

# Inferring the Demographic History of Aye-Ayes (*Daubentonia madagascariensis*) from High-Quality, Whole-Genome, Population-Level Data

John W. Terbot II <sup>1</sup>, Vivak Soni <sup>1</sup>, Cyril J. Versoza <sup>1</sup>, Susanne P. Pfeifer <sup>1,\*†</sup>, Jeffrey D. Jensen <sup>1,\*†</sup>

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

\*Corresponding authors: E-mails: [susanne@spfeiferlab.org](mailto:susanne@spfeiferlab.org); [jeffrey.d.jensen@asu.edu](mailto:jeffrey.d.jensen@asu.edu).

†These authors jointly supervised the project.

Accepted: December 30, 2024

## Abstract

The nocturnal aye-aye, *Daubentonia madagascariensis*, is one of the most elusive lemurs on the island of Madagascar. The timing of its activity and arboreal lifestyle has generally made it difficult to obtain accurate assessments of population size using traditional census methods. Therefore, alternative estimates provided by population genetic inference are essential for yielding much needed information for conservation measures and for enabling ecological and evolutionary studies of this species. Here, we utilize genomic data from 17 individuals—including 5 newly sequenced, high-coverage genomes—to estimate this history. Essential to this estimation are recently published annotations of the aye-aye genome which allow for variation at putatively neutral genomic regions to be included in the estimation procedures, and regions subject to selective constraints, or in linkage to such sites, to be excluded owing to the biasing effects of selection on demographic inference. By comparing a variety of demographic estimation tools to develop a well-supported model of population history, we find strong support for two demes, separating northern Madagascar from the rest of the island. Additionally, we find that the aye-aye has experienced two severe reductions in population size. The first occurred rapidly, ~3,000 to 5,000 years ago, and likely corresponded with the arrival of humans to Madagascar. The second occurred over the past few decades and is likely related to substantial habitat loss, suggesting that the species is still undergoing population decline and remains at great risk for extinction.

**Key words:** aye-aye, *Daubentonia madagascariensis*, primate, demographic inference, population history, population genetics.

## Significance

The inference of population history remains a central component of any understanding of the evolutionary history of a species, and this is of particular importance in highly endangered species such as the aye-aye. In this work, we implement several common demographic inference approaches, ultimately resulting in both an exceptionally well-fitting population history as well as in a valuable comparison of the relative strengths and weaknesses of these popular methodologies. Our presented inference describes a severe and relatively ancient population decline in aye-ayes likely associated with the first human colonization of Madagascar, as well as a far more recent decline likely associated with habitat destruction and fragmentation over only the past few decades. This work highlights the urgent need for continued conservation efforts in this species.

© The Author(s) 2025. 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](mailto: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](mailto:journals.permissions@oup.com).

## Introduction

The aye-aye, *Daubentonia madagascariensis* (Fig. 1a), is one of over 100 species of lemur endemic to the island of Madagascar (IUCN 2024). Initially described in the late 1780s, it is an elusive species in part due to its nocturnal lifestyle and was believed to be extinct until being rediscovered in the 1950s (Gotch 1995; Kay and Kirk 2000). As the sole extant member of the Daubentoniidae family, the aye-aye occupies a unique position as an early diverging lineage of the lemuriform clade (alongside Cheirogaleidae, Indridae, Lepilemuridae, and Lemuridae) having branched roughly 54.9 to 74.7 million years ago (Yoder 1997; Horvath et al. 2008; Perelman et al. 2011; Mclain et al. 2012).

The aye-aye's range spans from the north to the south of the island of Madagascar (Fig. 1b). Yet, as an arboreal primate, it is restricted to the western dry forests and the rainforests found in the east, central-west, and north of the island (Sterling 1994; Louis et al. 2020). Additional populations were introduced to the nearby islands of Ile Roger and Nosy Mangabe off the coast of Madagascar in the 1960s (Petter 1977). One of the greatest threats to the survival of the species is persistent deforestation further reducing and fragmenting the limited viable ecosystems for the aye-aye (Suzzi-Simmons 2023). Such risk factors have contributed to the aye-aye and most other Malagasy lemur species being assessed as vulnerable or endangered by the International Union for the Conservation of Nature (IUCN; IUCN 2024).

Aye-aye population densities are generally thought to be very low (Mittermeier et al. 2010); however, as direct sightings of aye-ayes are rare, population studies on wild census populations typically rely on trace evidence such as feeding damage and tree holes (Sterling 1993; Ancrenaz et al. 1994; Sterling and Richard 1995; Randimbiharirina et al. 2019). The IUCN has stressed an "urgent need for a systematic census of this important flagship species throughout its range" to allow for effective conservation efforts to be established (Louis et al. 2020). Provided that effective sampling and analyses are performed, population genomic assessments of the aye-aye's historical range and population history can thus be essential for filling in these knowledge gaps.

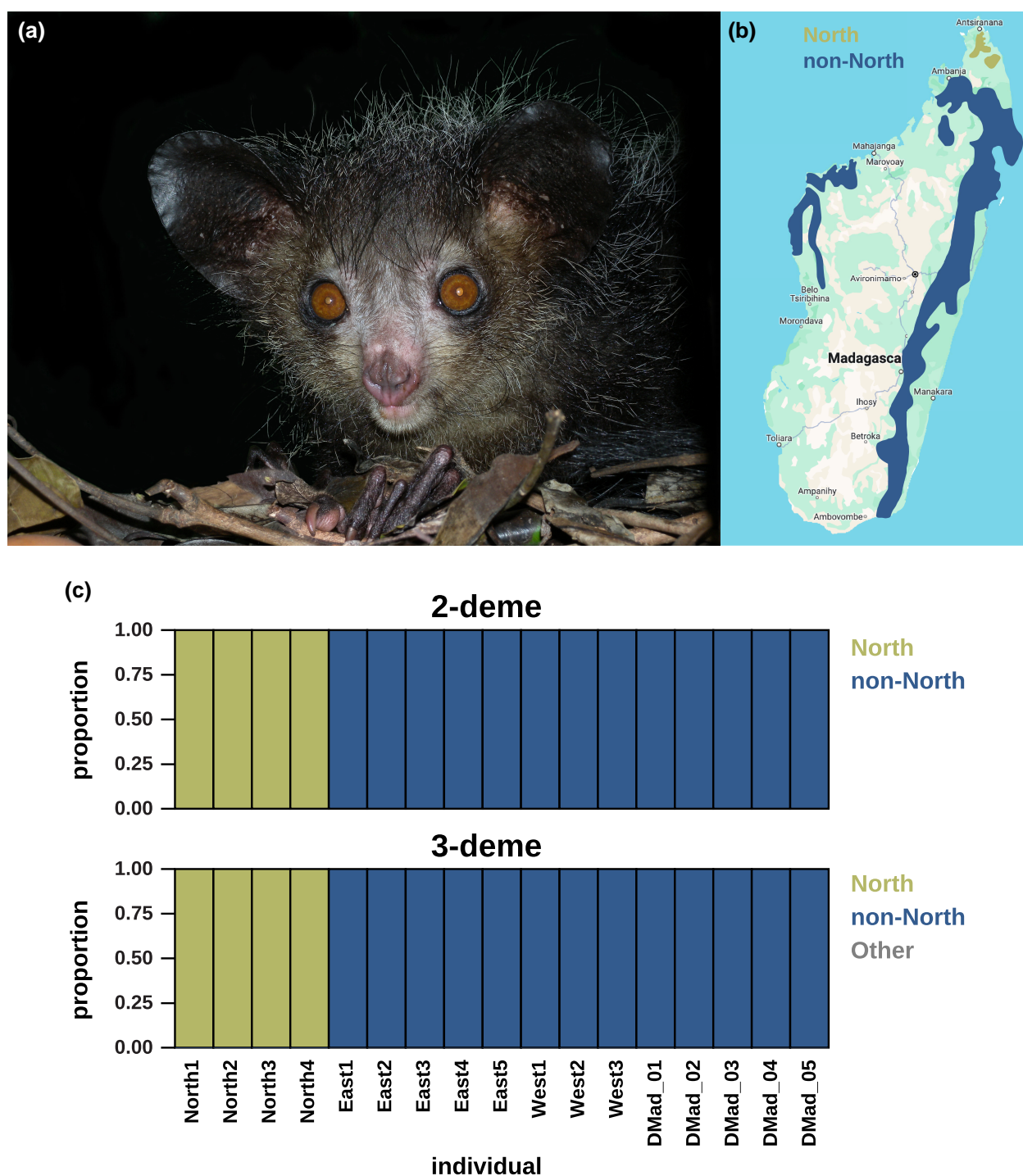
The first population genomic studies performed over a decade ago found that aye-ayes have the lowest genetic diversity among extant lemur taxa, with estimates for the species as a whole ranging from 0.051% based on heterozygous sites in a single individual's genome (Perry et al. 2012b) to 0.073% based on synonymous sites within transcriptome data of 1,175 genes from two individuals (Perry et al. 2012a). Genetic diversity within specific sub-populations corresponds to these estimated ranges with the Northern ( $n = 4$ ), Eastern ( $n = 5$ ), and Western ( $n = 3$ ) sub-populations being 0.054%, 0.057%, and 0.049%,

