Comparative genomic analysis of sifakas (Propithecus) reveals selection for folivory and high heterozygosity despite endangered status

Elaine E. Guevara 1,2*, Timothy H. Webster 3, Richard R. Lawler 4, Brenda J. Bradley 2, Lydia K. Greene 5,6, Jeaninn Ranaivosany 7, Joelisoa Ratsirarson 7, R. Alan Harris 8, Yue Liu 8, Shwetha Murali 8, Muthuswamy Raveendran 8, Daniel S. T. Hughes 9, Donna M. Muzny 8, Anne D. Yoder 10, Kim C. Worley 8, Jeffrey Rogers 8,10*

Sifakas (genus Propithecus) are critically endangered, large-bodied diurnal lemurs that eat leaf-based diets and show corresponding anatomical and microbial adaptations to folivory. We report on the genome assembly of Coquerel’s sifaka (P. coquereli) and the resequenced genomes of Verreaux’s (P. verreauxi), the golden-crowned (P. tattersalli), and the diademed (P. diadema) sifakas. We find high heterozygosity in all sifakas compared with other primates and endangered mammals. Demographic reconstructions nevertheless suggest declines in effective population size beginning before human arrival on Madagascar. Comparative genomic analyses indicate pervasive accelerated evolution in the ancestral sifaka lineage affecting genes in several complementary pathways relevant to folivory, including nutrient absorption and xenobiotic and fatty acid metabolism. Sifakas show convergent evolution at the level of the pathway, gene family, gene, and amino acid substitution with other folivores. Although sifakas have relatively generalized diets, the physiological challenges of habitual folivory likely led to strong selection.

INTRODUCTION

Sifakas (genus Propithecus) are lemurs of the Indriid family and represent a radiation of nine species that occupy various forested regions across Madagascar (1). Lemurs are a primate clade endemic to Madagascar that exhibit exceptional taxonomic and ecological diversity (1, 2). Lemurs’ lengthy geographic isolation (>50 million years) and their arrival earlier than the handful of other mammalian orders on Madagascar have resulted in a remarkable array of adaptations to diverse niches (2, 3). Madagascar is also characterized by unpredictable and hypervariable rainfall environments, marked seasonality, low primary productivity, and nitrogen-poor fruits, which are likely to have been important selective pressures shaping lemur evolution (4–6). Genomic analyses of lemurs are of interest for understanding the molecular bases for lemur traits; however, to date, genomic studies in primates have tended to focus on haplorrhines (7–9). Moreover, lemurs are among the mammalian groups most vulnerable to extinction, with >90% of species currently under threat (10). With one exception, all species within the genus Propithecus are classified as Critically Endangered (11), making it among the most threatened lemur genera. Hence, conservation efforts are urgently needed and genomic data can shed light on factors like inbreeding, genetic connectivity, historical population dynamics, and species delineation/discovery (12) that will be important in future assessments and planning.

Sifakas are diurnal, gregarious, and relatively large-bodied lemurs. They are considered anatomical folivores, showing phenotypic adaptations to folivory that include aspects of dental morphology, an elongated gastrointestinal (GI) tract, and a succulated cecum and gut microbial community that facilitate hindgut fermentation (13, 14). However, behaviorally, sifakas are less strict folivores compared with the members of the two other indriid genera, Indri and Avahi (15, 16), incorporating a sizeable portion of fruit in the diet when available (16–22) (fig. S1). Nevertheless, leaves comprise a considerable and crucial part of the diet in all populations studied, and anatomical adaptations and the ability to tolerate plant secondary compounds may enable sifakas to exploit foliage of many species as fallback foods (16, 17, 20, 21). Sifaka diets vary considerably by season, habitat, and species, and generally are highly diverse, including greater numbers of different plant species than sympatric diurnal lemurs (18–21).

Here, we report on a whole-genome reference assembly (Pcoq_1.0; GenBank accession GCA_000956105.1) that we generated for Coquerel’s sifaka (Propithecus coquereli), a species with a large but fragmented distribution in the dry, deciduous forests of northwestern Madagascar (24). We also generated whole-genome sequence data for three other sifaka species: Verreaux’s sifaka (Propithecus verreauxi) of the arid deciduous forests of southwestern Madagascar (25), the golden-crowned sifaka (Propithecus tattersalli) from the dry deciduous forests of the northern tip of Madagascar (26), and the diademed sifaka (Propithecus diadema) of the eastern rainforest (27) (Fig. 1). These species share a common ancestor estimated to have lived sometime between 6 and 16 million years ago (28).

