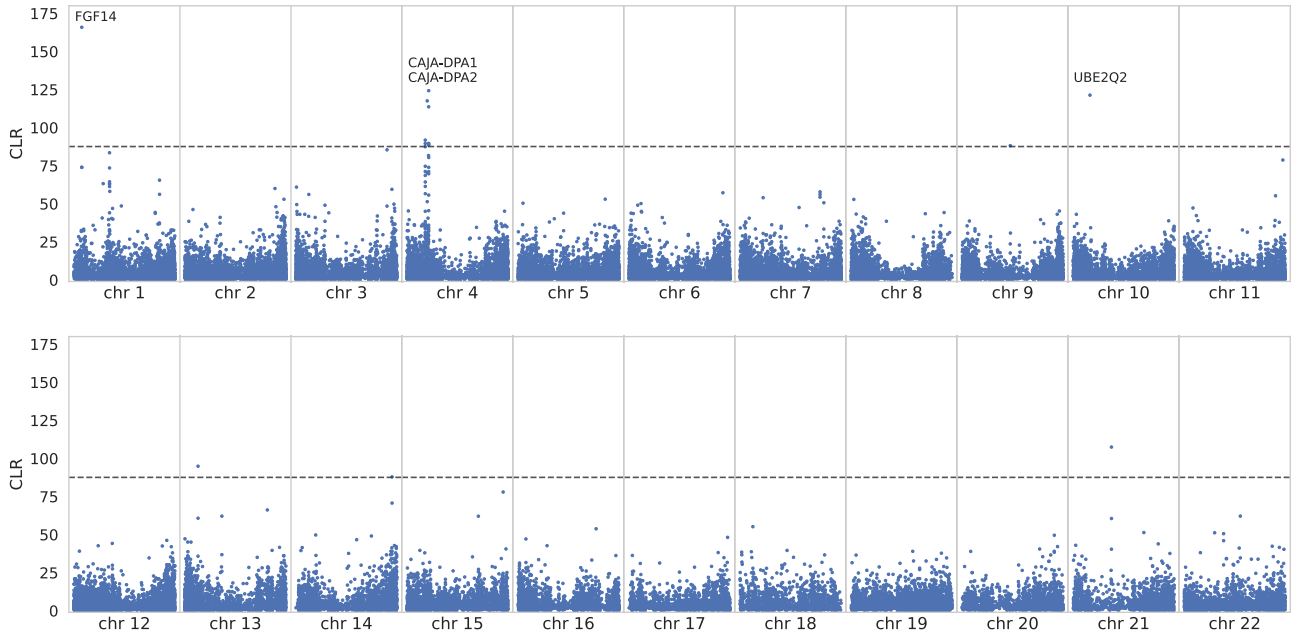

Fig. 6 Genome-wide balancing selection scan results using B_{0MAF}. The *x*-axis shows the position along each autosome (chromosomes 1–22) and the *y*-axis shows the composite likelihood ratio (CLR) value of the balancing selection statistic at each SNP. The horizontal dashed line represents the null threshold for detection.

female fertilization and the formation of the mammalian egg coat, has experienced positive selection across the mammalian clade (Swanson et al. 2001).

(4) A strongly enriched functional category arising from the joint candidates were genes related to calcium. As marmosets are characterized by a lack of readily available calcium in their diet and a poor ability to digest this mineral (Jarcho et al. 2013)—indeed, marmosets, and in particular lactating females, have been shown to exhibit a preference for calcium lactate solutions over plain water (Power et al. 1999)—this result may imply on-going selective pressures related to this mineral intake.

(5) Marmosets, who live in extended social family groups and engage in cooperative breeding, exhibit a rich set of vocalizations to identify and communicate information about predators and food, as well as to convey biologically important information such as identity, sex, and emotional states (Seyfarth and Cheney 2003; Agamaite et al. 2015; Lamothe et al. 2025). Notably in this regard, the gene TMEM145, which plays an important role in hearing, specifically in the structure and function of outer hair cell stereocilia in the inner ear (Roh et al. 2025), was both highly divergent (>75th percentile of neutral divergence) and a selective sweep candidate.