respectively, based on synonymous sites in low-coverage sequencing data (Perry et al. 2013). These low estimates of genetic diversity further highlight the threat to the long-term prospects of this species.

Contemporary methods of demographic estimation empower us to further evaluate the status of the aye-aye, particularly when combined with high-quality, population-level, whole-genome sequencing data, as we present here. Equally importantly, the recent publication of a fully annotated, chromosome-level genome assembly of the aye-aye (Versoza and Pfeifer 2024) allows us to separate genomic data between functional/constrained sites and putatively neutral sites. This division is essential for inferring population history whilst avoiding mis-inference owing to Hill–Robertson effects (Hill and Robertson 1966); in particular, given that the vast majority of fitness-impacting mutations are deleterious, background selection (BGS) effects (Charlesworth et al. 1993) are expected to be relatively common across the genome (Charlesworth and Jensen 2021).

Two broad types of approaches have been proposed for dealing with the confounding effects of direct selection and selection at linked sites on demographic inference. In coding-dense genomes (as in many primate pathogens, for example), one would expect overlapping effects of both demography and selection across entire genomes (e.g. Irwin et al. 2016; Sackman et al. 2019; Jensen 2021; Morales-Arce et al. 2022; Terbot et al. 2023a,2023b; Howell et al. 2023; Soni et al. 2024a). In other words, because of the large genomic proportion of directly selected sites, there are likely no, or few, genomic regions that are free from selective effects. In such species, the joint and simultaneous inference of population history with selection becomes necessary, and approximate Bayesian approaches have been developed for this purpose (Johri et al. 2020, 2021, 2023). However, owing to the complexity of this joint inference framework, these approaches remain limited to highly simplified demographic models.

Conversely, in coding-sparse genomes (such as those characterizing aye-ayes and other primates), so-called "two-step" inference approaches are possible (see Soni and Jensen 2024b). Here, by first focusing on neutral sites at sufficient distances from functional sites in order to avoid these BGS effects, one may infer a neutral population history. Multiple well-performing inference approaches have been developed for such neutral demographic estimation (e.g. Gutenkunst et al. 2009; Excoffier et al. 2013; and see the review of Beichman et al. 2018). In the second step, selective effects at and around functional sites may be quantified, within the context of this inferred demography. Importantly, this framework for developing an appropriate evolutionary baseline model—a key component of which is the history of population size change, structure, and gene flow characterizing the species in question—has been repeatedly shown to be essential for reducing false-positive



**Fig. 1.** a) An aye-aye from Madagascar; photo by Frank Vassen (Creative Commons CC BY 2.0 DEED). b) Geographic distribution of the aye-aye (*D. madagascariensis*) in Madagascar, with populations in the Northern range shown in green and in the non-Northern range in blue. Range information was obtained from the International Union for the Conservation of Nature Annual Report 2020 (Louis et al. 2020). c) Admixture plots based on the results obtained from fastSTRUCTURE. Each individual is represented by a colored bar representing the demes (with the North population shown in green, the non-North population in blue, and the “ghost” population in gray).

rates when scanning for loci targeted by recent positive or balancing selection (e.g. Barton 1998; Jensen et al. 2005; Poh et al. 2014; Matuszewski et al. 2018; Harris and

Jensen 2020; Charlesworth and Jensen 2022; Jensen 2023; Soni et al. 2023; Soni et al. 2024b; Soni and Jensen 2024a; and see the reviews of Johri et al. 2022a,2022b).

Improving our understanding of population dynamics in the aye-aye thus enables both a better assessment of their conservation status and evolutionary history, but will also allow for future study of the selective dynamics of the species as well. Additionally, the genomic resources resulting from these population-level sequence analyses will allow aye-ayes to serve as an invaluable outgroup for future primate analyses concerned with the study of virtually any evolutionary process or outcome, be it mutation, recombination, or distributions of fitness effects, to name but a few.

## Results and Discussion

### Whole-Genome, Population-Level Data

To study the demographic history of the aye-aye, we combined publicly available low-coverage (7.1 to 10.6×) whole-genome sequencing data for 12 wild-caught individuals (Perry et al. 2013) with the genomes of five individuals housed at the Duke Lemur Center (DLC) that were newly sequenced at high-coverage (>50×) for this study (supplementary table S1, Supplementary Material online). We mapped reads originating from each individual to the recently released high-quality, chromosome-level reference assembly for the aye-aye (Versoza and Pfeifer 2024), noting a much-improved mapping performance compared to the earlier reference assembly (Perry et al. 2012b) for the previously collected samples (94.5% compared to 84.9%; see supplementary table S1, Supplementary Material online in Perry et al. 2013). We subsequently called both variant and invariant sites according to the standard quality-control procedures outlined in the Genome Analysis Toolkit (GATK)'s Best Practices for non-model organisms (van der Auwera and O'Connor 2020). Following recommended guidelines in the field (Pfeifer 2017), we additionally excluded regions prone to mis-alignment and poor sequencing, yielding a data set of 550,630 biallelic, autosomal single nucleotide polymorphisms (SNPs) with a transition-transversion ratio of 2.47 in the accessible genome (supplementary table S2, Supplementary Material online). Lastly, as the effects of both direct selection and BGS have been shown to bias demographic inference (e.g. Ewing and Jensen 2014, 2016; Johri et al. 2020, 2021; and see the reviews of Charlesworth and Jensen 2021, 2024), we limited this data set to putatively neutral regions sufficiently distant from functional sites prior to any downstream analyses.