Our goals in this analysis are twofold: (i) to use genomic data to estimate key parameters that are relevant to conservation efforts and (ii) to assess patterns of selection on protein-coding genes in this genus. First, to better understand the conservation status and
historical population dynamics of these species, we assessed genomic diversity and reconstructed demographic change in each species. An important outstanding question in Madagascar conservation biology is the degree to which loss of forested habitats is recent and anthropogenic (29) or more ancient and driven by nonanthropogenic climate change (10). Analysis of multilocus microsatellite data of the golden-crowned sifaka, one of the species included here, indicated that population declines far predate human settlement of Madagascar and instead are more likely attributable to habitat change caused by Holocene droughts before human arrival (24). However, many lemurs, including some sifaka species, show evidence of more recent bottlenecks that likely coincide with anthropogenic disturbance (30–34). Different forest habitats and their resident lemurs are likely to have been influenced differently by climate and human activities across the island. Additional genomic analyses from sifaka species inhabiting ecologically distinct regions of Madagascar will expand our knowledge of historical population dynamics in lemurs and help determine whether population size changes occurred well before, during, or well after human settlement of the island.

Second, to gain insights into sifaka physiology, we assessed genome-wide accelerated protein-coding change in the ancestral sifaka lineage. The genomes of other foliage specialists, including other primates (colobine monkeys), giant pandas, red pandas, and koalas (fig. S1), have revealed positive selection and gene family expansion in pathways related to xenobiotic and nutrient metabolism, which could aid in detoxification of plant toxins and maximization of nutritional gain from leaves, respectively (35–37). Compared to other primate clades, folivores are overrepresented among lemurs (2, 6). However, molecular adaptation to folivory has not been extensively studied in lemurs [though see (38)]. The other lemurs represented by genomic data in our comparative analysis were the gray mouse lemur (Microcebus murinus) and the closely related black

Fig. 1. Sifaka samples and demographic history. The current known geographic distributions of the four sifaka species included in this study with sample origins. (Illustrations copyright 2020 Stephen D. Nash/IUCN SSC Primate Specialist Group. Used with permission.)
(Eulemur macaco) and blue-eyed black lemurs (Eulemur flavifrons), all of which are frugivorous/omnivorous (39–41). We also include several colobine species in our analysis and specifically compare our results to findings in this clade. The results presented here expand our knowledge of the genetic basis of mammalian folivore digestive physiology to an ecologically unique group of mammals, sifaka lemurs.

RESULTS
We sequenced and assembled the P. coquereli genome using DNA from a female sifaka that was born in Madagascar (Belalitra, Mahajanga Province) but housed at the Duke Lemur Center (DLC; see Materials and Methods; Fig. 1). We generated 104.7× whole-genome sequence coverage using 100–base pair (bp) paired-end sequence libraries and Illumina Hi-Seq 2000 and 2500 instruments. The assembly process, which used ALLPATHS-LG, Atlas-Link, and AtlasGapFill software (see Materials and Methods), produced a draft assembly (Pcoq_1.0; GenBank accession GCA_000956105.1) of total length 2.80 Gb, with contig N50 of 28.1 kb and scaffold N50 of 5.60 Mb (Table 1).

The annotation of Pcoq_1.0 was performed by the National Center for Biotechnology Information (NCBI) annotation team using their Eukaryotic Annotation Pipeline (www.ncbi.nlm.nih.gov/genome/annotation_euk/process). This draft assembly provides the sequences for 19,484 protein-coding genes, 408 noncoding genes, and 2867 pseudogenes. We assessed the completeness of this annotation by analyzing the BUSCO (benchmarking universal single-copy orthologs) euarchontoglires gene set of 6192 highly conserved protein-coding genes (42) and found that 91.25% of those genes are complete, while another 3.92% are fragmented. We additionally sequenced five other, unrelated, wild-born P. coquereli individuals originating from Ankijaibe, Belalitra, and Doanikey (Mahajanga Province) and housed at the DLC. We also sequenced two wild-born diademed sifakas (P. diadema) from near Andasibe (Toamasina), two wild-born golden-crowned sifakas (P. tattersalli) originating from two sites near Daraina (Antsiranana Province), and all housed at the DLC, as well as a single male Verreaux’s sifaka (P. verreauxi) born to wild-caught parents from Bevala (Toliara Province) at the DLC (Fig. 1). Last, we sequenced an additional 17 wild Verreaux’s sifaka from Bezà Mahafaly Special Reserve (Toliara Province; Fig. 1). All additional genomes were sequenced to an average of >23× coverage (Table 2). The sequencing data demonstrated a high mapping rate (>94%), low polymerase chain reaction (PCR) duplication rate (<15%), and high average base quality (Phred score > 35; Table 2).

