



**HAL**  
open science

## Genetic variation in *Loudetia simplex* supports the presence of ancient grasslands in Madagascar

George Tiley, Andrew Crowl, Tchana Almary, W. Luke, Cédrique Solofondranohatra, Guillaume Besnard, Caroline Lehmann, Anne Yoder, Maria Vorontsova

### ► To cite this version:

George Tiley, Andrew Crowl, Tchana Almary, W. Luke, Cédrique Solofondranohatra, et al.. Genetic variation in *Loudetia simplex* supports the presence of ancient grasslands in Madagascar. *Plants, People, Planet*, inPress, 6 (2), pp.315-329. 10.1002/ppp3.10437 . hal-04245276

**HAL Id: hal-04245276**

**<https://hal.science/hal-04245276>**










Submitted on 22 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## RESEARCH ARTICLE

# Genetic variation in *Loudetia simplex* supports the presence of ancient grasslands in Madagascar

George P. Tiley<sup>1</sup>  | Andrew A. Crowl<sup>2</sup>  | Tchana O. M. Almary<sup>3</sup>  |  
W. R. Quentin Luke<sup>4</sup>  | Cédrique L. Solofondranohatra<sup>5</sup>  | Guillaume Besnard<sup>6</sup>  |  
Caroline E. R. Lehmann<sup>7,8</sup>  | Anne D. Yoder<sup>2</sup>  | Maria S. Vorontsova<sup>1</sup> 

<sup>1</sup>Royal Botanic Gardens Kew, Richmond, UK

<sup>2</sup>Department of Biology, Duke University, Durham, North Carolina, USA

<sup>3</sup>Kew Madagascar Conservation Centre, Antananarivo, Madagascar

<sup>4</sup>East African Herbarium, National Museums of Kenya, Nairobi, Kenya

<sup>5</sup>Madagascar Biodiversity Center, Parc Botanique et Zoologique de Tsimbazaza, Antananarivo, Madagascar

<sup>6</sup>CNRS-UPS-IRD, UMR5174, Laboratoire Évolution & Diversité Biologique, Université Paul Sabatier, Toulouse, France

<sup>7</sup>School of GeoSciences, University of Edinburgh, Edinburgh, UK

<sup>8</sup>Royal Botanic Garden Edinburgh, Edinburgh, UK

## Correspondence

George P. Tiley, Royal Botanic Gardens Kew, Richmond, UK.  
Email: [g.tiley@kew.org](mailto:g.tiley@kew.org)

## Funding information

European Union's Horizon 2020, Grant/Award Number: 101026923; LabEx TULIP, Grant/Award Number: ANR-10-LABX-0041; The Royal Society International Collaboration Award; Duke University

## Societal Impact Statement

Recognizing *Loudetia*-dominated grasslands were widespread prior to human colonization highlights that open ecosystems were and continue to be an important component of Madagascar's biodiversity. A better understanding of the plant species that form grassland ecosystems is necessary for effective land management strategies that support livelihoods, but substantial financial and logistical barriers exist to implementing conservation genetic studies using contemporary genomic tools. Some challenges for population genetic analyses of non-model polyploids lacking reference genomes can be ameliorated by developing computational resources that leverage a cost-effective data generation strategy that requires no prior genetic knowledge of the target species. This may benefit conservation programs with small operating budgets while reducing uncertainty compared to status quo microsatellite assays.

## Summary

- The extent of Madagascar's grasslands prior to human colonization is unresolved. We used population genetic analyses of a broadly dominant C<sub>4</sub> fire-adapted grass, *Loudetia simplex*, as a proxy for estimating grassland change through time. We carefully examined the utility of target-enrichment data for population genetics to make recommendations for conservation genetics. We explored the potential of estimating individual ploidy levels from target-enrichment data and how assumptions about ploidy could affect analyses.
- We developed a novel bioinformatic pipeline to estimate ploidy and genotypes from target-enrichment data. We estimated standard population genetic summary statistics in addition to species trees and population structure. Extended Bayesian skyline plots provided estimates of population size through time for empirical and simulated data.
- All Malagasy *L. simplex* individuals sampled in this study formed a clade and possibly indicated an ancestral Central Highland distribution of 800 m in altitude and above. Demographic models suggested grassland expansions occurred prior to the Last Interglacial Period and supported extensive grasslands prior to human

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Plants, People, Planet* published by John Wiley & Sons Ltd on behalf of New Phytologist Foundation.

colonization. Though there are limitations to target-enrichment data for population genetic studies, we find that analyses of population structure are reliable.

- Genetic variation in *L. simplex* supports widespread grasslands in Madagascar prior to the more recent periods of notable paleoclimatic change. However, the methods explored here could not differentiate between paleoclimatic change near the Last Glacial Maximum and anthropogenic effects. Target-enrichment data can be a valuable tool for analyses of population structure in the absence a reference genome.

#### KEYWORDS

anthropogenic climate change, conservation genetics, effective population size, HybSeq, polyploidy, target-enrichment

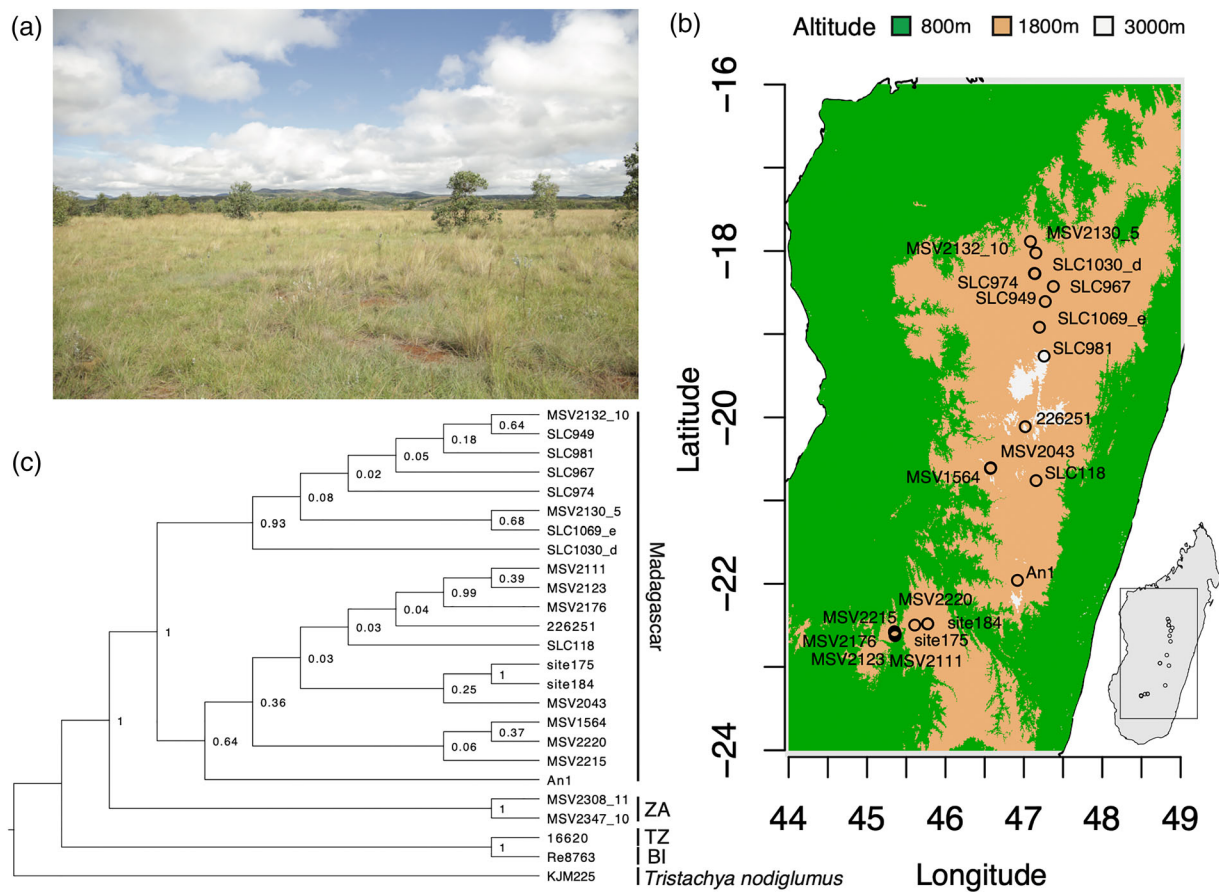
## 1 | INTRODUCTION

The degree to which anthropogenic activities have transformed the ancestral vegetative composition of Madagascar's ecosystems is heavily debated. While the narrative that most of central Madagascar was once covered in closed-canopy forests has been rejected (Bond et al., 2008, 2023; Hackel et al., 2018; Lehmann et al., 2022; Solofondranohatra et al., 2020; Vorontsova et al., 2016), the degree to which a former mosaic of grassland, shrubland, and forests has become more dominated by grasses is contentious (reviewed in Joseph et al., 2021). Understanding population dynamics of individual ecologically important species and plant communities will have profound effects on conservation policy decisions and the economic productivity of grasslands for local communities. However, it is difficult to determine where grasslands, which we consider to be an ecosystem where the ground layer is a majority of tussock-forming grass with little tree cover that experiences a dry season with regular fire regimes (e.g., Dixon et al., 2014), were in the past and how extensive they were. Stratigraphic sediment records from Central Madagascar indicate an increased relative abundance of grasses and fire through the Holocene (Burney, 1987; Gasse & Van Campo, 1998; Virah-Sawmy et al., 2010), especially near the expected age of human colonization (c. 2000 years ago [ka]; Dewar & Wright, 1993; Dewar et al., 2013; Pierron et al., 2017). Megafaunal isotope records also support Holocene expansions of  $C_4$  grasslands but show a vegetative mosaic of  $C_4$  grasses, and  $C_3$  vegetation was at least present (Crowley et al., 2021) near the last glacial maximum (LGM; c. 26.5–19 ka). Reconstructing spatio-temporal vegetative dynamics beyond the LGM begins to approach limitations of paleoecological data, but molecular evolutionary approaches can fill the gaps by providing probabilistic inferences through time. Population genetic approaches for reconstruct demographic histories have been successfully applied to lemurs to reconstruct patterns of past forest connectivity (Quéméré et al., 2012; Salmons et al., 2017; Teixeira et al., 2021; Tiley et al., 2022; Yoder et al., 2016) and may similarly provide insights into the Late Pleistocene dynamics of grasses between the LGM and Last Interglacial period (LIG; c. 132–112 ka) or older. This strategy has

been used to reconstruct vegetative dynamics of temperate Japanese grasslands from forb species (Yamaura et al., 2019).

