A whole-genome scan for evidence of recent positive and balancing selection in aye-ayes (*Daubentonia madagascariensis*) utilizing a well-fit evolutionary baseline model

Vivak Soni[1], John W. Terbot II[1], Cyril J. Versoza[1], Susanne P. Pfeifer[1,†], and Jeffrey D. Jensen[1,†]

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

[†] co-corresponding authors; jointly supervised the project

(susanne@spfeiferlab.org; jeffrey.d.jensen@asu.edu)

Keywords

primate; aye-aye; selective sweep; balancing selection; genome scan; demography

ABSTRACT

The aye-aye (*Daubentonia madagascariensis*) is one of the 25 most endangered primate species in the world, maintaining amongst the lowest genetic diversity of any primate measured to date. Characterizing patterns of genetic variation within aye-aye populations, and the relative influences of neutral and selective processes in shaping that variation, is thus important for future conservation efforts. In this study, we performed the first whole-genome scans for recent positive and balancing selection in the species, utilizing high-coverage population genomic data from newly sequenced individuals. We generated null thresholds for our genomic scans by creating an evolutionarily appropriate baseline model that incorporates the demographic history of this aye-aye population, and identified a small number of candidate genes. Most notably, a suite of genes involved in olfaction — a key trait in these nocturnal primates — were identified as experiencing long-term balancing selection. We also conducted analyses to quantify the expected statistical power to detect positive and balancing selection in this population using site frequency spectrum-based inference methods, once accounting for the potentially confounding contributions of population history, recombination and mutation rate variation, and purifying and background selection. This work, presenting the first high-quality, genome-wide polymorphism data across the functional regions of the aye-aye genome, thus provides important insights into the landscape of episodic selective forces in this highly endangered species.

INTRODUCTION

A strepsirrhine endemic to Madagascar and the world's largest nocturnal primate, the aye-aye (*Daubentonia madagascariensis*) is the only extant member of the *Daubentoniidae* family, exhibiting a geographical range wider than any other member of the Lemuroidea superfamily (Sterling 1993). However, rapid habitat destruction is thought to have contributed to a severe population decline over the past few decades (Louis et al. 2020; Suzzi-Simmons 2023), along with direct human predation at least partially owing to the regional Malagasy cultural belief that aye-ayes are a harbinger of illness and death (Andriamasimanana 1994). These ongoing trends – coupled with harboring amongst the lowest genetic diversity of any primate measured to date (Perry et al. 2012; Kuderna et al. 2023) – have placed the aye-aye on the list of the 25 most endangered primate species in the world, according to the International Union for Conservation of Nature and Natural Resources Species Survival Commission Primate Specialist Group (Schwitzer et al. 2013; Louis et al. 2020; and see the discussion in Gross 2017). As such, characterizing patterns of genetic variation within aye-aye populations, and the relative influences of neutral and selective processes in shaping that variation, will be important for future conservation efforts.

Signatures of episodic selection

Although it is well understood that the demographic history of a population together with the recurrent action of natural selection acts to shape patterns of polymorphism at the DNA sequence level, disentangling the effects of these processes remains an ongoing concern (e.g., Ewing and Jensen 2016; Charlesworth and Jensen 2022; Jensen 2023). Nevertheless, distinguishing these processes is fundamental for gaining an improved

understanding of general evolutionary dynamics, inasmuch as it would improve our understanding of the relative importance of adaptive and nonadaptive factors in shaping levels of genetic variation in natural aye-aye populations as well as facilitate the accurate identification of genomic regions that have experienced recent bouts of episodic selection. Existing methods for detecting recent beneficial fixations rely on the changes in patterns of variation at linked sites (i.e., by characterizing the resulting effects of the associated selective sweep), though the nature of these changes will naturally depend on the details of the selective pressure (see the review of Stephan 2019). The term "selective sweep" describes the process whereby a positively selected mutation rapidly increases in frequency and fixes within a population, with linked variation following the same trajectory to an extent determined by the level of linkage (Maynard Smith and Haigh 1974; and see the review of Charlesworth and Jensen 2021). The fixation of these linked variants is expected to temporally reduce local nucleotide diversity (Berry et al. 1991). Under a single selective sweep model with recombination, there is also an expectation of a skew in the site frequency spectrum (SFS) toward both high- and low-frequency derived alleles within the vicinity of the beneficial mutation (Braverman et al. 1995; Simonsen et al. 1995; Fay and Wu 2000). The theoretical expectations under this model of a single, recent selective sweep have been well described (e.g., Kim and Stephan 2002; Kim and Nielsen 2004), and a composite likelihood ratio (CLR) test was developed by Kim and Stephan (2002) based on these expectations that detects such local reductions in nucleotide diversity and skew in the SFS along the chromosome. This signature is used to identify candidate loci that have experienced the recent action of positive selection by comparing the probability of the observed SFS under the standard neutral model with that under the model of a selective sweep. Subsequent work demonstrated that certain demographic histories may be

problematic however, with, for example, severe population bottlenecks often replicating patterns of positive selection (Jensen et al. 2005). Thus, to help reduce this issue, Nielsen et al. (2005) adapted the CLR method for genome-wide data utilizing a null model instead derived from the empirically observed SFS, which they termed SweepFinder (along with a more recent implementation, SweepFinder2; DeGiorgio et al. 2016).

Unlike signatures of selective sweeps, those of balancing selection (see the reviews of Fijarczyk and Babik 2015; Bitarello et al. 2023) – a term that encapsulates a variety of selective processes that maintain genetic variability in populations – can potentially last for considerable timescales (Lewontin 1987). Indeed, the temporal history of a balanced allele has been split into multiple phases, with detectable genomic signatures varying in each phase. For example, Fijarczyk and Babik (2015) characterized these phases as recent (<0.4N_e generations), intermediate (0.4-4N_e generations), and ancient (>4N_e generations), where N_e is the effective population size. The initial trajectory of a newly introduced mutation under balancing selection is indistinguishable from that of a partial selective sweep (Soni and Jensen 2024a), whereby the newly arisen mutation rapidly increases to its balanced frequency, conditional on escaping stochastic loss. The signatures of these partial sweeps include a potential excess of intermediate frequency alleles, extended linkage disequilibrium (LD) owing to the associated genetic hitchhiking effects (see the reviews of Crisci et al. 2013; Charlesworth and Jensen 2021), and weaker genetic structure at genes experiencing balancing selection (Schierup et al. 2000). Once the balanced frequency is reached, the allele under balancing selection fluctuates about this frequency, with recombination breaking up the aforementioned LD patterns (Wiuf et al. 2004; Charlesworth 2006; Pavlidis et al. 2012). If balancing selection persists in species with a divergence time predating the expected coalescent time, the allele under selection may continue to segregate as a trans-species

polymorphism (Klein et al. 1998; Leffler et al. 2013). To facilitate detection of long-term balancing selection, Cheng and DeGiorgio (2020) developed a class of CLR-based methods for detecting balancing selection that utilizes a mixture model, combining the expectation of the SFS under neutrality with the expectation under balancing selection, to infer the expected SFS at both a putatively selected site and at increasing distances away from that site, released under the BalLeRMix software package (Cheng and DeGiorgio 2020). As with SweepFinder2, this class of methods utilizes a null model directly derived from the empirical SFS in an attempt to account for deviations from the standard neutral expectation in a model-free manner. Such approaches have been shown to be well-powered in detecting long-term balancing selection (>25N_e generations in age; Soni and Jensen 2024a), depending, as they do, on new mutations accruing on the balanced haplotype and thereby generating the expected skew in the SFS toward intermediate frequency alleles.

Inferring positive and balancing selection in non-human primates

As one might expect, the majority of scans for positive selection in primates have focused upon human population genomic data. However, a number of studies have found signals of putative positive selection in non-human primates, chiefly in the great apes (The Chimpanzee Sequencing and Analysis Consortium 2005; Enard et al. 2010; Locke et al. 2011; Prüfer et al. 2012; Scally et al. 2012; Bataillon et al. 2015; McManus et al. 2015; Cagan et al. 2016; Munch et al. 2016; Nam et al. 2017; Schmidt et al. 2019), as well as in biomedically-relevant species such as rhesus macaques (The Rhesus Macaque Genome Sequencing and Analysis Consortium et al. 2007) and vervet monkeys (Pfeifer 2017b), often with contradictory results, likely owing to differing methodological approaches as well as to high false-positive rates related to a neglect of demographic effects. For example, Enard et al.

(2010) proposed a variant of the Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) and applied it to chimpanzee, orangutan, and macaque population genomic data, finding a high number of orthologous genes exhibiting simultaneous signatures of selective sweeps; conversely, Cagan et al. (2016) utilized a variety of test statistics for detecting positive selection across differing time scales, finding relatively little overlap between species.

Although the first whole-genome, short-read assembly for the aye-aye was published over a decade ago (Perry et al. 2012), and a more recent long-read assembly was made available in 2023 (Shao et al. 2023), the lack of protein-coding gene annotations has greatly limited the ability to scan for recent, episodic selective events in the species. However, the release of a fully annotated, chromosome-level hybrid *de novo* assembly (based on a combination of Oxford Nanopore Technologies long-reads and Illumina short-reads, and scaffolded using genome-wide chromatin interaction data) earlier this year (Versoza and Pfeifer 2024) provides a unique opportunity to identify patterns of episodic selection in this endangered species. In this study, we have performed scans for selective sweeps and balancing selection using SweepFinder2 (DeGiorgio et al. 2016) and BalLeRMix (Cheng and DeGiorgio 2020), respectively, utilizing unique, high-quality, whole-genome, population-level data. Importantly, to account for the confounding effects of demography (Barton 1998; Ewing and Jensen 2014; Poh et al. 2014; Harris and Jensen 2020; and see Charlesworth and Jensen 2024), we used the demographic model of Terbot et al. (2024) which was generated using non-functional regions in the aye-aye genome that are free from the effects of background selection. This demographic history fits neutral population data exceedingly well, and suggests a history in which the aye-aye population size was greatly reduced with first human contact in Madagascar 3,000 - 5,000 years ago, with an additional decline owing to recent habitat loss over the past few decades. This well-fitting null model is thus utilized

here to determine thresholds for positive and balancing selection scans across functional regions in order to avoid extreme false-positive rates (Thornton and Jensen 2007; Poh et al. 2014; Johri et al. 2022a,b, 2023; Soni et al. 2023) – a particularly important feature in this application given the severe bottleneck history of the species (Terbot et al. 2024). Through this process, we have identified a number of candidate loci with evidence of positive and balancing selection effects, and discuss these results in light of the recently available gene annotations.