All sifakas exhibit high genome-wide heterozygosity (Fig. 2). We estimated genome-wide individual heterozygosity as the number of high confidence heterozygous sites divided by the number of bases in the genome for which variants could be reliably identified based on coverage and mapping quality. Heterozygosity ranged from 0.00569 to 0.00587 for the Coquerel’s sifakas, 0.00314 to 0.00468 for Verreaux’s sifaka, 0.00239 to 0.00249 for golden-crowned sifakas, and 0.00396 to 0.00397 for diademed sifakas (Fig. 2).

We used the pairwise sequentially Markovian coalescent [PSMC; (43)] to infer the inverse instantaneous coalescence rate (IICR) (44), which can reflect changes in effective population size ($N_e$) under panmictia where coalescent rate is directly inversely proportional to $N_e$, and/or changes in population connectivity under a scenario of population structure. We used several different generation time estimates, spanning 9 to 17.5 years, which represent estimates of the average age of female first reproduction (45) and the time required for the population to grow by the net reproductive rate (46), respectively, based on demographic analysis of a population of P. verreauxi studied long term (see section S2). The genomes of all individual sifakas analyzed indicated cycles of increasing and declining IICR over the past million years, with pronounced decreases in the past 100,000 years (Fig. 1). The two western species (P. coquereli and P. verreauxi) show higher peaks and sharper declines in IICR, whereas the two eastern species (P. diadema and P. tattersalli) showed more gradual declines. Different individuals of the same species generally show highly similar demographic reconstructions (Fig. 1); P. verreauxi showed the greatest differences between individuals. Estimates of IICR were highly sensitive to the mutation rate used (fig. S2), but marked declines in IICR, possibly reflecting $N_e$, before human arrival on Madagascar were detected regardless of the assumed mutation rate.

Sequencing and assembly of the P. coquereli genome revealed several colobine species in our analysis and specifically compare our results to findings in this clade. The results presented here expand our knowledge of the genetic basis of mammalian folivore digestive physiology to an ecologically unique group of mammals, sifaka lemurs.

**Table 1. Propithecus coquereli genome assembly description.**

<table>
<thead>
<tr>
<th>GC content</th>
<th>43.2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total size (Mbp)</td>
<td>2,798.15</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>299,070</td>
</tr>
<tr>
<td>Number of scaffolds</td>
<td>22,539</td>
</tr>
<tr>
<td>Contig N50</td>
<td>28,129</td>
</tr>
<tr>
<td>Contig L50</td>
<td>18,523</td>
</tr>
<tr>
<td>Scaffold N50</td>
<td>5.60 Mb</td>
</tr>
<tr>
<td>Scaffold L50</td>
<td>149</td>
</tr>
</tbody>
</table>

We also identified several bitter taste receptors that showed provisional evidence of positive selection, although none of the analyses reached significance after correcting for multiple testing ($q < 0.05$).
TAS2R40 had $d_{NS}/d_S > 1$ (sifaka $d_{NS}/d_S = 1.330$; background $d_{NS}/d_S = 0.494$; $P = 0.022$). Positively selected sites were identified within TAS2R7 (C60V, $P = 0.029$, Bayes empirical Bayes score = 0.816) and TAS2R16 (S70R, $P = 0.007$, Bayes empirical Bayes score = 0.95).

In comparing our data to previous studies, we found overlap between our results and those of the analysis of the snub-nosed monkey (genus *Propithecus*) genomes (35). Snub-nosed monkeys are folivorous colobine monkeys whose ancestors diverged from the ancestors of *Propithecus* approximately 70 million years ago (28). Parallel changes were observed at the levels of the genetic pathway, the gene, and particular amino acid substitution. Loci that show rapid evolution in both sifakas and snub-nosed monkeys included genes involved in both lipid metabolism and xenobiotic metabolism (table S4). One gene, *ACE* (angiotensin-converting enzyme), showed