An array of tools now exists to estimate the effective population size ( $N_e$ ) of species through time. As generating genomic data for non-model groups has become more accessible, population genomic methods that utilize precisely estimated site frequency spectra (SFS; Blischak et al., 2023; Excoffier et al., 2013; Gutenkunst et al., 2009; Liu & Fu, 2020; Nielsen, 2000) have become a valuable tool for reconstructing the demographic history of natural populations. There are still barriers to generating population genomic data though, either due to the lack of suitable tissue, an appropriate reference genome, or costs. Target-enrichment libraries have been transformative for the analysis of plant phylogenetic relationships (e.g., Baker et al., 2022; Breinholt et al., 2021), but the population genetic utility of such data is not well-explored. A survey of taxa with low sample sizes suggested population genetic parameters estimated from target-enrichment data could be unreliable, in addition to violating assumptions of the neutral coalescent model that underpins most tests of selection and demographic change (Sлимп et al., 2021). However, the relatively long loci recovered by target-enrichment (or HybSeq) methods could be appropriate for a demographic modeling approach, the extended Bayesian skyline plot (EBS; Heled & Drummond, 2008), which has been successfully applied to a range of conservation genetic questions despite some model violations. We are interested in exploring the potential of target-enrichment data for population or conservation genetic insights from plant groups that may be lacking genomic resources.

Here, we examined the population genetics of the grass species *Loudetia simplex* (Tristachyidae: Panicoideae: Poaceae). It is largely dominant in Madagascar's Central Highlands (Koechlin, 1993; Hagl et al., 2021; Figure 1a) and is notable for its fire- and grazing-adapted functional traits (Solofondranohatra et al., 2018). Population genetic analyses of microsatellites suggest Malagasy *L. simplex* has been isolated from mainland African populations with no detectable level of gene flow and that there is additional population structure between the northern and southern extents of the species range across the Central Highlands of Madagascar (Hagl et al., 2021). Here, we use a subset of individuals analyzed by Hagl et al. (2021) along with a new



**FIGURE 1** *Loudetia simplex* distribution and single origin in Madagascar. (a) A grassy ecosystem dominated by *L. simplex* near Antsirabe Madagascar planted with eucalyptus (photo credit: George P. Tiley). (b) Our sampling of *L. simplex* was based on collections made throughout Madagascar's Central Highland Plateau, which is typically considered an altitude between 800 and 1800 meters (m). (c) Phylogenetic relationships of individuals based on the phased (PH) data. Numbers next to nodes are posterior probabilities. ZA = South Africa, TZ = Tanzania, BI = Burundi.

sample from the center of the *L. simplex* distribution to explore the population genetic utility of target-enrichment data and its implications for the natural history of Madagascar's grasslands. Because Malagasy *L. simplex* are putative polyploids (tetraploids and hexaploids; Hagl et al., 2021), we developed novel bioinformatic tools for the processing and analysis of polyploid data, integrated into the PATÉ allele phasing pipeline (Tiley et al., 2021).

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling and sequencing

Target-enrichment data were generated for 24 individuals of *L. simplex* and one outgroup, *Tristachya nodiglumus* (Table S1). Twenty individuals were selected to best represent Madagascar from available collections as a single unstructured population and avoid complications of population substructure and low sample sizes in downstream analyses (Figure 1b). Genomic DNA samples were used from Hagl et al. (2021), and new extracts used a modified CTAB protocol (Doyle & Doyle, 1987) from silica-dried leaf tissue. Genomic DNA was then sent to Rapid

Genomics (Gainesville, FL, USA) where sample enrichment and sequencing was performed with a bait kit developed for angiosperms following Breinholt et al. (2021). All sample metadata are provided in Table S1.

### 2.2 | Sequence assembly and processing

Initial contigs were assembled with HybPiper (Johnson et al., 2016) with default settings. We then used HybPiper's supercontigs as reference sequences for PATÉ (Crowl et al., 2022; Tiley et al., 2021) to genotype individuals, filter low-quality variants with GATK v4.2.0 (DePristo et al., 2011; McKenna et al., 2010), and phase haplotype sequences using HPOP-G (Xie et al., 2016). Ploidy information for individuals (Table S1) was based on previous microsatellite allele count data (Hagl et al., 2021); however, we also implemented a statistical test to infer ploidy level in PATÉ based on mixtures of distributions from allele balance data, similar to other approaches that use the number of reads supporting an alternate allele over total reads to infer ploidy (Weiß et al., 2018). We extended the mixture models of Tiley et al. (2018) to determine the ploidy of a sample with the model weights as a selection criterion (Methods S1). PATÉ

control files with all filtering options for GATK are provided in the [Supporting Information](https://doi.org/10.5061/dryad.905qfttx) (<https://doi.org/10.5061/dryad.905qfttx>). Because target-enrichment probes may represent multiple exons from the same gene with overlapping flanking regions, we were concerned that some assembled loci might overlap and violate some independence assumptions in downstream analyses. Therefore, we performed an all-by-all blastn search (Camacho et al., 2009) with an e-value of  $1 \times 10^{-5}$  among the supercontig assemblies followed by single-linkage clustering (<https://github.com/gtiley/HybSeq-Cleaning>). Only the longest locus per cluster was retained for downstream analyses.

To investigate nucleotide variation and the potential population genetic value of non-coding versus coding regions of target-enrichment loci, we identified the following regions:

- *Left flank*—Non-coding sequence upstream of the 5' start codon or splice site.
- *Core*—The coding region targeted by probes for sample enrichment.
- *Right flank*—Non-coding sequence downstream of the 3' stop codon or splice site.

Identification of individual regions was automated (<https://github.com/gtiley/Locus-Splitter>) based on the tblastx program from BLAST v2.10.0 (Camacho et al., 2009) to find in-frame alignments between our assembled sequences and the original transcriptome loci used for probe development. Individual regions were aligned with MAFFT v7.471 (Katoh & Standley, 2013). The aligned regions were then concatenated together to make whole-locus alignments.

Because we are interested in investigating the analytic impact of ploidy in sequence data, we created three datasets from the phased ploidy-aware data:

- *One haplotype (OH)*—Only one phased haplotype sequence was randomly selected from each individual for each locus.
- *Ambiguous genotype (AG)*—The genotypic information for an individual is represented as a single sequence with IUPAC ambiguity codes for biallelic sites. Analyses that treat IUPAC codes as missing data were randomly resolved into two sequences.
- *Phased haplotypes (PHs)*—All PH sequences for an individual are used.

We expect the OH data to be representative of majority consensus sequences (e.g., HybPiper supercontigs) because our target-enrichment loci are anchored in single exons, and we anticipate that any risk of assembling chimeric sequences from such data without paralog warnings is low. Any site genotyped as polyallelic, having more than one alternate allele, was assumed to be an error and treated as missing sequence across all datasets. All expected copies of a locus are represented even if they are invariant. For example, some individuals in the PH data may have six identical sequences at a locus if they are hexaploids, and every site is homozygous. While the ploidy

at an individual locus may vary due to differential gene loss among chromosomes, this is a simplifying assumption to preserve expected allele frequency information.

## 2.3 | Phylogeny estimation

A species tree was estimated for all *L. simplex* individuals using whole-locus sequences for the OH, AG, and PH datasets with BPP v4.6.2 (Flouri et al., 2020). We chose a full-likelihood multispecies coalescent (MSC; Rannala & Yang, 2003) model because it is robust to low-information gene tree estimates (Xu & Yang, 2016). Markov chain Monte Carlo (MCMC) analyses collected 10,000 posterior samples, saving every 10 samples after a 10,000 generation burnin. Model and prior specifications are provided in [Methods S2](#). The starting tree for our Bayesian species tree analyses was based on a concatenated maximum likelihood (ML) estimate from the OH data with IQTREE v2.1.2 (Minh et al., 2020). We used unpartitioned analyses with ModelFinderPlus (Kalyaanamoorthy et al., 2017) for the OH, AG, and PH data to estimate ML evolutionary distances between individuals, with random concatenation of haplotypes within individuals.

## 2.4 | Population structure

Potential population structure within our sample of *L. simplex* from Madagascar was explored with STRUCTURE v2.3.4 (Pritchard et al., 2000). For each dataset, we sampled only one biallelic SNP per locus, if present, and required a minimum minor allele count of three with no more than 25% missing data. To explore uncertainty due to sampling error of sites in a target-enrichment dataset that is relatively small compared with contemporary population genomic data, we generated 10 datasets with random SNP sampling. The number of clusters that best satisfy Hardy–Weinberg assumptions was determined with the  $\Delta K$  method (Evanno et al., 2005). Principal component analyses (PCAs) were conducted on the OH, AG, and PH datasets with adegenet v2.1.3 (Jombart & Ahmed, 2011) by compressing allele counts to the same dimension as the lowest ploidy level present. For example, a hexaploid with three reference and three alternate alleles would be scored as a 2 on a scale from 0 to 4, to be comparable with tetraploids in the dataset. Additional details on allelic imbalances would be lost as, for example, hexaploids with either one or two alternate alleles would be represented as a 1 on the tetraploid scale. Scripts for generating PCA or other clustering-ready matrices from FASTA files are available on GPT's GitHub (<https://github.com/gtiley/fasta-conversion>). Additional details about compression of PCA data matrices can be found in [Methods S3](#). The presence of isolation-by-distance as an explanation for observed patterns of genetic structure and ordination was tested with the patristic distances from the ML analyses or the uncorrected pairwise distances (p-distances) regressed against the natural logarithm of the great circle distance between individuals. A Mantel test was performed for isolation-



by-distance analyses in the ape v5.6.2 R package (Paradis & Schliep, 2019).

## 2.5 | Population genetic summary statistics

The pairwise nucleotide diversity ( $\pi$ ), number of segregating sites ( $S$ ), and Tajima's  $D$  were calculated across locus regions and datasets with the PopGenome v2.7.5 R package (Pfeifer et al., 2014). The significance of Tajima's  $D$  statistics was determined by 100 standard coalescent simulations with ms (Hudson, 2002). We generated 200 bootstrap replicates of each alignment to calculate a percentile bootstrap 95% confidence interval and compare the relative estimation error between regions and datasets using the root mean square error (RMSE) from the bootstrap samples. The scaling and interpretation of parameters are described in Methods S4. We tested for differences in the distribution of summary statistics among locus regions (left flanks, cores, and right flanks) and data types (OH, AG, and PH) with Wilcoxon rank-sum tests. We tested for biases in Tajima's  $D$  results among locus regions and data types with  $\chi^2$  tests.

