

Inferring Patterns of Purifying, Positive, and Balancing Selection in the Coppery Titi Monkey (*Plecturocebus cupreus*) Utilizing a Well-Fit Evolutionary Baseline Model

Vivak Soni ¹, Cyril J. Versoza ¹, John W. Terbot II ¹, Gabriella J. Spatola ¹, Karen L. Bales ^{2,3}, Susanne P. Pfeifer ^{1,*†}, Jeffrey D. Jensen ^{1,*†}

¹Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA

²Department of Psychology, University of California, Davis, CA, USA

³California National Primate Research Center, University of California, Davis, CA, USA

[†]Jointly supervised the project.

*Corresponding authors: E-mails: susanne@spfeiferlab.org; jeffrey.d.jensen@asu.edu.

Accepted: May 07, 2026

Abstract

Despite the coppery titi monkey (*Plecturocebus cupreus*) being a model system for the study of neurodevelopment and behavior, the evolutionary forces shaping observed levels and patterns of genetic variation in the species have remained poorly studied. In order to illuminate the pervasive effects of purifying and background selection, we have fit a distribution of fitness effects of newly arising exonic mutations, utilizing patterns of polymorphism and divergence based on a recently published high-quality genome assembly. To further characterize episodically acting selective processes, we additionally performed the first whole-genome scans for recent positive and balancing selection in this species, reducing false-positive rates by incorporating the demographic history of the population into an evolutionary null model. These scans identified a small number of biomedically relevant genes with strong statistical support for having experienced recent selective sweeps or long-term balancing selection. In addition, we identified four genomic deletions bearing the signatures of balancing selection. Taken together, this study provides the first insights into patterns of persistent and episodic selective processes in this species.

Key words: primate, platyrrhine, distribution of fitness effects, positive selection, balancing selection, genome scans.

Significance

The coppery titi monkey (*Plecturocebus cupreus*) is a model system for the study of neurodevelopment and behavior. To date however, the selective forces shaping observed levels and patterns of exonic variation in the species have remained poorly characterized. In this study, we quantified the pervasive effects of purifying and background selection in shaping functional variation, and identified a number of biomedically relevant genes bearing signatures of recent positive selection as well as long-term balancing selection.

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Introduction

Native to the Amazonian forests of Peru and Brazil, the coppery titi monkey (*Plecturocebus cupreus*) is a platyrrhine that inhabits the upper and middle canopy of its forest habitat (Heymann et al. 2021). *P. cupreus* is one of 34 recognized species of the genus *Plecturocebus*, itself one of three titi monkey genera, and is a member of the Western Amazon clade of the *moloch* group that diverged ~4 million years ago, along with *P. moloch*, *P. brunneus*, *P. dubius*, and *P. caligatus* (Byrne et al. 2016, 2018). Coppery titi monkeys belong to the small number of primate species that are monogamous and pair bonded. Paternal care is common in family units (Mendoza and Mason 1986; Valeggia et al. 1999), which can be constituted by up to three generations of offspring (Kinzey 1997). As such, they have been the focus of much neurobiological research into social behavior, particularly on traits related to monogamy (Lukas and Clutton-Brock 2013), pair-bonding and male parenting (Bales et al. 2007, 2017, 2021; Lau et al. 2024), as well as general cognition (Bales et al. 2017; Lau et al. 2020). Yet, despite this focus on phenotypic traits, relatively little is known regarding the evolutionary forces shaping genomic variation in the species. Specifically, the relative contributions of purifying and background selection in dictating genomic variation as well as the episodic effects of positive and balancing selection, and any association between the action of these processes and well-studied phenotypes in this species, have yet to be characterized.

Inference of the Distribution of Fitness Effects: Long-term Selective Dynamics

Given that deleterious mutations make up the vast majority of fitness-impacting mutations, the purging of these variants by purifying selection is expected to be the dominant form of selection shaping genomic variation. Relatedly, the effect of this process on linked sites (ie background selection [BGS]; Charlesworth et al. 1993) additionally shapes both levels and patterns of genomic diversity in the vicinity of functional regions (see the review of Charlesworth and Jensen 2021). As the efficacy of selection is mediated by the effective population size (N_e), the proportion of constrained sites subject to these processes will generally be expected to be greater in populations with larger N_e . As posited by the Neutral and Nearly Neutral Theories of Molecular Evolution (Kimura 1968, 1983; Ohta 1973), what remains in terms of observed genomic polymorphism, as well as observed divergence, is likely dominated by the action of genetic drift, with the potential for relatively rare and episodic contributions from other forms of

selection including positive or balancing selection (and see Jensen et al. 2019).

The characterization of the distribution of selection coefficients associated with new mutations (termed the distribution of fitness effects [DFE]) is therefore crucial when developing evolutionary baseline models for population genomic inference (Comeron 2014, 2017; Ewing and Jensen 2014, 2016; Morales-Arce et al. 2022; Johri et al. 2022a; Howell et al. 2023; Terbot et al. 2023; Soni and Jensen 2025). By summarizing the relative proportion of effectively neutral, weakly deleterious, and strongly deleterious mutations in functional genomic regions, the DFE effectively describes the extent of selective constraint (Eyre-Walker and Keightley 2007; Keightley and Eyre-Walker 2010; Bank et al. 2014).

While initial DFE inference methods relied on a two-step approach firstly inferring a population history from putatively neutral synonymous sites prior to inferring a DFE from non-synonymous sites (Eyre-Walker and Keightley 2007, 2009; Keightley and Eyre-Walker 2007; Boyko et al. 2008; Schneider et al. 2011), recent studies have demonstrated that the underlying neglect of BGS effects (Johri et al. 2020, 2021), as well as mutation and recombination rate heterogeneity (Soni et al. 2024), can lead to potentially severe mis-inference under this class of approach. However, for species with coding-sparse genomes (such as primates) and available whole-genome polymorphism data, one can alternatively infer complex demographic models from intergenic, neutral regions unlinked to directly selected sites (and thus largely free from BGS effects), prior to performing DFE inference in functional regions conditional on that inferred population history (Soni and Jensen 2025). While similarly a two-step approach as in early implementations, the advantages of this alternative framework include the ability to estimate a highly complex demographic history, the ability to directly account for BGS effects when performing DFE inference, and the relaxation of the assumption of synonymous site neutrality.

Inference of Episodic Selective Events: Recent Positive Selection and Balancing Selection

Although genome-wide patterns of genetic variation can generally be explained without invoking more episodic selective processes such as selective sweeps or balancing selection, a number of approaches have been developed for the detection of these processes in localized genomic windows, generally relying on expected changes in patterns of variation at linked sites (see the reviews of Stephan 2019; Charlesworth and Jensen 2021 for sweep dynamics, and Soni and Jensen 2024 for balancing selection). Individual selective sweeps

are associated with a skew in the SFS toward rare and high-frequency alleles (Braverman et al. 1995; Simonsen et al. 1995; Fay and Wu 2000), as well as a local reduction in nucleotide diversity around the selected locus (Berry et al. 1991). The severity of the hitchhiking effect is determined by a number of factors, including the strength and age of the beneficial mutation, as well as the level of recombination in the given genomic region. Conversely, balancing selection maintains genetic variability in populations (see the reviews of Fijarczyk and Babik 2015; Bitarello et al. 2023). The genomic signatures of balancing selection will vary depending on the age of the segregating balanced allele, with recent balancing selection (in which the age of the balanced mutation is $<0.4N_e$ generations) generating extended LD due to hitchhiking effects as the mutation increases rapidly in frequency, as well as a relative increase in intermediate frequency alleles (Schierup et al. 2000; and see the reviews of Crisci et al. 2012; Charlesworth and Jensen 2021). The initial pattern of extended LD—which is indistinguishable from an incomplete selection sweep in this early phase—dissipates once the selected mutation reaches its equilibrium frequency, as recombination subsequently breaks up linkage (Wiuf et al. 2004; Charlesworth 2006; Pavlidis et al. 2012). However, the skew in the SFS toward intermediate frequency alleles becomes more pronounced as the segregation time of the balanced mutation increases, due to new mutations arising on the balanced haplotype (Soni and Jensen 2024). If the balanced mutation segregates for evolutionary timescales of $>4N_e$ generations (so-called ancient balancing selection), and for periods greater than the divergence time of the species in question, trans-species polymorphisms may be detectable (Klein et al. 1998; Leffler et al. 2013).

The theoretical expectations under a model of a single, recent selective sweep, and of balancing selection once the selected mutation has reached equilibrium frequency, provide the basis of composite likelihood ratio (CLR) tests developed for the detection of these processes. Kim and Stephan (2002) developed the first CLR-based method for the detection of selective sweeps, though subsequent studies have demonstrated that neutral evolutionary processes such as recent population bottlenecks can generate similar genomic signatures and thus result in mis-inference (eg Jensen et al. 2005; Charlesworth and Jensen 2022). Building on this method, the SweepFinder software (Nielsen et al. 2005; DeGiorgio et al. 2016) utilizes a null derived from the empirically observed SFS in order to model these neutral processes and partially address the problem of mis-inference. For the detection of long-term balancing selection, Cheng and DeGiorgio (2020) developed a mixed-model approach that combines

the expectation of the SFS under balancing selection and under neutrality, thereby modeling the change in the SFS shape as the genomic distance increases from the putatively selected site. This modeling approach underlies a suite of CLR-based methods released under the BalLeRMix software package (Cheng and DeGiorgio 2020, 2022). These methods have been shown to increase power to detect balancing selection as the balanced mutation segregates over longer timescales ($>25N_e$ generations; Soni and Jensen 2024). Similar to SweepFinder2, this approach derives a null from the empirical SFS in order to account for alternative and potentially biasing evolutionary processes.

DFE Inference and Genomic Scans for Recent Positive Selection and Balancing Selection in Non-Human Primates

The majority of DFE inference studies in non-human primates have applied variants of the Eyre-Walker and Keightley (2007) approach to exonic data from the great apes (eg Castellano et al. 2019; Tataru and Bataillon 2020), whilst others have focused on gene regulatory regions (eg Simkin et al. 2014; Anderson et al. 2020; Kuderna et al. 2024). More recent studies have harnessed the two-step framework of Soni and Jensen (2025) to infer the DFE in an outgroup species to the haplorhine lineage, the aye-aye (Soni et al. 2025a), as well as in a species of biomedical interest, the common marmoset (Soni et al. 2025b). Briefly, the DFEs were inferred via forward-in-time simulation under a demographic model previously inferred for aye-ayes (Terbot et al. 2025, and see the review of Soni et al. 2025c) and common marmosets (Soni et al. 2025d) from non-coding regions that are sufficiently distant from functional genomic regions such that BGS effects will not bias demographic inference. The inferred DFEs in both species were characterized by a greater proportion of deleterious variants relative to that inferred in humans by Johri et al. (2023), consistent with the larger long-term N_e in both aye-ayes and common marmosets relative to humans. As with DFE inference, the majority of genome scans in non-human primates have focused on the great apes (eg 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), though more recent studies have conducted scans for recent positive selection and balancing selection in the aforementioned aye-ayes (Soni et al. 2025e) and common marmosets (Soni et al. 2025b) as well.