RESULTS & DISCUSSION

We sequenced the genomes of five unrelated aye-aye (*D. madagascariensis*) individuals (three females and two males) housed at the Duke Lemur Center to an average coverage of >50x. After mapping reads to the recently published chromosome-level genome assembly for the species, we called variant and invariant sites following the best practices for non-model organisms (Pfeifer 2017a; van der Auwera and O'Connor 2020). We ran genomic scans across this newly generated population-level dataset, using the 14 autosomal scaffolds from the aye-aye genome assembly of Versoza and Pfeifer (2024) (excluding the X-chromosome, i.e., scaffold 9). The CLR methods implemented in SweepFinder2 and the B_{0MAF} statistic of the BalLeRMix software package were used to infer selective sweeps and balancing selection, respectively. Selective sweep inference was performed at each single nucleotide polymorphism (SNP), whilst balancing selection inference was performed in windows of size 10 and 100 SNPs. Figures 1 and 2 provide the results of the genome-wide scans for SweepFinder2 and B_{0MAF} based on 100 SNP windows, respectively (and see

Supplementary Figure S1 for the genome-wide scan results with B_{0MAF} on 10 SNP windows), highlighting a number of peaks along the likelihood surface for each analysis.

Although it is common to use outlier approaches to identify candidate regions experiencing positive selection, it has previously been shown that such approaches are often associated with extreme false-positive rates (Teshima et al. 2006; Thornton and Jensen 2007; Jensen et al. 2008; Jensen 2023; Soni et al. 2023). Furthermore, these approaches are problematic as any evolutionary model (including standard neutrality) will naturally have a 1% and 5% tail, and thus assuming that genes in these tails of an observed empirical distribution are likely sweep candidates is inherently flawed (see Harris et al. 2018). Moreover, it is not a given that recent sweeps, if they exist, will necessarily even appear in the tails of the empirical distribution under any given demographic model. Thus, we instead followed the recent recommendations of Johri et al. (2022b) in carefully constructing an evolutionarily appropriate baseline model accounting for commonly acting evolutionary processes, including a well-fit population history based on neutral genomic data, purifying and background selection acting on functional sites as modelled via a realistic distribution of fitness effects (DFE), as well as underlying mutation and recombination rate heterogeneity. Furthermore, we characterized the expected true- and false-positive rates of the utilized statistical approaches for our given genomic dataset, as positive and balancing selection will not necessarily be detectable within the context of any given baseline model (Barton 1998; Thornton and Jensen 2007; Poh et al. 2014; Harris and Jensen 2020); importantly, even if these events are not detectable, this approach remains necessary for managing false-positive rates.

We therefore simulated 100 replicates for each of the 14 autosomes using msprime (Baumdicker et al. 2022), under the aye-aye demographic model inferred by Terbot et al.

(2024), utilizing the genome assembly of Versoza and Pfeifer (2024), and modelling the specific data details of our newly presented whole-genome dataset. We performed SweepFinder2 and B_{0MAF} analyses on these baseline simulations, using the maximum CLR values across all simulation replicates as the null thresholds for positive and balancing selection inference, reasoning that these are the maximum values that can be generated in the absence of episodic selective processes under the baseline model considered, and any CLR values that exceeded the thresholds in our empirical analyses were considered to represent meaningful candidate regions. The identified threshold values under this model were 211.747 for SweepFinder2 inference at each SNP, 50.817 for B_{0MAF} inference on 10 SNP windows, and 244.382 for B_{0MAF} inference on 100 SNP windows.

Signatures of episodic selection in the aye-aye genome

A total of 3,462 loci met our null threshold for selective sweep inference using SweepFinder2, which mapped to 71 genes within the aye-aye genome. Scaffolds 1, 7, 10, and 11 contained at least one candidate region, although numerous regions on other scaffolds show peaks that were below our null threshold. For balancing selection inference with B_{0MAF}, no windows met our null threshold for windows of size 10 SNPs, and 163 windows met our null threshold for windows of size 100 SNPs. The latter windows mapped to 60 candidate genes, covering all autosomal scaffolds apart from scaffolds 11, 14, and 15. Supplementary File S1 provides tables of candidate regions overlapping genes exhibiting CLR values greater than the null thresholds for these analyses.

We manually curated the 71 sweep candidate genes, particularly noting those associated with the peak of each significant likelihood surface (e.g., see Figure 3 for scaffolds 1 and 7 for zoomed inset plots, and Supplementary Figures S2-S3 for additional scaffolds

containing candidate regions). Because balancing selection candidate genes were based on 100 SNP windows, these peaks were already quite localized by comparison, though we similarly manually curated this associated set of 60 candidate genes (see Figure 4 for the results in scaffold 1 with mapped genes marked on the plot, and Supplementary Figures S4-S12 for additional scaffolds containing candidate regions). The gene that exhibited the strongest signal of positive selection (i.e., the highest CLR value) was SMPD4 on scaffold 10, whilst the gene exhibiting the strongest signal of balancing selection was LRP1B on scaffold 6 for the 100 SNP window analysis. Biallelic loss-of-function variants in SMPD4 have been found to cause microcephaly, a rare and severe neurodevelopmental disorder with progressive congenital microcephaly and early death in humans (Magini et al. 2019; Smits et al. 2023). LRP1B is a putative tumor suppressor (Brown et al. 2021), and one of the most altered genes in human cancer (Principe et al. 2021). Mutations in LRP1B have been associated with an increased tumor mutation burden (Yu et al. 2022). In one of the earliest large-scale primate genome scans, Nielsen et al. (2005) found that a number of genes involved in tumor suppression were identified as positive selection candidates in humans, as well as genes involved in spermatogenesis, which our analysis in aye-ayes also identified (see below).

A single candidate region (scaffold 1:315,975,047-315,981,365) in the 100 SNP window balancing selection analysis was found to overlap with a 89,250bp inversion (scaffold 1:315,972,759-316,062,009). This region may represent a false positive, owing to the reduced recombination related to the inversion potentially generating long haplotype structure (Stevison et al. 2011). However, inversions may themselves be selectively maintained, particularly if they contain a beneficial combination of alleles (e.g., Hager et al.

2022; and see Villoutreix et al. 2021); thus, this candidate region will require future dissection.

Gene functional analysis

A gene function analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID; Ma et al. 2023) predicted an enrichment for 19 Gene Ontology (GO) terms in aye-ayes at a $p \leq 0.05$ for recent selective sweep candidate genes – however, no terms passed a false discovery rate (FDR) threshold of 0.05 or 0.1 for selective sweep candidate genes. Conversely, 14 of the 43 GO terms identified for the 100 SNP window analysis passed both the $p \leq 0.05$ and the FDR ≤ 0.05 thresholds for balancing selection. Table 1 provides the top 12 functional categories for balancing selection candidate genes for the 100 SNP analysis (and see Supplementary File S1 for all enriched categories). These categories show considerable enrichment and the majority are involved in olfaction.

Olfaction

Seven enriched gene functions for balancing selection (enrichment score >1) were related to olfaction, all with high-fold enrichment (ranging from 5.82 to 16.38). A number of OR genes were also found to meet our null threshold for selective sweeps. OR genes provide the basis for the sense of smell (e.g., Buck and Axel 1991), and comprise the largest gene superfamily in mammalian genomes (Glusman et al. 2001; Zozulya et al. 2001). Although olfaction plays a role in locating food, mating, and avoiding danger, the importance and sensitivity of smell varies significantly even amongst closely related species, and it has been suggested that this gene family is subject to a birth-and-death model of evolution, whereby new genes are formed by gene duplication and some of the duplicate genes differentiate in

function, whilst others become inactive or are removed from the genome (Niimura and Nei 2003, 2005a,b, 2007). Indeed, OR genes within primates show relaxed selective constraints in apes relative to Old and New World monkeys (Dong et al. 2009), and whilst patterns of variability in chimpanzees are consistent with purifying selection acting on intact OR genes, patterns in humans have been suggested to indicate the action of positive selection (Gilad et al. 2003a). These findings were partially corroborated by Williamson et al. (2007), who found evidence of recent selective sweeps in numerous OR genes in human populations. Gilad et al. (2003a) argued that the differing selective processes acting on OR genes in humans and chimpanzees are likely reflections of differences in lifestyles between humans and other great apes, resulting in distinct sensory needs. Further studies have suggested that both positive selection (Gilad et al. 2003b) and balancing selection (Alonso et al. 2008) are acting on OR genes in humans. Importantly, it has been proposed that heterozygosity in ORs can increase the number of different odorant-binding sites in the genome (Lancet 1994), and thus heterozygote advantage may be an important process. Given that aye-ayes have been shown to discriminate based on scent (Price and Feistner 1994) and use scent-marking behaviors to attract mates (Winn 1994), OR genes represent interesting candidate loci for having experienced on-going positive and balancing selection.

Rhodopsin

Six enriched gene functions for balancing selection (enrichment score >1) were related to G protein coupled receptors (GPCR), with fold enrichment scores ranging from 3.75 to 6.22. GPCRs are cell surface receptors responsible for detecting extracellular molecules and activating cellular responses; consequently, they are involved in numerous physiological processes. Two of the six enriched GPCR functions were specifically related to

rhodopsin, which is the opsin responsible for mediating dim light vision (Litman and Mitchell 1996). It has previously been shown that, despite being nocturnal, aye-ayes maintain dichromacy – potentially supporting previous work that found that the red/green opsin gene survived the long nocturnal phase of mammalian evolution, which has been hypothesized to relate to a role in setting biorhythms (Nei et al. 1997). These identified candidate regions may also lend credence to the speculation that dichromatic nocturnal primates may be able to perceive color while foraging under moonlight conditions (Perry et al. 2007).

PATE gene family

The PATE gene family has been shown to express in the testis, encoding a sperm-related protein (Bera et al. 2002). Of the four genes that make up the PATE gene family, three were found in candidate selective sweep regions in our analysis (Figure 3), whilst PATE1 was also found in our balancing selection scans based on 100 SNP windows (Figure 4). Soler-Garcia et al. (2005) found that PATE is highly expressed in the male genital tract, and that proteins are secreted into the semen, suggesting a potential in mammalian sperm maturation, whilst Margalit et al. (2012) found that PATE proteins are involved in sperm-oolemma fusion and penetration.