---

**Table 2. Additional *Propithecus* spp. genome sequence descriptions. SNVs, single nucleotide variants.**

<table>
<thead>
<tr>
<th>Measure</th>
<th>P. diadema</th>
<th>P. diadema</th>
<th>P. tattersalli</th>
<th>P. tattersalli</th>
<th>P. verreaux</th>
<th>P. verreaux</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads</td>
<td>765,447,178</td>
<td>749,239,612</td>
<td>717,945,044</td>
<td>716,893,458</td>
<td>954,251,816</td>
<td>790,776,474</td>
</tr>
<tr>
<td>Mapping rate</td>
<td>94.3%</td>
<td>94.5%</td>
<td>94.6%</td>
<td>95.1%</td>
<td>94.5%</td>
<td>94.6%</td>
</tr>
<tr>
<td>PCR duplication rate</td>
<td>4.0%</td>
<td>3.4%</td>
<td>3.2%</td>
<td>2.5%</td>
<td>15.0%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Average phred quality</td>
<td>36.1</td>
<td>36.1</td>
<td>36.3</td>
<td>36.8</td>
<td>38.9</td>
<td>35.8</td>
</tr>
<tr>
<td>Average depth coverage</td>
<td>23×</td>
<td>23×</td>
<td>23×</td>
<td>23×</td>
<td>39×</td>
<td>33×</td>
</tr>
<tr>
<td>Variants called (after filtering)</td>
<td>40,822,967</td>
<td>40,851,734</td>
<td>22,519,595</td>
<td>22,701,370</td>
<td>26,512,635</td>
<td>24,814,200</td>
</tr>
<tr>
<td>SNVs</td>
<td>37,479,654</td>
<td>37,516,145</td>
<td>20,714,023</td>
<td>20,865,986</td>
<td>24,376,639</td>
<td>22,875,767</td>
</tr>
<tr>
<td>Indels</td>
<td>3,343,313</td>
<td>3,335,589</td>
<td>1,805,572</td>
<td>1,835,384</td>
<td>2,281,800</td>
<td>1,938,433</td>
</tr>
</tbody>
</table>

---

**Fig. 2. Genome-wide heterozygosity compared across primates and other mammals.** IUCN “Red List” status: NE, not evaluated; LC, least concern; NT, near threatened; VU, vulnerable; EN, endangered; CR, critically endangered; EX, extinct. Mammalian families included represented by clip art. Tasmanian devil silhouette by Sarah Werning from PhyloPic.org. Data from (54, 114–117).
the same two convergent amino acid substitutions in both lineages (M109L and I869V).

We also found positive selection on the pancreatic ribonuclease (RNase) gene, RNASE1, which is thought to be important for breaking down bacterial RNA in herbivores (51). In primates, there have been independent duplications of RNASE1 in African and Asian colobine monkeys, including snub-nosed monkeys that, like bovine ruminants, have evolved foregut fermentation (51, 52).

RNASE1 exhibited an elevated \( d_\theta/d_\delta \) (sifaka \( d_\theta/d_\delta = 5.574 \), background \( d_\theta/d_\delta = 0.291 \)) and 11 positively selected sites on the ancestral sifaka branch. Sifakas showed amino acid substitutions at five of the same sites predicted to change the catalytic activity of the duplicated RNASE1B gene found in douc langurs (Pygathrix nemaus) (Fig. 3) (53). All mutations were confirmed with Sanger sequencing (section 56). Site 83, which is one of the two sites predicted to have the greatest influence on the enzymatic activity of RNASE1 based on site-directed mutagenesis experiments (53), shows the same amino acid substitution (D > E) in sifakas and duplicated colobine sequences. Sifakas also show the same substituted amino acid residue (S) at position 39 as two Asian colobine species (Semnopithecus entellus and Trachypithecus francoisi). As in colobines, substitutions in sifakas include many losses of arginines, although sifakas have also gained one arginine. The loss of positively charged arginine residues contributes to the change in enzymatic activity. The loss of arginines might also enhance the stability of the enzyme in the gut as arginines are especially vulnerable to hydrolysis by trypsin, one of the primary digestive enzymes active in the upper GI tract, and to alteration during bacterial fermentation (52).

P. diadema and P. tattersalli additionally show substitutions at position 1, another site predicted to change enzymatic activity in colobines. These two sifaka species show different substitutions: K1N and K1T, respectively.

In total, we found 10 gene families to have undergone rapid expansion and 33 gene families to have undergone rapid contraction on the ancestral sifaka branch (fig. S4). Gene families of ribosomal proteins, oxidases, and transcription factors showed both expansion and contraction. Families showing expansion were involved in response to oxidative damage, apoptosis, and DNA damage. Families of immune receptors, heat shock proteins, olfactory receptors, and proteins involved in translation and protein folding were among those showing contraction.