In this study, we characterize the DFE and perform genomic scans for selective sweeps and balancing selection in the coppery titi monkey. This quantification was

enabled by a number of recent genomic developments in the species. Firstly, the publication of a highly contiguous genome assembly, with gene-level annotations (Pfeifer et al. 2024), allowed for the necessary demarcation of neutral and functional genomic regions. Secondly, the generation of high-quality, whole-genome polymorphism data allowed for an accurate quantification of genome-wide variation in the species. Based on these developments, recent studies have provided the basis for an appropriate genomic null model for selection inference. For example, in quantifying the population history of the coppery titi monkey, Terbot et al. (2026) described a population size change history characterized by ancestral population growth followed by a strong and recent population contraction. These population size change events were found to correspond well with large-scale climatic changes impacting wetlands in the species' native region as governed by general glaciation patterns, and must be carefully accounted for when performing selection analyses. Furthermore, direct (pedigree-based) and indirect (divergence- and polymorphism-based) analyses have characterized both sex-specific and fine-scale patterns of mutation rate and recombination rate heterogeneity across the coppery titi monkey genome (Soni et al. 2026; Versoza et al. 2026a, 2026b), factors which are also critical for characterizing selective effects. However, these previous genomic studies in this species have focused upon neutral evolutionary processes, characterizing the demographic history of the species together with both broad- and fine-scale mutation and recombination landscapes. Here, we utilize these results to perform the first population-level selection analyses focused upon functional genomic regions, characterizing the strength of purifying selection acting on coding regions as well as identifying loci targeted by positive and balancing selection. Our results suggest that the DFE in coppery titi monkeys is characterized by an excess of effectively neutral mutations relative to humans and common marmosets, another platyrrhine. Power analyses demonstrate that there is limited power to detect episodic forms of selection in coppery titi monkeys, due to the recent and severe population bottleneck experienced by the sampled population. This dynamic is borne out in our genomic scans for selective sweeps and balancing selection, which revealed a limited number of candidate genes related to neurological development and ethanol metabolism.

Results and Discussion

Inferring the DFE in Coppery Titi Monkeys

To infer the DFE of new mutation in the coppery titi monkey, we first sought to characterize patterns of

exonic divergence between coppery titi monkeys and humans. For this purpose, we generated population-level polymorphism data by whole-genome sequencing six unrelated coppery titi monkeys to an average coverage of 43× per individual (range: 38×–57×). As information from polymorphism data is limited to relatively short evolutionary timescales (on the order of $4N_e$ generations), divergence is an important additional summary statistic to consider when performing DFE inference. To that end, we updated the coppery titi monkey genome in the 447-way mammalian multiple species alignment (Zoonomia Consortium 2020; Kuderna et al. 2024) with the high-quality, chromosome-level genome of Pfeifer et al. (2024), and retrieved substitutions between coppery titi monkeys and humans (see the “Materials and Methods” section for details). We found that no individual exon exhibited a higher divergence rate than the maximum neutral divergence observed in non-coding regions by Soni et al. (2026) in 1 kb windows (Fig. S1). This result is consistent with expectations, given that purifying selection depresses divergence in functional genomic regions (Charlesworth et al. 1993).

We ran forward-in-time simulations in SLiM (Haller et al. 2026) under the coppery titi monkey demographic model inferred by Terbot et al. (2026), illustrated in Fig. 1a, in order to fit empirical values of exonic divergence, Watterson's θ_w (Watterson 1975), Tajima's D (Tajima 1989), and the number of singletons per site with a DFE shape comprised of neutral, weakly/moderately deleterious, and strongly deleterious mutational classes. Thus, following a $10N_{ancestral}$ generation burn-in phase (where $N_{ancestral}$ is the initial coppery titi monkey population size), a divergence phase between *P. cupreus* and humans of 32 million years (Glazko and Nei 2003), and a generation time of 6 years (Pacifi et al. 2013; Perez et al. 2013), data was simulated concluding with the demographic model of Terbot et al. (2026), with six diploid individuals sampled as per the empirical data. To infer a well-fitting DFE, we began from DFEs previously inferred using a similar approach in humans (Johri et al. 2023) and common marmosets (Soni et al. 2025b), and performed manual grid step iterations from these DFEs based upon the fit to the data (see Fig. 1b for the best-fitting DFE, and Fig. 1c–f for the fits of the resulting summary statistics to the empirical data). Although the coppery titi monkey is more closely related to the common marmoset than to humans, the inferred DFE exhibits greater similarity to the latter. This is likely owing to the similar long-term N_e values of the coppery titi monkey and humans relative to the much larger N_e in the common marmoset. Specifically, a lower N_e is expected to reduce the efficacy of purifying selection given that the strength of selection acting on an individual mutation is the product of the selection coefficient, s , and N_e . Thus, purifying selection would be expected to

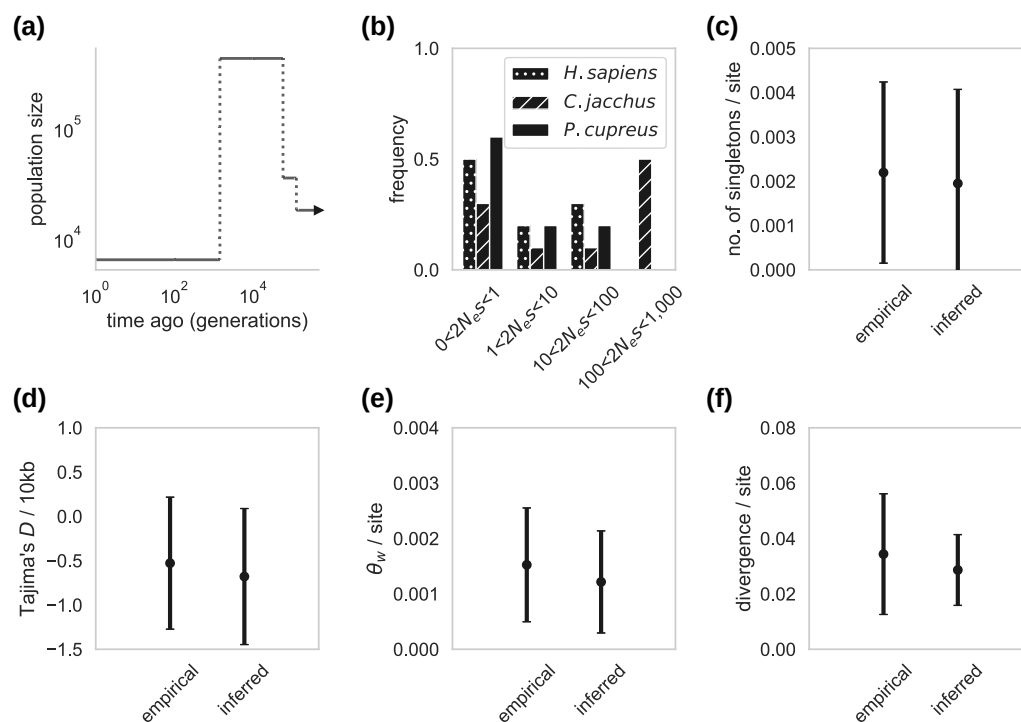


Fig. 1. DFE model-fitting results and summaries. a) The (high mutation rate) demographic history previously inferred for this population (re-drawn from Terbot et al. 2026). b) Comparison between the best-fitting discrete DFE in the coppery titi monkey (*P. cupreus*) inferred in this study (black filled bars), the DFE inferred in another platyrrhine, the common marmoset (*C. jacchus*), by Soni et al. (2025b) (striped bars), and the DFE in humans (*H. sapiens*) inferred by Johri et al. (2023) (dotted bars). Exonic mutations were drawn from a DFE comprised of three fixed classes: nearly neutral mutations (ie $2N_{\text{ancestral}} s < 10$), weakly/moderately deleterious mutations ($10 \leq 2N_{\text{ancestral}} s < 100$), and strongly deleterious mutations ($100 \leq 2N_{\text{ancestral}} s$). c–f) Comparisons 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.

be less efficient in coppery titi monkeys and humans relative to common marmosets—an expectation consistent with the inferred DFEs from each species.

Quantifying the Power to Detect Recent Positive Selection and Balancing Selection in Coppery Titi Monkeys

A number of factors impact the statistical power to detect positive and balancing selection, including the action of other neutral and selective processes, as well as the nature of the data itself (eg Johri et al. 2021, 2022a; Soni et al. 2023, 2024; Soni and Jensen 2024, 2025). It is therefore important to quantify the power to detect episodic processes such as recent positive selection and balancing selection within the context of the demographic history of the population in question, the effects of purifying selection and BGS, and patterns of mutation and recombination rate heterogeneity, all specifically accounting for the empirical sample size and data structure. We therefore simulated the

well-fitting demographic model inferred in coppery titi monkeys by Terbot et al. (2026), using the forward-in-time simulator SLiM (Haller et al. 2026). In order to model the effects of purifying selection and BGS, we utilized the DFE inferred in this study, whilst mutation rates were drawn from the empirical fine-scale mutation rate distribution inferred by Soni et al. (2026), and recombination rates were drawn from a normal distribution for each 1 kb window, such that the mean rate across each simulation replicate was equal to the mean genomic rate (Versoza et al. 2026b and see the “Materials and Methods” section for further details).