Although primate sperm displays a general uniformity, previous studies have found variations in sperm morphology (Cummins and Woodall 1985; Gage 1998), often predicated on the absence or presence of sperm competition. Indeed, the use of coagulated ejaculate that forms sperm plugs to avoid sperm competition and increase male fertilization success has been described in multiple species of primates, particularly those exhibiting polygynandrous mating systems (i.e., multi-male, multi-female mating systems; Dixson et al. 2002; Martinez and Garcia 2020). As aye-ayes are polygynandrous (Quinn and Wilson 2004),

these candidate loci may similarly be hypothesized to relate to mate competition / sexual selection shaping the PATE family of genes.

Zinc-finger genes

Multiple zinc-finger (ZNF) genes were identified as having undergone recent selective sweeps or balancing selection in our scans of the aye-aye genome. ZNF genes are the largest family of transcription factors in mammalian genomes and play an important role in gene regulation. Previous studies have identified a number of KRAB-ZNF genes – a sub-family of the deeply conserved *Kruppel*-type zinc-finger (KZNF) genes (Bellefroid et al. 1993; Looman et al. 2002; Huntley et al. 2006) – with evidence of positive selection in humans (Nielsen et al. 2005; Nowick et al. 2010, 2011; Jovanovic et al. 2021); additional research has suggested that KRAB-ZNF genes follow a species-specific – as opposed to tissue-specific – pattern of expression (Kapopoulou et al. 2016), suggesting that these genes have different tissue preferences in different species and thus have functionally diversified across the primate lineage (Liu et al. 2014). In addition, numerous gene regulatory factors were identified as putative candidate regions, likely related to the well-described roles of these genes in modifying expression patterns (Berrio et al. 2020; Liu and Robinson-Rechavi 2020; Jovanovic et al. 2021; and see the review of McDonald and Reed 2023).

Quantifying power to detect recent positive and balancing selection in aye-ayes

It has previously been demonstrated that the statistical power to detect positive selection in any given population will depend on a variety of factors, ranging from the details of the population history, to the amount and configuration of the analyzed data itself (e.g., Johri et al. 2021, 2022b; Soni et al. 2023; Soni and Jensen 2024a; Soni et al. 2024b). In order

to quantify the specific power of this analysis to detect candidate regions in aye-ayes, we ran forward-in-time simulations in SLiM (Haller and Messer 2023), utilizing the well-fitting demographic history presented in Terbot et al. (2024). As the underlying DFE in aye-ayes remains uncharacterized, functional regions were simulated using the discrete DFE inferred for humans by Johri et al. (2023) and verified by Soni and Jensen (2024b), in order to provide a reasonable proxy of purifying and background selection effects. Additionally, without access to mutation and recombination maps that are available in heavily studied species, we were unable to directly simulate these effects. However, in order to model the impact of this uncertainly as well as likely rate heterogeneity (Johri et al. 2022a; Soni et al. 2024), we drew rates from a uniform distribution for each 1kb window, such that the mean across each simulation replicate was equal to the mean genomic rate (see the "Materials and Methods" section for more details).

In each simulation replicate, a single beneficial mutation was introduced. For selective sweep models, three different selection regimes were considered, with population-scaled strengths of selection, $2N_es$, of 100, 1,000, and 10,000, where N_e is the ancestral population size and s is the strength of selection acting on the beneficial mutation. Five different introduction times of the beneficial mutation were considered, τ = 0.1, 0.2, 0.5, 1, and 2, where τ is the time before sampling in N generations. These values were chosen as there is not expected to be power to detect selective sweeps beyond 4N generations, though power generally decays much more rapidly (Kim and Stephan 2002; Przeworski 2002, 2003; Ormond et al. 2016). Only simulation replicates in which the beneficial mutation fixed were retained. For balancing selection, the beneficial mutation was modelled as experiencing negative frequency-dependent selection, and introduced at τ = 10N, 50N, and 75N generations, as it has been shown that SFS-based methods have little power to detect

recent balancing selection (Soni and Jensen 2024a). Only simulation replicates in which the balanced mutation was still segregating at the time of sampling were retained.

Figures 5 and 6 provide ROC plots for selective sweep inference with SweepFinder2 and B_{0MAF}, respectively. As shown, the power to detect positive selection in aye-ayes is expected to be reasonably poor in all cases. This is unsurprising, given that aye-ayes have undergone a recent bottleneck, an event which may replicate patterns of variation consistent with a selective sweep (e.g., Barton 1998; Poh et al. 2014; Harris and Jensen 2020), followed by a period of further decline, whilst balancing selection inference is expected to be confounded by the resulting skew in the SFS toward intermediate frequency alleles (Soni and Jensen 2024a). Additionally, identifying the episodic and locus-specific patterns of positive selection is more challenging in populations that have experienced bottlenecks owing to the large genealogical variance generated by this event, leading to widely dispersed test statistics across the genome which can give the illusion of a locus-specific pattern (Thornton and Jensen 2007). While power to detect selective sweeps generally related with strength as expected, this relationship was partially off-set by the fact that at higher strengths of selection the beneficial mutation fixed more rapidly in the population, and thus the time since fixation was longer thereby reducing inference power, as has been previously described analytically (Kim and Stephan 2000). By contrast, power to detect balancing selection increased with time since the introduction of the balanced mutation, as has been previously shown (e.g., Soni and Jensen 2024a). In summary, these power analyses provide a key for interpreting our empirical analysis, in demonstrating that any statistically detectable selective sweep would need be both strong and recent, while any statistically detectable loci experiencing balancing selection would need to be relatively ancient.

Concluding thoughts

We ran the first large-scale scans for loci having experienced selective sweeps and balancing selection in the aye-aye genome, using newly generated, high-quality, whole-genome, population-level data, and utilizing the recent fully-annotated genome assembly of Versoza and Pfeifer (2024) together with the well-fitting demographic model inferred from non-functional genomic regions of Terbot et al. (2024). By simulating an evolutionarily appropriate baseline model employing these details (Johri et al. 2022a,b), we were able to generate conservative null thresholds for these scans, thereby greatly reducing the types of false-positive rates associated with outlier approaches (Teshima et al. 2006; Thornton and Jensen 2007; Jensen et al. 2008; Soni et al. 2023). The differences between our conservative baseline model approach and a traditional genomic outlier approach are considerable (and see the discussions in Howell et al. 2023; Jensen 2023; Johri et al. 2023; Terbot et al. 2023). For example, selection inference with B_{0MAF} on 10 SNP windows yielded no candidate windows, and inference on 100 SNP windows yielded 163 candidate windows. An outlier approach interpreting the (commonly used) 5% tail of the empirical distribution of CLR values as candidate loci would instead give 22,971 candidate windows at 10 SNP windows, and 2,280 for 100 SNP windows. More to the point, however, as demonstrated in our power analyses, the great majority of density in these tail distributions may be readily generated by the population history alone, and thus need not invoke positive or balancing selection as an explanation.

Through this more thorough and conservative approach, we have identified a number of promising candidate genes with evidence of having been episodically impacted by positive and balancing selection during the recent evolutionary history of the species,

with particularly notable examples being those involved in olfaction and spermatogenesis. Given the on-going destruction of aye-aye habitats, it may well be hypothesized that these vital functions are experiencing changing selection pressures, particularly in light of their solitary lifestyle and polygynandrous mating system.

MATERIALS AND METHODS

Animal subjects

This study was approved by the Duke Lemur Center's Research Committee (protocol BS-3-22-6) and Duke University's Institutional Animal Care and Use Committee (protocol A216-20-11). The study was performed 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.

Samples and whole-genome sequencing

We sequenced five unrelated aye-aye (*D. madagascariensis*) individuals (three females and two males) using DNA extracted from whole blood samples previously collected at the Duke Lemur Center (Durham, NC, USA). A 150-bp paired-end library was prepared for each sample using the NEBNext Ultra II DNA PCR-free Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and sequenced at high-coverage (>50-fold) on the Illumina NovaSeq platform (Illumina, San Diego, CA, USA) (Supplementary Table S1).

Calling variant and invariant sites

Calling of variant and invariant sites followed the best practices for non-model organisms as described in Pfeifer (2017a) and van der Auwera and O'Connor (2020). In brief, we pre-processed raw reads by removing adapters and low-quality bases from the read-ends using the Genome Analysis Toolkit (GATK) *MarkIlluminaAdapters* v.4.2.6.1 (van der Auwera

and O'Connor 2020) and TrimGalore v.0.6.10 (https://github.com/FelixKrueger/TrimGalore), respectively. We mapped the prepared reads to the aye-aye reference assembly (DMad_hybrid; GenBank accession number: JBFSEQ000000000; Versoza and Pfeifer 2024) using the Burrows Wheeler Aligner (BWA-MEM) v.0.7.17 (Li and Durbin 2009) and marked duplicates using GATK's *MarkDuplicates* v.4.2.6.1. We further refined the read mappings by performing multiple sequence realignments using GATK's *RealignerTargetCreator* and *IndelRealigner* v.3.8, recalibrating base quality scores using GATK's *BaseRecalibrator* and *ApplyBQSR* v.4.2.6.1, and conducting another round of duplicate removal using GATK's *MarkDuplicates* v.4.2.6.1. As population genomic resources for aye-ayes are limited, no "gold standard" dataset exists for this recalibration step; instead, we utilized a high-confidence training dataset obtained from pedigreed individuals (see Versoza et al. 2024a for details). We then called variant and invariant sites from high-quality read alignments (' *--minimum-mapping-quality* 40 ') using GATK's *HaplotypeCaller* v.4.2.6.1 with the ' *pcr_indel_model* ' parameter set to NONE as a PCR-free protocol was used during library preparation, the ' *--heterozygosity* ' parameter set to 0.0005 to reflect species-specific levels of heterozygosity (Perry et al. 2013), and the ' *-ERC* ' parameter set to BP_RESOLUTION to output both variant and invariant sites. Next, we jointly assigned genotype likelihoods across all five individuals at all sites (' *-all-sites* ') using GATK's *GenotypeGVCFs* v.4.2.6.1, accounting for species-specific levels of heterozygosity as detailed above. Following the GATK Best Practices, we applied a set of site-level "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) to the sites genotyped in all individuals (AN = 34), and applied upper and lower cutoffs on the individual depth of coverage ($0.5 \times DP_{ind}$ and $2 \times DP_{ind}$) to remove regions with an unusual read depth indicative

of erroneous calls. The resulting dataset was limited to the autosomes and divided into variant (i.e., SNPs) and invariant sites for downstream analyses (Supplementary Table S2).