**DISCUSSION**

Detailed genomic and genetic analysis of sifakas and other strepsirrhines can provide critical insight into their phylogenetic, demographic, and adaptive histories. Our genome assembly (Pcoq_1.0) is comparable to other strepsirrhine assemblies, with global assembly statistics (Table 1) similar to those of the small-eared greater galago (Otolemur garnetti) draft assembly (OtoGar3; Contig N50: 27,100; Contig L50: 21,634; Scaffold N50: 13,852,661; Scaffold L50: 47). Although not a chromosome-level assembly and not as complete as the most recent gray mouse lemur (M. murinus) reference assembly (Mmur_3.0; Contig N50: 210,702; Contig L50: 2987; Scaffold N50: 108,171,978; Scaffold L50: 10), the annotation pipeline resulted in the identification of 19,484 protein-coding genes, 408 noncoding genes, and 2867 pseudogenes. For comparison, Mmur_3.0 contains 20,671 protein-coding and 7142 noncoding genes, and OtoGar3 contains 19,558 protein-coding and 4827 noncoding genes. Thus, our assembly is of sufficient quality to assess protein evolution. This new sifaka reference assembly extends our knowledge of lemur genomics to an ecologically unique branch of strepsirrhine evolution.

Although ecological and population pressures currently face sifaka species, we found that standing genetic variation remains high. Three of the four sifaka species analyzed showed higher genome-wide heterozygosity than any of the other primates for which comparable data are available (Fig. 2). Our findings are consistent with the results of several previous genetic diversity assessments of various sifaka species, including P. coquereli and P. tattersalli, studied here (30, 33, 54–56). However, despite evidence for relatively high genetic diversity in sifakas, several studies have found evidence of population bottlenecks or declines in various species (30, 33, 34). There is some evidence that heterozygosity can temporarily increase following a population bottleneck (57), as may be the case here. Our finding of long-term declines in IICR in all sifaka species would seem to be at odds with the high estimates of genome-wide heterozygosity. A slow and steady population decline need not be accompanied by a reduction in heterozygosity, however, as numerous demographic and genetic factors (mutation rate, dispersal distance, overlapping generations, subpopulation connectivity, and admixture) can blur the relationship between past demographic events and genetic diversity (58). This trend of steady declines in IICR during the late Pleistocene is consistent with some demographic reconstructions of other lemur species using whole-genome sequence data (59–61). These declines have been attributed to a drying climate and resulting loss of prime forest habitat (59, 60, 62). Lake drill cores offer evidence of severe drying in East Africa from ~145,000 to 60,000 years ago (63), a time range overlapping the period when a marked decline in the four species’ effective population sizes seems to have occurred. The demographic history inferred here indicates that these sifaka populations had already declined markedly before the arrival of people in Madagascar, even when considering recent, earlier estimates of human arrivals (64). This result, however, does not rule out more recent declines driven by anthropogenic disturbance and/or Holocene droughts (34, 65), as detected in other studies of sifakas (30, 33, 34) and other lemurs (31, 32). A paucity of recombination events in the last several thousand years precludes resolution at this time scale using the IICR approach (43).

We also note that there are other limitations to PSMC reconstruction. Parameter choice can have a large influence on PSMC results (43). Generation time estimates used in the model have a sizable influence on the estimated timing of demographic events.
while mutation rate influences the inferred magnitude of change (Fig. 1 and fig. S2). Empirical estimates of per generation mutation rate for sifaka species would be desirable given the large difference in estimates of IICR depending on whether the human mutation rate or a faster mutation rate similar to that estimated for mouse lemurs is used (Fig. 1 and fig. S2) (66). Population structure may also confound interpretation of the demographic analyses. The PSMC assumes panmixia, which is often violated in real-world cases. Changes in population structure may produce the same signal in an individual’s genome as would historical change in population size (44), which has led to criticism of simplistic interpretations about species’ past population sizes. Thus, it has been argued that $N_e$ estimates from PSMC analyses are better termed IICR (44). We sought to obtain more robust estimates of IICR by including multiple individuals per species and particularly individuals from more than one geographic locale when possible (Fig. 1). Within a species, we find that individuals’ genomes produce highly similar past IICR trajectories. A greater intraspecific difference is apparent for P. verreauxi, providing possible evidence for population structure (Fig. 1). Nevertheless, we are unable to distinguish between changes in population size, changes in population structure, or the combined effects of both. Future work sequencing more genomes of each species to determine the pattern of rare versus intermediate frequency variants combined with improved paleoecological reconstructions of changes in forest cover will aid in modeling sifaka demographic history.

The maintenance of genetic diversity by these sifaka populations has several possible explanations. It may be that sifakas are relatively resilient to habitat fragmentation and able to disperse across deforested areas. This is consistent with evidence for continued gene flow in crowned sifakas (P. coronatus) (55) and P. verreauxi (67) despite uneven forest cover, as well as patterns of isolation-by-distance in

Table 4. REGs involved in digestion and xenobiotic metabolism.