## 2.6 | Demographic history

We chose the EBSP (Heled & Drummond, 2008) as the most appropriate model of  $N_e$  change through time for our sample of target enrichment data. Here, we assumed a strict molecular clock underlying all lineages but allowed rates of evolution to vary among loci. One locus was fixed to the clock rate, and all other locus rates were relative. Rates of molecular evolution among plants are especially complicated, as there can be life histories with variable reproductive strategies, unknown generation times, and overlapping generations. Thus, we relied on an estimate of the neutral substitution rate at third-position synonymous sites from grass genomes (Christin et al., 2014) and constructed a vague prior around this mean rate, such that any biologically plausible uncertainty in the rate of evolution is reflected in our posterior sample. It is a tenuous rate estimate at best, because it remains unknown whether polyploid *L. simplex* largely propagates clonally, through apomixis, or through sexual reproduction. The MCMC analyses for the OH, AG, and PH data collected 20,000 samples after a burnin of 50,000,000 and sample interval of 10,000. Two replicates were run for each analysis to evaluate convergence, and mixing was assessed with Tracer v1.7 (Rambaut et al., 2018). All model and prior specifications are provided in Methods S5 and are replicable from XML files on Dryad (<https://doi.org/10.5061/dryad.905qfttqx>).

## 2.7 | Adequacy of demographic models

To explore if EBSP models could reliably differentiate between Late Pleistocene versus Holocene population size increases, we evaluated the accuracy of EBSP inferences with simulations that captured some features of our target-enrichment data. We simulated 57 loci that

were 1000 bp in length for 20 haploid individuals 100 times with ms (Hudson, 2002) and seq-gen (Rambaut & Grass, 1997) for four scenarios: (1) a constant  $N_e$  of 100,000, (2) an  $N_e$  increase from 10,000 to 100,000 individuals roughly coinciding with the LGM at 20,000 years in the past, (3) the same 10-fold increase but at 2000 years in the past to reflect anthropogenic expansion, and (4) a scenario combining both LGM and anthropogenic effects by exponentially increasing  $N_e$  from 10,000 to 100,000 between 20,000 years ago and the present. Each simulated dataset was then evaluated with the same EBSP model using a fixed clock used for our empirical analyses, except the MCMC algorithm collected 10,000 posterior samples with a sample interval of 1000 after a 100,000 sample burnin. Commands for simulations are given in Methods S6.

## 3 | RESULTS

### 3.1 | Sequencing statistics

The probe set targeted 408 loci from which an average of 327 loci were assembled by HybPiper (Table S2). These were supported by an average of 217 read pairs per locus. Although there was variable read depth across loci and individuals, the number of assembled loci was significantly correlated with read depth (Pearson's correlation coefficient = .57,  $p$  value = .002). After joint-genotyping all individuals as diploids, normal mixture models suggested all individuals were hexaploids, with the exception of two individuals from South Africa (Table S3). Some individuals were inferred as pentaploids, but because this was unexpected based on previous observations from microsatellites (Table S1), these cases were assumed to be hexaploids. Although available evidence leaves the ploidy of some individuals ambiguous and raises skepticism of the inference methods, genotyping and phasing of individuals proceeded with the ploidy suggested by our mixture models. We inferred an average of 247 loci with allelic variation, which had a heterozygosity of 0.012 variants per base. After reducing the number of loci with BLAST clustering to avoid tight linkage, a final 301 loci were retained from the OH, AG, and PH datasets for downstream analyses. The average number of haplotype blocks across all loci was 1.005, meaning that in very few cases, there was insufficient data to phase across the entire locus (Table S2). In cases where multiple variants within a locus could not be phased, only the variants from the longest block were retained, and variants on the shorter block were replaced with missing data.

### 3.2 | Phylogeny estimation

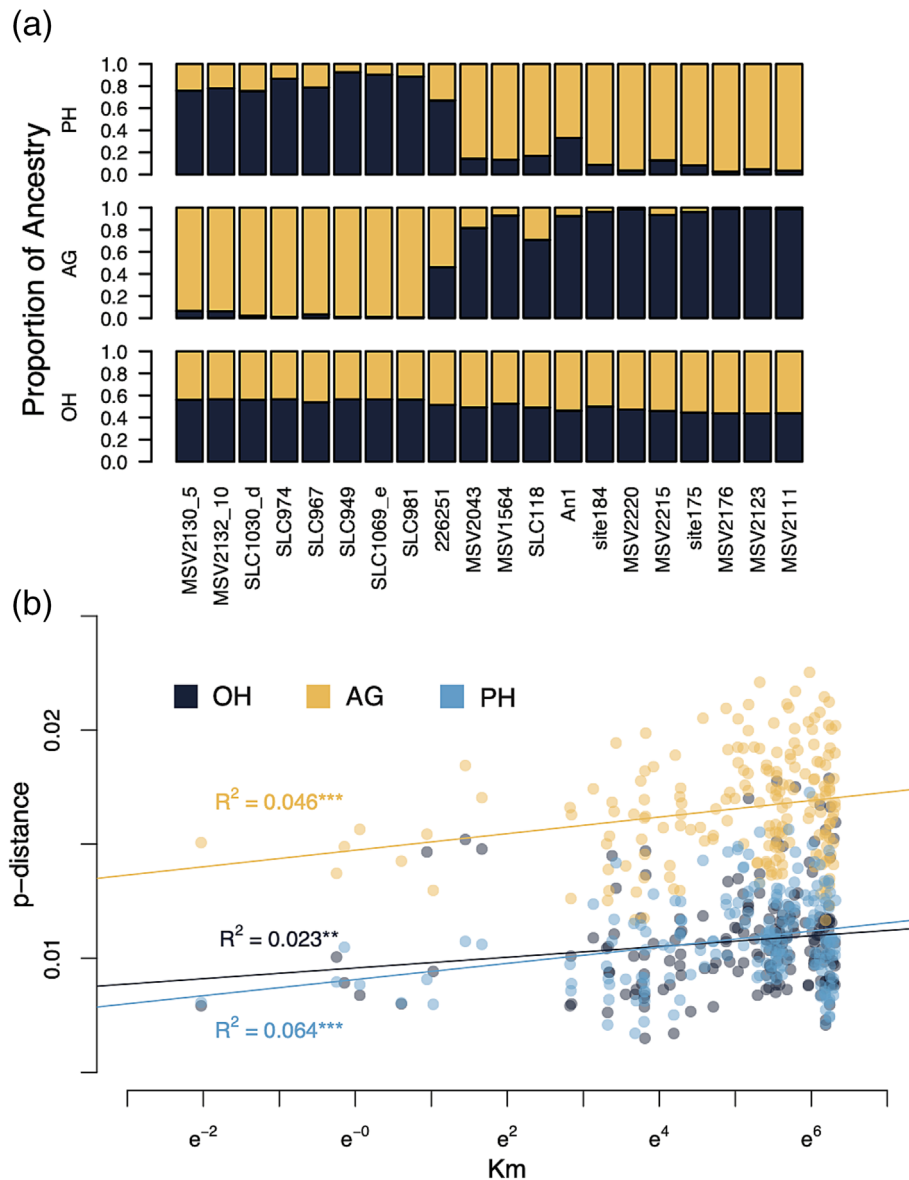
Species trees revealed all Malagasy *L. simplex* form a well-supported clade (Figure S1), but there was disagreement in the biogeographic history within Madagascar based on the treatment of ploidy. Individuals in northern Madagascar formed a clade with a posterior probability of 0.9 or greater across the OH, AG, and PH analyses. However, the individuals from central and southern Madagascar were

paraphyletic in the OH and AG analyses (Figure S1a,b). A clade of central and southern individuals was recovered in the PH analysis but with a weak posterior probability of 0.64 (Figures 1c and S1c). Individuals from the center of the *L. simplex* distribution appear to be the earliest diverging lineages in the OH and AG analyses (Figure S1a,b) or with respect to the sample of central and southern individuals in the PH analysis (Figures 1c and S1c).

### 3.3 | Population sub-structure in Madagascar

STRUCTURE analyses recovered a variable number of Hardy-Weinberg groups across the 10 datasets that randomly sampled one non-singleton biallelic SNP per locus (Table S4). For the OH data, we found a  $K$  of two, three, or four in eight, one, and one of the respective replicates. For AG data, the optimal number of clusters was more

ambiguous with a  $K$  of two, three, and four recovered for four, five, and one replicates. PH data found  $K$  of two and three for eight and two replicates. We note that the method of Evanno et al. (2005) cannot infer if one panmictic population is the best model, and while there is reasonable geographic structure at  $K$  of two weighted across biological and technical replicates (Figure 2a), the nonzero admixture between individuals at the most extreme ends could indicate a single population with a cline of genetic variation. This is further supported by the lack of distinct cluster assignments at higher values of  $K$  despite some evidence for additional population structure (Figure S2). A weak signal of isolation-by-distance was detected when regressing genetic distance against spatial distance (Figure 2b). Notably, the AG data inferred much higher pairwise distances between individuals; although, the relative differences and overall trend were similar. The same patterns can be observed when using patristic distances from ML phylogenetic estimates, although the disparity



**FIGURE 2** Patterns of genetic structure and isolation by distance in *Loudetia simplex* from Madagascar across data types. (a) Aggregated STRUCTURE results across runs for an a priori number of clusters of two ( $K = 2$ ) for the one haplotype (OH), ambiguous genotype (AG), and phased haplotypes (PH) data types. (b) A plot of natural-log-scaled geographic distance against the pairwise distance ( $p$ -distance). Lines are from a least-squares linear regression with the coefficient of determination ( $R^2$ ) shown for each data type. Statistical significance was determined with a Mantel test using 9999 permutations. Significance levels are  $p$  value  $< .05$  (\*),  $p$  value  $< .01$  (\*\*), and  $p$  value  $< .001$  (\*\*\*). Colors represent different data types: one haplotype (OH), ambiguous genotypes (AG), and phased haplotypes (PH).

between AG versus OH or PH distances is reduced, as is the percent of genetic variation explained by spatial distances (Figure S3).

PCAs reflected findings from STRUCTURE analyses and reinforce the presence of a genetic cline (Figure 3). For OH data, where only one reference or alternate allele was observed per individual, Malagasy *L. simplex* is clearly distinguished from mainland African individuals (Figure 3a), but there is no clear separation of individuals based on geographic location (Figure 3b). PCAs of data that account for heterozygosity but not dosage recovered a single Malagasy versus mainland African pattern (Figure 3c), and PCAs reflect a north-to-south gradient of genetic variation regardless of whether allelic dosage was treated as hexaploid (Figure 3d), tetraploid (Figure 3e), or diploid (Figure 3f). The total percent of variance explained by principal component axis 1 (PC1) and axis 2 (PC2) increased as allelic dosage was compressed from a hexaploid to a diploid state.

### 3.4 | Variation in summary statistics

There were significant differences in the distributions of  $\pi$ ,  $S$ , and Tajima's  $D$  among locus regions and between the OH, AG, and PH data (Figure 4). For  $\pi$ , there was always a difference between flanking regions and the exon core (Table S5), with the cores surprisingly having more variation. Differences were also observed between the flanking regions and cores for  $S$  and Tajima's  $D$ , except when comparing the left flanks and cores for OH data (Table S5). The only case of a difference between the left and right flanks was when estimating  $\pi$  for AG data (Table S5). The OH and PH data resulted in similar estimates of  $\pi$ ; however, the AG and PH data had similar distributions of  $S$  (Table S5). Any differences in Tajima's  $D$  statistics were largely explained by an over-representation of values less than negative two, indicating positive selection or population size expansion, in the OH data and among core regions across datasets (Table 1).