Signatures of balancing selection in the common marmoset genome

By contrast, only 14 strongly supported candidate regions met our null threshold for balancing selection inference, mapping to four genes within the common marmoset genome. Figure 6 provides the genome-wide scan results for B_{0MAF} (and see Supplementary Figs. S24 to S45 for the scan results along individual chromosomes, and Supplementary File S1 for candidate windows exhibiting CLR values greater than the null threshold). We were unable to perform a gene functional analysis with such a small number of candidate genes, and manually curated this set as an alternative, identifying two MHC genes: *CAJA-DPA1* and *CAJA-DPA2*. MHC loci have frequently been implicated in balancing selection in multiple species (see the review by Radwan et al. 2020), and the sharing of polymorphism is a signature of balancing selection that has been identified both between human populations (Soni et al. 2022) and across apes (Leffler et al. 2013; Teixeira et al. 2015). Additional support for balancing selection in -DPA1 genes was recently argued in the form of trans-species polymorphisms shared among the African apes (Fortier and Pritchard 2025). Notably, Antunes et al. (1998) found that the *CAJA-DP* region may be altered in appearance in common marmosets, given that other higher primate species have three distinct functional MHC class II regions (*-DR*, *-DQ*, and *-DP*), whilst *CAJA-DPA1* could only be detected in low quantities of PCR product in the common marmoset, suggesting that *-DR* and *-DQ* represent the main functional MHC class II regions in this species. Given the general reduction in power, a fruitful future approach for the detection of balancing selection in marmosets could include the search for maintained trans-species polymorphisms (e.g., Leffler et al. 2013), once high-quality, well-annotated genomes and population-level sequence data of more closely related species become available.

CONCLUDING THOUGHTS

In this study, we have inferred both recent and long-term patterns of natural selection in the common marmoset genome. We have estimated a well-fitting DFE to model the effects of purifying and background selection, which additionally accounts for observed patterns of exonic divergence. We found evidence of an increased proportion of newly arising strongly deleterious variants in marmosets relative to humans, potentially related to their larger estimated long-term effective population size.

Additionally, in order to perform the first large-scale scans for loci having experienced selective sweeps and/or balancing selection in the marmoset genome, we generated an evolutionary baseline model utilizing the recently estimated population history of Soni et al. (2025d), our inferred DFE, as well as existing knowledge regarding mutation and recombination rate heterogeneity, taking into account the reproductive dynamics of common twin-births and chimeric sampling particular to this species in this model construction. Utilizing this conservative approach to reduce FPRs, a number of genes with relevance to

both biomedical and evolutionary interest were identified, including genes related to immune, sensory, and reproductive functions. Notably, these gene sets particularly highlighted MHC genes as having strong evidence for experiencing long-term balancing selection, consistent with an accumulating body of work across the primate clade.

DATA AVAILABILITY

This study was based on sequence data available under NCBI BioProject PRJNA1215741. All scripts to generate and analyze simulated data, as well as results from selection scans, are available at the GitHub repository: https://github.com/vivaksoni/marmoset_selection.

REFERENCES

Agamaite JA, Chang C-J, Osmanski MS, Wang X (2015) A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). J Acoust Soc Am 138(5):2906–2908. https://doi.org/10.1121/1.4934268.

Aisha J, Yenugu S (2023) Characterization of SPINK2, SPACA7 and PDCL2: effect of immunization on fecundity, sperm function and testicular transcriptome. Reprod Biol 23(1):100711. https://doi.org/10.1016/j.repbio.2022.100711.

Alvarenga DAM, Pina-Costa A, Bianco JrC, Moreira SB, Brasil P, Pissinatti A et al. (2017) New potential *Plasmodium brasilianum* hosts: tamarin and marmoset monkeys (family Callitrichidae). Malar J 16(1):71. https://doi.org/10.1186/s12936-017-1724-0.

Anderson JA, Vilgalys TP, Tung J (2020) Broadening primate genomics: new insights into the ecology and evolution of primate gene regulation. Curr Opin Genet Dev 62:16–22. https://doi.org/10.1016/j.gde.2020.05.009.