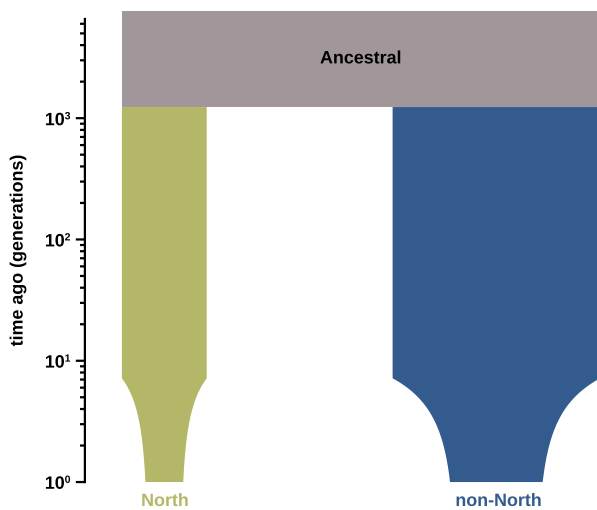
### Preliminary Estimation of Population Structure and Demography

Using marginal likelihood values produced by fast STRUCTURE, population structure between two demes was found to be most likely ( $k=1$ ,  $L=-0.92860$ ;  $k=2$ ,  $L=-0.90683$ ;  $k=3$ ,  $L=-0.90684$ ;  $k=4$ ,  $L=-0.90873$ ;  $k=5$ ,  $L=-0.90685$ , with  $k$  being the number of demes and

$L$  the marginal likelihood). The assignment of individuals to their respective demes based on admixture proportions was almost complete (Fig. 1c), with all individuals having a 0.999998 to 0.000002 deme assignment. Even for the second most likely structure of three demes, this level of assignment was maintained, with only two demes having individuals assigned to them by majority admixture proportion and the third deme constituting a “ghost” population. The assignment of individuals used in Perry et al. (2013) for two demes matched those previously reported in constituting “North” and “non-North” demes, and all newly sequenced individuals were assigned to the non-North deme. While previous studies found support for a separation between Eastern and Western populations, this study did not find sufficient evidence to support such a division based on this higher quality sequencing data.

Preliminary demographic history estimated using MSMC2 was characterized by a severe bottleneck deep in the aye-aye's past, accompanied by a population split and recent rapid population growth, based upon the inclusion of either all invariant sites or only invariant sites within putatively neutral regions (supplementary fig. S1, Supplementary Material online). The estimated ancestral population size of this ultimately poorly fitting model was ~340,000 when all invariant sites were used and 650,000 when only putatively neutral invariant sites were included. The estimated time of the bottleneck and population split occurred ~8,370 generations ago, resulting in a North deme of ~7,400 individuals and a non-North deme of ~4,000 individuals, based upon putatively neutral invariant sites. The timing of this bottleneck when using all invariant sites was ~6,000 generations ago and resulted in a North deme of ~1,000 individuals and a non-North deme of ~2,000 individuals. Under this inferred model, the population size of both demes would further reduce over time and then experience recent rapid growth (within the past ~3 generations) bringing population sizes back into the hundreds of thousands. This estimated rapid growth with large contemporary population sizes does not correspond well with our current understanding of aye-aye demographics, which is thought to be characterized by small and possibly declining population sizes (Louis et al. 2020). Importantly however, this U-shaped demographic pattern of an ancient bottleneck and recent, explosive growth is a known mis-inference problem with MSMC2 and has been observed in a wide variety of organisms including vervet monkeys, passenger pigeons, elephants, *Arabidopsis*, and grapevines (see discussion of Johri et al. 2021 and additional performance issues resulting in erroneous population histories described in Mazet et al. 2016; Chikhi et al. 2018; Beichman et al. 2017). Notably, this MSMC2 estimated model indeed does not recapitulate our observed empirical data (see Assessment of Estimations Using Simulations).





**Fig. 2.** Diagram of the best-fitting demographic model produced by fastsimcoal2. The North population is shown in green, the non-North population in blue, and the ancestral population in taupe.

### Demographic Estimation Using fastsimcoal2 and $\delta\text{a}\delta\text{i}$

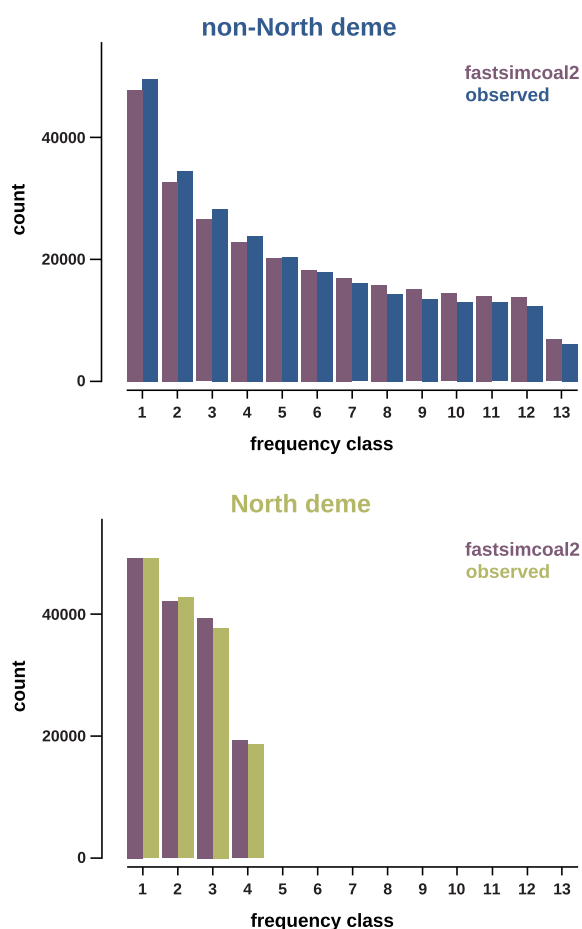
Of the 16 models tested using fastsimcoal2, by far the most likely, based on maximum likelihood comparisons, was the bottleneck + decline model without migration (maximum likelihood =  $-3,406,939$ ; Fig. 2; [supplementary table S3, Supplementary Material](#) online). The next most likely models were the bottleneck + growth without migration model (maximum likelihood =  $-3,407,054$ ; [supplementary fig. S2, Supplementary Material](#) online) and the bottleneck without migration model (maximum likelihood =  $-3,407,079$ ; [supplementary fig. S3, Supplementary Material](#) online). These results remained consistent when model comparisons were performed using the Akaike information criterion (AIC; Akaike 1974), which uses maximum likelihood scores alongside a penalty for the number of parameters in each model ( $\Delta\text{AIC}$  for bottleneck + decline = 0.0,  $\Delta\text{AIC}$  for bottleneck + growth = 229.1, and  $\Delta\text{AIC}$  for bottleneck = 275.9). Despite seemingly large differences in the probabilities of each of these models, the estimated parameters for the bottleneck were surprisingly consistent: from an ancestral population size of  $\sim 23,500$  haploid genomes (i.e. 11,750 diploid individuals), which were reduced to roughly 38% of that size  $\sim 1,100$  generations prior. More specifically, the best parameters for the bottleneck + decline model were an ancestral haploid effective population size of 23,389, experiencing a bottleneck 1,133 generations prior, reducing the total haploid size to 9,276, with 29% of the population in the North deme and the remainder in the non-North deme; more recently, these demes experienced further population decline at a constant rate resulting in an additional reduction in population sizes (to 525 diploid individuals in the North deme and 1,285 diploid individuals

in the non-North deme). Notably, based on a generation time of 5 years (Louis et al. 2020), this would place the initial bottleneck as occurring  $\sim 5,500$  years before present, corresponding with the estimated arrival of humans to Madagascar (Dewar and Richard 2012; Salmona et al. 2017; Hansford et al. 2020; Balboa et al. 2024, but see Mitchell 2020). Moreover, the recent population decline in the past 35 years corresponds to observational reports of a  $\sim 50\%$  reduction in aye-aye populations during the past 20 to 40 years (Louis et al. 2020).