We attempted to explore biases in summary statistic estimates due to sampling error of sites through bootstrapping the OH (Figure S4), AG (Figure S5), and PH data (Figure S6). Our observed summary statistics typically fell outside of their percentile bootstrap 95% confidence intervals, but other confidence interval approximation techniques (e.g., Efron & Narasimhan, 2020) may be more appropriate for population genetic summary statistics (Leblois et al., 2003). However, we used the bias, difference between the bootstrap mean and observed estimate, and the RMSE of bootstrap replicates with respect to the observed estimate, to gain some general insights. Notably, the bias of  $\pi$  is larger and much noisier in the AG data (Figure S5) compared with the OH (Figure S4) and PH (Figure S6) data. Even if the bias for a whole-locus is low, there can be high variability between regions. For the observed values, estimates are not always proportional among regions either. The RMSE of an estimate tends to decrease along with genetic variation, implying some role of genotyping error or rare variants and singletons inflating values at the higher end of the spectrum. Loci with very low levels of genetic variation can also be problematic though, as the RMSE for flanking regions increases towards the tail of the distribution, creating opportunity for

sampling error to wildly swing estimates regardless of ploidy (Figures S4–S6). The number of segregating sites increased with ploidy of the data, but  $\pi$  was relatively stable. However, differences between  $\pi$  and  $S$  will affect estimates of Tajima's  $D$ , and this could affect interpretations of demography or selection acting on loci; although, the majority of loci and their regions were not significantly different from neutral expectations given the simulation-based test statistics (Table S6).

### 3.5 | Population size change through time

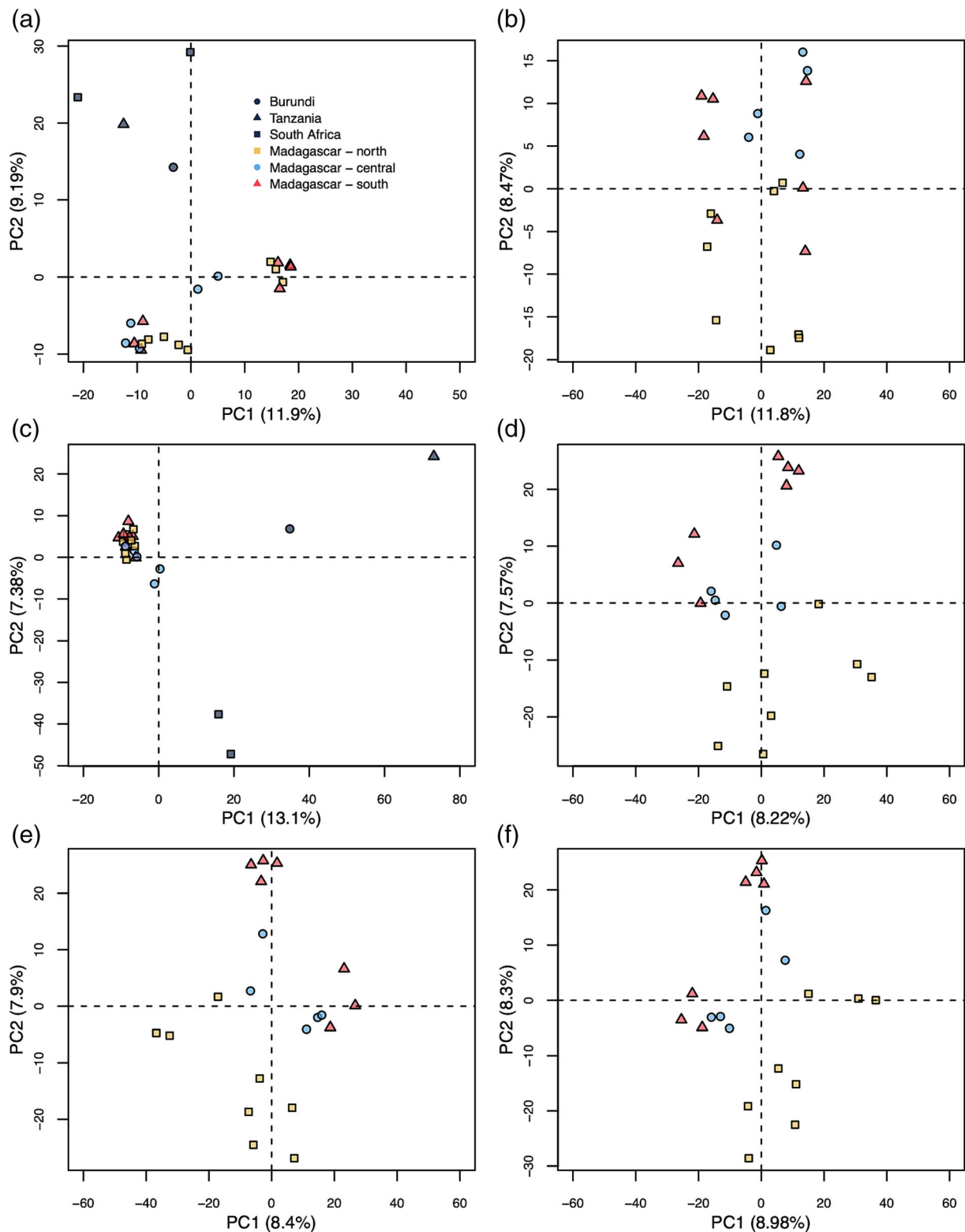
After filtering for longer loci with low amounts of missing data, 57 loci were retained for EBSP analyses. Reducing our data to a subset of more informative loci was done for computational convenience, because EBSP models are parameter-rich and require long MCMC runs for convergence. Runs converged for each data type for both analyses with a fixed rate of molecular evolution (Figure 5) and variable rate (Figure S7). All analyses suggested the *L. simplex*  $N_e$  increased through the Late Pleistocene, well before the LIG, approximately one million years ago. Between the LGM and period of potential human colonization of Madagascar in the Holocene,  $N_e$  had already reached present-day estimates.  $N_e$  estimates through the Holocene were only possible for the PH data, as accounting for the within-individual variation should have captured more recent coalescences. Simulations cast doubt on the seriousness with which EBSP results can be interpreted through the Holocene though. For example, when  $N_e$  was truly constant through time, some individual EBSP analyses would suggest recent  $N_e$  increases; although, the mean estimate across simulations recovered the constant population scenario (Figure 6a). This problem was exaggerated in the presence of model over-fitting (Figure S8). In scenarios with recent  $N_e$  increases, the contemporary  $N_e$  was underestimated, but a trend indicating a recent  $N_e$  increase was detectable (Figure 6b–d). However, simulations indicated it would not be possible to differentiate between  $N_e$  increases that occurred during the LGM (Figure 6b) versus the Late Holocene (Figure 6c), let alone some combination of paleoclimate and anthropogenic effects (Figure 6d). Interpreting these recent  $N_e$  increases from EBSP analyses becomes impossible with model over-fitting (Figure S8).

## 4 | DISCUSSION

### 4.1 | Robustness of analyses across loci and ploidy

Although target-enrichment data have been transformative for species-level phylogenomic studies (Baker et al., 2022; Breinholt et al., 2021), the robustness of population genetic estimates for a sample of closely related individuals was largely unexplored (but see Slimp et al., 2021). Some of the most important measures of genetic variation for populations were susceptible to sampling error, especially in the flanking intron regions (Figures S4–S6). Although estimates from cores showed less error, they will lead to some violations of



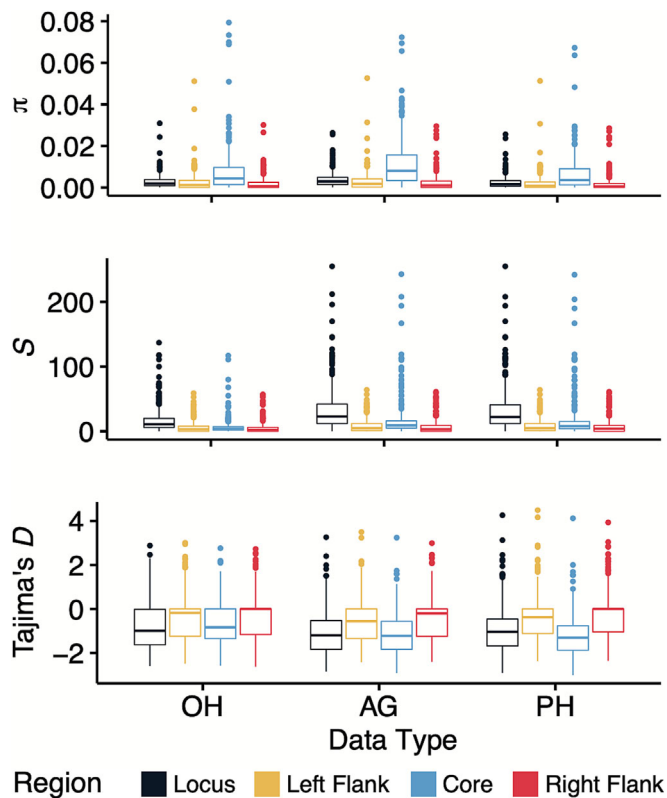


**FIGURE 3** Principal component analyses of genetic variation in *Loudetia simplex* across data types. (a) All individuals using one haplotype (OH) data. (b) Only Malagasy individuals using OH data. (c) All individuals treating all sites as diploid with ambiguous genotype (AG) data. (d) Only Malagasy individuals treating all sites as hexaploid with phased haplotype (PH) data. (e) Only Malagasy individuals compressing unbalanced allelic dosage to single categories. (f) Only Malagasy individuals treating all sites as diploid by considering all heterozygosity regardless of dosage to be equivalent.

population genetic methods that assume evolution consistent with neutral theory, given potential signatures of purifying selection (Table 1). This creates some conflict, because most of the information is contained within the exon cores of target-enrichment loci (Figure 4), presumably due to read coverage. Because Tajima's  $D$  alone

cannot disentangle demography from selection, some judgment calls and tolerance for downstream model violations are necessary. The majority of loci examined were consistent with neutral expectations based on simulations (Table S6), and the proportion of loci with significant Tajima's  $D$  statistics were exaggerated when ignoring heterozygosity (Table 1). Although, the estimates of nucleotide diversity and simulations under the standard neutral coalescent used here should be reasonable for polyploids in cases of perfect multisomic inheritance in outcrossers (Arnold et al., 2012; Blischak et al., 2023), we note that the presence of subgenome structure and selfing could cause deviations in expectations of allele frequencies (e.g., Blischak et al., 2023; Meirmans & Van Tienderen, 2013; Ronfort et al., 1998). The mechanism of polyploidization is unknown in *L. simplex*, but our results do provide some expectations of how genotyping choices for polyploids can impact standard summary statistic estimates.