Antunes SG, de Groot NG, Brok H, Doxiadis G, Menezes AA, Otting N et al. (1998) The common marmoset: a New World primate species with limited MHC class II variability. Proc Natl Acad Sci USA 95(20):11745e50. https://doi.org/10.1073/pnas.95.20.11745.

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A et al. (2020) Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587(7833):246–251. https://doi.org/10.1038/s41586-020-2871-y.

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

Barton NH (1998) The effect of hitch-hiking on neutral genealogies. Genet Res 72(2):123–133. https://doi.org/10.1017/S0016672398003462.

Bataillon T, Duan J, Hvilsom C, Jin X, Li Y, Skov L et al. (2015) Inference of purifying and positive selection in three subspecies of chimpanzees (*Pan troglodytes*) from exome sequencing. Genome Biol Evol 7(4):1122–1132. https://doi.org/10.1093/gbe/evv058.

Benirschke K, Anderson JM, Brownhill LE (1962) Marrow chimerism in marmosets. Science 138(3539):513–515. https://doi.org/10.1126/science.138.3539.513.

Berry AJ, Ajioka JW, Kreitman M (1991) Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics 129(4):1111–1117. https://doi.org/10.1093/genetics/129.4.1111.

Bitarello BD, Brandt DYC, Meyer D, Andrés AM (2023) Inferring balancing selection from genome-scale data. Genome Biol Evol 15(3):evad032. https://doi.org/10.1093/gbe/evad032.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4(5):e1000083. https://doi.org/10.1371/journal.pgen.1000083.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140(2):783–796. https://doi.org/10.1093/genetics/140.2.783.

Cagan A, Theunert C, Laayouni H, Santpere G, Pybus M, Casals F et al. (2016) Natural selection in the great apes. Mol Biol Evol 33(12):3268–3283. https://doi.org/10.1093/molbev/msw215.

Carrion JrR, Patterson JL (2012) An animal model that reflects human disease: the common marmoset (*Callithrix jacchus*). Curr Opin Virol 2(3):357e62. https://doi.org/10.1016/j.coviro.2012.02.007.

Castellano D, Macià MC, Tataru P, Bataillon T, Munch K (2019) Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. Genetics 213(3):953–966. https://doi.org/10.1534/genetics.119.302494.

Charlesworth B, Jensen JD (2021) Effects of selection at linked sites on patterns of genetic variability. Annu Rev Ecol Evol Syst 52:177–197. https://doi.org/10.1146/annurev-ecolsys-010621-044528.

Charlesworth B, Jensen JD (2022) Some complexities in interpreting apparent effects of hitchhiking: a commentary on Gompert et al. Mol Ecol 31(17):4440–4443. https://doi.org/10.1111/mec.16573.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134(4):1289–1303. https://doi.org/10.1093/genetics/134.4.1289.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2(4):e64. https://doi.org/10.1371/journal.pgen.0020064.

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S et al. (2018) SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience 7(1):1–6. https://doi.org/10.1093/gigascience/gix120.

Cheng X, DeGiorgio M (2020) Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. Mol Biol Evol 37(11):3267–3291. https://doi.org/10.1093/molbev/msaa134.

Chiou K, Snyder-Mackler N (2024) Marmosets contain multitudes. eLife 13:e97866. https://doi.org/10.7554/eLife.97866.

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

Comeron JM (2014) Background selection as baseline for nucleotide variation across the *Drosophila* genome. PLoS Genet 10(6):e1004434. https://doi.org/10.1371/journal.pgen.1004434.

Comeron JM (2017) Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. Phil Trans R Soc B 372(1736):20160471. https://doi.org/10.1098/rstb.2016.0471.

Crisci JL, Poh Y-P, Mahajan S, Jensen JD (2013) The impact of equilibrium assumptions on tests of selection. Front Genet 4:235. https://doi.org/10.3389/fgene.2013.00235.

DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R (2016) SweepFinder2: increased sensitivity, robustness and flexibility. Bioinformatics 32(12):1895–1897. https://doi.org/10.1093/bioinformatics/btw051.

del Rosario RCH, Krienen FM, Zhang Q, Goldman M, Mello C, Lutservitz A et al. (2024) Sibling chimerism among microglia in marmosets. eLife 13:RP93640. https://doi.org/10.7554/eLife.93640.

Enard D, Depaulis F, Roest Crollius H (2010) Human and non-human primate genomes share hotspots of positive selection. PLoS Genet 6(2):e1000840. https://doi.org/10.1371/journal.pgen.1000840.

Ewing GB, Jensen JD (2014) Distinguishing neutral from deleterious mutations in growing populations. Front Genet 5:684–687. https://doi.org/10.3389/fgene.2014.00007.

Ewing GB, Jensen JD (2016) The consequences of not accounting for background selection in demographic inference. Mol Ecol 25(1):135–141. https://doi.org/10.1111/mec.13390.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. Nat Rev Genet 8(8):610–618. https://doi.org/10.1038/nrg2146.

Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol 26(9):2097–2108. https://doi.org/10.1093/molbev/msp119.

Faulkes CG, Arruda MF, Monteiro Da Cruz MA (2003) Matrilineal genetic structure within and among populations of the cooperatively breeding common marmoset, *Callithrix jacchus*. Mol Ecol 12(4):1101e8. https://doi.org/10.1046/j.1365-294x.2003.01809.x.

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. Genetics 155(3):1405–1413. https://doi.org/10.1093/genetics/155.3.1405.

Fijarczyk A, Babik W (2015) Detecting balancing selection in genomes: limits and prospects. Mol Ecol 24(14):3529–3545. https://doi.org/10.1111/mec.13226.

Fortier AL, Pritchard JK (2025) Ancient trans-species polymorphism at the major histocompatibility complex in primates. eLife 14:RP103547. https://doi.org/10.7554/eLife.103547.1.

Garber PA, Caselli CB, McKenney AC, Abreu F, De la Fuente MF, Araujo A et al. (2019) Trait variation and trait stability in common marmosets (*Callithrix jacchus*) inhabiting ecologically distinct habitats in northeastern Brazil. Am J Primatol 81(7):e23018. q0.1002/ajp.23018.

Gengozian N, Batson JS, Greene CT, Gosslee DG (1969) Hemopoietic chimerism in imported and laboratory-bred marmosets. Transplantation 8(5):633–652. https://doi.org/10.1097/00007890-196911000-00009.

George RD, McVicker G, Diederich R, Ng SB, MacKenzie AP, Swanson WJ et al. (2011) *Trans* genomic capture and sequencing of primate exomes reveals new targets of positive selection. Genome Res 21(10):1686–1694. https://doi.org/10.1101/gr.121327.111.

Gopal P, Sarihan EI, Chie EK, Kuzmishin G, Doken S, Pennell NA et al. (2019) Clonal selection confers distinct evolutionary trajectories in BRAF-driven cancers. Nat Commun 10(1):5143. https://doi.org/10.1038/s41467-019-13161-x.

Haller BC, Messer PW (2023) SLiM 4: multispecies eco-evolutionary modeling. Am Nat 201(5):E127–E139. https://doi.org/10.1086/723601.

Han H-J, Powers SJ, Gabrielson KL (2022) The common marmoset—biomedical research animal model applications and common spontaneous diseases. Toxicol Pathol 50(5):628–637. https://doi.org/10.1177/01926233221095449.

Harris RA, Tardif SD, Vinar T, Wildman DE, Rutherford JN, Rogers J et al. (2014) Evolutionary genetics and implications of small size and twinning in callitrichine primates. Proc Natl Acad Sci USA 111(4):1467–1472. https://doi.org/10.1073/pnas.1316037111.

Harris RA, Raveendran M, Warren W, LaDeana HW, Tomlinson C, Graves-Lindsay T et al. (2023) Whole genome analysis of SNV and indel polymorphism in common marmosets (*Callithrix jacchus*). Genes 14(12):2185. https://doi.org/10.3390/genes14122185.