Figure 2 depicts receiver operating characteristic (ROC) plots across 100 simulated replicates, for selective sweep inference conducted using SweepFinder2 (DeGiorgio et al. 2016), as applied to simulations utilizing two different strengths of selection ($2N_e s = 100$ and 1,000), and four different times since the introduction of the beneficial mutation ($\tau = 0.1, 0.2, 0.5$, and 1.0 N generations). Notably, it is only at the very large

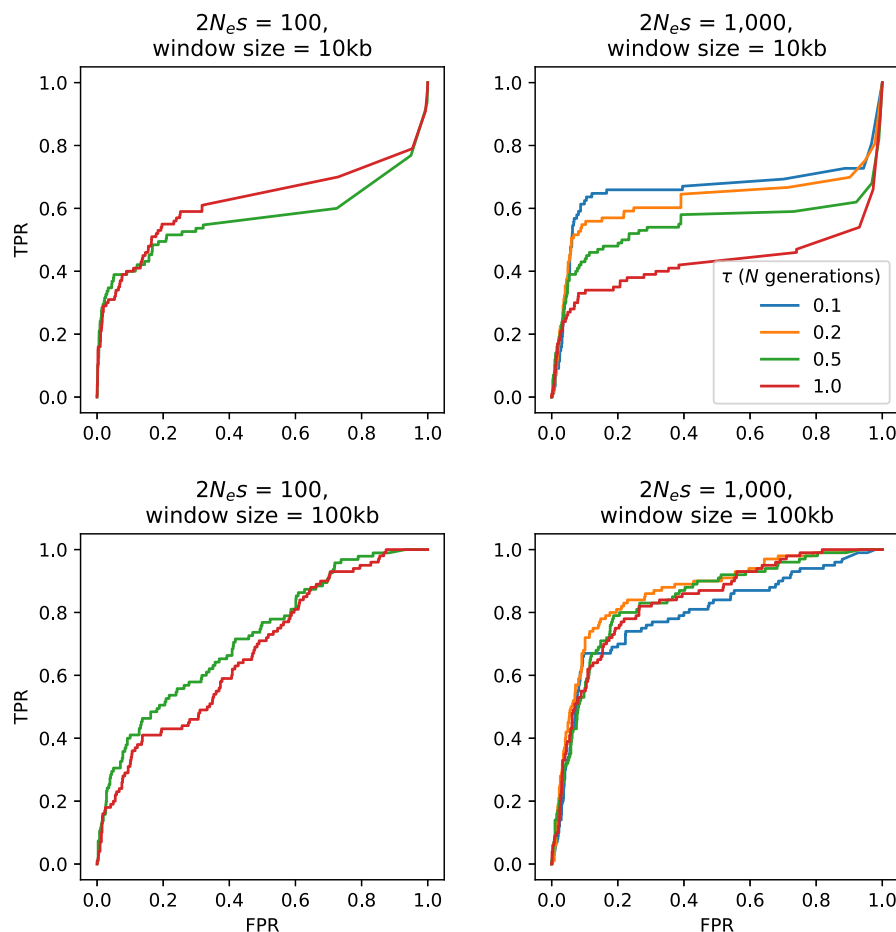


Fig. 2. ROC plots based on 100 simulated replicates under the coppery titi monkey demographic model inferred by Terbot et al. (2026) for selective sweep inference using the SweepFinder2 method. The FPR and the TPR are provided on the x-axis and y-axis of the ROC plots, respectively. Power analyses were conducted across two selection regimes—population-scaled strengths of selection of $2N_e s = 100$ and 1000 ; four times of introduction of the beneficial mutation ($\tau = 0.1, 0.2, 0.5$, and $1.0 N$ generations ago), and window sizes of 10 and 100 kb. Note that no ROC plots could be plotted for simulated scenarios in which no beneficial mutations reached fixation at the time of sampling.

window size of 100 kb that there is reasonable power to detect selective sweeps, and only at the stronger selection value of $2N_e s = 1,000$; at smaller window sizes there is little power (see Fig. S2), owing to the limited variation observed. These results relate to the demographic history of coppery titi monkeys, in which the population has been inferred to experience a severe population size collapse $\sim 0.08N_{\text{ancestral}}$ generations prior to sampling, resulting in a current population size $\sim 1.5\%$ of the size pre-bottleneck (Terbot et al. 2026). Furthermore, only extremely strong selective sweeps would be expected to be identifiable on this bottlenecked genetic background, thereby reducing inference power, as has previously been described (Barton 1998; Jensen et al. 2005; Thornton and Jensen 2007; Poh et al. 2014; Harris and Jensen 2020).

To quantify the statistical power for the detection of balancing selection using the B_{OMAF} method

implemented in the BalLeRMix package (Cheng and DeGiorgio 2020) under our coppery titi monkey evolutionary baseline model, we introduced a mutation experiencing negative frequency-dependent selection at three different times prior to sampling ($\tau_B = 10, 50$, and $75 N$ generations). As shown in Fig. 3, power increases with τ_B ; however, power remains relatively low regardless of window size, even when the balanced mutation has been segregating for extremely long evolutionary timescales. As severe population size reductions will similarly reduce the power to detect balancing selection (Soni and Jensen 2024), this result is consistent with expectation.

Taken together, these power analyses demonstrate the limitations present when attempting to infer these episodic selective processes in recently bottlenecked populations, such as the coppery titi monkey. Moreover, these results highlight the potential perils of more

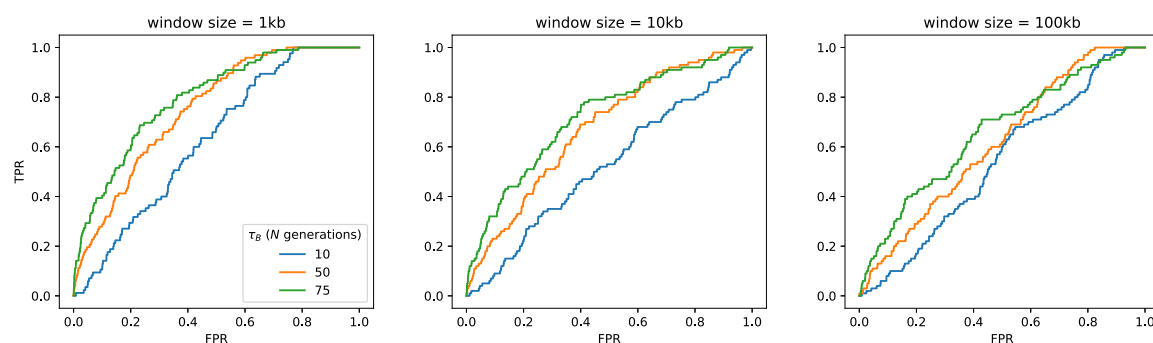


Fig. 3. ROC plots based on 100 simulated replicates under the coppery titi monkey demographic model inferred by Terbot et al. (2026) for balancing selection inference using the B_{OMAF} method. The FPR and the TPR are provided on the x-axis and y-axis of the ROC plots, respectively. Power analyses were conducted across three times of introduction of the balanced mutation ($\tau_B = 10, 50$ and 75 N generations), and window sizes of 1, 10, and 100 kb.

commonly utilized outlier approaches, whereby an arbitrarily chosen threshold is set (eg assuming the 1% or 5% tail of loci in the empirical distribution are selection candidates), which have been demonstrated to result in often exorbitant false-positive rates (Teshima et al. 2006; Thornton and Jensen 2007; Jensen et al. 2008; Jensen 2023; Soni et al. 2023). Instead, following the recommendations of Johri et al. (2022a, 2022b), we constructed an evolutionarily appropriate baseline model accounting for commonly operating evolutionary processes that have been previously quantified in coppery titi monkeys. By performing inference with SweepFinder2 (DeGiorgio et al. 2016) and B_{OMAF} (Cheng and DeGiorgio 2020) on simulated replicates incorporating these species-specific details, we quantified the highest inferred CLR values observed in the absence of positive or balancing selection, and utilized those values as empirical null thresholds (112.3 and 113.8 for the low mutation and high mutation rate demographic models in Sweepfinder2). This in turn resulted in ~4,000 candidates (~0.04% of all tested loci), when performing inference at each single-nucleotide polymorphism (SNP) (Fig. 4, and see Figs. S3 to S24 for chromosome-specific plots). The inferred null thresholds for B_{OMAF} were 236.6 for the low mutation rate model and 280.2 for the high mutation rate model, yielding 3,080 (0.15% of all tested loci) and 1,954 (0.09% of all tested loci) candidates, respectively, with inference performed at every fifth SNP (Fig. 5, and see Figs. S3 to S24 for chromosome-specific plots). For comparison, a naïve outlier approach with a 5% tail threshold would result in 528,301 sweep candidate loci and 105,661 balancing selection candidate loci, whilst a 1% threshold would yield 105,660 sweep candidates and 21,132 balancing selection candidates. However, as demonstrated in our simulations, the vast majority of these “outlier loci” residing in the tails of the

distribution are fully consistent with the baseline model alone, and thus need not invoke positive or balancing selection in order to explain the data (and see Jensen 2023).

Genomic Signatures of Positive Selection in Coppery Titi Monkeys

The identified selective sweep candidate loci mapped to 47 genes (see [Supplementary File S1](#) for max CLR values per candidate gene). Of these, only 19 were considered well-annotated, based on the percent sequence identity, conserved domains and synteny with humans (see the “Materials and Methods” section for further details). Performing gene functional analysis of these 19 candidate genes using DAVID (Ma et al. 2023) yielded no significant GO terms (see [Table S1](#)), nor did analysis based upon the most rapidly diverging genes (see [Table S2](#)).

The gene with the highest CLR values was *PTPN2* (CLR=388.6), identified as a cancer immunotherapy target by Manguso et al. (2017), and involved in the regulation of developmental differentiation of immune cells (Song et al. 2022). Although no previous signatures of positive selection in *PTPN2* have been described in primates, Bello et al. (2023) identified this gene as a selective sweep candidate in an avian species. Given that coppery titi monkeys are a model for neurological development, a notable selective sweep candidate is the gene *MDN1*, a susceptibility gene for epilepsy (Wen et al. 2025), and a previously described sweep candidate in cattle, hypothesized to be related to domestication (Velayudhan et al. 2023). The candidate gene *TMPRSS11D* is also of interest due to high expression levels of this gene in human airways, activating the SARS-CoV-2 spike protein and influenza viruses, a notable result given hypothesized susceptibility amongst non-human primates (Melin et al. 2020); this gene has