Predicting the effects of assumptions about genotype on estimates of genetic variation and downstream analyses are not always straightforward. For example, the pairwise distances (Figure 2b) or evolutionary distances (Figure S3) between individuals were higher when assuming all individuals were diploid (AG) compared with assuming all individuals were haploid (OH) or genotyped at their anticipated ploidy level (PH). If we were willing to ignore the absolute estimates of genetic distance, the interpretation of isolation-by-distance analyses was the same across the OH, AG, and PH data. This implies that biologically meaningful conclusions can arise even with uncertainty in ploidy and phasing. Interestingly, treating all individuals as diploid was sufficient for STRUCTURE to recover northern and southern clusters with some admixed central individuals of varying proportions (Figures 2a and S2). Similarly, PCAs reflected the geographic distribution of our *L. simplex* samples in the AG and PH data with, surprisingly, the AG data explaining more of the observed genetic variance (Figure 3f) than the PH data (Figure 3d). This may be due to genotyping error at higher ploidy levels caused by read stochasticity and a flat genotype likelihood surface (Blischak et al., 2016) or caused by incorrect ploidy inferences. A similar mixture model approach to the one implemented here (Weiß et al., 2018) has been shown to have good performance based on comparisons of allele balance distributions from target-enrichment data and flow cytometry-based ploidy estimates (Viruel et al., 2019). We are



**FIGURE 4** Distributions of population genetic summary statistics for different locus regions across data types. Box-and-whisker plots show distribution medians and quartiles for the expected pairwise nucleotide divergence ( $\pi$ ), the number of segregating sites ( $S$ ), and a normalized measure of the difference between two estimators of nucleotide divergence, indicating deviations from neutral expectations (Tajima's  $D$ ). Plots show distributions of summary statistics for the whole locus, left flanking intron, exon core, or right flanking intron regions, grouped by the one haplotype (OH), ambiguous genotype (AG), or phased haplotype (PH) data types.

**TABLE 1** Differences in the distribution of Tajima's  $D$  with respect to ploidy and region along target-enrichment loci.  $\chi^2$  tests that categorized Tajima's  $D$  statistics by their deviation from neutral expectations and significance based on simulations were applied across data types and locus regions. When testing for over- or under-representation of Tajima's  $D$  among locus regions, the data type effects were fixed by analyzing the one haplotype (OH), ambiguous genotype (AG), or phase haplotypes (PH) data alone. When testing for differences among the data types, the gene regions were fixed by testing the left flanking intron, right flanking intron, or exon core alone.

Tested effect	Fixed effect	$\chi^2$	$p$ value	Largest residual
Locus region	OH	7.72	.27	Under-representation of Tajima's $D < -2$ in core regions
	AG	37.77	.0001	Over-representation of Tajima's $D < -2$ in core regions
	PH	83.47	.0001	Over-representation of Tajima's $D < -2$ in core regions
Data type	Left flank	12.65	.046	Over-representation of Tajima's $D < -2$ in OH dataset
	Right flank	8.61	.19	Over-representation of Tajima's $D < -2$ in OH dataset
	Core	57.15	.0001	Over-representation of Tajima's $D < -2$ in OH dataset

cautiously optimistic about the prospects of direct ploidy estimation from sequence data, because likelihood-based criteria can be prone to overfitting (Tiley et al., 2018). However, our ploidy estimates are generally consistent with the previous inferences from microsatellite allele counts (Hagl et al., 2021; Tables S1 and S3). A positive insight from our comparisons across ploidy levels is that if unknown individuals are genotyped as diploid, there will be sufficient information about population structure and that ploidy estimation should not be a barrier to investigations of natural plant populations with little a priori cytological or genetic knowledge.

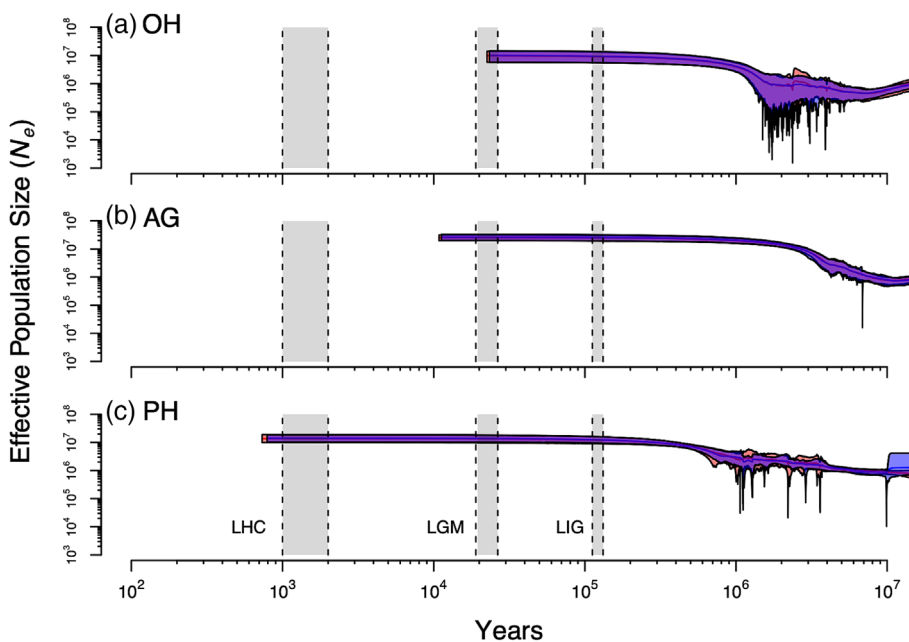
## 4.2 | Limits to differentiating anthropogenic and paleoclimate $N_e$ change

Large population genomic studies in plants that can leverage demographic estimates from SFS (e.g., Blischak et al., 2023; Excoffier et al., 2013; Liu & Fu, 2020) are commonly focused on temperate regions (e.g., Yamaura et al., 2019) and species exploited for energy or agriculture (e.g., Huang et al., 2014; Wang et al., 2017). Thus, it is tempting to leverage models that use gene trees as data such as the EBSP (Heled & Drummond, 2008) or other extensions of skyline plots (Pybus et al., 2000), as they could be paired with relatively low-cost target-enrichment strategies that yield multi-locus nuclear data while stretching material quality.

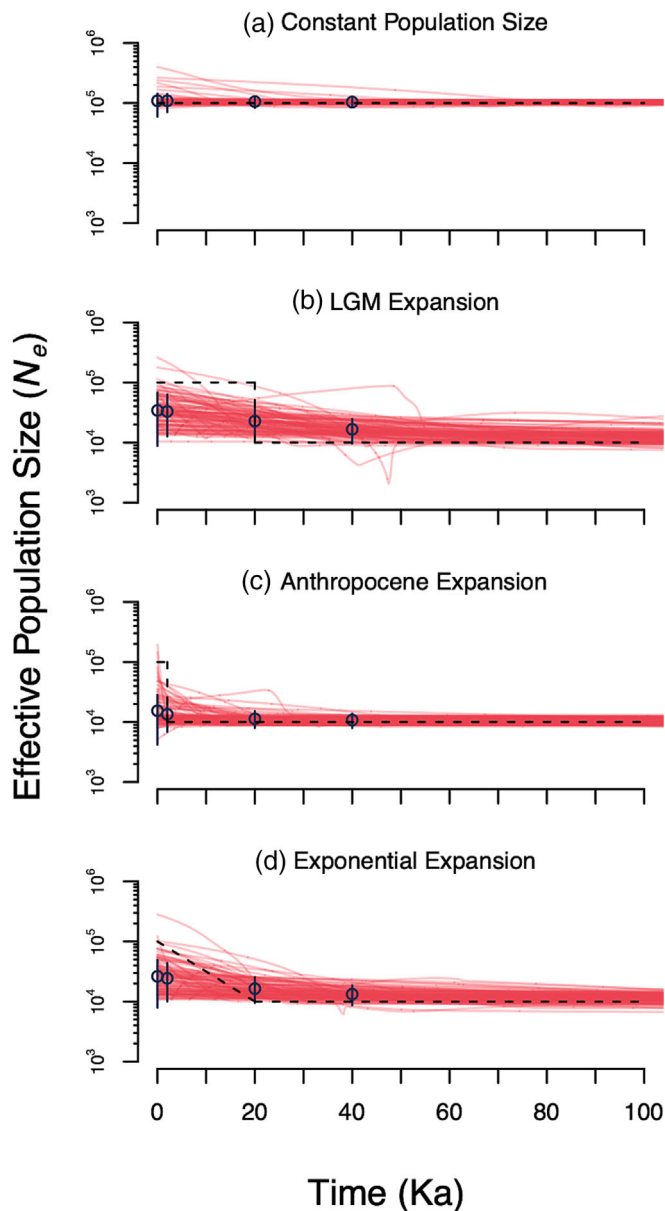
While our analyses provide valuable insights into the ancient origins of *L. simplex* dominated grasslands, both our empirical and simulated analyses show we cannot differentiate between  $N_e$  increases that occurred during human colonization or paleoclimate change near the LGM. The change in vegetative composition towards more  $C_4$  grass species in Central Madagascar is well-documented (Burney, 1987; Crowley et al., 2021; Gasse & Van Campo, 1998; Virah-Sawmy et al., 2010). It is likely that these increases in  $C_4$  grasses

included *L. simplex*; although, an increase in census size may not be reflected by  $N_e$  (Figures 5 and S7), the expected population size under a neutral coalescent model. Our simulations showed that detecting these recent increases is difficult and that we cannot differentiate between a 10-fold increase that occurred 200 (2 ka) or 2000 generations ago (20 ka; Figure 6). Diagnosing a true  $N_e$  change becomes impossible in the presence of model overfitting (Figure S8), and diagnosing these problems in empirical data can be difficult aside from convergence assessments.

These limitations do not necessarily disagree with an assessment of EBSP models from a larger genomic dataset of endemic palm species in Madagascar (Helmstetter et al., 2021). Helmstetter et al. (2021) found that the estimated  $N_e$  trends were valuable predictors of IUCN risk assessments and that population decreases could be detected at anthropogenic time scales. However, it should be easier to detect population declines than population increases in a system that already had a large degree of standing genetic variation, such as *L. simplex*. This is because the time until the most recent common ancestor of a genealogy will be much more recent in a small population than a large one (e.g., see eq. 2.13 from Gillespie, 2004). Analyses of demographic models with genomic SFS should be able to increase resolution at recent anthropogenic time scales (Patton et al., 2019) and differentiate anthropogenic from paleoclimatic demographic effects (Tiley et al., 2022), but this would require more extensive sampling and sequencing. Additionally, our analyses treat all individuals as a single population, and because we detected some level of population structure, we cannot rule out their confounding effects on analyses of  $N_e$  through time (Heller et al., 2013; Mazet et al., 2016) or other confounding effects such as clonal growth and overlapping generations. Population-level sampling of multiple sites throughout a species range should be used to control for population structure, while accounting for localized effects such as agriculture and grazing versus regional climate effects.