Harris RB, Jensen JD (2020) Considering genomic scans for selection as coalescent model choice. Genome Biol Evol 12(6):871–877. https://doi.org/10.1093/gbe/evaa093.

Harris RB, Sackman A, Jensen JD (2018) On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. PLoS Genet 14(12):e1007859. https://doi.org/10.1371/journal.pgen.1007859.

Hickey G, Paten B, Earl D, Zerbino D, Haussler D (2013) HAL: a hierarchical format for storing and analyzing multiple genome alignments. Bioinformatics 29(10):1341–1342. https://doi.org/10.1093/bioinformatics/btt128.

Hill JP (1932) II. Croonian lecture.—The developmental history of the primates. Phil Trans R Soc B 221(474–482):45–178. https://doi.org/10.1098/rstb.1932.0002.

Hoh B-P, Zhang X, Deng L, Yuan K, Yew C-W, Saw W-Y et al. (2020) Shared signature of recent positive selection on the *TSBP1–BTNL2–HLA-DRA* genes in five native populations from North Borneo. Genome Biol Evol 12(12):2245–2257. https://doi.org/10.1093/gbe/evaa207.

Howell AA, Terbot JW, Soni V, Johri P, Jensen JD, Pfeifer SP (2023) Developing an appropriate evolutionary baseline model for the study of human cytomegalovirus. Genome Biol Evol 15(4):evad059. https://doi.org/10.1093/gbe/evad059.

Huber CD, Kim BY, Marsden CD, Lohmueller KE (2017) Determining the factors driving selective effects of new nonsynonymous mutations. Proc Natl Acad Sci USA 114(17):4465–4470. https://doi.org/10.1073/pnas.1619508114.

Irwin KK, Laurent S, Matuszewski S, Vuilleumier S, Ormond L, Shim H et al. (2016) On the importance of skewed offspring distributions and background selection in virus population genetics. Heredity 117(6):393–399. https://doi.org/10.1038/hdy.2016.58.

Jaquish CE, Tardif SD, Toal RL, Carson RL (1996) Patterns of prenatal survival in the common marmoset (*Callithrix jacchus*). J Med Primatol 25(1):57–63. https://doi.org/10.1111/j.1600-0684.1996.tb00194.x.

Jarcho MR, Power ML, Layne-Colon DG, Tardif SD (2013) Digestive efficiency mediated by serum calcium predicts bone mineral density in the common marmoset (*Callithrix jacchus*). Am J Primatol 75(2):153–160. https://doi.org/10.1002/ajp.22093.

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

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

Jensen JD, Thornton KR, Andolfatto P (2008) An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. PLoS Genet 4(9):e1000198. https://doi.org/10.1371/journal.pgen.1000198.

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

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

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

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

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

Keightley PD, Eyre-Walker A (2007) Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177(4):2251–2261. https://doi.org/10.1534/genetics.107.080663.

Keightley PD, Eyre-Walker A (2010) What can we learn about the distribution of fitness effects of new mutations from DNA sequence data?. Phil Trans R Soc B 365(1544):1187–1193. https://doi.org/10.1098/rstb.2009.0266.

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160(2):765. https://doi.org/10.1093/genetics/160.2.765.

Klein J, Sato A, Nagl S, O'hUigín C (1998) Molecular trans-species polymorphism. Ann Rev Ecol Evol Syst 29(1):1–21. https://doi.org/10.1146/annurev.ecolsys.29.1.1.

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

Lamothe C, Obliger-Debouce M, Best P, Trapeau R, Ravel S, Artières T et al. (2025) A large annotated dataset of vocalizations by common marmosets. Sci Data 12(1):782. https://doi.org/10.1038/s41597-025-04951-8.

Leffler EM, Gao Z, Pfeifer S, Segurel L, Auton A, Venn O et al. (2013) Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science 339(6127):1578–1582. https://doi.org/10.1126/science.1234070.

Lewontin RC (1987) Polymorphism and heterosis: old wine in new bottles and vice versa. J Hist Biol 20(3):337–349. https://doi.org/10.1007/BF00139459.

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

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM et al. (2011) Comparative and demographic analysis of orang-utan genomes. Nature 469(7331):529–533. https://doi.org/10.1038/nature09687.

Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A et al. (2023) Database Commons: a catalog of worldwide biological databases. Genomics Proteom Bioinforma 21(5):1054–1058. https://doi.org/10.1016/j.gpb.2022.12.004.

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O et al. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 50(W1):W276–W279. https://doi.org/10.1093/nar/gkac240.

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

Malukiewicz J, Cartwright RA, Curi NHA, Dergam JA, Igayara CS, Moreira SB et al. (2021) Mitogenomic phylogeny of *Callithrix* with special focus on human transferred taxa. BMC Genomics 22(1):239. https://doi.org/10.1186/s12864-021-07533-1.

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

Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H (2005) Emergence of young human genes after a burst of retroposition in primates. PLoS Biol 3(11):e357. https://doi.org/10.1371/journal.pbio.0030357.

McManus KF, Kelley JL, Song S, Veeramah KR, Woerner AE, Stevison LS et al. (2015) Inference of gorilla demographic and selective history from whole-genome sequence data. Mol Biol Evol 32(3):600–612. https://doi.org/10.1093/molbev/msu394.

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

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

Munch K, Nam K, Schierup MH, Mailund T (2016) Selective sweeps across twenty millions years of primate evolution. Mol Biol Evol 33(12):3065–3074. https://doi.org/10.1093/molbev/msw199.

Nam K, Munch K, Mailund T, Nater A, Greminger MP, Krützen M et al. (2017) Evidence that the rate of strong selective sweeps increases with population size in the great apes. Proc Natl Acad Sci USA 114(7):1613–1618. https://doi.org/10.1073/pnas.1605660114.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. Genome Res 15(11):1566–1575. https://doi.org/10.1101/gr.4252305.

Okano H, Hikishima K, Iriki A, Sasaki E (2012) The common marmoset as a novel animal model system for biomedical and neuroscience research applications. Semin Fetal Neonatal Med 17(6):336–340. https://doi.org/10.1016/j.siny.2012.07.002.

Otsubo C, Otomo R, Miyazaki M, Matsushima-Hibiya Y, Kohno T, Iwakawa R et al. (2014) *TSPAN2* is involved in cell invasion and motility during lung cancer progression. Cell Rep 7(2):527–538. https://doi.org/10.1016/j.celrep.2014.03.027.

Pavlidis P, Metzler D, Stephan W (2012) Selective sweeps in multilocus models of quantitative traits. Genetics 192(1):225–239. https://doi.org/10.1534/genetics.112.142547.

Perez-Cruz C, Rodriguez-Callejas JDD (2023) The common marmoset as a model of neurodegeneration. Trends Neuros 46(5):394–409. https://doi.org/10.1016/j.tins.2023.02.002.

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

Pfeifer SP (2017b) The demographic and adaptive history of the African green monkey. Mol Biol Evol 34(5):1055–1065. https://doi.org/10.1093/molbev/msx056.

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

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

Power ML, Tardif SD, Layne DG, Schulkin J (1999) Ingestion of calcium solutions by common marmosets (*Callithrix jacchus*). Am J Primatol 47(3):255–261. 10.1002/(sici)1098-2345(1999)47:3%3C255::aid-ajp7%3E3.0.co;2-w.

Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B et al. (2012) The bonobo genome compared with the chimpanzee and human genomes. Nature 486(7404):527–531. https://doi.org/10.1038/nature11128.

Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J (2020) Advances in the evolutionary understanding of MHC polymorphism. Trends Genet 36(4):298–311. https://doi.org/10.1016/j.tig.2020.01.008.

Rausell A, Telenti A (2014) Genomics of host–pathogen interactions. Curr Op Immunol 30:32–38. https://doi.org/10.1016/j.coi.2014.06.001.

Roh JW, Oh KS, Lee J, Choi Y, Kim S, Hong JW, et al. (2025) TMEM145 is a key component in stereociliary link structures of outer hair cells. BioRxiv https://doi.org/10.1101/2025.02.10.637577.