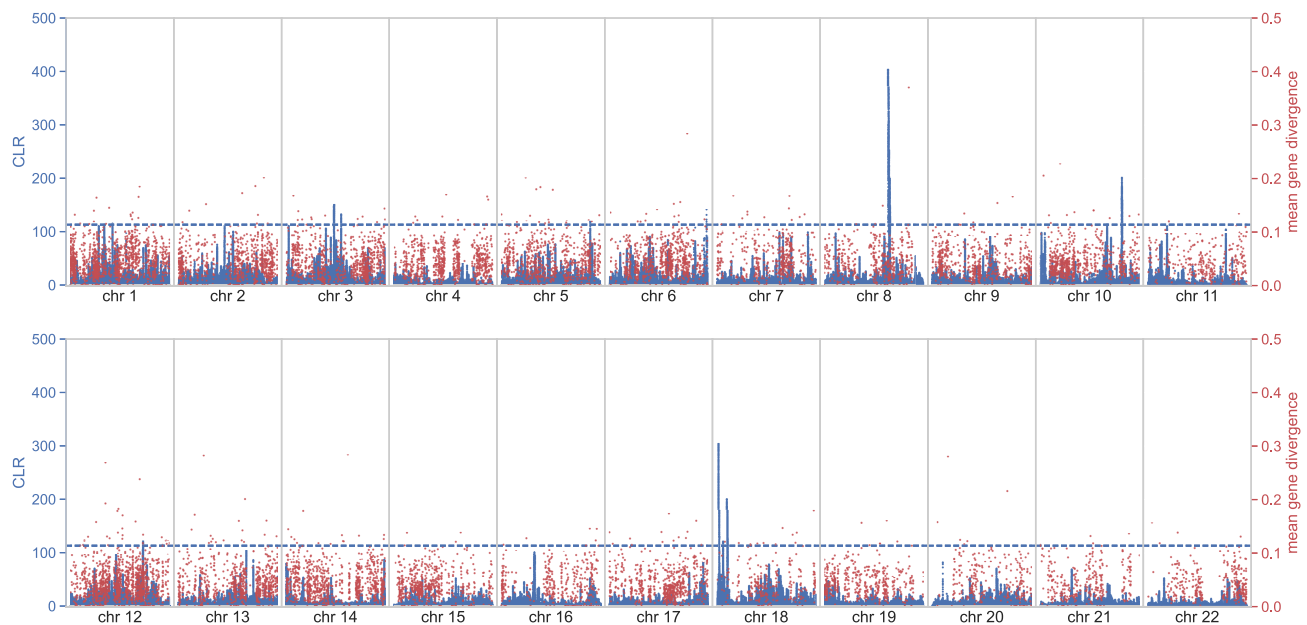


Fig. 4. Genome-wide scans for recent positive selection using SweepFinder2 (shown in blue) relative to empirical exonic divergence (red). The x-axis shows the position along the chromosome, the left y-axis shows the CLR value of the sweep statistic at each SNP, and the right y-axis provides the mean gene divergence. The horizontal blue dashed line represents the null threshold for sweep detection (note that the null thresholds for the low and high mutation rate coppery titi monkey demographic models (Terbot et al. 2026) are extremely similar—112.3 and 113.8, respectively—and are therefore indistinguishable in this figure; see the “Materials and Methods” section for details).

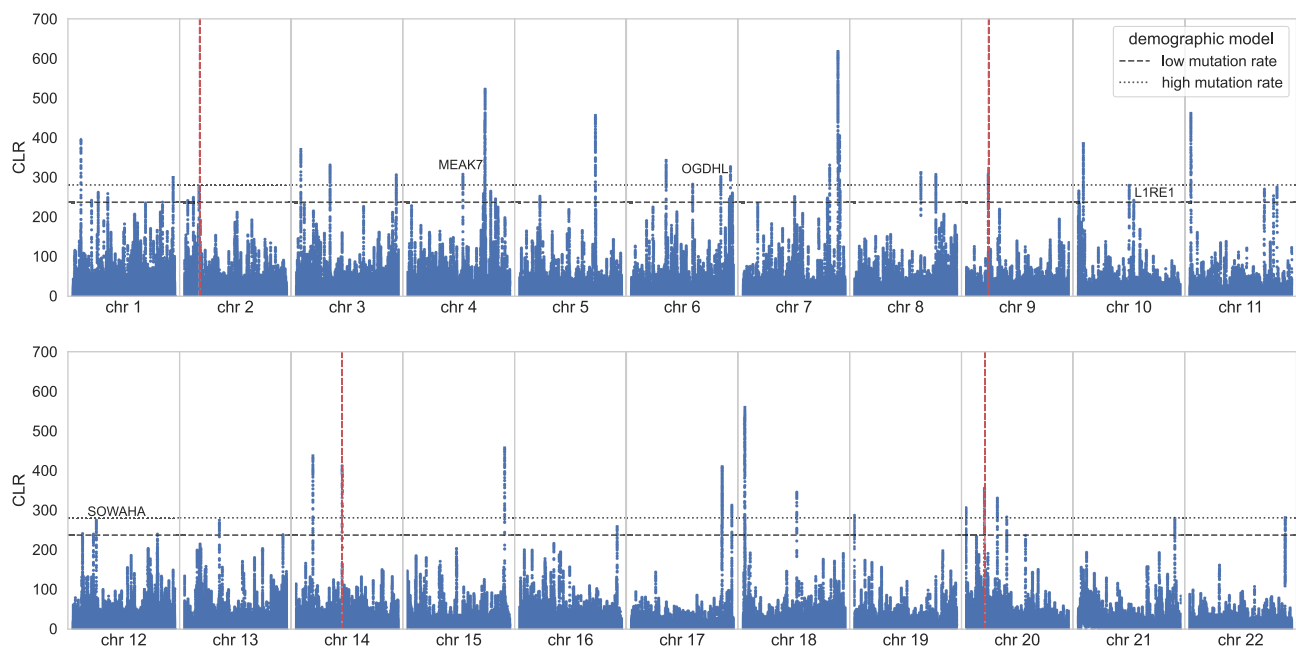


Fig. 5. Genome-wide scans for balancing selection using B_{OMAF} . The x-axis shows the position along the chromosome and the y-axis the CLR value of the balancing selection statistic at each fifth SNP. The horizontal dashed lines represent the null thresholds for the detection of balancing selection, based on the low and high mutation rate coppery titi monkey demographic models (Terbot et al. 2026) (see the “Materials and Methods” section for details). The positions of statistically significant deletions are marked by vertical dashed red lines.

Table 1 Genomic coordinates of deletions identified as balancing selection candidates

Chromosome	Start	End	Length	Maximum CLR	Mean recombination rate (cM/Mb)
2	23,269,995	23,273,368	3,373	240.28	0.426
9	21,882,705	21,885,776	3,071	300.37	0.035
14	58,818,610	58,819,850	1,240	329.20	0.467
20	16,496,560	16,497,174	614	296.54	8.087

The maximum CLR value is the highest CLR inferred by B_{OMAF} for SNPs that are located within the deletion. Recombination rates were calculated as the mean rate across each deletion, from fine-scale recombination maps inferred by Soni et al. (2026).

additionally been recently described as experiencing a hard selective sweep in African populations of humans (Cross et al. 2025).

The candidate gene *ALDH2* plays a key role in the second step of ethanol metabolism, by converting acetaldehyde into acetate (Rodriguez-Zavala et al. 2019). Accumulation of high levels of acetaldehyde have been shown to be both carcinogenic and mutagenic, having adverse effects on the immune system (Ceni et al. 2014; Chen et al. 2014; Zhou et al. 2016; Lee et al. 2017). *ALDH2* has been described as experiencing positive selection in humans (Schaschl et al. 2022), and one may hypothesize that the frugivorous lifestyle of *P. cupreus* may potentially result in the consumption of fermented fruit. Several genome-wide association studies have shown that the genomic region around *ALDH2* is associated with multiple human diseases including rheumatoid arthritis (Coenen et al. 2009), systemic lupus erythematosus (Bentham et al. 2015), type 1 diabetes (Auburger et al. 2014), hypertension (Yasukochi et al. 2017), and coronary artery disease (Wild et al. 2017). This region is therefore of considerable biomedical interest.

Genomic Signatures of Balancing Selection in Coppery Titi Monkeys

Although the candidate regions from the balancing selection scan mapped to 20 genes (see [Supplementary File S1](#) for the maximum CLR values per candidate gene), only four were considered well-annotated based on the percent sequence identity, conserved domains, and synteny with humans (see the “Materials and Methods” section for further details). Of these, two met the CLR null thresholds for both the low and high mutation rate demographic models. As balancing selection is expected to maintain high levels of haplotype structure, we visualized the candidate regions using LDBlockShow (Dong et al. 2021) to confirm this signature, finding strong LD haplotype blocks maintained in all four of our candidate genes ([Fig. S25](#)). Of the four candidate genes, both *MEAK7* and *SOWAHA* have been identified as selective sweep candidates in other species (Ben-Jemaa et al. 2020), with *MEAK7* being primarily involved in cell proliferation and migration

(Nguyen et al. 2018) (note that this gene is also overexpressed in numerous forms of cancer; Nguyen et al. 2019; Chang et al. 2024), whilst *SOWAHA* is a biomarker for colorectal and epithelial cancers (Yi et al. 2024; Yin et al. 2024).

In the context of the coppery titi monkey’s role as a model for neurodevelopment, the candidate gene *OGDHL* is noteworthy. Biallelic variants in *OGDHL* have been proposed to be associated with a number of neurodevelopmental disorders that result in variable epilepsy, ataxia, spasticity, growth abnormalities, dysmorphism, hearing, and visual impairments, and a number of missense variants have been observed to be at intermediate allele frequencies in humans (Yoon et al. 2017; Yap et al. 2021; Lin et al. 2023)—a pattern potentially consistent with the action of balancing selection (as well as consistent with the observed CLR value).

Genomic segments of ≥ 50 bp in length that alter genomic organization are known collectively as structural variants. Recent work by Versoza et al. (2026c) characterized the landscape of structural variation in the coppery titi monkey genomes, identifying over 13,000 such variants, including insertions, duplications, inversions, and deletions. We identified four of these deletions as strong balancing selection candidates based on our genome scan results (see [Table 1](#) and [Fig. 5](#)). All four deletions reside in genomic regions lacking functional annotation; moreover, three reside in regions of the genome with relatively reduced recombination rates (Soni et al. 2026), though that observation may itself owe to the presence of the segregating variant. Thus, while this is an intriguing observation, insufficient evidence is available to further interpret these structural variant results.

Conclusion

In this study, we have characterized long-term selective dynamics in the coppery titi monkey genome by fitting a DFE to observed patterns of polymorphism and divergence, finding evidence of an increased proportion of effectively neutral mutations relative to both humans and common marmosets. Additionally, applying the first genome-wide scans for positive and balancing selection

in the coppery titi monkey genome based on an evolutionary baseline model incorporating the species-specific population history, mutation rates, and recombination rates, we identified a small number of promising candidate loci, several of which are of notable biomedical significance. Notably, in-depth power analyses also provided a necessary framework in which to interpret these results, demonstrating limited power to detect episodic selective events. Thus, at tolerable false-positive rates, the population genetic environment of this species is such that only relatively recent and strong selective sweeps, and relatively ancient balancing selection, would be expected to be identifiable.