<table>
<thead>
<tr>
<th>Functional category</th>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digestion</td>
<td>PNLIP</td>
<td>Pancreatic lipase</td>
<td>The main digestive enzyme in triglyceride hydrolysis</td>
</tr>
<tr>
<td>Digestion</td>
<td>CLPS</td>
<td>Pancreatic colipase</td>
<td>Necessary pancreatic lipase cofactor</td>
</tr>
<tr>
<td>Digestion</td>
<td>UCN3</td>
<td>Urocin 3</td>
<td>Hormone regulating gut passage and energy homeostasis</td>
</tr>
<tr>
<td>Digestion</td>
<td>CAPN9</td>
<td>Calpain 9</td>
<td>Protease</td>
</tr>
<tr>
<td>Digestion</td>
<td>CTRB1</td>
<td>Chymotrypsinogen b1</td>
<td>Protease</td>
</tr>
<tr>
<td>Digestion</td>
<td>CTRB2</td>
<td>Chymotrypsinogen b2</td>
<td>Protease</td>
</tr>
<tr>
<td>Digestion</td>
<td>PGA3</td>
<td>Pepsinogen 3</td>
<td>Protease</td>
</tr>
<tr>
<td>Digestion</td>
<td>VI1</td>
<td>Villin-1</td>
<td>Regulatory of microvilli morphology</td>
</tr>
<tr>
<td>Digestion</td>
<td>PL51</td>
<td>Plastin-1</td>
<td>Regulatory of microvilli morphology</td>
</tr>
<tr>
<td>Xenobiotic metabolism</td>
<td>CYP2B6</td>
<td>Cytochrome P450 family subfamily B member 6</td>
<td>Cytochrome P450 monoxygenase</td>
</tr>
<tr>
<td>Xenobiotic metabolism</td>
<td>CYP2A13</td>
<td>Cytochrome P450 family subfamily A member 13</td>
<td>Cytochrome P450 monoxygenase</td>
</tr>
<tr>
<td>Xenobiotic metabolism</td>
<td>SLC22A2</td>
<td>Solute carrier family 22 member 2</td>
<td>Cation transporter</td>
</tr>
<tr>
<td>Xenobiotic metabolism</td>
<td>ABCG2</td>
<td>ATP binding cassette subfamily G member 2</td>
<td>Transmembrane transporter</td>
</tr>
<tr>
<td>Xenobiotic metabolism</td>
<td>UGT1A8</td>
<td>UDP-glucuronosyltransferase 1 family, polypeptide a8</td>
<td>UDP-glucuronosyltransferase</td>
</tr>
</tbody>
</table>

Fig. 3. Sequence alignment of RNASE1 sites predicted to affect enzymatic activity including duplicated colobine sequences. Positions of positively selected sites in sifakas highlighted in yellow. Duplicated colobine RNASE1 genes and sifaka species bolded.
genetic diversity observed across localities in *P. tattersalli* (56), Perrier’s sifaka (*P. perrieri*) (33), and *P. verreauxi* (67). This would seem to be in contrast to some species, like black-and-white ruffed lemurs (*Varecia variegata*), for which habitat cover is a major predictor of genetic structure (68). Specifically, the capacity for seasonally flexible folivory may buffer sifakas from energetic deficits as leaves remain relatively abundant even as habitat quality diminishes, aiding them in comparison to more specialized, dedicated folivores or frugivores (69–71). *P. tattersalli* appear to exhibit a high level of dietary flexibility that may allow it to tolerate disturbed and/or degraded habitat (23), and sifakas generally have diverse diets (18), including when compared with other sympatric diurnal lemurs (19, 21). Our findings of positive selection on pathways that may be involved in nutrient absorption and xenobiotic metabolism further lend support to this proposition. This finding is consistent with observations that other folivorous primate taxa, like howler monkeys (genus *Alouatta*), which also show extensive dietary flexibility, are relatively resistant to anthropogenic disturbance compared to sympatric frugivores (72, 73). In addition, the degree of disturbance, rather than disturbance alone, may be crucial to understand the impact on population dynamics. For example, there is evidence that some lemurs may benefit nutritionally from “edge effects” in landscapes where discontinuous tree cover allows greater sunlight exposure and leaves with higher protein content as a result (74). Thus, other lemurs in these same and other fragmented and/or degraded habitats in Madagascar may be more affected.

Moreover, it is unclear to what degree any ability of sifakas to cope with fragmentation is long term. There may be lag times between habitat disturbance and/or fragmentation and population declines (75–77). In particular, the long generation times observed in sifakas (45, 46) may obscure the impact of relatively recent connectivity loss. Further, there is evidence that habitat fragmentation leads to declines in diet quality, energy intake, physiological health measures that may reflect immune function, and body condition in *P. diadema* (78–80), indicating that habitat fragmentation is not without harm to these lemurs.