Generating null thresholds for selection inference

We simulated the demographic model of Terbot et al. (2024) to generate null thresholds for the inference of selection using the coalescent simulator msprime v1.3.2. (Baumdicker et al. 2022). Although the Terbot et al. (2024) model involves two populations – one population from northern Madagascar, consisting of four individuals previously sequenced at low coverage (Perry et al. 2013), and one population from the rest of the island, consisting of eight individuals previously sequenced at low coverage (Perry et al. 2013) as well as five individuals newly sequenced at high coverage (Terbot et al. 2024) – we here only consider the single population of newly sequenced individuals (n = 5), given the much higher data quality of this sample. As such, while the Terbot et al. (2024) work includes only the non-coding regions of these newly sequenced individuals, we here present the full genome data including functional regions as well. The demographic history of this population involves a relatively ancient size reduction (likely associated with human colonization of Madagascar), followed by a period of recent decline up to the current day (likely related to recent habitat loss). Based on this demographic history, we simulated 10 replicates for each of the 14 autosomal chromosomes included in the most recent aye-aye genome assembly (Versoza and Pfeifer 2024), modelling a mutation rate of 1.52e-8 per base pair per generation (i.e., the mutation rate previously reported in another lemur species; Campbell et al. 2021), and a recombination rate of 1e-8 cM/Mb (as recently inferred from pedigree data; Versoza, Lloret-Villas et al. 2024).

To generate null thresholds, we ran both SweepFinder2 v1.0. (DeGiorgio et al. 2016) and the B_{0MAF} method of Cheng and DeGiorgio (2020) on the allele frequency files generated from our simulated demographic data. In brief, we performed inference at each SNP with SweepFinder2 using the following command: `SweepFinder2 –lu GridFile FreqFile SpectFile OutFile`. Additionally, we utilized two inference schemes in B_{0MAF}: (1) windows containing 10 SNPs with a 5 SNP step size and (2) windows containing 100 SNPs with a 50 SNP step size, using the following command: `python3 BalLeRMix+_v1.py -I FreqFile --spect SfsFile -o OutFile -w W -s S –usePhysPos –noSub –MAF –rec 1e-8`, where *W* is the window size and *S* is the step size, both based on number of SNPs. Because we lacked information on the polarization of SNPs, allele frequencies were folded and only polymorphic sites were included in the analyses. Notably, the highest CLR value across all null model simulations was set as the null threshold for inference, under the assumption that this is the highest value that can be generated in the absence of positive or balancing selection, thereby providing a conservative scan to reduce false positive rates.

Inferring recent positive and balancing selection in the aye-aye genome

We ran SweepFinder2 and B_{0MAF} on the 14 aye-aye autosomes using the same inference schema discussed above. Only those inference values greater than the null threshold values were considered as putatively experiencing positive or balancing selection. The genes in which these selected sites were located were identified using the genome annotations of Versoza and Pfeifer (2024), whilst overlaps with structural variants were also identified, based on those described by Versoza et al. (2024b). Because the number of identified genes was relatively small (<200 for both the sweep and balancing selection scans), we manually curated our candidates. In brief, for sweep candidates, we first

identified genes under the significant likelihood surface. These candidate genes were then run through the NCBI database (Sayers et al. 2022) and Expression Atlas (Madeira et al. 2022) in order to identify function and expression patterns in other primate species. Additionally, we performed a Gene Ontology analysis (The Gene Ontology Consortium 2023) using the Database for Annotation, Visualization, and Integrated Discovery (Ma et al. 2023) on our candidate genes (conducted for all candidates together, and also separately for selective sweep and balancing selection candidates).

Power analyses

To assess how much statistical power exists to detect episodic selection in this aye-aye population given the details of both the demographic history and of the dataset itself, we simulated the Terbot et al. (2024) demographic model forward-in-time in SLiM v.4.0.1 (Haller and Messer 2023). Thereby, the simulated region was comprised of three functional regions, separated by intergenic regions of length 16,489bp. Each functional region contained nine exons of length 130bp, separated by introns of length 1,591bp, for a total region length of 91,161bp. These details were estimated from the Versoza and Pfeifer (2024) genome annotations to represent common aye-aye genomic architecture. Mutations in intronic and intergenic regions were modelled as effectively neutral, while exonic mutations were drawn from a DFE comprised of four fixed classes (Johri et al. 2020), whose frequencies are denoted by *fi*: *f0* with $0 \leq 2N_e s < 1$ (i.e., effectively neutral mutations), *f1* with $1 \leq 2N_e s < 10$ (i.e., weakly deleterious mutations), *f2* with $10 \leq 2N_e s < 100$ (i.e., moderately deleterious mutations), and *f3* with $100 \leq 2N_e s < 2N_e$ (i.e., strongly deleterious mutations), where N_e is the effective population size and s is the reduction in fitness of the mutant homozygote relative to wild-type. Within each bin, s was drawn from a uniform distribution. We utilized

the DFE inferred from human population genomic data by Johri et al. (2023), and verified in Soni and Jensen (2024b), as a proxy. This modelling of a realistic DFE in functional regions enabled us to account for the effects of purifying and background selection, in addition to population history, when assessing a baseline model of commonly acting evolutionary processes in the species.

To model uncertainty and heterogeneity in the underlying mutation and recombination rate, each 1kb region of the simulated chromosome was assigned a different rate. Rates were drawn from a uniform distribution such that the chromosome-wide average was approximately the fixed rate previously observed in pedigree data (i.e., 1 cM/Mb; Versoza, Lloret-Villas, et al. 2024). For variable recombination rates, the minimum and maximum parameters of the uniform distribution were 0.01 and 10 cM/Mb, respectively (i.e., a 100-fold decrease and a 10-fold increase on the fixed rate). For variable mutation rates the minimum and maximum parameters of the uniform distribution were set at 0.61e-8 and 3.8e-8 (i.e., 0.5x and 2.5x the fixed rate previously reported in lemurs, respectively).

Simulations had a 10*N* generation burn-in time, where *N* is the ancestral population size of 23,706. A further 10*N* generations were then simulated for the sweep analysis, whilst a further 85*N* generations were simulated for the balancing selection analysis. In each simulation replicate, a single positively selected mutation was introduced. For the selective sweep analysis, the beneficial mutation was introduced at τ = [0.1, 0.2, 0.5, 1, 2], where τ is the time before sampling in *N* generations. Three different beneficial selection coefficients were simulated: $2N_e s$ = [100, 1,000, 10,000], where N_e is equal to the ancestral population size of 23,706. For the balancing selection analysis, the balanced mutation was introduced at τ = [10*N*, 50*N*, 75*N*]. The balanced mutation experienced negative frequency-dependent selection, which was modelled such that the selection coefficient of the balanced mutation

was dependent on its frequency in the population: $S_{bp} = F_{eq} - F_{bp}$, where S_{bp} is the selection coefficient of the balanced mutation, F_{eq} is the equilibrium frequency of the balanced mutation (here set to 0.5), and F_{bp} is the frequency of the balanced mutation in the population. Simulations were structured such that if the selective sweep failed to fix, or the balanced mutation was either fixed or lost from the population, the simulation would restart at the point of introduction of the selected mutation.

Scans for selective sweeps and balancing selection were then performed on the simulated data, as per the procedure discussed above. ROC plots were generated in order to summarize expected performance. Because selective sweep inference was performed on each SNP, 100bp, 1kb and 10kb windows were generated for creating ROC plots for SweepFinder2 results (note that this was not necessary for B_{0MAF} as SNP-based windows were used for inference).

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 helpful discussion. DNA extraction, library preparation, and Illumina sequencing was conducted at Azenta Life Sciences (South Plainfield, NJ, USA). Computations were performed on the Sol supercomputer at Arizona State University (Jennewein et al. 2023) and on the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. This is Duke Lemur Center publication # XXXX.

FUNDING

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM151008 to SPP and the National Science Foundation under Award Number DBI-2012668 to the Duke Lemur Center. VS, JT, and JDJ were supported by National Institutes of Health Award Number R35GM139383 to JDJ. CJV was supported by the National Science Foundation CAREER Award DEB-2045343 to SPP. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

REFERENCES

Alonso S, Lopez S, Izagirre N, De La Rua C. 2008. Overdominance in the human genome and olfactory receptor activity. *Mol Biol Evol*. 25(5):997–1001. https://doi.org/10.1093/molbev/msn049

Andriamasimanana M. 1994. Ecoethological study of free-ranging aye-ayes (*Daubentonia madagascariensis*) in Madagascar. *Folia Primatologica*, *62*(1–3):37–45. https://doi.org/10.1159/000156761

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

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

Baumdicker F, Bisschop G, Goldstein D, Gower G, Ragsdale AP, Tsambos G, Zhu S, Eldon B, Ellerman EC, Galloway JG, et al. 2022. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, *220*(3):iyab229. https://doi.org/10.1093/genetics/iyab229

Bellefroid EJ, Marine JC, Ried T, Lecocq PJ, Rivière M, Amemiya C, Poncelet DA, Coulie PG, De Jong P, Szpirer C. 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *The EMBO Journal*, *12*(4):1363–1374. https://doi.org/10.1002/j.1460-2075.1993.tb05781.x

Bera TK, Maitra R, Iavarone C, Salvatore G, Kumar V, Vincent JJ, Sathyanarayana BK, Duray P, Lee BK, Pastan I. 2002. *PATE,* a gene expressed in prostate cancer, normal prostate, and testis, identified by a functional genomic approach. *Proc Nat Acad Sci USA*, *99*(5):3058–3063. https://doi.org/10.1073/Proc Nat Acad Sci USA.052713699

Berrio A, Haygood R, Wray GA. 2020. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. *BMC Genomics*, *21*(1):359. https://doi.org/10.1186/s12864-020-6752-4

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

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

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

Brown LC, Tucker MD, Sedhom R, Schwartz EB, Zhu J, Kao C, Labriola MK, Gupta RT, Marin D, Wu Y, et al. 2021. *LRP1B* mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. *Journal for ImmunoTherapy of Cancer*, *9*(3):e001792. https://doi.org/10.1136/jitc-2020-001792

Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell*, *65*(1):175–187. https://doi.org/10.1016/0092-8674(91)90418-X

Campbell CR, Tiley GP, Poelstra JW, Hunnicutt KE, Larsen PA, Lee HJ, Thorne JL, Dos Reis M, Yoder AD. 2021. Pedigree-based and phylogenetic methods support surprising patterns of mutation rate and spectrum in the gray mouse lemur. *Heredity* (Edinb), *127*(2):233-244. https://doi.org/10.1038/s41437-021-00446-5

Cagan A, Theunert C, Laayouni H, Santpere G, Pybus M, Casals F, Prüfer K, Navarro A, Marques-Bonet T, Bertranpetit J, et al. 2016. Natural Selection in the great apes. *Mol Biol Evol*, *33*(12):3268–3283. https://doi.org/10.1093/molbev/msw215

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

Charlesworth B, Jensen JD. 2021. Effects of selection at linked sites on patterns of genetic variability. *Annu Rev Ecol Evol Syst, 52*(1):177–197. https://doi.org/10.1146/annurev-ecolsys-010621-044528

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

Charlesworth B, Jensen JD. 2024. Population genetics. In: Encyclopedia of Biodiversity, 3rd edition. Vol. 7:467-483. Elsevier. https://doi.org/10.1016/B978-0-12-822562-2.00021-9

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

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

Cummins JM, Woodall PF. 1985. On mammalian sperm dimensions. *Reproduction*, *75*(1):153–175. https://doi.org/10.1530/jrf.0.0750153

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

Dixson A, Dixson B, Anderson M. 2005. Sexual selection and the evolution of visually conspicuous sexually dimorphic traits in male monkeys, apes, and human beings. *Annu Rev Sex Res*, *16*:1–19.