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

Rosso L, Marques AC, Weier M, Lambert N, Lambot M-A, Vanderhaeghen P et al. (2008) Birth and rapid subcellular adaptation of a hominoid-specific CDC14 protein. PLoS Biol 6(6):e140. https://doi.org/10.1371/journal.pbio.0060140.

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

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

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC et al. (2022) Database resources of the national center for biotechnology information. Nucleic Acids Res 50(D1):D20–D26. https://doi.org/10.1093/nar/gkab1112.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J et al. (2012) Insights into hominid evolution from the gorilla genome sequence. Nature 483(7388):169–175. https://doi.org/10.1038/nature10842.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision on variation at multi-allelic loci under balancing selection. Genet Res 76(1):51–62. https://doi.org/10.1017/S0016672300004535.

Schmidt JM, De Manuel M, Marques-Bonet T, Castellano S, Andrés AM (2019) The impact of genetic adaptation on chimpanzee subspecies differentiation. PLoS Genet 15(11):e1008485. https://doi.org/10.1371/journal.pgen.1008485.

Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD (2011) A method for inferring the rate of occurrence and fitness effects of advantageous mutations. Genetics 189(4):1427–1437. https://doi.org/10.1534/genetics.111.131730.

Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen H-C, Kitts PA et al. (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res 27(5):849–864. https://doi.org/10.1101/gr.213611.116.

Seyfarth RM, Cheney DL (2003) Signalers and receivers in animal communication. Annu Rev Psychol 54:145–173. https://doi.org/10.1146/annurev.psych.54.101601.145121.

Simkin A, Bailey JA, Gao FB, Jensen JD (2014) Inferring the evolutionary history of primate microRNA binding sites: overcoming motif counting biases. Mol Biol Evol 31(7):1894–1901. https://doi.org/10.1093/molbev/msu129.

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141(1):413–429. https://doi.org/10.1093/genetics/141.1.413.

Soni V, Jensen JD (2024) Temporal challenges in detecting balancing selection from population genomic data. G3 14(6):1–14. https://doi.org/10.1093/g3journal/jkae069.

Soni V, Jensen JD (2025) Inferring demographic and selective histories from population genomic data using a 2-step approach in species with coding-sparse genomes: an application to human data. G3 15(4):jkaf019. https://doi.org/10.1093/g3journal/jkaf019.

Soni V, Vos M, Eyre-Walker A (2022) A new test suggests hundreds of amino acid polymorphisms in humans are subject to balancing selection. PLoS Biol 20(6):e3001645. https://doi.org/10.1371/journal.pbio.3001645.

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

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

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

Soni V, Versoza CJ, Jensen JD, Pfeifer SP (2025b) Inferring the landscape of mutation and recombination in the common marmoset (*Callithrix jacchus*) in the presence of twinning and hematopoietic chimerism. BioRxiv https://doi.org/10.1101/2025.07.01.662565.

Soni V, Versoza CJ, Pfeifer SP, Jensen JD (2025c) 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 87(7):e70058. https://doi.org/10.1002/ajp.70058.

Soni V, Versoza CJ, Vallender EJ, Jensen JD, Pfeifer SP (2025d) Accounting for chimerism in demographic inference: reconstructing the history of common marmosets (*Callithrix jacchus*) from high-quality, whole-genome, population-level data. Mol Biol Evol 42(6):msaf119. https://doi.org/10.1093/molbev/msaf119.

Stephan W (2019) Selective sweeps. Genetics 211(1):5–13. https://doi.org/10.1534/genetics.118.301319.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA 98(5):2509–2514. https://doi.org/10.1073/pnas.051605998.

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

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

Tataru P, Bataillon T (2020) polyDFE: inferring the distribution of fitness effects and properties of beneficial mutations from polymorphism data. In: Dutheil JY (ed) Statistical population genomics, vol 2090. Methods in molecular biology. Springer, New York, NY, pp 125–146

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

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

Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps?. Genome Res 16(6):702–712. https://doi.org/10.1101/gr.5105206.

Teixeira JC, De Filippo C, Weihmann A, Meneu JR, Racimo F, Dannemann M et al. (2015) Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees, and bonobos. Mol Biol Evol 32(5):1186–1196. https://doi.org/10.1093/molbev/msv007.