Materials and Methods

Animal Subjects

Animals were maintained at the California National Primate Research Center (Davis, CA). This study was performed in compliance with all regulations regarding the care and use of captive primates, including the NIH Guidelines for the Care and Use of Animals and the American Society of Primatologists' Guidelines for the Ethical Treatment of Nonhuman Primates. Procedures were approved by the UC-Davis Institutional Animal Care and Use Committee (protocol 22523).

Whole Genome, Population-level Data

We isolated genomic DNA from peripheral blood samples of six unrelated coppery titi monkey individuals, obtained during routine veterinary examinations. For each sample, we prepared a 150 bp paired-end sequencing library using a PCR-free protocol, and sequenced the resulting libraries on an Illumina NovaSeq 6000 system (Illumina, San Diego, CA, USA). We trimmed remnant adapter sequences from the 3'-end of the raw reads using fastp v.0.24.0 (Chen et al. 2018). Using BWA-MEM v.0.7.15 (Li 2013), we aligned the adapter-trimmed reads to the coppery titi monkey genome assembly (GenBank #: GCA_040437455.1; Pfeifer et al. 2024), marking split reads (-M). In line with established best-practice workflows (Pfeifer 2017), we first identified and flagged read duplicates using the *MarkDuplicates* module available via the Genome Analysis Toolkit (GATK) v.4.4.0 (van der Auwera and O'Connor 2020) to eliminate high-coverage support arising from optical and sampling biases, and then performed base quality score recalibration using the *BaseRecalibrator* and *ApplyBQSR* modules to correct for systematic biases in machine-reported base quality scores. Recalibrations were based on a high-confidence training dataset of graph-based, high-precision genotypes previously obtained from the colony

(Versoza et al. 2026a). Using these pre-processed read mappings, we called both polymorphic and monomorphic sites in each sample using the *HaplotypeCaller* module (with "-ERC BP_RESOLUTION" enabled), applying a minimum mapping quality threshold of 40 ("-minimum-mapping-quality 40") and disabling PCR error correction ("-pcr-indel-model NONE"). Afterward, we combined per-sample outputs using the *CombineGVCFs* module and performed joint genotyping across all samples using the *GenotypeGVCFs* module (with the "-all-sites" option enabled). To improve genotyping accuracy, polymorphic sites were re-genotyped by generating local variation graphs with GraphTyper v.2.7.2 (Eggertsson et al. 2017).

We focused our downstream analyses on autosomal, biallelic SNPs and monoallelic sites genotyped in every sample that passed built-in quality criteria and that were located outside of regions exhibiting an unusual depth of coverage (defined here as less than half (<0.5x), or more than twice (>2x), the mean depth per sample), harboring nearby insertions or deletions (within 5 bp), or showing a low mappability (based on a 150bp-read SNPable mask with a stringency parameter of 1; <https://lh3lh3.users.sourceforge.net/snpable.shtml>).

Calculating Exonic Divergence Between Coppery Titi Monkeys and Humans

To calculate divergence between *P. cupreus* and *H. sapiens*, we replaced the scaffold-level *P. cupreus* genome assembly contained in the 447-way multiple species alignment (Zoonomia Consortium 2020; Kuderna et al. 2024) with the highly contiguous, chromosome-level, annotated genome assembly of Pfeifer et al. (2024), following the approach of Soni, Versoza et al. (2026) via HAL v.2.2 (Hickey et al. 2013) and Cactus v.2.9.2 (Armstrong et al. 2020). We then extracted the sub-alignment containing the updated coppery titi monkey genome, the human genome, and the reconstructed ancestor (PrimatesAnc003), identifying fixed differences between coppery titi monkeys and humans via HAL's *halSnps* function. Specifically, we focused on those fixed differences where coppery titi monkeys exhibited a different allele than humans and PrimatesAnc003, with the latter two sharing the same allele. We excluded sites containing gaps or missing nucleotides or that were polymorphic in either species, and limited our analyses to regions in which we could confidently map the short-read data (see the "Whole genome, population-level data" section for details). These cautionary steps are necessary due to the relatively long divergence time between humans and coppery titi monkeys. Finally, we extracted exonic sites in order to calculate divergence per exon.

DFE Inference

We ran forward-in-time simulations in SLiM v.5.0.1 (Haller et al. 2026), under the demographic model of Terbot et al. (2026) inferred using a mutation rate of 1.07×10^{-8} per base pair per generation (/bp/gen) (referred to as the “high mutation rate model” throughout this study), sampling six diploid individuals (ie $n = 12$) to match the empirical data. Briefly, an ancestral population of size $N_{\text{ancestral}} = 17,807$ individuals expanded to $\sim 35,000$ individuals 122,126 generations ago, before expanding again to $\sim 425,000$ individuals 55,881 generations ago. Finally, the population underwent a severe contraction 1,406 generations ago, to the current population size of 6,330 individuals (for additional details, see Terbot et al. 2026). Following a $10 N_{\text{ancestral}}$ generation burn-in, we simulated a divergence phase between *P. cupreus* and humans of 32 million years (Glazko and Nei 2003)—assuming a generation time of 6 years (Pacifi et al. 2013; Perez et al. 2013)—after which the demographic model was simulated. In brief, 100 replicates of a 618 bp region were simulated; this region length is the mean empirical exon length, when limiting to those exons > 500 bp in length. We drew mutation and recombination rates from a normal distribution for each simulated exon, using a mean mutation rate of 1.07×10^{-8} /bp/gen and a mean recombination rate of 1.00×10^{-8} /bp/gen (ie the rates used by Terbot et al. (2026) to infer the *P. cupreus* demographic model), such that the mean rate across all 100 simulation replicates was equal to these mean rates. We drew mutations from a DFE comprised of three fixed classes (Johri et al. 2020): nearly neutral mutations ($2N_{\text{ancestral}} s < 10$, where $N_{\text{ancestral}}$ is the ancestral population size, and s is the reduction in fitness of the mutant homozygote relative to the wild-type), weakly/moderately deleterious mutations ($10 \leq 2N_{\text{ancestral}} s < 100$), and strongly deleterious mutations ($100 \leq 2N_{\text{ancestral}} s$).

To infer the best-fitting DFE, we performed a grid search by simulating under varying proportions of mutations in each DFE category and comparing the fit of the mean and variance of four summary statistics between our empirical and 100 simulation replicates: 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 pylib-seq v.1.8.3 (Thornton 2003) to calculate all summary statistics except exonic divergence, which was calculated as the number of fixations that occurred in our simulated population post-burn-in. This approach enabled direct comparison to the empirically observed number of substitutions along the *P. cupreus* branch, following the split from humans. Starting with the DFEs inferred by Johri et al. (2023) for humans and Soni et al. (2025c)

for the common marmoset (ie DFEs inferred using a similar approach in a species with a similar effective population size as well as another platyrrhine, respectively), we performed manual grid step iterations based upon the fit of the DFE to the data, until a visually consistent fit of the mean and variance of our chosen summary statistics between our empirical and simulated data was identified.

Power Analyses

To evaluate how much statistical power exists to detect episodic selective events in our *P. cupreus* population, we performed 100 simulation replicates under the high mutation rate demographic model inferred by Terbot et al. (2026) (see the “DFE inference” section for details), using a 10-fold scaling reduction in population size, sampling six diploid individuals. We simulated 33 functional regions—comprised of nine 130 bp exons, separated by introns of size 1,591 bp. Functional regions were flanked by intergenic regions of length 16,489 bp (following Soni et al. 2025b). The number of functional regions chosen were such that the total simulated length was ~ 1 Mb. Exonic mutations were drawn from the DFE inferred in this study, whilst intronic and intergenic mutations were selectively neutral. In order to model mutation rate heterogeneity, mutation rates were drawn from the empirical fine-scale mutation rate distribution inferred by Soni et al. (2026) for each 1 kb window, such that the mean rate across each simulated replicate was equal to the (10-fold rescaled) mean rate of 1.07×10^{-7} /bp/gen. Recombination rates were drawn from a normal distribution with a mean of 1.00×10^{-7} /bp/gen, the mean rescaled rate across all windows per simulation replicate. In each replicate, a single positively selected mutation was introduced, with two possible beneficial population-scaled selection coefficients $2N_{\text{ancestral}} s = [100, 1000]$, and four times of introduction of the beneficial mutation in $N_{\text{ancestral}}$ generations $\tau = [0.1, 0.2, 0.5, 1.0]$, for selective sweep analyses. For balancing selection analysis, we introduced the balanced mutation at $\tau_b = [10, 50, 75] N_{\text{ancestral}}$ generations prior to the end of the simulation (ie representing time segregating prior to sampling). The balanced mutation experienced negative-frequency dependent selection, with the selection coefficient of the balanced mutation, S_{bp} , dependent on its frequency in the population (F_{eq}) and its equilibrium frequency (here modeled as 0.5): $S_{bp} = 0.5 - F_{eq}$. Simulations were restarted if the beneficial mutation failed to fix or was lost from the population in the case of selective sweeps, or if the balanced mutation was fixed or lost from the population (restarting at the time of introduction of the balanced mutation).

To quantify statistical power, we performed genome scans on this simulated data using SweepFinder2 (DeGiorgio et al. 2016) and B_{OMAF} (Cheng and DeGiorgio 2020). SweepFinder2 inference was performed at each SNP, whilst B_{OMAF} inference was performed at every fifth SNP. The following SweepFinder2 command was used: *SweepFinder2 -lu GridFile FreqFile SpectFile OutFile*. For B_{OMAF} the following command was used: *python BallerMix+_v1.py -i InputFile -spect SfsFile -o OutFile -s 5 -usePhysPos -MAF-noSub -rec 1e-5*

ROC plots were generated by combining loci into genomic windows of sizes [100, 1,000, 10,000, 100,000] bp for selective sweep inference, and sizes [1,000, 10,000, 100,000] bp for balancing selection inference, with the highest CLR per window utilized for assessing whether the window was a true positive, false positive, true negative, or false negative.