For comparative inference with  $\delta\text{a}\delta\text{i}$ , we ran the Portik et al. (2017) dadi\_pipeline across 17 demographic models to identify the most appropriate (see the [Materials and Methods](#) section for more details). The *bottleneck\_split\_sizeChange* model had the highest mean log-likelihood ( $-1,420.676$ ) and AIC (2,853.352) scores across five rounds of inference using dadi\_pipeline. Notably, this model is similar to the best-fitting model described above as identified by fastsimcoal2. To infer the best-fitting parameters of this model, an additional 1,000 inference rounds were run. The inferred parameter values were broadly in agreement with the fastsimcoal2 inference results, with an ancestral haploid population size of 23,000 (11,500 diploid individuals) and a bottleneck occurring 1,263 generations ago, resulting in daughter populations of size 3,152 haploid genomes (1,576 diploid individuals) in the North population and 8,174 haploid genomes (4,087 diploid individuals) in the non-North population. These populations then underwent a gradual population decline starting 40 generations ago, to estimated current day haploid genome sizes of 1,622 (811 diploid individuals) and 1,344 (672 diploid individuals) for the non-North and North populations, respectively. [Supplementary figs. S4 and S5, Supplementary Material](#) online, provide a schematic of the best-fitting model and parameters, the comparative fit of the empirical and inferred two-dimensional site-frequency spectrum (SFS) and the residuals.

### Assessment of Estimations Using Simulations

The fit of the SFS resulting from the estimated MSMC2 model ([supplementary fig. S6, Supplementary Material](#) online) to the observed data was extremely poor ([supplementary fig. S7, Supplementary Material](#) online), and this estimate can thus comfortably be disregarded as a viable model. By contrast, the resulting SFS for both the best fastsimcoal2 and  $\delta\text{a}\delta\text{i}$  models corresponded quite well to the observed SFS (Fig. 3; [supplementary fig. S8, Supplementary Material](#) online). Furthermore, re-estimated parameters using the simulated SFS as input corresponded well to the distribution of estimated parameters using the observed SFS ([supplementary figs. S9 and S10, Supplementary Material](#) online), confirming model identifiability. For all estimated parameters, the ranges of those



**Fig. 3.** Folded SFS of the best-fitting demographic models produced by fastsimcoal2, compared to the empirically observed SFS.

based on observed compared with those based on simulated data overlapped heavily. Together, this indicates that the estimated model and parameters are robust and consistent with observed genetic variation.

Overall, the similar results between fastsimcoal2 and  $\delta\delta i$  are reassuring. Both approaches inferred models of population splits via bottleneck of approximately the same time period, followed by a recent population decline. This is particularly encouraging given that the  $\delta\delta i$  model was uniquely able to include the possibility of exponential decline or growth. Both methods also inferred similar ancestral and post-bottleneck population sizes, as well as the timing of the population split.

## Conclusions

Inference suggests that the demographic history of the aye-aye is largely tied to the impacts of human activity. This study is consistent with field-based estimates suggesting a very recent decline of ~50% in aye-aye populations

occurring within the past 30 to 50 years. Additionally, we have found evidence for a prior bottleneck of ~50% from the ancestral size that occurred 1,000 to 1,100 generations in the past. A population split into two demes (North and non-North) further accompanied this ancestral bottleneck and is consistent with previously described population structuring (Perry et al. 2013). With an aye-aye generation time of 3 to 5 years, this corresponds well with the estimated timing of the first human settlements on the island of Madagascar, though that colonization date is still a topic of considerable uncertainty with estimates ranging from 1,000 to 10,000 years before present (Hansford et al. 2020). Notably, similar population declines and fragmentations for at least two other species of lemur on Madagascar have been dated to the same general period: the golden crowned sifaka (*Propithecus tattersalli*) and Perrier's sifaka (*Propithecus perrieri*) (Salmona et al. 2017). In addition, demographic estimates of Malagasy populations of the African bush pig (*Potamochoerus larvatus*) include a founding bottleneck also dating to the same period, indicating a possible introduction to the island by humans (Balboa et al. 2024). While climatic changes not driven by human land use may have contributed to the population fragmentation and decline of lemur species on Madagascar over the past several thousand years (Salmona et al. 2017), the impact of humans on Malagasy wildlife, particularly in the past century, is increasingly well supported by such estimates of demographic history. Given that human arrival is generally accepted to impact the population sizes of local megafauna, the incidental evidence from these demographic studies may even be interpreted as providing support for dating the arrival of humans to Madagascar to 3,000 to 5,000 years before present.

Regarding future work on the aye-aye, this study supports two main points. First, conservation measures are at a critical point as the population size of the aye-aye appears to be in decline, with the total effective population size now estimated to be below 2,000 individuals. Second, the well-fitting demographic history inferred here will prove valuable in all future population genomic studies in the species, as this neutral baseline model—accounting for the dominant role of population history in shaping allele frequencies—is a necessary pre-requisite for further inference ranging from the detection of selective sweeps to population-level estimates of genome-wide recombination rates (Johri et al. 2022a).

## Materials and Methods

### Animal Subjects

The DLC's Research Committee (protocol BS-3-22-6) and Duke University's Institutional Animal Care and Use Committee (protocol A216-20-11) approved this study.

We performed this study in compliance with all regulations regarding the care and use of captive primates, including the U.S. National Research Council's Guide for the Care and Use of Laboratory Animals and the U.S. Public Health Service's Policy on Human Care and Use of Laboratory Animals.

### Sample Collection, Preparation, and Sequencing

We extracted genomic DNA (gDNA) from whole blood samples obtained from five aye-ayes (*D. madagascariensis*) housed at the DLC (Durham, NC, USA)—two wild-born females, two wild-born males, and one colony-born female of wild-born parents—using the PureLink Genomic DNA Mini Kit. For each sample, we prepared a 150 bp paired-end library using the NEBNext Ultra II DNA PCR-free Library Prep Kit (New England Biolabs, Ipswich, MA, USA). We checked qualities and quantities of libraries using a High Sensitivity D1000 ScreenTape on an Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA), a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and real-time PCR (Applied Biosystems, Carlsbad, CA, USA) prior to sequencing them on the Illumina NovaSeq platform (Illumina, San Diego, CA, USA) to >50× coverage each (supplementary table S1, Supplementary Material online).

Additionally, we downloaded publicly available, low-coverage (7.1 to 10.6×) whole-genome sequencing data for 12 wild-caught individuals from the NCBI Sequence Read Archive (BioProject PRJNA189299). These data were previously generated by extracting gDNA from samples of liver tissue and whole blood using the Illumina TruSeq DNA Sample Preparation v2 Low Throughput Protocol, shearing the DNA to ~300 bp fragments using a Covaris S2 instrument, enriching the fragments via PCR, and 101 bp paired-end sequencing on the Illumina HiSeq 2000 platform (Illumina, San Diego, CA, USA) (for details, see Perry et al. 2013).

### Variant Calling, Genotyping, and Filtering

To avoid technical artifacts, we first trimmed raw reads using TrimGalore v.0.6.10 (<https://github.com/FelixKrueger/TrimGalore>) with a default quality trimming threshold of 20, and we marked the locations of adapter sequences in both the novel and public whole-genome data using GATK *MarkIlluminaAdapters* v.4.1.8.1 and v.4.2.6.1 (van der Auwera and O'Connor 2020), respectively. We then used the Burrows–Wheeler Aligner (BWA-MEM) v.0.7.17 (Li and Durbin 2009) to map the pre-processed reads to the aye-aye reference assembly (DMad\_hybrid; GenBank accession number: JBFSEQ000000000; Versoza and Pfeifer 2024) and identified duplicate reads using GATK's *MarkDuplicates* v.4.1.8.1 and v.4.2.6.1. We refined read alignments by performing multiple sequence realignments around insertions and deletions using GATK's *RealignerTargetCreator* and *IndelRealigner* v.3.8, recalibrating base quality scores using

GATK's *BaseRecalibrator* and *ApplyBQSR* v.4.2.6.1 together with high-confidence training data obtained from pedigreed individuals housed at the DLC (Versoza et al. 2024a, 2024b), and, finally, carrying out a further round of duplicate removal using GATK's *MarkDuplicates* v.4.2.6.1. We then used high-quality read mappings ("*--minimum-mapping-quality* 40") to call variant and invariant sites using GATK's *HaplotypeCaller* v.4.2.6.1 in the "*-ERC BP\_RESOLUTION*" mode, with the "*--heterozygosity*" parameter set to 0.0005 to reflect the level of genetic diversity previously observed in the species (Perry et al. 2013). To account for differences in the library preparation protocols (see Sample Collection, Preparation, and Sequencing), we set the "*--pcr-indel-model*" to NONE for the novel data (five individuals) and CONSERVATIVE for the previously published data (12 individuals). We merged individual calls using GATK's *CombineGVCFs* v.4.2.6.1 and jointly genotyped them at all sites ("*-all-sites*") using GATK's *GenotypeGVCFs* v.4.2.6.1.