The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ et al. (2023) The Gene Ontology knowledgebase in 2023. Genetics 224(1):iyad031. https://doi.org/10.1093/genetics/iyad031.

The Marmoset Genome Sequencing and Analysis Consortium (2014) The common marmoset genome provides insight into primate biology and evolution. Nat Genet 46(8):850e7. https://doi.org/10.1038/ng.3042.

The Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316(5822):222–234. https://doi.org/10.1126/science.1139247.

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

Thornton KR, Jensen JD (2007) Controlling the false-positive rate in multilocus genome scans for selection. Genetics 175(2):737–750. https://doi.org/10.1534/genetics.106.064642.

Van der Auwera G, O'Connor B (2020) Genomics in the cloud, 1st edn. O'Reilly Media. Sebastopol, California, USA.

van der Lee R, Wiel L, van Dam TJP, Huynen MA (2017) Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. Nucleic Acids Res 45(18):10634–10648. https://doi.org/10.1093/nar/gkx704.

Versoza CJ, Pfeifer SP (2024) A hybrid genome assembly of the endangered aye-aye (*Daubentonia madagascariensis*). G3 14(10):jkae185. https://doi.org/10.1093/g3journal/jkae185.

Ward JM, Buslov AM, Vallender EJ (2014) Twinning and survivorship of captive common marmosets (*Callithrix jacchus*) and cotton-top tamarins (*Saguinus oedipus*). J Am Assoc Lab Anim Sci 53(1):7–11.

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

Wei Z, Peddibhotla S, Lin H, Fang X, Li M, Rosen JM et al. (2011) Early-onset aging and defective DNA damage response in *Cdc14b* -deficient mice. Mol Cell Biol 31(7):1470–1477. https://doi.org/10.1128/MCB.01330-10.

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

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

Wiuf C, Zhao K, Innan H, Nordborg M (2004) The probability and chromosomal extent of trans-specific polymorphism. Genetics 168(4):2363–2372. https://doi.org/10.1534/genetics.104.029488.

Wu MS, Tani K, Sugiyama H, Hibino H, Izawa K, Tanabe T et al. (2000) MHC (major histocompatibility complex)-DRB genes and polymorphisms in common marmoset. J Mol Evol 51(3):214e22. https://doi.org/10.1007/s002390010083.

Xie Y, Zhang L, Wang L, Chen B, Guo X, Yang Y et al. (2024) EphB1 promotes the differentiation and maturation of dendritic cells in non-small cell lung cancer. Cancer Lett 582:216567. https://doi.org/10.1016/j.canlet.2023.216567.

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

Yang X, Mao Y, Wang X-K, Ma D-N, Xu Z, Gong N et al. (2023) Population genetics of marmosets in Asian primate research centers and loci associated with epileptic risk revealed by whole-genome sequencing. Zoo Res 44(5):837–847. https://doi.org/10.24272/j.issn.2095-8137.2022.514.

Yaseen IH, Monk PN, Partridge LJ (2017) Tspan2: a tetraspanin protein involved in oligodendrogenesis and cancer metastasis. Biochem Soc Trans 45(2):465–475. https://doi.org/10.1042/BST20160022.

Zoonomia Consortium (2020) A comparative genomics multitool for scientific discovery and conservation. Nature 587(7833):240–245. https://doi.org/10.1038/s41586-020-2876-6.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM151008 to SPP and R35GM139383 to JDJ, and by National Science Foundation CAREER award DEB-2045343 to SPP. 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.

AUTHOR CONTRIBUTIONS

VS, CV, SP, and JJ conceived of and conducted the study, interpreted results, and wrote the manuscript. SP and JJ obtained research funding.

COMPETING INTERESTS

The authors declare no competing interests.

RESEARCH ETHICS STATEMENT

Animals previously housed at the New England Primate Research Center 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. All samples were collected during routine veterinary care under approved protocols.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-025-00804-7.

Correspondence and requests for materials should be addressed to Susanne P. Pfeifer or Jeffrey D. Jensen.