**FIGURE 5** Extended Bayesian skyline plot results of effective population size change through time with fixed clock rates for *Loudetia simplex* from Madagascar. Two independent runs (one in red and the other in blue) are shown with mean population sizes shown by solid lines. Polygons show the 95% highest posterior density (HPD) intervals with the overlap between runs in purple. Results are shown for the (a) one haplotype (OH), (b) ambiguous genotype (AG), and (c) phased haplotype (PH) data. Relevant time periods are shown in gray, which includes likely human colonization (LHC), the last glacial maximum (LGM), and last interglacial period (LIG).



**FIGURE 6** Simulation results from extended Bayesian skyline plot analyses for an effective population size increase given an appropriate number of change points. Individual red lines are the mean effective population size ( $N_e$ ) estimates from 100 simulated datasets. Blue points and vertical lines are the mean of mean  $N_e$  estimates and their 95% highest posterior density (HPD) intervals. The dashed black line is the true demographic history of the population.

### 4.3 | New insights into *L. simplex* evolution and policy implications

Our analyses yielded new findings about the evolution of *L. simplex* that are potentially consequential for the broader understanding of Malagasy grasslands. First, our species trees (Figure S1) indicate that individuals from what we call Central Madagascar—a loose term meant only to indicate individuals from the center of the sampled *L. simplex* distribution—represent the ancestral distribution. It is likely

that *L. simplex* dominated grasslands have origins in some of the higher elevation regions of the Central Highlands. For example, MSV1564 comes from the Itremo protected area (Figure 1b; Table S1), and further investigations of this region should be productive for characterizing natural grasslands in Madagascar. The hypothesis of Central Highland origins, which we consider 800 m in altitude and above, is also supported by our analyses of population structure (Figures 2a and S2) and the weak but significant pattern of isolation-by-distance (Figures 2b and S3).

Our models of  $N_e$  change through time suggest that *L. simplex* grasslands were already extensive (Figures 5 and S7) well before any hypothesized human presence in Madagascar (c. 10 ka; Hansford et al., 2018) let alone human colonization (c. 2 ka; Dewar & Wright, 1993; Dewar et al., 2013; Pierron et al., 2017). We found an approximate 10-fold increase between one MA and the LIG, which is beyond the limitations of available paleoecological records (Burney, 1987; Crowley et al., 2021; Gasse & Van Campo, 1998; Railsback et al., 2020; Virah-Sawmy et al., 2010). Strong assumptions about the mutation rate and generation time to calibrate analyses to absolute time and  $N_e$  were necessary, but our interpretations of results are within the biologically plausible uncertainty of these parameters. Even if the per-year substitution rate was overestimated by a factor of 10, the  $N_e$  increase would still be constrained to the Late Pleistocene, occurring between the LIG and LGM. These analyses, especially in the context of a previous investigation of population structure (Hagl et al., 2021), suggest that *L. simplex* dominated grasslands have pre-human origins and likely experienced some  $N_e$  increase coincident with Pleistocene climate change. Tree planting programs have been deployed to combat the devastating habitat loss and deforestation in Madagascar since the 1950s (Green & Sussman, 1990; Harper et al., 2007; Vieilledent et al., 2018; Schüßler et al., 2018), with country-wide forest loss from 1953 to 2014 at 44% (Vieilledent et al., 2018). The results of our study strongly imply that tree planting efforts should be cautious of treating *L. simplex* dominated grasslands as degraded forests, and we suggest that focusing resources on forest edges with native species will be more productive (Miandrimanana et al., 2019).

### 4.4 | Potential use cases of target-enrichment data for conservation genetics

One application of target-enrichment data going forward could be analyses of population structure and population assignment of unknown individuals. For example, many conservation programs performing long-term biodiversity monitoring rely on reduced-SNP panels to determine provenance of an individual or sample (e.g., Bourgeois et al., 2018; Stronen et al., 2022). Developing SNP panels requires at least some a priori genomic resources, so target-enrichment libraries could provide feasible approaches in the absence of such references and practical improvements over existing barcoding strategies for identifying, for example, Madagascar's rosewoods (Hassold et al., 2016); although, acquiring viable DNA samples from

woody tissue would remain a barrier (Jiao et al., 2018). We anticipate limited use cases of target-enrichment data in population genetics going forward as reference genome assembly is becoming more accessible through decreasing long-read costs, and less complicated library preparation protocols for whole-genome resequencing will become preferential to any target-enrichment strategy. However, the population genetic inferences from target-enrichment data should be biologically meaningful when available.

#### 4.5 | A practical need for plant population genomics in Madagascar

There are few studies of plant population genetics from the Central Highlands, and those are limited to AFLP (Gardiner et al., 2017) or microsatellite data (Hagl et al., 2021; Salmons et al., 2020). Demographic analyses from the endemic olive tree *Noronhia lowryi* suggested recent population declines of the woody species, but it is unclear if these declines represent anthropogenic or paleoclimatic change near the LGM (Salmons et al., 2020). While our analyses of *L. simplex* are also ambiguous regarding this window of anthropogenic effects on the extent of grasslands, we can at least conclude that the *L. simplex*  $N_e$  was quite large. Interpreting what a large  $N_e$  means in the context of the *L. simplex* species distribution is difficult though; it could be a single large distribution or multiple smaller isolated distributions that yield the expected measures of genetic variation. Contemporary population genomic analyses will be necessary to provide statistical power to differentiate between competing demographic hypotheses, not only for grasses such as *L. simplex* but also species representing plant diversity across families and life history. A better reconstruction of the natural history of Madagascar's grasslands is not only necessary for understanding diversity within and outside of the Central Highlands (Burbrink et al., 2019; Everson et al., 2020; Yoder et al., 2016) but also to make effective land management decisions for tree planting programs.

#### AUTHOR CONTRIBUTIONS

George P. Tiley, Anne D. Yoder, and Maria S. Vorontsova conceived the study. W. R. Quentin Luke, Cédrique L. Solofondranohatra, Caroline E. R. Lehmann, Guillaume Besnard, and Maria S. Vorontsova collected samples. George P. Tiley and Tchana O. M. Almary curated data. George P. Tiley and Andrew A. Crowl analyzed data. George P. Tiley wrote the first draft. Tchana O. M. Almary translated the abstract to French and Malagasy. All authors revised the manuscript and approved the final version.

#### ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101026923, awarded to GPT and MSV. GB is supported by LabEx TULIP (ANR-10-LABX-0041). CERL was supported by a The Royal Society International Collaboration Award. Thank you to MEDD Madagascar, Madagascar National Parks, Isalo

MNP, and Itremo NPA for supporting our research and exportation permits in 2014 and 2017 and also to COSTECH Tanzania for their support for the 2014 collection permits for MSV and WRQL. Sequence data were generated with research funds from Duke University to ADY. This work would not have been possible without Stuart Cable and everyone at the Kew Madagascar Conservation Centre. We thank Emily Lockwood for DNA extractions and Lorena Endara and J. Gordon Burleigh for coordinating library preparation and sequencing. Colleen Seymour provided helpful feedback on an earlier version of this manuscript. We thank two anonymous reviewers and the editor Andrew Young for helpful comments that improved the manuscript.

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

#### DATA AVAILABILITY STATEMENT

Scripts for analyses and input files are available through the Dryad Digital Repository: <https://doi.org/10.5061/dryad.905qfttqx>. Additional scripts are available on GPT's GitHub for identifying exon-intron boundaries from target enrichment data (<https://github.com/gtiley/Locus-Splitter>) generating matrices for PCAs from fasta files (<https://github.com/gtiley/fasta-conversion>) and the ploidy test implemented through PATÉ (<https://github.com/gtiley/Phasing>). Raw sequence data are available via the NCBI short read archive (SRA) database, with individual SRA and BioSample identifiers in Table S1. Raw sequence data are associated with NCBI BioProject PRJNA952819.

#### ORCID

George P. Tiley  <https://orcid.org/0000-0003-0053-0207>

Andrew A. Crowl  <https://orcid.org/0000-0002-1745-0687>

Tchana O. M. Almary  <https://orcid.org/0009-0007-0723-1919>


W. R. Quentin Luke  <https://orcid.org/0000-0002-7256-031X>

Cédrique L. Solofondranohatra  <https://orcid.org/0000-0003-1808-4954>

Guillaume Besnard  <https://orcid.org/0000-0003-2275-6012>

Caroline E. R. Lehmann  <https://orcid.org/0000-0002-6825-124X>

Anne D. Yoder  <https://orcid.org/0000-0002-1781-9552>

Maria S. Vorontsova  <https://orcid.org/0000-0003-0899-1120>

#### REFERENCES

- Arnold, B., Bomblies, K., & Wakeley, J. (2012). Extending coalescent theory to autotetraploids. *Genetics*, 192(1), 195–204. <https://doi.org/10.1534/genetics.112.140582>
- Baker, W. J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L. R., Brewer, G., Carruthers, T., Clarkson, J. J., Cook, J., Cowan, R. S., Dodsworth, S., Epitawalage, N., François, E., Gallego, B., Johnson, M. G., Kim, J. T., Leempoel, K., ... Forest, F. (2022). A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology*, 71(2), 301–319. <https://doi.org/10.1093/sysbio/syab035>
- Blischak, P. D., Kubatko, L. S., & Wolfe, A. D. (2016). Accounting for genotype uncertainty in the estimation of allele frequencies in autopolyploids. *Molecular Ecology Resources*, 16(3), 742–754. <https://doi.org/10.1111/1755-0998.12493>