There is also evidence that regional *fadys*, or taboos, against hunting and consuming sifakas have offered a degree of protection from human hunting pressure (81–83). Nevertheless, hunting of *P. verreauxi* has been observed in the southwest (82), some surveys report relatively high rates of consumption of both *P. diadema* and *P. verreauxi* (12 to 29% of respondents had eaten sifaka in the last year) (84), and there is some indication that the strength of traditional *fadys* may be diminishing in some regions (85).

A relatively large geographic distribution, even in a fragmented habitat where populations are still connected by dispersal and gene flow, may contribute to high heterozygosity in *P. coquereli*, *P. verreauxi*, and *P. diadema* (24, 25, 27). The congenor *P. coronatus* also shows a large geographic distribution and high genetic diversity (55). Although census population size estimates are lacking for *P. verreauxi* and *P. diadema* and population size is thought to be decreasing in all four species, the *P. coquereli* population size was nevertheless recently estimated to be around 47,000 in Ankarafantsika National Park alone (25–27, 86), which is consistent with the especially high estimates of genome-wide heterozygosity in this species. Last, social structure may maintain heterozygosity (87). Sifakas exhibit male-biased dispersal, although some females may also leave their natal group to establish a new group (88). Males may disperse multiple times over the course of their lives, usually among neighboring groups (88), and social group boundaries tend to become porous during the very short breeding season, during which time males may sire offspring in neighboring groups (89). These dynamics lead to a “neighborhood”-like intermediate level of hierarchical social organization (88). *P. tattersalli* and *P. verreauxi* social groups have been found to exhibit excess heterozygosity (89, 90), indicating that relatively small-scale social structure can generate and maintain genetic diversity, although the relationship between heterozygosity maintained at the social group level and at a larger spatial scale, such as a metapopulation, is not straightforward (90).

Our results do offer a glimmer of optimism. They confirm that considerable genetic diversity still persists in these endangered sifaka species despite habitat fragmentation, likely due to sifakas’ dietary and behavioral flexibility, including the ability to disperse across open landscapes (33, 55, 56, 67). Although there may be a lag time between anthropogenic disturbance and population declines, our results, along with those of others (30, 33, 54–56), indicate that genetically viable populations of these species may be maintained if further habitat loss or fragmentation is prevented.

The results of our comparative genomic analyses indicate accelerated genetic change in sifakas in several pathways that may play a role in folivory. Evolutionary changes in genes involved in digestion and intestinal absorption (Table 4) likely contribute to maximizing nutrient acquisition. Similarly, some of the genes in the enriched pathways of small-molecule and biosynthetic processes may also aid in nutrient absorption and utilization. Enhanced nutrient uptake capacity would be advantageous under conditions of the unpredictable food resource availability and low primary productivity that characterize Madagascar (2, 4). This nutritional edge might be particularly beneficial when transitioning to a leaf-based diet. For example, genes involved in fatty acid, amino acid, and vitamin processing have been found to show adaptive convergence in giant and red pandas, two carnivorans specialized for a leaf-based diet. In sifakas, *SLC44A4* may contribute to absorption of vitamin B1 (thiamine) produced by microbiota (91), while *SLC52A1* may aid in efficient vitamin B2 (riboflavin) transport (92), *SLC23A1* in vitamin C uptake (92), and *SLC6A20* in amino acid uptake (93). We identified rapid evolution in many genes involved in lipid metabolism (Tables S3 and S5), some of which may enable sifakas to maximize the capture of lipids from a lipid-poor diet. All but one of the genes that display overlapping rapid evolution in sifakas and snub-nosed monkeys also show accelerated change in cattle (35), further implicating these genes in adaptation to herbivory or folivory (Table S4).

Enhanced detoxification capacity, indicated by enriched GO terms related to drug metabolism, could confer tolerance of toxic plant compounds like alkaloids (94, 95), which are particularly abundant in the diets of sifakas compared to other folivorous inirds (16). These xenobiotic molecules are small enough to be absorbed by the host GI tract, after which they may have harmful effects or be energetically expensive to neutralize (94). In particular, two genes in the cytochrome P450 family 2 (*CYP2B6* and *CYP2A13*) showed accelerated change on the ancestral sifaka branch. Gonzalez and Nebert (47) argued that this gene family has expanded in particular species to aid in detoxification of plant secondary compounds as part of an animal-plant arms race. In koalas, one subfamily (cytochrome P450 family 2 subfamily C) in this group expanded and some genes show evidence of selection, which has been proposed can help detoxify eucalyptus leaves (36). Convergent amino acid substitutions in genes in the cytochrome P450 family 2 have also
been identified in giant and red pandas (37), suggesting a common role for these genes in adaptation to a foliage-based diet in mammals.