We partitioned the call set into autosomal, biallelic SNPs and invariant sites genotyped in all individuals (AN = 34) and filtered sites according to the GATK Best Practices "hard filter" criteria (i.e. QD < 2.0, QUAL < 30.0, SOR > 3.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0, with acronyms as defined by the GATK package) using GATK's *SelectVariants* and *VariantFiltration* v.4.2.6.1, respectively. Additionally, we excluded regions prone to mis-alignment and/or poor sequencing by removing repetitive elements annotated in the DMad\_hybrid genome assembly (Versoza and Pfeifer 2024) and applying upper and lower cutoffs on the individual depth of coverage ( $0.5 \times DP_{ind}$  and  $2 \times DP_{ind}$ ).

### Putatively Neutral Regions

Prior to performing demographic inference, we limited the data set to putatively neutral regions in order to avoid the confounding effects of direct selection, or selection at linked sites. To this end, following Johri et al.'s (2020, 2023) recommendations for constructing a species-specific evolutionarily appropriate null model, we used the DMad\_hybrid genome annotation (Versoza and Pfeifer 2024), containing a total of 33,346 genes (18,858 of which protein-coding), to remove any sites falling within 10 kb of exons. This trimming was intended to largely eliminate the effects of BGS, which has been shown to bias demographic inference (e.g. Ewing and Jensen 2014, 2016; Johri et al. 2020, 2021; and see the reviews of Charlesworth and Jensen 2021, 2024). Moreover, as additional sites may be experiencing purifying selection (e.g. binding sites; Simkin et al. 2014), sequence elements previously identified as constrained across 239 primate genomes (Kuderna et al. 2024) were excluded from further analyses. For simplicity, we will

refer to the excluded sites as “selectively constrained” and the included sites as “putatively neutral” elsewhere in this study.

### Preliminary Estimation of Population Structure and Demography

As our initial step in estimating the demographic history of the aye-aye, we determined the extent of population structure within our sample. To this end, we first limited our data set to putatively neutral SNPs using VCFtools v.0.1.16 (Danecek et al. 2011) and converted the output to the format required by fastSTRUCTURE (Raj et al. 2014) using PGDSpider v.2.1.1.5 (Lischer and Excoffier 2012). Using this data set, we assessed population structure using the independent-loci model implemented in fastSTRUCTURE v.1.0, thereby allowing individuals to exhibit ancestry from one to five demes (i.e. deme counts ranged from  $k = 1$  to  $k = 5$ ). From these results, we then selected the number of demes and deme assignment of individuals based on the value of  $k$  (deme count) that had the greatest maximum likelihood.

Following this, we processed the sequence information along with filters detailing selectively constrained regions and total site counts which passed quality filtering using a custom Perl script to generate input files for MSMC2 (Schiffels and Durbin 2014) which contained information on putatively neutral SNPs and the number of intervening invariant sites. Invariant sites within selectively constrained regions were alternatively included or silenced from analyses for comparison; in addition, only variant sites within the putatively neutral regions that were either phased for all individuals or homozygous within all individuals were included. Using the deme assignment recommended by fastSTRUCTURE, we analyzed these input files with MSMC2 v.2.1.4 to provide an initial estimation of demographic history.

### Demographic Estimation Using fastsimcoal2

Based on the information obtained from the preliminary analyses using fastSTRUCTURE and MSMC2, we wrote several demographic models as input files for fastsimcoal2 (Excoffier et al. 2013, 2021; Marchi et al. 2024). Specifically, we tested four model sets, all with an asymmetric population split occurring in the past. One model set included a constant rate of population decline since the population split; others allowed a population bottleneck to occur alongside the population split followed by either no subsequent change in population size, a constant rate of recent population growth, or a constant rate of recent population decline. For all model sets, both demes shared the bottleneck ratio, growth rates, and decline rates when relevant. All sets included models with and without migration since the population split; we tested migration between the two demes at both asymmetric and symmetric rates. Note that in the context of fastsimcoal2, “population

size” refers to the effective population size of haploid genomes and not precise census counts of diploid individuals.

The parameter search method used by fastsimcoal2 uses an initial range with a fixed lower bound as a minimum value and a flexible upper bound that can be extended unless explicitly told to use the provided upper bound as a fixed maximum value. Across the model sets tested, nine different parameters were estimated: current population size of the North deme (1,000 to 100,000, all models), current population size of the non-North deme (1,000 to 100,000, all models), time since recent recovery growth or population decline began (1 to 15 generations, bottleneck + growth and bottleneck + decline models), ratio between current and pre-growth or pre-decline population sizes (0.01 to 1 bounded, bottleneck + growth models; 1 to 3 bounded, bottleneck + decline models), time since bottleneck from last event (5 to 250 generations, bottleneck models), ratio between post- and pre-bottleneck population sizes (0.0001 to 0.5 bounded, bottleneck models), time since population decline began (1 to 250 generations, continuous decline models), ratio between pre-decline and current population sizes (1.1 to 100,000, continuous decline models), and migration rates (0.0001 to 0.1, migration models; two parameters with same initial ranges were used for models with asymmetric migration).

Once we prepared model input files, we produced the multidimensional SFS for the putatively neutral SNPs using easySFS v. 0.0.1 (Coffman et al. 2016) and the sequence data prepared previously. From this, we estimated model parameters using fastsimcoal2 v.28 with the following settings: 150,000 simulations per tested parameter, 50 maximization cycles, and resetting parameter searches after 10 consecutive failed cycles (-n 150000 -L 50 -y 10). We estimated parameters with 250 replicates for each model. We then selected the best performing model and parameter combinations based on maximum likelihood and minimizing the difference between estimated and observed likelihoods.

### Demographic Estimation Using $\delta\delta i$

To further validate our demographic estimations, we additionally performed parameter estimation using  $\delta\delta i$  (Gutenkunst et al. 2009). Specifically, we fit 17 two-population demographic models using dadi\_pipeline v.3.1.5 (Portik et al. 2017). These models included simple models, simple models plus instantaneous size change, ancient migration or secondary contact, ancient migration or secondary contact plus instantaneous size change, and two-epoch models with continuous migration, as described in the dadi\_pipeline documentation. Additionally, we added a model of our own, based on the best-fitting fastsimcoal2 model. This model—named *bottleneck\_split\_sizeChange*—consists of a two-population split from the ancestral population via a bottleneck, followed by



exponential size change in each population. Where the model differs from the best-fitting fastsimcoal2 model is that each daughter population was given an individual size change parameter, allowing for either exponential growth or decline. This model consisted of six parameters: *nuB1* (ratio of the population size of the non-North population relative to the ancestral population after the population split), *nuB2* (ratio of the population size of the North population relative to the ancestral population after the population split), *Ts* (time at which the population split occurred, measured in units of  $2N_a$  generations, where  $N_a$  is the ancestral population size), *Tg* (time at which exponential growth/decline began in both daughter populations, measured in units of  $2N_a$  generations), *nuF1* (ratio of the population size of the non-North population relative to the ancestral population at the time of sampling), and *nuF2* (ratio of the population size of the North population relative to the ancestral population at the time of sampling). Following the dadi\_pipeline guidelines, the best-fitting model was inferred based on the AIC score (Akaike 1974).