Generating Null Thresholds for Genome Scans of Recent Positive Selection and Balancing Selection

We simulated the two *P. cupreus* demographic models inferred by Terbot et al. (2026) in order to generate genome-wide null thresholds for selection inference. For details of the demographic model inferred using the higher mutation rate of 1.07×10^{-8} /bp/gen, see the section above entitled “DFE inference.” We again sampled six diploid individuals (ie $n = 12$), to match the empirical data. The demographic model inferred using the lower end of mutation rate estimates (0.488×10^{-8} /bp/gen; referred to as the “low mutation rate model” throughout this study)—generated for comparison and to account for underlying uncertainty—also involves three instantaneous population size changes. However, the ancestral population size in this model is $N_{ancestral} = 169,079$, with a population contraction to $\sim 45,000$ individuals occurring 351,168 generations ago, followed by an expansion to almost 2 million individuals that occurred 131,015 generations ago. Finally, under this model, the population undergoes a severe population contraction 3,160 generations ago, to the present-day size of 12,338 individuals (for additional details, see Terbot et al. 2026). These demographic models were inferred on a dataset in which functional regions, as well as the 10 kb flanking functional regions, were masked to avoid the biasing effects of purifying selection and BGS (Charlesworth et al. 1993; Johri et al. 2020). We therefore simulated the full masked length of each chromosome, with 10 replicates simulated for each chromosome. We drew mutation and recombination rates from a uniform distribution for each simulated replicate, using a mean mutation rate of 1.07×10^{-8} /bp/gen or 0.488×10^{-8} /bp/gen for the respective

demographic model, and a mean recombination rate of 1.00×10^{-8} /bp/gen, such that the mean rate per simulation replicate was equal to these mean rates.

To generate genome-wide null thresholds, we ran SweepFinder2 (DeGiorgio et al. 2016) and the B_{OMAF} balancing selection approach (Cheng and Degiorgio 2020) on our simulated data. We ran each software using the same commands outlined in the preceding section (see the section entitled “Power analyses”). Because we lacked information on the polarization of SNPs, allele frequencies were folded, and only polymorphic sites were considered. The highest CLR value across all null model simulations was set as the null threshold for selection inference, given that this observation represents the highest value generated in the absence of selective sweeps or balancing selection, thereby minimizing the downstream number of false positives.

Inferring Recent Positive Selection and Balancing Selection in the *P. cupreus* Genome

We applied the inference approach discussed in the preceding section on our empirical data, utilizing the same commands used for the generation of null thresholds and for power analyses. Loci at which the empirical CLR was greater than the null threshold value were considered putative selection candidates. We identified candidate genes in which these loci were located. To verify the quality of the gene annotations for these candidate genes, we ran a BLASTx (Altschul et al. 1990) search on the coding sequences of our candidate genes to identify the best hits in humans. We then ran BLASTp on the best hits to verify whether the best hit in coppery titi monkeys returned the original sequence. We also compared the number of conserved domains across all protein coding sequences in a given gene, to the number of conserved domains in the best hits in humans, using CD-Search (Marchler-Bauer and Bryant 2004). Finally, we checked for synteny by comparing the flanking genes for each candidate gene between coppery titi monkeys and humans. Only genes with an amino acid identity $>85\%$, sequence divergence $<15\%$, close alignment between the number of conserved domains between coppery titi monkeys and humans, and synteny were considered for functional analysis.

We manually curated gene functions and expressions via the NCBI database (Sayers et al. 2022) and the Expression Atlas (Madeira et al. 2022), respectively. Finally, we performed gene ontology analysis (The Gene Ontology Consortium et al. 2023) via the Database for Annotation, Visualization, and Integrated Discovery (DAVID; Ma et al. 2023).

Visualizing Haplotype Structure

We used the LDBlockShow software (Dong et al. 2021) to generate haplotype structure plots for our balancing selection candidate genes. First, we identified the candidate locus with the highest CLR value for the candidate gene based on the B_{OMAF} genome scans. We then ran LDBlockShow on a region that contained the identified site together with 10 flanking SNPs on each side.

Supplementary Material

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

Acknowledgments

We would like to thank the California National Primate Research Center for providing the coppery titi monkey samples used in this study. Computations were performed on the Open Science Grid (OSG 2015) supported by National Science Foundation awards 2030508 and 2323298, the JetStream2 supercomputer at Indiana University through allocation BIO250421 from the ACCESS program (Boerner et al. 2023) supported by National Science Foundation awards 2138259, 2138286, 2138307, 2137603, and 2138296, and 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 Number R35GM151008 to S.P.P., Award Number R35GM139383 to J.D.J., and the California National Primate Research Center base grant P51OD011107. CJV was supported by the National Science Foundation CAREER Award DEB-2045343 to S.P.P. K.L.B. was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Mental Health of the National Institutes of Health under Award Numbers R01HD092055 and MH125411, and by the Good Nature Institute. 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 has been deposited under NCBI BioProject PRJNA1356510.

Scripts to perform analyses from this study, as well as results files can be found at https://github.com/vivaksoni/titi_selection

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Anderson JA, Vilgalys TP, Tung J. Broadening primate genomics: new insights into the ecology and evolution of primate gene regulation. *Curr Opin Genet Dev.* 2020;62:16–22. <https://doi.org/10.1016/j.gde.2020.05.009>.
- Armstrong J, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature.* 2020;587:246–251. <https://doi.org/10.1038/s41586-020-2871-y>.
- Auburger G. 12q24 locus association with type 1 diabetes: *SH2B3* or *ATXN2*? *World J Diabetes.* 2014;5:316–327. <https://doi.org/10.4239/wjdv5.i3.316>.
- Bales KL, et al. Titi monkeys as a novel non-human primate model for the neurobiology of pair bonding. *Yale J Biol Med.* 2017;90:373–387.
- Bales KL, et al. What is a pair bond? *Horm Behav.* 2021;136:105062. <https://doi.org/10.1016/j.yhbeh.2021.105062>.
- Bales KL, Mason WA, Catana C, Cherry SR, Mendoza SP. Neural correlates of pair-bonding in a monogamous primate. *Brain Res.* 2007;1184:245–253. <https://doi.org/10.1016/j.brainres.2007.09.087>.
- Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. *Genetics.* 2014;196:841–852. <https://doi.org/10.1534/genetics.113.156190>.
- Barton NH. The effect of hitch-hiking on neutral genealogies. *Genet Res (Camb).* 1998;72:123–133. <https://doi.org/10.1017/S0016672398003462>.
- Bataillon T, et al. Inference of purifying and positive selection in three subspecies of chimpanzees (*Pan troglodytes*) from exome sequencing. *Genome Biol Evol.* 2015;7:1122–1132. <https://doi.org/10.1093/gbe/evw058>.
- Bello SF, Lawal RA, Adeola AC, Nie Q. The study of selection signature and its applications on identification of candidate genes using whole genome sequencing data in chicken—a review. *Poult Sci.* 2023;102:102657. <https://doi.org/10.1016/j.psj.2023.102657>.
- Ben-Jemaa S, Mastrangelo S, Lee S-H, Lee JH, Boussaha M. Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. *Sci Rep.* 2020;10:19466. <https://doi.org/10.1038/s41598-020-76576-3>.
- Bentham J, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet.* 2015;47:1457–1464. <https://doi.org/10.1038/ng.3434>.
- Berry AJ, Ajioka JW, Kreitman M. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics.* 1991;129:1111–1117. <https://doi.org/10.1093/genetics/129.4.1111>.

- Bitarello BD, Brandt DYC, Meyer D, Andrés AM. Inferring balancing selection from genome-scale data. *Genome Biol Evol.* 2023;15: evad032. <https://doi.org/10.1093/gbe/evad032>.
- Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. 2023. ACCESS: advancing innovation: NSF's Advanced cyberinfrastructure co-ordination ecosystem: services & support. In: Practice and experience in advanced research computing (PEARC '23), July 23–27, 2023, Portland, OR, USA. ACM, New York, NY, USA.
- Boyko AR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 2008;4: e1000083. <https://doi.org/10.1371/journal.pgen.1000083>.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics.* 1995;140:783–796. <https://doi.org/10.1093/genetics/140.2.783>.
- Byrne H, et al. Phylogenetic relationships of the New World titi monkeys (*Callicebus*): first appraisal of taxonomy based on molecular evidence. *Front Zool.* 2016;13:10. <https://doi.org/10.1186/s12983-016-0142-4>.
- Byrne H, et al. Titi monkey biogeography: parallel Pleistocene spread by *Plecturocebus* and *Cheracebus* into a Post-Pebas Western Amazon. *Zool Scr.* 2018;47:499–517. <https://doi.org/10.1111/zsc.12300>.
- Cagan A, et al. Natural selection in the great apes. *Mol Biol Evol.* 2016;33:3268–3283. <https://doi.org/10.1093/molbev/msw215>.
- Castellano D, Macià MC, Tataru P, Bataillon T, Munch K. Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. *Genetics.* 2019;213: 953–966. <https://doi.org/10.1534/genetics.119.302494>.
- Ceni E, Mello T, Galli A. Pathogenesis of alcoholic liver disease: role of oxidative metabolism. *World J Gastroenterol.* 2014;20: 17756–17772. <https://doi.org/10.3748/wjg.v20.i47.17756>.
- Chang I, et al. Targeting of lysosomal-bound protein mEAK-7 for cancer therapy. *Front Oncol.* 2024;14:1375498. <https://doi.org/10.3389/fonc.2024.1375498>.
- Charlesworth B, Jensen JD. Effects of selection at linked sites on patterns of genetic variability. *Annu Rev Ecol Evol Syst.* 2021;52:177–197. <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 et al. (2022). *Mol Ecol.* 2022;31:4440–4443. <https://doi.org/10.1111/mec.16573>.
- Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics.* 1993;134:1289–1303. <https://doi.org/10.1093/genetics/134.4.1289>.
- Charlesworth D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2006;2:e64. <https://doi.org/10.1371/journal.pgen.0020064>.
- Chen C-H, Ferreira JCB, Gross ER, Mochly-Rosen D. Targeting aldehyde dehydrogenase 2: new therapeutic opportunities. *Physiol Rev.* 2014;94:1–34. <https://doi.org/10.1152/physrev.00017.2013>.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Cheng X, DeGiorgio M. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *Mol Biol Evol.* 2020;37:3267–3291. <https://doi.org/10.1093/molbev/msaa134>.
- Cheng X, DeGiorgio M. BalLeRMix+: mixture model approaches for robust joint identification of both positive selection and long-term balancing selection. *Bioinformatics.* 2022;38: 861–863. <https://doi.org/10.1093/bioinformatics/btab720>.
- Coenen MJH, et al. Common and different genetic background for rheumatoid arthritis and coeliac disease. *Hum Mol Genet.* 2009;18:4195–4203. <https://doi.org/10.1093/hmg/ddp365>.
- Cameron JM. Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genet.* 2014;10: e1004434. <https://doi.org/10.1371/journal.pgen.1004434>.
- Cameron JM. Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philosophical Transactions of the Royal Society B.* 2017;372:20160471. <https://doi.org/10.1098/rstb.2016.0471>.
- Crisi JL, Poh Y-P, Bean A, Simkin A, Jensen JD. Recent progress in polymorphism-based population genetic inference. *J Hered.* 2012;103:287–296. <https://doi.org/10.1093/jhered/esr128>.
- Cross CN, et al. 2025. The evolutionary landscape of host immunity genes involved in respiratory and other immune-related diseases, and their association with severe COVID-19 outcomes [Preprint]. *medRxiv*:2025.08.28.25334640. <https://doi.org/10.1101/2025.08.28.25334640>.
- DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics.* 2016;32:1895–1897. <https://doi.org/10.1093/bioinformatics/btw051>.
- Dong S-S, et al. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefing in Bioinformatics.* 2021;22: bbaa227. <https://doi.org/10.1093/bib/bbaa227>.
- Eggertsson HP, et al. GraphTyper enables population-scale genotyping using pangenome graphs. *Nat Genet.* 2017;49: 1654–1660. <https://doi.org/10.1038/ng.3964>.
- Enard D, Depaulis F, Roest Crollius H. Human and non-human primate genomes share hotspots of positive selection. *PLoS Genet.* 2010;6:e1000840. <https://doi.org/10.1371/journal.pgen.1000840>.
- Ewing GB, Jensen JD. Distinguishing neutral from deleterious mutations in growing populations. *Front Genet.* 2014;5:7. <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;25:135–141. <https://doi.org/10.1111/mec.13390>.
- Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 2007;8:610–618. <https://doi.org/10.1038/nrg2146>.
- Eyre-Walker A, Keightley PD. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol.* 2009;26: 2097–2108. <https://doi.org/10.1093/molbev/msp119>.
- Fay JC, Wu C-I. Hitchhiking under positive Darwinian selection. *Genetics.* 2000;155:1405–1413. <https://doi.org/10.1093/genetics/155.3.1405>.
- Fijarczyk A, Babik W. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol.* 2015;24:3529–3545. <https://doi.org/10.1111/mec.13226>.
- Glazko GV, Nei MM. Estimation of divergence times for major lineages of primate species. *Mol Biol Evol.* 2003;20:424–434. <https://doi.org/10.1093/molbev/msg050>.
- Haller BC, Ralph PL, Messer PW. SLiM 5: eco-evolutionary simulations across multiple chromosomes and full genomes. *Mol Biol Evol.* 2026;43:msaf313. <https://doi.org/10.1093/molbev/msaf313>.
- Harris RB, Jensen JD. Considering genomic scans for selection as co-alescent model choice. *Genome Biol Evol.* 2020;12:871–877. <https://doi.org/10.1093/gbe/evaa093>.