Dong D, He G, Zhang S, Zhang Z. 2009. Evolution of olfactory receptor genes in primates dominated by birth-and-death process. *Genome Biol Evol*, *1*:258–264. https://doi.org/10.1093/gbe/evp026

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

Ewing GB, Jensen JD. 2014. Distinguishing neutral from deleterious mutations in growing populations. *Front Genet*, *5*:7. https://doi.org/10.3389/fgene.2014.00007

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

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

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

Gage MJG. 1998. Mammalian sperm morphometry. *Proc R Soc B*, *265*(1391):97–103. https://doi.org/10.1098/rspb.1998.0269

Gilad Y, Bustamante CD, Lancet D, Pääbo S. 2003a. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet*, *73*(3):489–501. https://doi.org/10.1086/378132

Gilad Y, Man O, Pääbo S, Lancet D. 2003b. Human specific loss of olfactory receptor genes. *Proc Nat Acad Sci USA*, *100*(6):3324–3327. https://doi.org/10.1073/pnas.0535697100

Glusman G, Yanai I, Rubin I, Lancet D. 2001. The complete human olfactory subgenome. *Genome Res*, *11*(5):685–702. https://doi.org/10.1101/gr.171001

Gross M. 2017. Primates in peril. *Curr Biol*, *27*(12):R573–R576. https://doi.org/10.1016/j.cub.2017.06.002

Hager ER, Harringmeyer OS, Wooldridge TB, Theingi S, Gable JT, McFadden S, Neugeboren B, Turner KM, Jensen JD, Hoekstra HE. 2022. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, *377*(6604):399–405. https://doi.org/10.1126/science.abg0718

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

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

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

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

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics*, *116*(1):153–159.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res*, *16*(5):669–677. https://doi.org/10.1101/gr.4842106

Jennewein DM, Lee J, Kurtz C, Dizon W, Shaeffer I, Chapman A, Chiquete A, Burks J, Carlson A, Mason N, et al. 2023. The Sol Supercomputer at Arizona State University. *Practice and Experience in Advanced Research Computing*, 296–301. https://doi.org/10.1145/3569951.3597573

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

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

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

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

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

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

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

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

Jovanovic VM, Sarfert M, Reyna-Blanco CS, Indrischek H, Valdivia D I, Shelest E, Nowick K. 2021. Positive selection in gene regulatory factors suggests adaptive pleiotropic changes during human evolution. *Front Genet*, *12*:662239. https://doi.org/10.3389/fgene.2021.662239

Kapopoulou A, Mathew L, Wong A, Trono D, Jensen JD. 2016. The evolution of gene expression and binding specificity of the largest transcription factor family in primates: population genetics of KRAB-ZF genes. *Evolution*, *70*(1):167–180. https://doi.org/10.1111/evo.12819

Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics*, *167*(3):1513–1524. https://doi.org/10.1534/genetics.103.025387

Kim Y, Stephan W. 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, *155*(3):1415–1427. https://doi.org/10.1093/genetics/155.3.1415

Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, *160*(2):765.

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

Kuderna LFK, Gao H, Janiak MC, Kuhlwilm M, Orkin JD, Bataillon T, Manu S, Valenzuela A, Bergman J, Rousselle M, et al. 2023. A global catalog of whole-genome diversity from 233 primate species. *Science*, *380*(6648):906-913. https://doi.org/10.1126/science.abn7829

Lancet D. 1994. Olfaction: exclusive receptors. *Nature*, *372*(6504):321–322. https://doi.org/10.1038/372321a0

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

Lewontin RC. 1987. Polymorphism and heterosis: old wine in new bottles and vice versa. *Journal of the History of Biology*, *20*(3):337–349. https://doi.org/10.1007/BF00139459

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics,* 25:1754–1760.

Litman BJ, Mitchell CD. 1996. Rhodopsin structure and function. In *Biomembranes: A Multi-Volume Treatise* (Vol. 2, pp. 1–32). Elsevier. https://doi.org/10.1016/S1874-5342(07):80004-3

Liu H, Chang L-H, Sun Y, Lu X, Stubbs L. 2014. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol Evol*, *6*(3):510–525. https://doi.org/10.1093/gbe/evu030

Liu J, Robinson-Rechavi M. 2020. Robust inference of positive selection on regulatory sequences in the human brain. *Science Advances*, *6*(48):eabc9863. https://doi.org/10.1126/sciadv.abc9863

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

Looman C, Åbrink M, Mark C, Hellman L. 2002. KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol Biol Evol*, *19*(12):2118–2130. https://doi.org/10.1093/oxfordjournals.molbev.a004037

Louis EE, Sefczek TM, Randimbiharinirina DR, Raharivololona B, Rakotondrazandry JN, Manjary D, Aylward M, Ravelomandrato F. 2020. *Daubentonia madagascariensis*. *The IUCN Red List of Threatened Species* 2020: e.T6302A115560793. https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T6302A115560793.en.

Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A, Xiao J, Zhao W, Bao Y, Zhang Z. 2023. Database Commons: a catalog of worldwide biological databases. *Genomics, Proteomics Bioinformatics*, *21*(5):1054–1058. https://doi.org/10.1016/j.gpb.2022.12.004

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

Magini P, Smits DJ, Vandervore L, Schot R, Columbaro M, Kasteleijn E, Van Der Ent M, Palombo F, Lequin MH, Dremmen M, et al. 2019. Loss of SMPD4 causes a developmental disorder characterized by microcephaly and congenital arthrogryposis. *Am J Hum Gen*, *105*(4):689–705. https://doi.org/10.1016/j.ajhg.2019.08.006

Margalit M, Yogev L, Yavetz H, Lehavi O, Hauser R, Botchan A, Barda S, Levitin F, Weiss M, Pastan I, et al. 2012. Involvement of the prostate and testis expression (PATE)-like proteins in

sperm-oocyte interaction. *Human Reproduction*, *27*(5):1238–1248. https://doi.org/10.1093/humrep/des064

Martinez G, Garcia C. 2020. Sexual selection and sperm diversity in primates. *Molecular and Cellular Endocrinology*, *518*:110974. https://doi.org/10.1016/j.mce.2020.110974

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

McDonald JMC, Reed RD. 2023. Patterns of selection across gene regulatory networks. *Seminars in Cell Developmental Biology*, *145*:60–67. https://doi.org/10.1016/j.semcdb.2022.03.029

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

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

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

Nei M, Zhang J, Yokoyama S. 1997. Color vision of ancestral organisms of higher primates. *Mol Biol Evol*, *14*(6):611–618. https://doi.org/10.1093/oxfordjournals.molbev.a025800

Nielsen R, Bustamante C, Clark A G, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*, *3*(6):e170. https://doi.org/10.1371/journal.pbio.0030170

Niimura Y, Nei M. 2003. Evolution of olfactory receptor genes in the human genome. *Proc Nat Acad Sci USA*, *100*(21):12235–12240. https://doi.org/10.1073/pnas.1635157100

Niimura Y, Nei M. 2005a. Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene*, *346*:13–21. https://doi.org/10.1016/j.gene.2004.09.025

Niimura Y, Nei M. 2005b. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Nat Acad Sci USA*, *102*(17):6039–6044. https://doi.org/10.1073/pnas.0501922102

Niimura Y, Nei M. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE*, *2*(8):e708. https://doi.org/10.1371/journal.pone.0000708

Nowick K, Fields C, Gernat T, Caetano-Anolles D, Kholina N, Stubbs L. 2011. Gain, loss and divergence in primate zinc-finger genes: a rich resource for evolution of gene regulatory differences between species. *PLoS ONE*, *6*(6):e21553. https://doi.org/10.1371/journal.pone.0021553

Nowick K, Hamilton A T, Zhang H, Stubbs L. 2010. Rapid sequence and expression divergence suggest selection for novel function in primate-specific KRAB-ZNF genes. *Mol Biol Evol*, *27*(11):2606–2617. https://doi.org/10.1093/molbev/msq157

Ormond L, Foll M, Ewing GB, Pfeifer SP, Jensen JD. 2016. Inferring the age of a fixed beneficial allele. *Mol Ecol*, *25*(1):157–169. https://doi.org/10.1111/mec.13478

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

Perry GH, Martin RD, Verrelli BC. 2007. Signatures of functional constraint at aye-aye opsin genes: the potential of adaptive color vision in a nocturnal primate. *Mol Biol Evol*, *24*(9):1963–1970. https://doi.org/10.1093/molbev/msm124

Perry GH, Reeves D, Melsted P, Ratan A, Miller W, Michelini K, Louis EE, Pritchard JK, Mason CE, Gilad Y. 2012. A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biol Evol*, *4*(2):126–135. https://doi.org/10.1093/gbe/evr132

Perry GH, Louis EE, Ratan A, Bedoya-Reina OC, Burhans RC, Lei R, Johnson SE, Schuster SC, Miller W. 2013. Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar. *Proc Natl Acad Sci USA, 110*(15):5823-5828. https://doi.org/10.1073/pnas.1211990110

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

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

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

Price EC, Feistner ATC. 1994. Responses of captive aye-ayes (*Daubentonia madagascariensis*) to the scent of conspecifics: a preliminary investigation. *Folia Primatologica*, *62*(1–3):170–174. https://doi.org/10.1159/000156774

Príncipe C, Dionísio De Sousa IJ, Prazeres H, Soares P, Lima RT. 2021. LRP1B: A giant lost in cancer translation. *Pharmaceuticals*, *14*(9):836. https://doi.org/10.3390/ph14090836

Prüfer K, Munch K, Hellmann I, Akagi K, Miller J R, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature*, *486*(7404), 527–531. https://doi.org/10.1038/nature11128

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

Przeworski M. 2003. Estimating the time since the fixation of a beneficial allele. *Genetics*, *164*(4):1667–1676. https://doi.org/10.1093/genetics/164.4.1667

Quinn AD, Wilson D. 2004. *Daubentonia madagascariensis*. *Mammalian Species*, vol. 710:1-6.