Once the best-fitting model was inferred (*bottleneck\_split\_sizeChange*—based on consistently lowest AIC values across multiple runs), we reran parameter inference on this model across 1,000 inference replicates, with the optimization function set to 300 maximum iterations, using starting parameter values of *nuB1* = 0.1, *nuB2* = 0.1, *Ts* = 0.05, *Tg* = 0.005, *nuF1* = 0.1, and *nuF2* = 0.1 and using *δaδi*'s *perturb\_params* function to adjust these parameters two-fold up or down, within the following lower and upper parameter bounds:  $0.01 \leq nuB1 \leq 0.5$ ;  $0.01 \leq nuB2 \leq 0.5$ ;  $0.01 \leq Ts \leq 0.1$ ;  $0.001 \leq Tg \leq 0.01$ ;  $0.01 \leq nuF1 \leq 0.5$ ; and  $0.01 \leq nuF2 \leq 0.5$ . The best-fit parameter combination was that which yielded the highest log-likelihood score.

### Assessment of Inference Using Simulations

To assess the fit of our model selection and parameter estimation, we simulated 100 replicates of the best-fitting models and their parameters and visually compared the simulated to the observed SFS. For the best model-parameter combinations from fastsimcoal2 and MSMC2, we simulated SFS with fastsimcoal2; for the best model-parameter combination from *δaδi*, we performed simulations using msprime v. 1.3.2 (Baumdicker et al. 2022). Recombination rates were simulated to match previously observed pedigree data (1e-8 cM/Mb, Versoza et al. 2024c), and mutation rates were set to that previously reported in another lemur species (1.52e-8 per base pair per generation, Campbell et al. 2021; and as modeled in Soni et al. 2024c). For each of the simulated SFS, we estimated parameters using fastsimcoal2 and *δaδi* in a manner similar to the empirical data but performing only 100 (rather than 250) replicates per simulated SFS in fastsimcoal2. We then visually compared the distribution of estimated parameters determined using the observed SFS to

the distribution of best estimated parameters based on the simulated SFS.

### Supplementary Material

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

### Acknowledgements

We would like to thank Erin Ehmke, Kay Welser, and the Duke Lemur Center for providing the aye-aye samples used in this study and members of the Jensen Lab and Pfeifer Lab for their helpful comments and discussion. DNA extraction, library preparation, and Illumina sequencing were conducted at Azenta Life Sciences (South Plainfield, NJ, USA). Parameter estimations using fastsimcoal2 were performed using resources from the Open Science Pool (OSG 2006; Pordes et al. 2007; Sfiligoi et al. 2009; OSG 2015). Computations were performed on the Sol supercomputer at Arizona State University (Jennewein et al. 2023). This is Duke Lemur Center publication #1603.

### 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. and the National Science Foundation under award number DBI-2012668 to the Duke Lemur Center. J.W.T., V.S., and J.D.J. were supported by the National Institutes of Health award number R35GM139383 to J.D.J. C.J.V. was supported by the National Science Foundation CAREER award DEB-2045343 to SPP. A subset of analyses was performed using services provided by the OSG Consortium, which is supported by the National Science Foundation awards OAC-2030508 and OAC-1836650. 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

Sequence data are available under NCBI BioProjects PRJNA189299, PRJNA1085541, and PRJNA1179987. Scripts used for this study are available at [https://github.com/jwterbot2/ayeaye\\_demography/tree/main](https://github.com/jwterbot2/ayeaye_demography/tree/main); the versions, settings, and parameters of the software used in this study are described in the Materials and Methods section.

## Literature Cited

- Akaike H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*. 1974;19(6):716–723. <https://doi.org/10.1109/TAC.1974.1100705>.
- Ancrenaz M, Lackman-Ancrenaz I, Mundy N. Field observations of aye-ayes (*Daubentonia madagascariensis*) in Madagascar. *Folia Primatol*. 1994;62(1-3):22–36. <https://doi.org/10.1159/000156760>.
- Balboa RF, Bertola LD, Brüniche-Olsen A, Rasmussen MS, Liu X, Besnard G, Salmons J, Santander CG, He S, Zinner D, et al. African bushpigs exhibit porous species boundaries and appeared in Madagascar concurrently with human arrival. *Nat Commun*. 2024;15(1):172. <https://doi.org/10.1038/s41467-023-44105-1>.
- Barton NH. The effect of hitch-hiking on neutral genealogies. *Genet Res*. 1998;72(2):123–133. <https://doi.org/10.1017/S0016672398003462>.
- Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, Zhu S, Eldon B, Ellerman EC, Galloway JG, et al. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*. 2022;220(3):iyab229. <https://doi.org/10.1093/genetics/iyab229>.
- Beichman AC, Huerta-Sanchez E, Lohmueller KE. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu Rev Ecol Evol Syst*. 2018;49(1):433–456. <https://doi.org/10.1146/annurev-ecolsys-110617-062431>.
- Beichman AC, Phung TN, Lohmueller KE. Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3 (Bethesda)*. 2017;7(11):3605–3620. <https://doi.org/10.1534/g3.117.300259>.
- Campbell CR, Tiley GP, Poelstra JW, Hunnicutt KE, Larsen PA, Lee HJ, Thorne JL, Dos Reis M, Yoder AD. Pedigree-based and phylogenetic methods support surprising patterns of mutation rate and spectrum in the gray mouse lemur. *Heredity (Edinb)*. 2021;127(2):233–244. <https://doi.org/10.1038/s41437-021-00446-5>.
- Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993;134(4):1289–1303. <https://doi.org/10.1093/genetics/134.4.1289>.
- Charlesworth B, Jensen JD. Effects of selection at linked sites on patterns of genetic variability. *Annu Rev Ecol Evol Syst*. 2021;52(1):177–197. <https://doi.org/10.1146/annurev-ecolsys-010621-044528>.
- Charlesworth B, Jensen JD. Some complexities in interpreting apparent effects of hitchhiking: a commentary on Gompert et al. (2022). *Mol Ecol*. 2022;31(17):4440–4443. <https://doi.org/10.1111/mec.16573>.
- Charlesworth B, Jensen JD. Population genetics. In: Scheiner SM, editor. *Encyclopedia of biodiversity*. 3rd ed. Amsterdam: Elsevier; 2024. p. 467–483.
- Chikhi L, Rodríguez W, Grusea S, Santos P, Boitard S, Mazet O. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity (Edinb)*. 2018;120(1):13–24. <https://doi.org/10.1038/s41437-017-0005-6>.
- Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. Computationally efficient composite likelihood statistics for demographic inference. *Mol Biol Evol*. 2016;33(2):591–593. <https://doi.org/10.1093/molbev/msv255>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Dewar RE, Richard AF. Madagascar: a history of arrivals, what happened, and will happen next. *Annu. Rev. Anthropol*. 2012;41(1):495–517. <https://doi.org/10.1146/annurev-anthro-092611-145758>.
- 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(1):135–141. <https://doi.org/10.1111/mec.13390>.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 2013;9(10):e1003905. <https://doi.org/10.1371/journal.pgen.1003905>.
- Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. Fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics*. 2021;37(24):4882–4885. <https://doi.org/10.1093/bioinformatics/btab468>.
- Gotch A. *Latin Names explained: a guide to the scientific classification of reptiles, birds & mammals*. London, UK: Blandford Press; 1995.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009;5(10):e1000695. <https://doi.org/10.1371/journal.pgen.1000695>.
- Hansford JP, Wright PC, Pérez VR, Muldoon KM, Turvey ST, Godfrey LR. Evidence for early human arrival in Madagascar is robust: a response to Mitchell. *J Isl Coast Archaeol*. 2020;15(4):596–602. <https://doi.org/10.1080/15564894.2020.1771482>.
- Harris RB, Jensen JD. Considering genomic scans for selection as coalescent model choice. *Genome Biol Evol*. 2020;12(6):871–877. <https://doi.org/10.1093/gbe/evaa093>.
- Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res*. 1966;8(3):269–294. <https://doi.org/10.1017/S0016672300010156>.
- Horvath JE, Weisrock DW, Embry SL, Fiorentino I, Balhoff JP, Kappeler P, Wray GA, Willard HF, Yoder AD. Development and application of a phylogenomic toolkit: resolving the evolutionary history of Madagascar's lemurs. *Genome Res*. 2008;18(3):489–499. <https://doi.org/10.1101/gr.7265208>.
- Howell AA, Terbot JW, Soni V, Johri P, Jensen JD, Pfeifer SP. Developing an appropriate evolutionary baseline model for the study of human cytomegalovirus. *Genome Biol Evol*. 2023;15(4):1–12. <https://doi.org/10.1093/gbe/evad059>.
- Irwin KK, Laurent S, Matuszewski S, Vuilleumier S, Ormond L, Shim H, Bank C, Jensen JD. On the importance of skewed offspring distributions and background selection in virus population genetics. *Heredity (Edinb)*. 2016;117(6):393–399. <https://doi.org/10.1038/hdy.2016.58>.
- IUCN. The IUCN red list of threatened species. Version 2024-1; 2024. Available from: <https://www.iucnredlist.org>
- Jennwein DM, Lee J, Kurtz C, Dizon W, Shaeffer I, Chapman A, Chiquete A, Burks J, Carlson A, Mason N, et al. The Sol supercomputer at Arizona State University. In: Practice and experience in advanced research computing 2023: computing for the common good (PEARC '23). New York (NY), USA: Association for Computing Machinery; 2023. p. 296–301.
- Jensen JD. 2021. Studying population genetic processes in viruses: from drug-resistance evolution to patient infection dynamics. In: Bamford DH, Zuckerman M, editors. *Encyclopedia of virology*. Amsterdam: Elsevier. p. 227–232.
- Jensen JD. Population genetic concerns related to the interpretation of empirical outliers and the neglect of common evolutionary processes. *Heredity (Edinb)*. 2023;130(3):109–110. <https://doi.org/10.1038/s41437-022-00575-5>.
- Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. Distinguishing between selective sweeps and demography using