- Heymann EW, Calouro AM, de la Torre S, Vermeer J. *Plecturocebus cupreus* (amended version of 2018 assessment). IUCN Red List Threat Species; 2021.
- Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*. 2013;29:1341–1342. <https://doi.org/10.1093/bioinformatics/btt128>.
- Howell AA, et al. Developing an appropriate evolutionary baseline model for the study of human cytomegalovirus. *Genome Biol Evol*. 2023;15:evad059. <https://doi.org/10.1093/gbe/evad059>.
- Jennewein DM, et al. 2023. The Sol supercomputer at Arizona state university. In: Practice and experience in advanced research computing. Portland OR USA: ACM. p. 296–301. Available from: <https://dl.acm.org/doi/10.1145/3569951.3597573>
- Jensen JD, et al. The importance of the neutral theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution*. 2019;73:111–114. <https://doi.org/10.1111/evo.13650>.
- Jensen JD. Population genetic concerns related to the interpretation of empirical outliers and the neglect of common evolutionary processes. *Heredity (Edinb)*. 2023;130:109–110. <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;170:1401–1410. <https://doi.org/10.1534/genetics.104.038224>.
- Jensen JD, Thornton KR, Andolfatto P. An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. *PLoS Genet*. 2008;4:e1000198. <https://doi.org/10.1371/journal.pgen.1000198>.
- Johri P, et al. The impact of purifying and background selection on the inference of population history: problems and prospects. *Mol Biol Evol*. 2021;38:2986–3003. <https://doi.org/10.1093/molbev/msab050>.
- Johri P, et al. Recommendations for improving statistical inference in population genomics. *PLoS Biol*. 2022a;20: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;215:173–192. <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: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:msad100. <https://doi.org/10.1093/molbev/msad100>.
- Keightley PD, Eyre-Walker A. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*. 2007;177:2251–2261. <https://doi.org/10.1534/genetics.107.080663>.
- Keightley PD, Eyre-Walker A. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B*. 2010;365:1187–1193. <https://doi.org/10.1098/rstb.2009.0266>.
- Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*. 2002;160:765. <https://doi.org/10.1093/genetics/160.2.765>.
- Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217:624–626. <https://doi.org/10.1038/217624a0>.
- Kimura M. The neutral theory of molecular evolution. 1st ed Cambridge University Press; 1983.
- Kinzey W. New world primates: ecology, evolution, and behavior. Aldine de Gruyter; 1997.
- Klein J, Sato A, Nagl S, O'hUigin C. Molecular trans-species polymorphism. *Annu Rev Ecol Evol Syst*. 1998;29:1–21. <https://doi.org/10.1146/annurev.ecolsys.29.1.1>.
- Kuderna LFK, et al. Identification of constrained sequence elements across 239 primate genomes. *Nature*. 2024;625:735–742. <https://doi.org/10.1038/s41586-023-06798-8>.
- Lau AR, et al. Age, pair tenure and parenting, but not face identity, predict looking behaviour in a pair-bonded South American primate. *Animal Behavior*. 2024;217:53–63. <https://doi.org/10.1016/j.anbehav.2024.08.015>.
- Lau AR, Clink DJ, Bales KL. Individuality in the vocalizations of infant and adult coppery titi monkeys (*Plecturocebus cupreus*). *Am J Primatol*. 2020;82:e23134. <https://doi.org/10.1002/ajp.23134>.
- Lee DJ, Lee HM, Kim JH, Park IS, Rho YS. Heavy alcohol drinking downregulates ALDH2 gene expression but heavy smoking up-regulates SOD2 gene expression in head and neck squamous cell carcinoma. *World J Surg Oncol*. 2017;15:163. <https://doi.org/10.1186/s12957-017-1225-1>.
- Leffler EM, et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*. 2013;339:1578–1582. <https://doi.org/10.1126/science.1234070>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [preprint]. *arXiv:1303.3997*. <https://doi.org/10.48550/arXiv.1303.3997>
- Lin S-J, et al. Evaluating the association of biallelic OGDHL variants with significant phenotypic heterogeneity. *Genome Med*. 2023;15:102. <https://doi.org/10.1186/s13073-023-01258-4>.
- Locke DP, et al. Comparative and demographic analysis of orangutan genomes. *Nature*. 2011;469:529–533. <https://doi.org/10.1038/nature09687>.
- Lukas D, Clutton-Brock TH. The evolution of social monogamy in mammals. *Science*. 2013;341:526–530. <https://doi.org/10.1126/science.1238677>.
- Ma L, et al. Database commons: a catalog of worldwide biological databases. *Genomics Proteomics Bioinformatics*. 2023;21:1054–1058. <https://doi.org/10.1016/j.gpb.2022.12.004>.
- Madeira F, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*. 2022;50:W276–W279. <https://doi.org/10.1093/nar/gkac240>.
- Manguso RT, et al. In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature*. 2017;547:413–418. <https://doi.org/10.1038/nature23270>.
- Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res*. 2004;32:W327–W331. <https://doi.org/10.1093/nar/gkh454>.
- McManus KF, et al. Inference of gorilla demographic and selective history from whole-genome sequence data. *Mol Biol Evol*. 2015;32:600–612. <https://doi.org/10.1093/molbev/msu394>.
- Melin AD, Janiak MC, Marrone IIF, Arora PS, Higham JP. Comparative ACE2 variation and primate COVID-19 risk. *Commun Biol*. 2020;3:641. <https://doi.org/10.1038/s42003-020-01370-w>.
- Mendoza SP, Mason WA. Parental division of labour and differentiation of attachments in a monogamous primate (*Callicebus moloch*). *Anim Behav*. 1986;34:1336–1347. [https://doi.org/10.1016/S0003-3472\(86\)80205-6](https://doi.org/10.1016/S0003-3472(86)80205-6).
- 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;128:79–87. <https://doi.org/10.1038/s41437-021-00493-y>.