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

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

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

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

Schwitzer C, Mittermeier RA, Davies N, Johnson S, Ratsimbazafy J, Razafindramanana J, Louis EE Jr, Rajaobelina SS, editors. 2013. Lemurs of Madagascar: A strategy for their conservation 2013–2016. Bristol: IUCN SSC Primate specialist group, Bristol Conservation and Science Foundation, and Conservation International. 185 pp. ISBN: 978-1-934151-62-4

Shao Y, Zhou L, Li F, Zhao L, Zhang B-L, Shao F, Chen J-W, Chen C-Y, Bi X, Zhuang X-L, et al. 2023. Phylogenomic analyses provide insights into primate evolution. *Science*, *380*(6648):913–924. https://doi.org/10.1126/science.abn6919

Simonsen KL, Churchill GA, Aquadro CF. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, *141*(1), 413–429.

Smits DJ, Schot R, Krusy N, Wiegmann K, Utermöhlen O, Mulder MT, Den Hoedt S, Yoon G, Deshwar AR, Kresge C, et al. 2023. SMPD4 regulates mitotic nuclear envelope dynamics and its loss causes microcephaly and diabetes. *Brain*, *146*(8):3528–3541. https://doi.org/10.1093/brain/awad033

Soler-García ÁA, Maitra R, Kumar V, Ise T, Nagata S, Beers R, Bera TK, Pastan I. 2005. The PATE gene is expressed in the accessory tissues of the human male genital tract and encodes a secreted sperm-associated protein. *Reproduction*, *129*(4):515–524. https://doi.org/10.1530/rep.1.00576

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

Soni V, Jensen JD. 2024b. 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, preprint. https://doi.org/10.1101/2024.09.19.613979

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

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

Stephan W. 2019. Selective sweeps. *Genetics, 211*:5-13. https://doi.org/10.1534/genetics.118.301319

Sterling EJ. 1993. Patterns of range use and social organization in aye-ayes (*Daubentonia Madagascariensis*) on Nosy Mangabe. In P. M. Kappeler J. U. Ganzhorn, editors. *Lemur Social Systems and Their Ecological Basis.* Springer US. p. 1–10. https://doi.org/10.1007/978-1-4899-2412-4_1

Stevison LS, Hoehn KB, Noor MAF. 2011. Effects of inversions on within- and between-species recombination and divergence. *Genome Biol Evol*, *3*:830–841. https://doi.org/10.1093/gbe/evr081

Suzzi-Simmons A. 2023. Status of deforestation of Madagascar. *Global Ecology and Conservation*, *42*:e02389. https://doi.org/10.1016/j.gecco.2023.e02389

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

Terbot JW, Soni V, Versoza CJ, Pfeifer SP, Jensen JD. 2024. Inferring the demographic history of aye-ayes (*Daubentonia madagascariensis)* from high-quality, whole-genome, population-level data. biorxiv, preprint.

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

The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, *437*(7055):69–87. https://doi.org/10.1038/nature04072

The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, et al. 2023. The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1):iyad031. https://doi.org/10.1093/genetics/iyad031

The Rhesus Macaque Genome Sequencing and Analysis Consortium; Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science, 316*(5822):222-234. https://doi.org/10.1126/science.1139247

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

Van der Auwera G, O'Connor B. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. Sebastopol, CA: O'Reilly Media

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

Versoza CJ, Jensen JD, Pfeifer SP. 2024a. Characterizing the rates and patterns of *de novo* germline mutations in the aye-aye (*Daubentonia madagascariensis*). biorxiv, preprint.

Versoza CJ, Jensen JD, Pfeifer SP. 2024b. The landscape of structural variation in aye-ayes (*Daubentonia madagascariensis)*. biorxiv, preprint.

Versoza CJ, Lloret-Villas A, Jensen JD, Pfeifer SP. 2024. A pedigree-based map of crossovers and non-crossovers in aye-ayes (*Daubentonia madagascariensis)*. biorxiv, preprint.

Villoutreix R, Ayala D, Joron M, Gompert Z, Feder JL, Nosil P. 2021. Inversion breakpoints and the evolution of supergenes. *Mol Ecol*, *30*(12):2738–2755. https://doi.org/10.1111/mec.15907

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Gen*, *3*(6):e90. https://doi.org/10.1371/journal.pgen.0030090

Winn RM. 1994. Preliminary study of the sexual behaviour of three aye-ayes (*Daubentonia madagascariensis*) in captivity. *Folia Primatologica*, *62*(1–3):63–73. https://doi.org/10.1159/000156764

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

Yu G, Mu H, Fang F, Zhou H, Li H, Wu Q, Xiong Q, Cui Y. 2022. LRP1B mutation associates with increased tumor mutation burden and inferior prognosis in liver hepatocellular carcinoma. *Medicine*, *101*(26):e29763. https://doi.org/10.1097/MD.0000000000029763

Zozulya S, Echeverri F, Nguyen T. 2001. The human olfactory receptor repertoire. *Genome Biology*, *2*(6), research0018.1. https://doi.org/10.1186/gb-2001-2-6-research0018

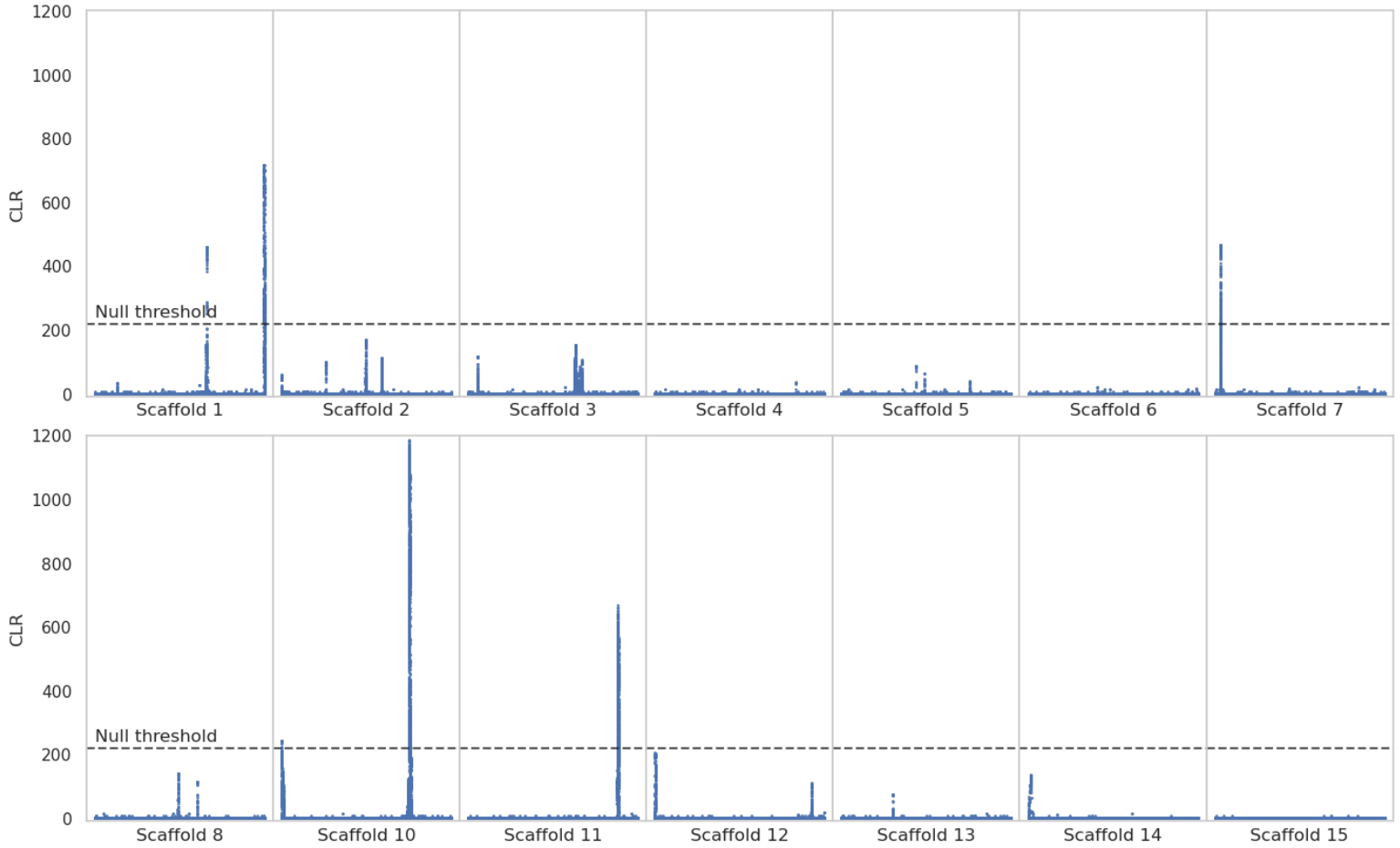

Figure 1: Genome scans for selective sweeps using SweepFinder2. Blue data points are CLR values inferred at each SNP. The dashed line is the threshold for sweep detection, determined by the highest CLR value across 100 simulated replicates of each of the 14 autosomal scaffolds (see "Materials and Methods" section for further details). The x-axis represents the position along the scaffold, and the y-axis represents the CLR value at each SNP.