- Blischak, P. D., Sajjan, M., Barker, M. S., & Gutenkunst, R. N. (2023). Demographic history inference and the polyploid continuum. *Genetics*, 224, iyad107. <https://doi.org/10.1093/genetics/iyad107>
- Bond, W. J., Silander, J. A. Jr., Ranaivonasy, J., & Ratsirarson, J. (2008). The antiquity of Madagascar's grasslands and the rise of C<sub>4</sub> grassy biomes. *Journal of Biogeography*, 35(10), 1743–1758. <https://doi.org/10.1111/j.1365-2699.2008.01923.x>
- Bond, W. J., Silander, J. A., & Ratsirarson, J. (2023). Madagascar's grassy biomes are ancient and there is much to learn about their ecology and evolution. *Journal of Biogeography*, 50(3), 614–621. <https://doi.org/10.1111/jbi.14494>
- Bourgeois, S., Senn, H., Kaden, J., Taggart, J. B., Ogden, R., Jeffery, K. J., Bunnefeld, N., Abernethy, K., & McEwing, R. (2018). Single-nucleotide polymorphism discovery and panel characterization in the African forest elephant. *Ecology and Evolution*, 8(4), 2207–2217. <https://doi.org/10.1002/ece3.3854>
- Breinholt, J. W., Carey, S. B., Tiley, G. P., Davis, E. C., Endara, L., McDaniel, S. F., Neves, L. G., Sessa, E. B., von Konrat, M., Chantanaorrapint, S., Fawcett, S., Ickert-Bond, S. M., Labiak, P. H., Larrain, J., Lehnert, M., Lewis, L. R., Nagalingum, N. S., Patel, N., Rensing, S. A., ... Burleigh, J. G. (2021). A target enrichment probe set for resolving the flagellate land plant tree of life. *Applications in Plant Sciences*, 9(1), e11406. <https://doi.org/10.1002/aps3.11406>
- Burbrink, F. T., Ruane, S., Kuhn, A., Rabibisoa, N., Randriamahatantsoa, B., Raselimanana, A. P., Andrianarimalala, M. S. M., Cadle, J. E., Lemmon, A. R., Lemmon, E. M., Nussbaum, R. A., Jones, L. N., Pearson, R., & Raxworthy, C. J. (2019). The origins and diversification of the exceptionally rich gemsnakes (Colubroidea: Lamprophiidae: Pseudoxyrhophiinae) in Madagascar. *Systematic Biology*, 68(6), 918–936. <https://doi.org/10.1093/sysbio/syz026>
- Burney, D. A. (1987). Late Holocene vegetational change in Central Madagascar. *Quaternary Research*, 28(1), 130–143. [https://doi.org/10.1016/0033-5894\(87\)90038-X](https://doi.org/10.1016/0033-5894(87)90038-X)
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Christin, P.-A., Spriggs, E., Osborne, C. P., Strömberg, C. A., Salamin, N., & Edwards, E. J. (2014). Molecular dating, evolutionary rates, and the age of the grasses. *Systematic Biology*, 63(2), 153–165. <https://doi.org/10.1093/sysbio/syt072>
- Crowl, A. A., Fritsch, P. W., Tiley, G. P., Lynch, N. P., Ranney, T. G., Ashrafi, H., & Manos, P. S. (2022). A first complete phylogenomic hypothesis for diploid blueberries (*Vaccinium* section *Cyanococcus*). *American Journal of Botany*, 109(10), 1596–1606. <https://doi.org/10.1002/ajb2.16065>
- Crowley, B. E., Godfrey, L. R., Hansford, J. P., & Samonds, K. E. (2021). Seeing the forest for the trees—And the grasses: Revisiting the evidence for grazer-maintained grasslands in Madagascar's central highlands. *Proceedings of the Royal Society B: Biological Sciences*, 288(1950), 20201785. <https://doi.org/10.1098/rspb.2020.1785>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. <https://doi.org/10.1038/ng.806>
- Dewar, R. E., Radmilahy, C., Wright, H. T., Jacobs, Z., Kelly, G. O., & Berna, F. (2013). Stone tools and foraging in northern Madagascar challenge Holocene extinction models. *Proceedings of the National Academy of Sciences*, 110(31), 12583–12588. <https://doi.org/10.1073/pnas.1306100110>
- Dewar, R. E., & Wright, H. T. (1993). The culture history of Madagascar. *Journal of World Prehistory*, 7, 417–466. <https://doi.org/10.1007/BF00997802>
- Dixon, A. P., Faber-Langerdoen, D., Josse, C., Morrison, J., & Loucks, C. J. (2014). Distribution mapping of world grassland types. *Journal of Biogeography*, 41(11), 2003–2019. <https://doi.org/10.1111/jbi.12381>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19(1), 11–15.
- Efron, B., & Narasimhan, B. (2020). The automatic construction of bootstrap confidence intervals. *Journal of Computational and Graphical Statistics*, 29(3), 608–619. <https://doi.org/10.1080/10618600.2020.1714633>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/J.1365-294X.2005.02553.X>
- Everson, K. M., Jansa, S. A., Goodman, S. M., & Olson, L. E. (2020). Montane regions shape patterns of diversification in small mammals and reptiles from Madagascar's moist evergreen forest. *Journal of Biogeography*, 47(10), 2059–2072. <https://doi.org/10.1111/jbi.13945>
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9(10), e1003905. <https://doi.org/10.1371/journal.pgen.1003905>
- Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2020). A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, 37(4), 1211–1223. <https://doi.org/10.1093/molbev/msz296>
- Gardiner, L. M., Rakotoarivivo, M., Rajaovelona, L. R., & Clubbe, C. (2017). Population genetics data help to guide the conservation of palm species with small population sizes and fragmented habitats in Madagascar. *PeerJ*, 5, e3248. <https://doi.org/10.7717/peerj.3248>
- Gasse, F., & Van Campo, E. (1998). A 40,000-yr pollen and diatom record from Lake Tritrivakely, Madagascar, in the southern tropics. *Quaternary Research*, 49(3), 299–311. <https://doi.org/10.1006/qres.1998.1967>
- Gillespie, J. H. (2004). The coalescent. In *Population genetics: A concise guide* (pp. 40–47). John Hopkins University Press. <https://doi.org/10.5602/9780801880087>
- Green, G. M., & Sussman, R. W. (1990). Deforestation history of the eastern rain forests of Madagascar from satellite images. *Science*, 248(4952), 212–215. <https://doi.org/10.1126/science.248.4952.212>
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Hackel, J., Vorontsova, M. S., Nanjarisoa, O. P., Hall, R. C., Razanatsoa, J., Malakasi, P., & Besnard, G. (2018). Grass diversification in Madagascar: In situ radiation of two large C<sub>3</sub> shade clades and support for a Miocene to Pliocene origin of C<sub>4</sub> grassy biomes. *Journal of Biogeography*, 45(4), 750–761. <https://doi.org/10.1111/jbi.13147>
- Hagl, P. A., Gargiulo, R., Fay, M. F., Solofondranohatra, C., Salmons, J., Suescun, U., Rakotomalala, N., Lehmann, C. E. R., Besnard, G., Papadopoulos, A. S. T., & Vorontsova, M. S. (2021). Geographical structure of genetic diversity in *Loudetia simplex* (Poaceae) in Madagascar and South Africa. *Botanical Journal of the Linnean Society*, 196(1), 81–99. <https://doi.org/10.1093/botlinnean/boaa098>
- Hansford, J., Wright, P. C., Rasoamiaramana, A., Pérez, V. R., Godfrey, L. R., Erickson, D., ... Turvey, S. T. (2018). Early Holocene human presence in Madagascar evidenced by exploitation of avian megafauna. *Science Advances*, 4(9), eaat6925. <https://doi.org/10.1126/sciadv.aat6925>
- Harper, G. J., Steininger, M. K., Tucker, C. J., Juhn, D., & Hawkins, F. (2007). Fifty years of deforestation and forest fragmentation in Madagascar. *Environmental Conservation*, 34(4), 325–333. <https://doi.org/10.1017/S0376892907004262>
- Hassold, S., Lowry, P. P., Bauert, M. R., Razafintsalama, A., Ramamonjisoa, L., & Widmer, A. (2016). DNA barcoding of Malagasy Rosewoods: Towards a molecular identification of CITES-listed

- dalbergia species. *PLOS ONE*, 11(6), e0157881. <https://doi.org/10.1371/journal.pone.0157881>
- Heled, J., & Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8(1), 289. <https://doi.org/10.1186/1471-2148-8-289>
- Heller, R., Chikhi, L., & Siegmund, H. R. (2013). The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE*, 8(5), e62992. <https://doi.org/10.1371/journal.pone.0062992>
- Helmstetter, A. J., Cable, S., Rakotonasolo, F., Rabarijaona, R., Rakotoarinivo, M., Eiserhardt, W. L., Baker, W. J., & Papadopoulos, A. S. (2021). The demographic history of Madagascar micro-endemics: Have rare species always been rare? *Proceedings of the Royal Society B: Biological Sciences*, 288(1959), 20210957. <https://doi.org/10.1098/rspb.2021.0957>
- Huang, P., Feldman, M., Schroder, S., Bahri, B. A., Diao, X., Zhi, H., Estep, M., Baxter, I., Devos, K. M., & Kellogg, E. A. (2014). Population genetics of *Setaria viridis*, a new model system. *Molecular Ecology*, 23(20), 4912–4925. <https://doi.org/10.1111/mec.12907>
- Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, 18(2), 337–338. <https://doi.org/10.1093/bioinformatics/18.2.337>
- Jiao, L., Yu, M., Wiedenhoef, A. C., He, T., Li, J., Liu, B., Jiang, X., & Yin, Y. (2018). DNA barcode authentication and library development for the wood of six commercial *Pterocarpus* species: The critical role of Xylarium specimens. *Scientific Reports*, 8(1), 1945. <https://doi.org/10.1038/s41598-018-20381-6>
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, 4(7), 1600016. <https://doi.org/10.3732/apps.1600016>
- Jombart, T., & Ahmed, I. (2011). ADEGENET 1.3–1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Joseph, G. S., Rakotoarivelo, A. R., & Seymour, C. L. (2021). How expansive were Malagasy Central Highland forests, ericoids, woodlands and grasslands? A multidisciplinary approach to a conservation conundrum. *Biological Conservation*, 261, 109282. <https://doi.org/10.1016/j.biocon.2021.109282>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Koechlin, J. (1993). Grasslands of Madagascar. In R. T. Coupland (Ed.), *Natural grasslands: Eastern hemisphere and résumé* (pp. 291–301). Elsevier.
- Leblois, R., Estoup, A., & Rousset, F. (2003). Influence of mutational and sampling factors on the estimation of demographic parameters in a “continuous” population under isolation by distance. *Molecular Biology and Evolution*, 20(4), 491–502. <https://doi.org/10.1093/molbev/msg034>
- Lehmann, C. E. R., Solofondranohatra, C., Morton, J. A., Phelps, L. N., Ralimanana, H., Razanatoa, J., Rakotoarimanana, V., & Vorontsova, M. S. (2022). The grassy ecosystems of Madagascar. In S. M. Goodman (Ed.), *The new natural history of Madagascar* (pp. 152–168). Princeton University Press. <https://doi.org/10.2307/j.ctv2ks6tbb.27>
- Liu, X., & Fu, Y. X. (2020). Stairway plot 2: Demographic history inference with folded SNP frequency spectra. *Genome Biology*, 21(1), 280. <https://doi.org/10.1186/s13059-020-02196-9>
- Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution—Lessons for ancestral population size inference? *Heredity*, 116(4), 362–371. <https://doi.org/10.1038/hdy.2015.104>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Meirman, P. G., & Van Tienderen, P. H. (2013). The effects of inheritance in tetraploids on genetic diversity and population divergence. *Heredity*, 110(2), 131–137. <https://doi.org/10.1038/hdy.2012.80>
- Miandrimanana, C., Reid, J. L., Rivoaharison, T., & Birkinshaw, C. (2019). Planting position and shade enhance native seedling performance in forest restoration for an endangered Malagasy plant. *Plant Diversity*, 41(2), 118–123. <https://doi.org/10.1016/j.pld.2018.09.005>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Nielsen, R. (2000). Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*, 154(2), 931–942. <https://doi.org/10.1093/genetics/154.2.931>
- Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Patton, A. H., Margres, M. J., Stahlke, A. R., Hendricks, S., Lewallen, K., Hamede, R. K., ... Storfer, A. (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian devils. *Molecular Biology and Evolution*, 36(12), 2906–2921. <https://doi.org/10.1093/molbev/msz191>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- Pierron, D., Heiske, M., Razafindrazaka, H., Rakoto, I., Rabetokotany, N., Ravololomanga, B., Rakotozafy, L. M., Rakotomalala, M. M., Razafiarivony, M., Rasoarifetra, B., Raharijy, M. A., Razafindralambo, L., Ramilisonina, F. F., Lejambre, S., Thomas, O., Mohamed Abdallah, A., Rocher, C., Arachiche, A., Tonaso, L., ... Letellier, T. (2017). Genomic landscape of human diversity across Madagascar. *Proceedings of the National Academy of Sciences*, 114(32), E6498–E6506. <https://doi.org/10.1073/pnas.1704906114>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Pybus, O. G., Rambaut, A., & Harvey, P. H. (2000). An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155(3), 1429–1437. <https://doi.org/10.1093/genetics/155.3.1429>
- Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., & Chikhi, L. (2012). Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. *Proceedings of the National Academy of Sciences*, 109(32), 13028–13033. <https://doi.org/10.1073/pnas.1200153109>
- Railsback, L. B., Dupont, L. A., Liang, F., Brook, G. A., Burney, D. A., Cheng, H., & Edwards, R. L. (2020). Relationships between climate change, human environmental impact, and megafaunal extinction inferred from a 4000-year multi-proxy record from a stalagmite from northwestern Madagascar. *Quaternary Science Reviews*, 234, 106244. <https://doi.org/10.1016/j.quascirev.2020.106244>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. <https://doi.org/10.1093/sysbio/syy032>
- Rambaut, A., & Grass, N. C. (1997). Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees.