- Munch K, Nam K, Schierup MH, Mailund T. Selective sweeps across twenty millions years of primate evolution. *Mol Biol Evol*. 2016;33:3065–3074. <https://doi.org/10.1093/molbev/msw199>.
- Nam K, et al. Evidence that the rate of strong selective sweeps increases with population size in the great apes. *Proceedings of the National Academy of Sciences U.S.A.* 2017;114:1613–1618. <https://doi.org/10.1073/pnas.1605660114>.
- Nguyen JT, et al. Mammalian EAK-7 activates alternative mTOR signaling to regulate cell proliferation and migration. *Sci Adv*. 2018;4:eaao5838. <https://doi.org/10.1126/sciadv.aao5838>.
- Nguyen JT, et al. mEAK-7 forms an alternative mTOR complex with DNA-PKcs in human cancer. *iScience*. 2019;17:190–207. <https://doi.org/10.1016/j.isci.2019.06.029>.
- Nielsen R, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*. 2005;3:e170. <https://doi.org/10.1371/journal.pbio.0030170>.
- Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature*. 1973;246:96–98. <https://doi.org/10.1038/246096a0>.
- OSG. 2015. Open Science Data Federation. Available from: <https://osdf.osg-htc.org/>
- Pacifici M, et al. Generation length for mammals. *Nature Conservation*. 2013;5:89–94. <https://doi.org/10.3897/natureconservation.5.5734>.
- Pavlidis P, Metzler D, Stephan W. Selective sweeps in multilocus models of quantitative traits. *Genetics*. 2012;192:225–239. <https://doi.org/10.1534/genetics.112.142547>.
- Perez SI, Tejedor MF, Novo NM, Aristide L. Divergence times and the evolutionary radiation of new world monkeys (Platyrrhini, primates): an analysis of fossil and molecular data. *PLoS One*. 2013;8:e68029. <https://doi.org/10.1371/journal.pone.0068029>.
- Pfeifer SP. From next-generation resequencing reads to a high-quality variant data set. *Heredity (Edinb)*. 2017;118:111–124. <https://doi.org/10.1038/hdy.2016.102>.
- Pfeifer SP, Baxter A, Savidge LE, Sedlazeck FJ, Bales KL. *De novo* genome assembly for the coppery titi monkey (*Plecturocebus cupreus*): an emerging nonhuman primate model for behavioral research. *Genome Biol Evol*. 2024;16:evae108. <https://doi.org/10.1093/gbe/evae108>.
- Poh Y-P, Domingues VS, Hoekstra HE, Jensen JD. On the prospect of identifying adaptive loci in recently bottlenecked populations. *PLoS One*. 2014;9:e110579. <https://doi.org/10.1371/journal.pone.0110579>.
- Prüfer K, et al. The bonobo genome compared with the chimpanzee and human genomes. *Nature*. 2012;486:527–531. <https://doi.org/10.1038/nature11128>.
- Rodríguez-Zavala JS, Calleja LF, Moreno-Sánchez R, Yoval-Sánchez B. Role of aldehyde dehydrogenases in physiopathological processes. *Chem Res Toxicol*. 2019;32:405–420. <https://doi.org/10.1021/acs.chemrestox.8b00256>.
- Sayers EW, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. 2022;50:D20–D26. <https://doi.org/10.1093/nar/gkab1112>.
- Sally A, et al. Insights into hominid evolution from the gorilla genome sequence. *Nature*. 2012;483:169–175. <https://doi.org/10.1038/nature10842>.
- Schaschl H, Gollner T, Morris DL. Positive selection acts on regulatory genetic variants in populations of European ancestry that affect ALDH2 gene expression. *Science Reports*. 2022;12:4563. <https://doi.org/10.1038/s41598-022-08588-0>.
- Schierup MH, Vekemans X, Charlesworth D. The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet Res (Camb)*. 2000;76:51–62. <https://doi.org/10.1017/S0016672300004535>.
- Schmidt JM, De Manuel M, Marques-Bonet T, Castellano S, Andrés AM. The impact of genetic adaptation on chimpanzee subspecies differentiation. *PLoS Genet*. 2019;15:e1008485. <https://doi.org/10.1371/journal.pgen.1008485>.
- Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics*. 2011;189:1427–1437. <https://doi.org/10.1534/genetics.111.131730>.
- Simkin AT, Bailey JA, Gao F-B, Jensen JD. Inferring the evolutionary history of primate microRNA binding sites: overcoming motif counting biases. *Mol Biol Evol*. 2014;31:1894–1901. <https://doi.org/10.1093/molbev/msu129>.
- Simonsen KL, Churchill GA, Aquadro CF. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*. 1995;141:413–429. <https://doi.org/10.1093/genetics/141.1.413>.
- Song J, Lan J, Tang J, Luo N. PTPN2 in the immunity and tumor immunotherapy: a concise review. *Int J Mol Sci*. 2022;23:10025. <https://doi.org/10.3390/ijms231710025>.
- Soni V, et al. 2026. Inferring fine-scale rates of mutation and recombination in the coppery titi monkey (*Plecturocebus cupreus*) [Preprint]. *bioRxiv*:2026.01.13.699361. <https://doi.org/10.64898/2026.01.13.699361>
- Soni V, Jensen JD. Temporal challenges in detecting balancing selection from population genomic data. *G3 (Bethesda)*. 2024;14:jkae069. <https://doi.org/10.1093/g3journal/jkae069>.
- Soni V, Jensen JD. 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 (Bethesda)*. 2025;15:jka019. <https://doi.org/10.1093/g3journal/jka019>.
- Soni V, Johri P, Jensen JD. Evaluating power to detect recurrent selective sweeps under increasingly realistic evolutionary null models. *Evolution*. 2023;77:2113–2127. <https://doi.org/10.1093/evolut/qpaa120>.
- 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*. 2024;16:evae004. <https://doi.org/10.1093/gbe/evae004>.
- Soni V, Pfeifer SP, Jensen JD. Recent insights into the evolutionary genomics of the critically endangered aye-aye (*Daubentonia madagascariensis*). *Am J Primatol*. 2025c;87:e70105. <https://doi.org/10.1002/ajp.70105>.
- Soni V, Terbot JW, 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)*. 2025e;15:jka078. <https://doi.org/10.1093/g3journal/jka078>.
- Soni V, Versoza CJ, 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*. 2025a;87:e70058. <https://doi.org/10.1002/ajp.70058>.
- Soni V, Versoza CJ, Pfeifer SP, Jensen JD. Investigating the effects of chimerism on the inference of selection: quantifying genomic targets of purifying, positive, and balancing selection in common marmosets (*Callithrix jacchus*). *Heredity (Edinb)*. 2025b;134:645–657. <https://doi.org/10.1038/s41437-025-00804-7>.
- Soni V, Versoza CJ, Vallender EJ, Jensen JD, Pfeifer SP. 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*. 2025d;42:msaf119. <https://doi.org/10.1093/molbev/msaf119>.

- Stephan W. Selective sweeps. *Genetics*. 2019;211:5–13. <https://doi.org/10.1534/genetics.118.301319>.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Tataru P, Bataillon T. polyDFE: inferring the distribution of fitness effects and properties of beneficial mutations from polymorphism data. In: Dutheil JY, editors. *Statistical population genomics*. Vol. 2090: Springer US; 2020. p. 125–146. *Methods in Molecular Biology*. Available from: https://link.springer.com/10.1007/978-1-0716-0199-0_6
- Terbot JW, et al. Developing an appropriate evolutionary baseline model for the study of SARS-CoV-2 patient samples. *PLoS Pathog*. 2023;19:e1011265. <https://doi.org/10.1371/journal.ppat.1011265>.
- Terbot JW, et al. 2026. Inferring the demographic history of coppery titi monkeys (*Plecturocebus cupreus*) from high-quality, whole-genome, population-level data [preprint]. *Am J Primatol*. bioRxiv:2026.01.09.698678. <https://doi.org/10.64898/2026.01.09.698678>.
- 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:evae281. <https://doi.org/10.1093/gbe/evae281>.
- Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? *Genome Res*. 2006;16:702–712. <https://doi.org/10.1101/gr.5105206>.
- The Gene Ontology Consortium, et al. The gene ontology knowledgebase in 2023. *Genetics*. 2023;224:iyad031. <https://doi.org/10.1093/genetics/iyad031>.
- Thornton K. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics*. 2003;19:2325–2327. <https://doi.org/10.1093/bioinformatics/btg316>.
- Thornton KR, Jensen JD. Controlling the false-positive rate in multi-locus genome scans for selection. *Genetics*. 2007;175:737–750. <https://doi.org/10.1534/genetics.106.064642>.
- Valeggia CR, Mendoza SP, Fernandez-Duque E, Mason WA, Lasley B. Reproductive biology of female titi monkeys (*Callicebus moloch*) in captivity. *Am J Primatol*. 1999;47:183–195. [https://doi.org/10.1002/\(SICI\)1098-2345\(1999\)47:3<183::AID-AJP1>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1098-2345(1999)47:3<183::AID-AJP1>3.0.CO;2-J).
- van der Auwera G, O'Connor B. *Genomics in the cloud: using Docker, GATK, and WDL in terra*. O'Reilly Media; 2020.
- Velayudhan SM, et al. Selective sweeps in cattle genomes in response to the influence of urbanization and environmental contamination. *Genes (Basel)*. 2023;14:2083. <https://doi.org/10.3390/genes14112083>.
- Versoza CJ, Bales KL, Jensen JD, Pfeifer SP. Characterization of *de novo* germline mutations suggests a strong male mutation bias in coppery titi monkeys (*Plecturocebus cupreus*). *Mol Biol Evol*. 2026a;43:msag112. <https://doi.org/10.1093/molbev/msag112>.
- Versoza CJ, Bales KL, Jensen JD, Pfeifer SP. 2026b. Sex-specific landscapes of crossover and non-crossover recombination in coppery titi monkeys (*Plecturocebus cupreus*) [preprint]. bioRxiv:2026.01.11.698870. <https://doi.org/10.64898/2026.01.11.698870>.
- Versoza CJ, Bales KL, Jensen JD, Pfeifer SP. 2026c. The landscape of structural variation in coppery titi monkeys (*Plecturocebus cupreus*) [Preprint]. bioRxiv:2026.01.13.699302. <https://doi.org/10.64898/2026.01.13.699302>.
- Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 1975;7:256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9).
- Wen Q, et al. *MDN1* variants cause susceptibility to epilepsy. *Acta Epileptol*. 2025;7:17. <https://doi.org/10.1186/s42494-025-00209-3>.
- Wild PS, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J Clin Invest*. 2017;127:1798–1812. <https://doi.org/10.1172/JCI84840>.
- Wu C, Zhao K, Innan H, Nordborg M. The probability and chromosomal extent of trans-specific polymorphism. *Genetics*. 2004;168:2363–2372. <https://doi.org/10.1534/genetics.104.029488>.
- Yap ZY, et al. Bi-allelic variants in OGDHL cause a neurodevelopmental spectrum disease featuring epilepsy, hearing loss, visual impairment, and ataxia. *Am J Hum Genet*. 2021;108:2368–2384. <https://doi.org/10.1016/j.ajhg.2021.11.003>.
- Yasukochi Y, et al. Longitudinal exome-wide association study to identify genetic susceptibility loci for hypertension in a Japanese population. *Exp Mol Med*. 2017;49:e409–e409. <https://doi.org/10.1038/emmm.2017.209>.
- Yi X, et al. 2024. SOWAHA as a cancer suppressor gene influence metabolic reprogramming [preprint]. arXiv:2409.16332. <https://doi.org/10.48550/arXiv.2409.16332>.
- Yin Y, et al. RNA methylation-related genes INHBB and SOWAHA are associated with MSI status in colorectal cancer patients and may serve as prognostic markers for predicting immunotherapy efficacy. *Carcinogenesis*. 2024;45:337–350. <https://doi.org/10.1093/carcin/bgae004>.
- Yoon WH, et al. Loss of nardilysin, a mitochondrial co-chaperone for α -ketoglutarate dehydrogenase, promotes mTORC1 activation and neurodegeneration. *Neuron*. 2017;93:115–131. <https://doi.org/10.1016/j.neuron.2016.11.038>.
- Zhou Y, et al. Alcoholic beverage consumption and chronic diseases. *Int J Environ Res Public Health*. 2016;13:522. <https://doi.org/10.3390/ijerph13060522>.
- Zoonomia Consortium. A comparative genomics multitool for scientific discovery and conservation. *Nature*. 2020;587:240–245. <https://doi.org/10.1038/s41586-020-2876-6>.

Associate editor: Toni Gossmann