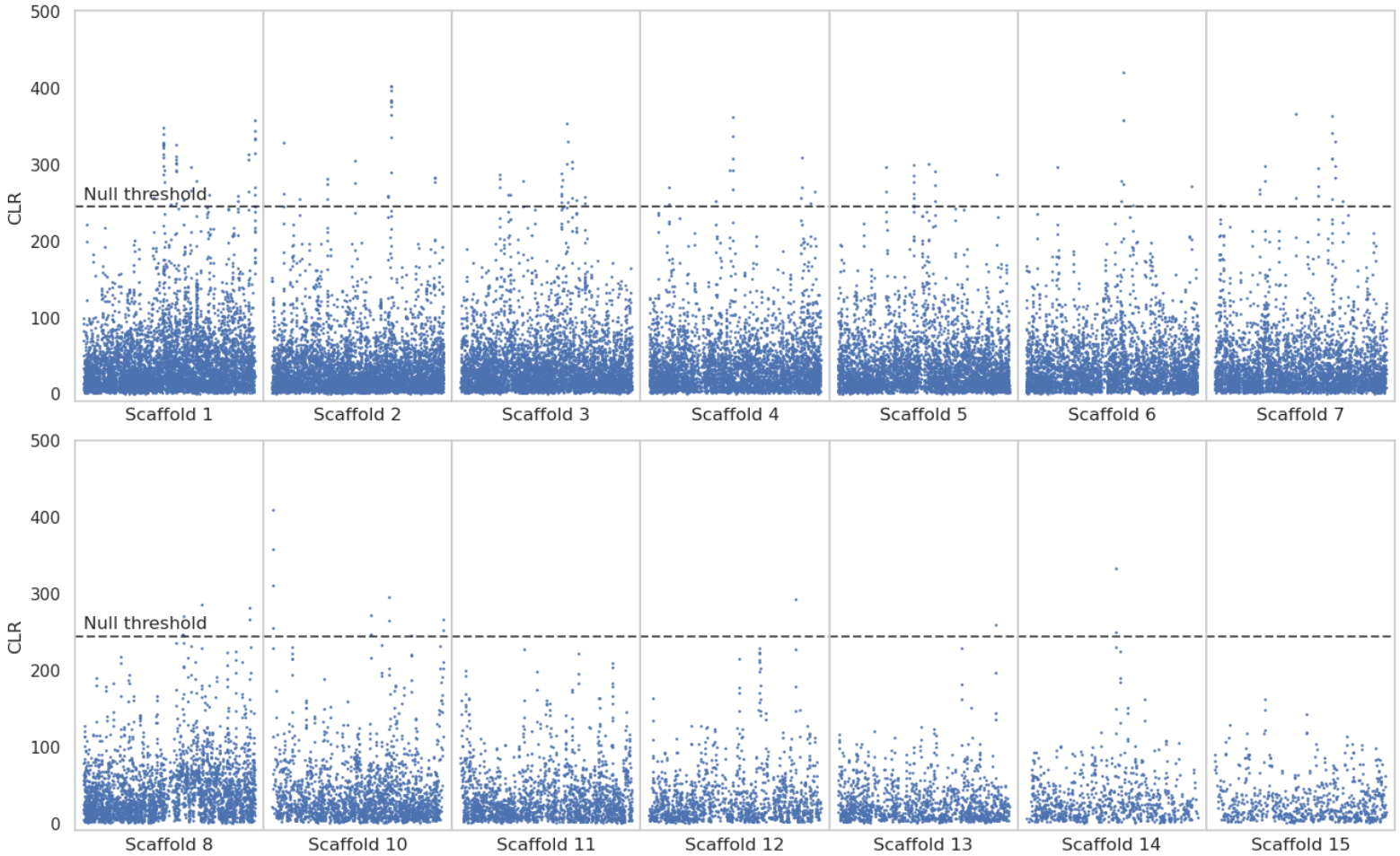

Figure 2: Genome scans for balancing selection using the B_{0MAF} method. Blue data points are CLR values inferred over windows of length 100 SNPs. The dashed line is the threshold for detection, determined by the highest CLR value across 100 simulated replicates of each of the 14 autosomal scaffolds (see "Materials and Methods" section for further details). The x-axis represents the position along the scaffold, and the y-axis represents the CLR value at each window.

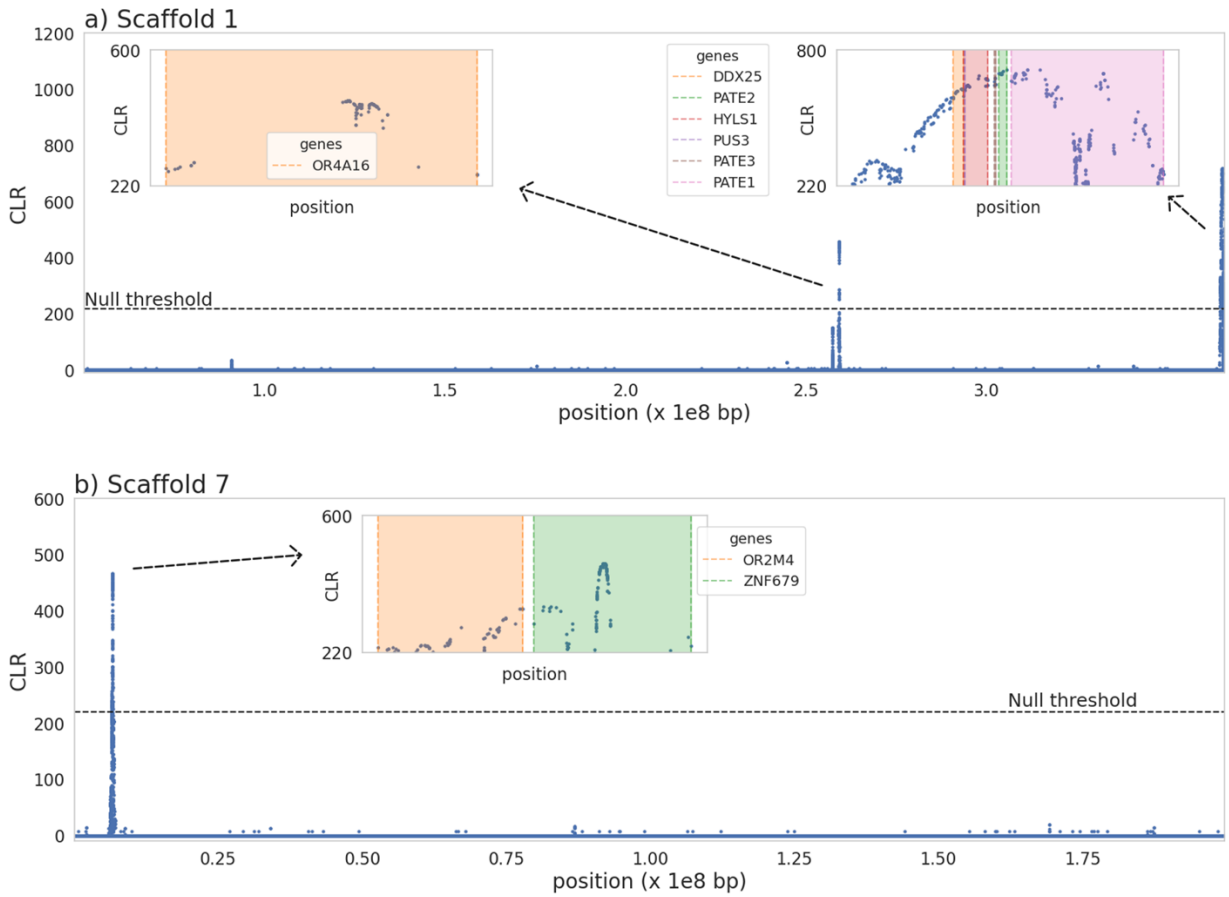

Figure 3: SweepFinder2 selective sweep scan results for a) scaffold 1 and b) scaffold 7. Inset plots zoom in on likelihood surface peaks, with genes in these regions highlighted. The x-axis represents the position along the scaffold, and the y-axis represents the CLR value at each SNP.

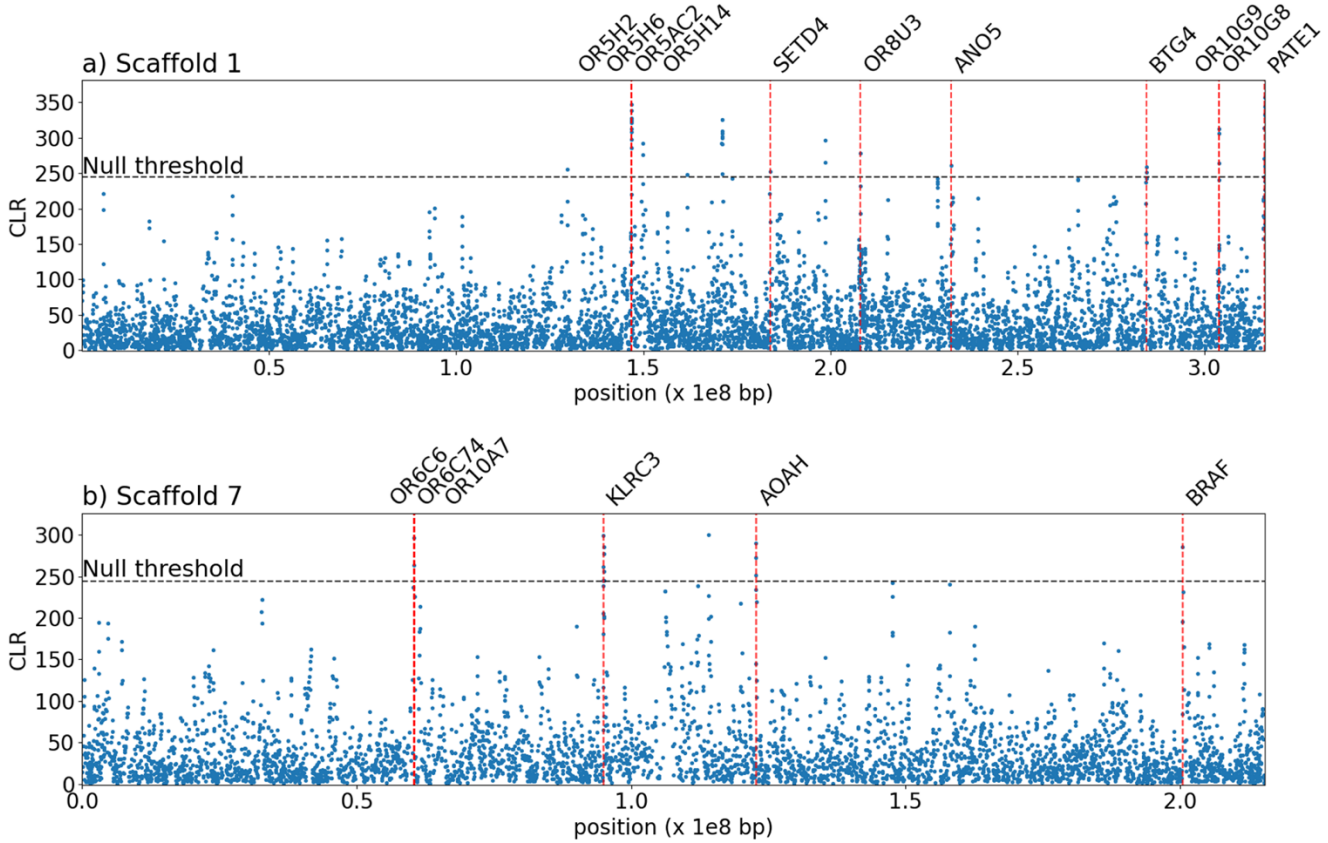

Figure 4: B_{0MAF} balancing selection scan results for 100 SNP window analysis on a) scaffold 1 and b) scaffold 7. Red vertical lines map to candidate genes. Instances where CLR values meet the null threshold, but no gene is denoted, indicates that no gene overlap was found. The x-axis represents the position along the scaffold, and the y-axis represents the CLR value of each 100 SNP window.