- DNA polymorphism data. *Genetics*. 2005;170(3):1401–1410. <https://doi.org/10.1534/genetics.104.038224>.
- Johri P, Aquadro CF, Beaumont M, Charlesworth B, Excoffier L, Eyre-Walker A, Keightley PD, Lynch M, McVean G, Payseur BA, et al. Recommendations for improving statistical inference in population genomics. *PLOS Biol*. 2022a;20(5):e3001669. <https://doi.org/10.1371/journal.pbio.3001669>.
- Johri P, Charlesworth B, Jensen JD. Toward an evolutionarily appropriate null model: jointly inferring demography and purifying selection. *Genetics*. 2020;215(1):173–192. <https://doi.org/10.1534/genetics.119.303002>.
- Johri P, Eyre-Walker A, Gutenkunst RN, Lohmueller KE, Jensen JD. On the prospect of achieving accurate joint estimation of selection with population history. *Genome Biol Evol*. 2022b;14(7):1–10. <https://doi.org/10.1093/gbe/evac088>.
- Johri P, Pfeifer SP, Jensen JD. Developing an evolutionary baseline model for humans: jointly inferring purifying selection with population history. *Mol Biol Evol*. 2023;40(5):1–14. <https://doi.org/10.1093/molbev/msad100>.
- Johri P, Riall K, Becher H, Excoffier L, Charlesworth B, Jensen JD. The impact of purifying and background selection on the inference of population history: problems and prospects. *Mol Biol Evol*. 2021;38(7):2986–3003. <https://doi.org/10.1093/molbev/msab050>.
- Kay RF, Kirk EC. Osteological evidence for the evolution of activity pattern and visual acuity in primates. *Am J Phys Anthropol*. 2000;113(2):235–262. [https://doi.org/10.1002/1096-8644\(200010\)113:2<235::AID-AJPA7>3.0.CO;2-9](https://doi.org/10.1002/1096-8644(200010)113:2<235::AID-AJPA7>3.0.CO;2-9).
- Kuderna LFK, Ulirsch JC, Rashid S, Ameen M, Sundaram L, Hickey G, Cox AJ, Gao H, Kumar A, Aguet F, et al. Identification of constrained sequence elements across 239 primate genomes. *Nature*. 2024;625(7996):735–742. <https://doi.org/10.1038/s41586-023-06798-8>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 2012;28(2):298–299. <https://doi.org/10.1093/bioinformatics/btr642>.
- Louis E, Sefczek T, Randimbiharirina D, Raharivololona B, Raktondrandry J, Manjary D, Aylward M, Ravelomandrato F. *Daubentonia madagascariensis*. IUCN Red List Threat. Species; 2020.
- Marchi N, Kapopoulou A, Excoffier L. Demogenomic inference from spatially and temporally heterogeneous samples. *Mol Ecol Resour*. 2024;24(1):1–11. <https://doi.org/10.1111/1755-0998.13877>.
- Matuszewski S, Hildebrandt ME, Achaz G, Jensen JD. Coalescent processes with skewed offspring distributions and nonequilibrium demography. *Genetics*. 2018;208(1):323–338. <https://doi.org/10.1534/genetics.117.300499>.
- Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity (Edinb)*. 2016;116(4):362–371. <https://doi.org/10.1038/hdy.2015.104>.
- McLain AT, Meyer TJ, Faulk C, Herke SW, Oldenburg JM, Bourgeois MG, Abshire CF, Roos C, Batzer MA. An Alu-based phylogeny of lemurs (Infraorder: Lemniformes). *PLoS One*. 2012;7(8):e44035. <https://doi.org/10.1371/journal.pone.0044035>.
- Mitchell P. Settling Madagascar: when did people first colonize the world's largest island? *J Isl Coast Archaeol*. 2020;15(4):576–595. <https://doi.org/10.1080/15564894.2019.1582567>.
- Mittermeier R, Louis E Jr, Richardson M, Schwitzer C, Langrand O, Rylands A, Hawkins F, Rajaobelina S, Ratsimbazafy J, Rasoloarison R, et al. *Lemurs of Madagascar*. 3rd ed. Crystal City (VA): Conservation International; 2010.
- Morales-Arce AY, Johri P, Jensen JD. Inferring the distribution of fitness effects in patient-sampled and experimental virus populations: two case studies. *Heredity (Edinb)*. 2022;128(2):79–87. <https://doi.org/10.1038/s41437-021-00493-y>.
- OSG. 2006. OSPool. OSG. <https://doi.org/10.21231/906P-4D78>.
- OSG. 2015. Open science data federation. OSG. <https://doi.org/10.21231/0KVZ-VE57>.
- Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MAM, Kessing B, Pontius J, Roelke M, Rumpel Y, et al. A molecular phylogeny of living primates. *PLoS Genet*. 2011;7(3):e1001342. <https://doi.org/10.1371/journal.pgen.1001342>.
- Perry GH, Louis EE, Ratan A, Bedoya-Reina OC, Burhans RC, Lei R, Johnson SE, Schuster SC, Miller W. Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar. *Proc Natl Acad Sci U S A*. 2013;110(15):5823–5828. <https://doi.org/10.1073/pnas.1211990110>.
- Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, et al. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res*. 2012a;22(4):602–610. <https://doi.org/10.1101/gr.130468.111>.
- Perry GH, Reeves D, Melsted P, Ratan A, Miller W, Michelini K, Louis EE, Pritchard JK, Mason CE, Gilad Y. A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biol Evol*. 2012b;4(2):126–135. <https://doi.org/10.1093/gbe/evr132>.
- Petter J. 1977. The aye-aye. In: His serene highness prince rainer III, Bourne G, editors. *Primate conservation*. New York, NY: Academic Press. p. 37–57.
- Pfeifer SP. From next-generation resequencing reads to a high-quality variant data set. *Heredity (Edinb)*. 2017;118(2):111–124. <https://doi.org/10.1038/hdy.2016.102>.
- Poh Y-P, Domingues VS, Hoekstra HE, Jensen JD. On the prospect of identifying adaptive loci in recently bottlenecked populations. *PLoS One*. 2014;9(11):e110579. <https://doi.org/10.1371/journal.pone.0110579>.
- Pordes R, Petravick D, Kramer B, Olson D, Livny M, Roy A, Avery P, Blackburn K, Wenaus T, Würthwein F, et al. The open science grid. *J Phys Conf Ser*. 2007;78:012057. <https://doi.org/10.1088/1742-6596/78/1/012057>.
- Portik DM, Leache AD, Rivera D, Blackburn DC, Rodel M, Barej MF, Hirschfeld M, Burger M, Fujita MK. Evaluating mechanisms of diversification in a Guineo-Congolian forest frog using demographic model selection. *Mol Ecol*. 2017;26(19):5245–5263. <https://doi.org/10.1111/mec.14266>.
- Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*. 2014;197(2):573–589. <https://doi.org/10.1534/genetics.114.164350>.
- Randimbiharirina D, Sefczek T, Raharivololona B, Andriamalala Y, Ratsimbazafy J, Louis E Jr. Aye-aye (*Daubentonia madagascariensis*). In: Schwitzer C, Mittermeier R, Rylands A, Chiozza F, Williamson E, Byler D, Wich S, Humle T, Johnson C, Mynott H, editors. *Primates in peril: the world's 25 Most endangered primates 2018–2020*. editors. Washington (DC): IUCN SSC Primate Specialist Group, International Primatological Society, Global Wildlife Conservation, and Bristol Zoological Society; 2019. p. 19–21.
- Sackman AM, Harris RB, Jensen JD. Inferring demography and selection in organisms characterized by skewed offspring distributions. *Genetics*. 2019;211(3):1019–1028. <https://doi.org/10.1534/genetics.118.301684>.