- Bioinformatics*, 13(3), 235–238. <https://doi.org/10.1093/bioinformatics/13.3.235>
- Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, 164(4), 1645–1656. <https://doi.org/10.1093/genetics/164.4.1645>
- Ronfort, J., Jenczewski, E., Bataillon, T., & Rousset, F. (1998). Analysis of population structure in autotetraploid species. *Genetics*, 150(2), 921–930. <https://doi.org/10.1093/genetics/150.2.921>
- Salmona, J., Heller, R., Quéméré, E., & Chikhi, L. (2017). Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. *Molecular Ecology*, 26(19), 5203–5222. <https://doi.org/10.1111/mec.14173>
- Salmona, J., Olofsson, J. K., Hong-Wa, C., Razanatsoa, J., Rakotonasolo, F., Ralimanana, H., Randriamboavonjy, T., Suescun, U., Vorontsova, M. S., & Besnard, G. (2020). Late Miocene origin and recent population collapse of the Malagasy savanna olive tree (*Noronhia lowryi*). *Biological Journal of the Linnean Society*, 129(1), 227–243. <https://doi.org/10.1093/biolinnean/blz164>
- Schüßler, D., Radespiel, U., Ratsimbazafy, J. H., & Mantilla-Contreras, J. (2018). Lemurs in a dying forest: Factors influencing lemur diversity and distribution in forest remnants of north-eastern Madagascar. *Biological Conservation*, 228, 17–26. <https://doi.org/10.1016/j.biocon.2018.10.008>
- Slimp, M., Williams, L. D., Hale, H., & Johnson, M. G. (2021). On the potential of Angiosperms353 for population genomic studies. *Applications in Plant Sciences*, 9(7), e11419. <https://doi.org/10.1002/aps.3.11419>
- Solofondranohatra, C. L., Vorontsova, M. S., Hackel, J., Besnard, G., Cable, S., Williams, J., Jeannoda, V., & Lehmann, C. E. (2018). Grass functional traits differentiate forest and savanna in the Madagascar central highlands. *Frontiers in Ecology and Evolution*, 6, 184. <https://doi.org/10.3389/fevo.2018.00184>
- Solofondranohatra, C. L., Vorontsova, M. S., Hempson, G. P., Hackel, J., Cable, S., Vololoniaina, J., & Lehmann, C. E. (2020). Fire and grazing determined grasslands of central Madagascar represent ancient assemblages. *Proceedings of the Royal Society B: Biological Sciences*, 287(1927), 20200598. <https://doi.org/10.1098/rspb.2020.0598>
- Stronen, A. V., Mattucci, F., Fabbri, E., Galaverni, M., Cocchiara, B., Nowak, C., Godinho, R., Ruiz-González, A., Kusak, J., Skrbinišek, T., Randi, E., Vlasheva, A., Mucci, N., & Caniglia, R. (2022). A reduced SNP panel to trace gene flow across southern European wolf populations and detect hybridization with other *Canis* taxa. *Scientific Reports*, 12(1), 4195. <https://doi.org/10.1038/s41598-022-08132-0>
- Teixeira, H., Montade, V., Salmona, J., Metzger, J., Bremond, L., Kasper, T., Daut, G., Rouland, S., & Radespiel, U. (2021). Past environmental changes affected lemur population dynamics prior to human impact in Madagascar. *Communications Biology*, 4(1), 1084. <https://doi.org/10.1038/s42003-021-02620-1>
- Tiley, G. P., Barker, M. S., & Burleigh, J. G. (2018). Assessing the performance of Ks plots for detecting ancient whole genome duplications. *Genome Biology and Evolution*, 10(11), 2882–2898. <https://doi.org/10.1093/gbe/evy200>
- Tiley, G. P., Crowl, A. A., Manos, P. S., Sessa, E. B., Solís-Lemus, C., Yoder, A. D., & Burleigh, J. G. (2021). Phasing alleles improves network inference with allopolyploids. *bioRxiv*. <https://doi.org/10.1101/2021.05.04.442457>
- Tiley, G. P., van Elst, T., Teixeira, H., Schüßler, D., Salmona, J., Blanco, M. B., Ralison, J. M., Randrianambinina, B., Rasoloarison, R. M., Stahlke, A. R., Hohenlohe, P. A., Chikhi, L., Louis, E. E., Radespiel, U., & Yoder, A. D. (2022). Population genomic structure in Goodman's mouse lemur reveals long-standing separation of Madagascar's central highlands and eastern rainforests. *Molecular Ecology*, 31(19), 4901–4918. <https://doi.org/10.1111/mec.16632>
- Vieilledent, G., Grinand, C., Rakotomalala, F. A., Ranaivosoa, R., Rakotoarijaona, J. R., Allnut, T. F., & Achard, F. (2018). Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation*, 222, 189–197. <https://doi.org/10.1016/j.biocon.2018.04.008>
- Virah-Sawmy, M., Willis, K. J., & Gillson, L. (2010). Evidence for drought and forest declines during the recent megafaunal extinctions in Madagascar. *Journal of Biogeography*, 37(3), 506–519. <https://doi.org/10.1111/j.1365-2699.2009.02203.x>
- Viruel, J., Conejero, M., Hidalgo, O., Pokorny, L., Powell, R. F., Forest, F., Kantar, M. B., Soto Gomez, M., Graham, S. W., Gravendeel, B., Wilkin, P., & Leitch, I. J. (2019). A target capture-based method to estimate ploidy from herbarium specimens. *Frontiers in Plant Science*, 937, 937. <https://doi.org/10.3389/fpls.2019.00937>
- Vorontsova, M. S., Besnard, G., Forest, F., Malakasi, P., Moat, J., Clayton, W. D., Ficinski, P., Savva, G. M., Nanjarisoa, O. P., Razanatsoa, J., Randriatsara, F. O., Kimeu, J. M., Luke, W. R., Kayombo, C., & Linder, H. P. (2016). Madagascar's grasses and grasslands: Anthropogenic or natural? *Proceedings of the Royal Society B: Biological Sciences*, 283(1823), 20152262. <https://doi.org/10.1098/rspb.2015.2262>
- Wang, L., Beissinger, T. M., Lorient, A., Ross-Ibarra, C., Ross-Ibarra, J., & Hufford, M. B. (2017). The interplay of demography and selection during maize domestication and expansion. *Genome Biology*, 18(1), 215. <https://doi.org/10.1186/s13059-017-1346-4>
- Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., & Burbano, H. A. (2018). nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*, 19(1), 122. <https://doi.org/10.1186/s12859-018-2128-z>
- Xie, M., Wu, Q., Wang, J., & Jiang, T. (2016). H-PoP and H-PoPG: Heuristic partitioning algorithms for single individual haplotyping of polyploids. *Bioinformatics*, 32(24), 3735–3744. <https://doi.org/10.1093/bioinformatics/btw537>
- Xu, B., & Yang, Z. (2016). Challenges in species tree estimation under the multispecies coalescent model. *Genetics*, 204(4), 1353–1368. <https://doi.org/10.1534/genetics.116.190173>
- Yamaura, Y., Narita, A., Kusumoto, Y., Nagano, A. J., Tezuka, A., Okamoto, T., Takahara, H., Nakamura, F., Isagi, Y., & Lindenmayer, D. (2019). Genomic reconstruction of 100 000-year grassland history in a forested country: Population dynamics of specialist forbs. *Biology Letters*, 15(5), 20180577. <https://doi.org/10.1098/rsbl.2018.0577>
- Yoder, A. D., Campbell, C. R., Blanco, M. B., Dos Reis, M., Ganzhorn, J. U., Goodman, S. M., Hunnicutt, K. E., Larsen, P. A., Kappeler, P. M., Rasoloarison, R. M., Ralison, J. M., Swofford, D. L., & Weisrock, D. W. (2016). Geogenetic patterns in mouse lemurs (genus *Microcebus*) reveal the ghosts of Madagascar's forests past. *Proceedings of the National Academy of Sciences*, 113(29), 8049–8056. <https://doi.org/10.1073/pnas.1601081113>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Tiley, G. P., Crowl, A. A., Almary, T. O. M., Luke, W. R. Q., Solofondranohatra, C. L., Besnard, G., Lehmann, C. E. R., Yoder, A. D., & Vorontsova, M. S. (2024). Genetic variation in *Loudetia simplex* supports the presence of ancient grasslands in Madagascar. *Plants, People, Planet*, 6(2), 315–329. <https://doi.org/10.1002/ppp3.10437>