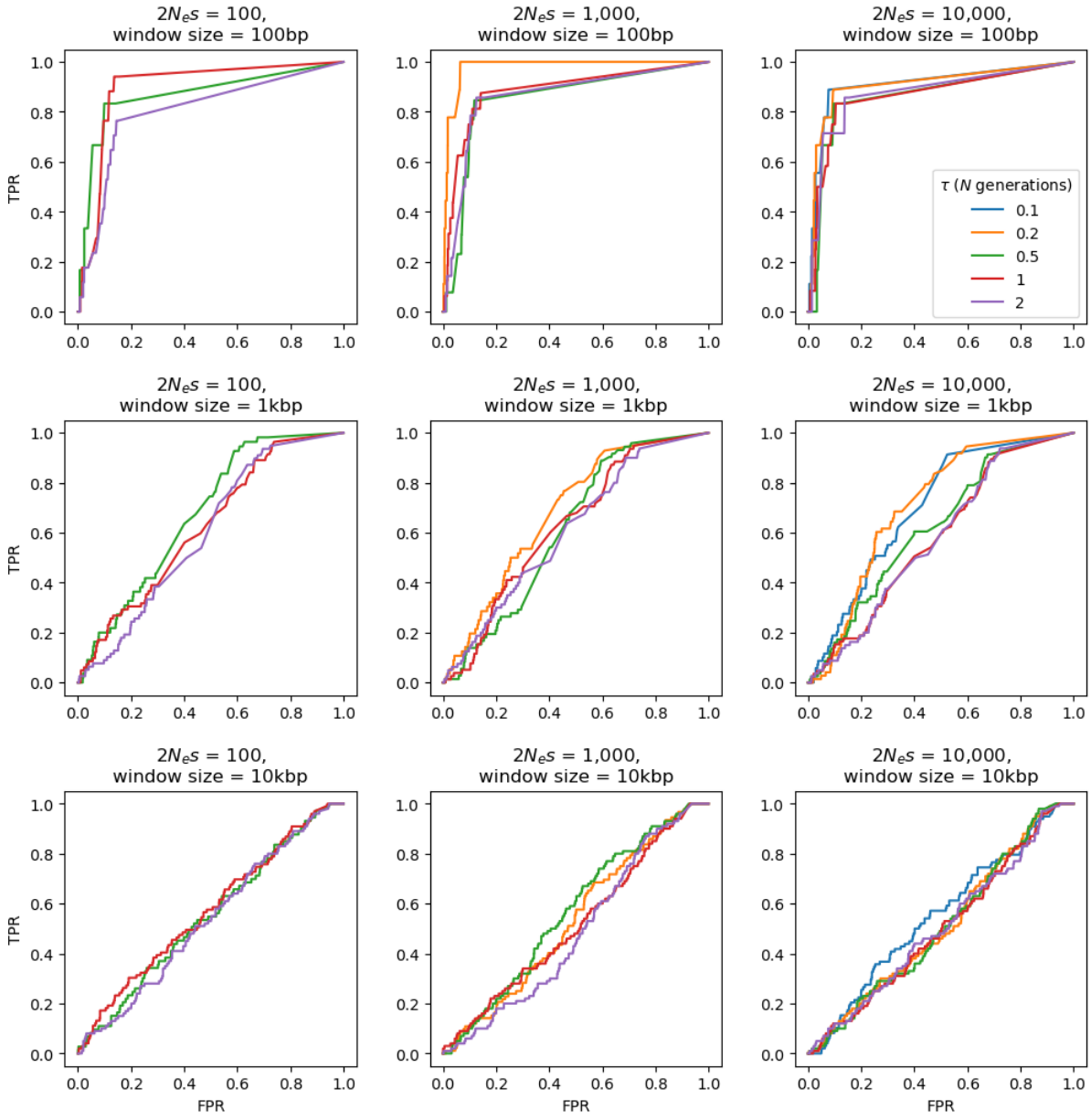

Figure 5: ROC plots for SweepFinder2 showing the change in true-positive rate (TPR) as the false-positive rate (FPR) increases, for sweep inference in aye-ayes across 100 simulation replicates under the Terbot et al. (2024) demographic model, with mutation and recombination rates drawn from a uniform distribution such that the mean rate per simulation rate is equal to the fixed rate (see "Materials and Methods" section). Power analysis was conducted across three selection regimes (population-scaled strengths of selection of $2N_es$ = 100, 1,000, and 10,000), five different times of introduction of the beneficial mutation (τ = 0.1, 0.2, 0.5, 1, and 2, in N generations), and three window sizes (100bp, 1kb, and 10kb). If no ROC is plotted, this is a case in which the beneficial mutation was unable to fix prior to the sampling time in any of the simulation replicates (e.g., at $2N_es$ = 100 and τ = 0.1).

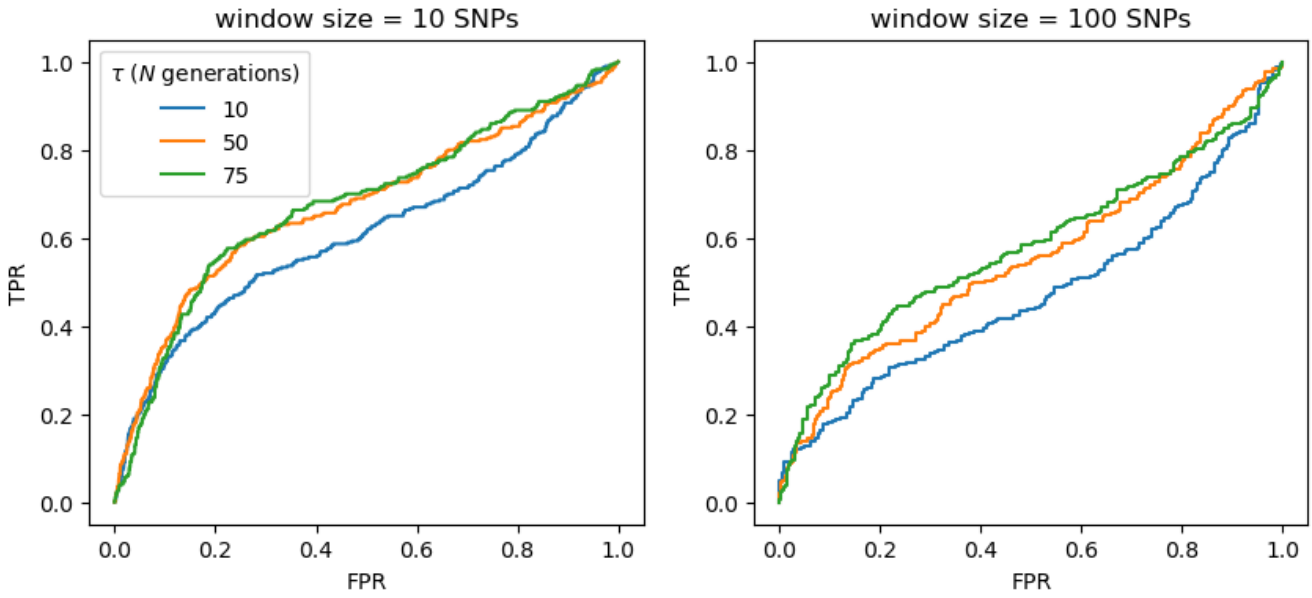

Figure 6: ROC plots for B_{0MAF} showing the change in true-positive rate (TPR) as the false-positive rate (FPR) increases, for balancing selection inference in aye-ayes across 100 simulation replicates under the Terbot et al. (2024) demographic model, with mutation and recombination rates drawn from a uniform distribution such that the mean rate per simulation rate is equal to the fixed rate (see "Materials and Methods" section). Power analyses were conducted across three different times of introduction of the balanced mutation (τ = 10N, 50N, and 75N generations).

Category	Term	P-Value	Fold Enrichment	FDR
KEGG_PATHWAY	hsa04740:Olfactory transduction	3.22E-06	7.23	2.96E-04
GOTERM_BP_DIRECT	GO:0050911~detection of chemical stimulus involved in sensory perception of smell	3.72E-06	9.32	8.16E-04
GOTERM_MF_DIRECT	GO:0004984~olfactory receptor activity	5.06E-06	8.92	4.71E-04
INTERPRO	IPR000725:Olfact_rcpt	5.07E-06	8.98	8.17E-04
UP_KW_BIOLOGICAL_PROCESS	KW-0552~Olfaction	6.24E-06	8.16	1.44E-04
GOTERM_MF_DIRECT	GO:0004930~G protein-coupled receptor activity	2.00E-05	6.22	9.32E-04
INTERPRO	IPR000276:GPCR_Rhodpsn	2.63E-05	6.05	1.70E-03
INTERPRO	IPR017452:GPCR_Rhodpsn_7TM	3.17E-05	5.91	1.70E-03
UP_SEQ_FEATURE	DOMAIN:G-protein coupled receptors family 1 profile	6.06E-05	6.36	2.12E-02
UP_KW_BIOLOGICAL_PROCESS	KW-0716~Sensory transduction	7.17E-05	5.82	8.25E-04
GOTERM_BP_DIRECT	GO:0007186~G protein-coupled receptor signalling pathway	1.62E-04	4.76	1.78E-02
GOTERM_MF_DIRECT	GO:0005549~odorant binding	2.20E-04	16.38	6.83E-03

Table 1: Top 12 hits from gene functional analysis with DAVID on balancing selection candidate genes from 100 SNP window analysis