- Salmona J, Heller R, Quéméré E, Chikhi L. Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. *Mol Ecol*. 2017;26(19):5203–5222. <https://doi.org/10.1111/mec.14173>.
- Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014;46(8):919–925. <https://doi.org/10.1038/ng.3015>.
- Sfiligoi I, Bradley DC, Holzman B, Mhashilkar P, Padhi S, Wurthwein F. 2009. The pilot way to grid resources using glideinWMS. In: 2009 WRI world congress on computer science and information engineering. Washington, DC: IEEE. p. 428–432.
- 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(7):1894–1901. <https://doi.org/10.1093/molbev/msu129>.
- Soni V, Jensen JD. Temporal challenges in detecting balancing selection from population genomic data. *G3 (Bethesda)*. 2024a;14(6):1–14. <https://doi.org/10.1093/g3journal/jkae069>.
- Soni V, Jensen JD. Inferring demographic and selective histories from population genomic data using a two-step approach in species with coding-sparse genomes: an application to human data. *bioRxiv* 613979. <https://doi.org/10.1101/2024.09.19.613979>, 21 November 2024b, preprint: not peer reviewed.
- Soni V, Johri P, Jensen JD. Evaluating power to detect recurrent selective sweeps under increasingly realistic evolutionary null models. *Evolution (N Y)*. 2023;77:2113–2127. <https://doi.org/10.1093/evolut/qpad120>.
- Soni V, Pfeifer SP, Jensen JD. The effects of mutation and recombination rate heterogeneity on the inference of demography and the distribution of fitness effects. *Genome Biol Evol*. 2024a;16(2):1–13. <https://doi.org/10.1093/gbe/evae004>.
- Soni V, Terbot JW, Jensen JD. Population genetic considerations regarding the interpretation of within-patient SARS-CoV-2 polymorphism data. *Nat Commun*. 2024b;15(1):3240. <https://doi.org/10.1038/s41467-024-46261-4>.
- Soni V, Terbot JW, Versoza CJ, Pfeifer SP, Jensen JD. A whole-genome scan for evidence of recent positive and balancing selection in aye-ayes (*Daubentonia madagascariensis*) utilizing a well-fit evolutionary baseline model. *bioRxiv* 622667. <https://doi.org/10.1101/2024.11.08.622667>, 11 November 2024c, preprint: not peer reviewed.
- Sterling EJ. Patterns of range use and social organization in aye-ayes (*Daubentonia madagascariensis*) on Nosy Mangabe. In: Kappeler PM, Ganzhorn JU, editors. *Lemur social systems and their ecological basis*. editors. Boston (MA): Springer US; 1993. p. 1–10.
- Sterling EJ. Taxonomy and distribution of *Daubentonia*: a historical perspective. *Folia Primatol*. 1994;62(1-3):8–13. <https://doi.org/10.1159/000156758>.
- Sterling EJ, Richard AF. Social organization in the aye-aye (*Daubentonia madagascariensis*) and the perceived distinctiveness of nocturnal primates. In: Alterman L, Doyle GA, Izard MK, editors. *Creatures of the dark: the nocturnal prosimians*. editors. New York (NY): Plenum Press; 1995. p. 439–451.
- Suzzi-Simmons A. Status of deforestation of Madagascar. *Glob Ecol Conserv*. 2023;42:e02389. <https://doi.org/10.1016/j.gecco.2023.e02389>.
- Terbot JW, Cooper BS, Good JM, Jensen JD. A simulation framework for modeling the within-patient evolutionary dynamics of SARS-CoV-2. *Genome Biol Evol*. 2023a;15(11):1–10. <https://doi.org/10.1093/gbe/evad204>.
- Terbot JW, Johri P, Liphardt SW, Soni V, Pfeifer SP, Cooper BS, Good JM, Jensen JD. Developing an appropriate evolutionary baseline model for the study of SARS-CoV-2 patient samples. *PLOS Pathog*. 2023b;19(4):e1011265. <https://doi.org/10.1371/journal.ppat.1011265>.
- Van der Auwera G, O'Connor B. *Genomics in the cloud: using Docker, GATK, and WDL in terra*. Sebastopol (CA): O'Reilly Media; 2020.
- Versoza CJ, Ehmke E, Jensen JD, Pfeifer SP. Characterizing the rates and patterns of de novo germline mutations in the aye-aye (*Daubentonia madagascariensis*). *bioRxiv* 622690. <https://doi.org/10.1101/2024.11.08.622690>, 11 November 2024a, preprint: not peer reviewed.
- Versoza CJ, Jensen JD, Pfeifer SP. The landscape of structural variation in aye-ayes (*Daubentonia madagascariensis*). *bioRxiv* 622672. <https://doi.org/10.1101/2024.11.08.622672>, 11 November 2024b, preprint: not peer reviewed.
- Versoza CJ, Lloret-Villas A, Jensen JD, Pfeifer SP. A pedigree-based map of crossovers and non-crossovers in aye-ayes (*Daubentonia madagascariensis*). *bioRxiv* 622675. <https://doi.org/10.1101/2024.11.08.622675>, 11 November 2024c, preprint: not peer reviewed.
- Versoza CJ, Pfeifer SP. A hybrid genome assembly of the endangered aye-aye (*Daubentonia madagascariensis*). *G3 (Bethesda)*. 2024;14(10):jkae185. <https://doi.org/10.1093/g3journal/jkae185>.
- Yoder AD. Back to the future: a synthesis of strepsirrhine systematics. *Evol. Anthropol. Issues, News, Rev*. 1997;6(1):11–22. [https://doi.org/10.1002/\(SICI\)1520-6505\(1997\)6:1<11::AID-EVAN6>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1520-6505(1997)6:1<11::AID-EVAN6>3.0.CO;2-Z).

Associate editor: Laura Katz