Overall, these findings also raise the interesting question of how host and microbial genes and metabolism may operate synergistically to mediate folivory across evolutionary and ecological time scales. Gut microbiota are fundamental to folivore and herbivore nutrition, contributing to plant fiber and secondary compound degradation, essential amino acid, vitamin, and nutrient biosynthesis (96). Previous work has shown the gut microbiota of sifakas to comprise taxa and metabolic processes tuned to their hosts’ specialized diets (14) and that respond, in real time, to changing diets (97). The ecological and metabolic success of folivory in sifakas suggests that these lemurs could be prime candidates in which to investigate the coevolution of the genomes of the host organism and its associated microbiota to better understand the shift to novel, digestively challenging feeding strategies.

We also found evidence for positive selection on several bitter taste receptors on the ancestral sifaka branch. This finding is interesting given that Verreaux’s sifakas demonstrate a higher threshold for avoiding bitter-tasting tannins than do sympatric Lemur catta (95). Genetic variation in the binding domains of these genes is thought to be shaped by environmental fine-tuning involving a tradeoff between the conflicting needs to avoid toxins versus preventing the unwarranted exclusion of valuable food resources (98). One of the genes identified, TAS2R16, is one of the best-studied human bitter taste receptor genes. It primarily binds β-glucopyranosides, compounds produced by many plants and insects as deterrents, including several common crops like cruciferous vegetables (e.g., Brussels sprouts and broccoli) and manioc. In the latter, β-glucopyranosides contribute to the production of cyanide when the root is consumed (99). There is potential evidence for functional and possibly adaptive variation in this gene in humans [discussed in (98)]. Sifakas showed amino acid substitutions at three sites within this protein that may be important for signal transduction (V45L, L185V, and L244V) (99).

It is possible that greater tolerance for bitter compounds could coevolve with greater xenobiotic neutralization capacity. Analyses of the koala genome identified, for example, concomitant evolution of bitter taste receptors and cytochrome P450 genes (36). Verreaux’s sifakas consume more alkaloids than do sympatric Lemur catta at Berenty Reserve and Bezà Mahafaly Special Reserve (21, 95). Some bitter taste receptor ligands are also targets of cytochrome P450 enzymes and/or solute carriers, including one, SLC22A2, identified specifically in the present study (100). Moreover, bitter taste receptors are expressed throughout the GI tract, and detection in the gut may actually stimulate the expression of other xenobiotic metabolizers, like ATP-binding cassette transporters (101) and UDP-glucuronoyltransferases (50), many of which were identified in this study.

It is worth noting that we observe in sifakas positive selection on RNase 1 and substitutions at five amino acid sites within it that were experimentally shown to alter enzymatic activity. Selection on duplicated RNASE1 genes in colobine and bovine foregut fermenters is thought to be related to the breakdown of bacterial RNA from large bacterial communities living in the foregut. This breakdown facilitates absorption of important nutrients, like nitrogen and phosphorus (52). Independent amino acid substitutions in colobines and bovine ruminants both result in a lowered optimal pH for RNase 1 enzymatic activity, presumably to boost enzymatic efficiency in the acidic environment of the foregut (51).

A recent study reported that the duplication of the RNASE1 gene in howler monkeys (Alouatta spp.) was followed by amino acid changes paralleling those in colobines (102). Howler monkeys, like sifakas, rely on hindgut fermentation to obtain energy from foliage (94). Although colobines have ruminant capacity (94), they also have an enlarged cecum and colon, which enable an additional round of fermentation and energy capture in the hindgut (94). These collective findings suggest a role for RNASE1 in hindgut and foregut fermentation. Nevertheless, the lack of evidence for RNASE1 duplications in sifakas is notable and warrants further study given the mode of RNASE1 evolution observed in these other lineages. Last, because the RNase1 enzymes play nondigestive roles in other tissues, in particular in the immune system (22), we cannot rule out alternative explanations for selection on RNASE1; however, that alternate pressures could lead to parallel gene changes between sifakas and other folivores seems less likely than parallelism due to dietary adaptation.

Although a smaller proportion of the sifaka diet is composed of leaves compared with colobines and other indriids (16, 94), our findings indicate extensive selection for the capacity for folivory in sifakas. These signatures in the genome may reflect a more folivorous indriid ancestor and/or a scenario in which leaves represent an essential fallback food. Because we currently lack data for other indriid genera, we cannot rule out the possibility that many of the protein-coding changes we assign to the sifaka branch actually lie deeper within indriids. Future genomic analyses of the indri (I. indri) and woolly lemurs (Avahi spp.) are warranted to answer this question. Data for members of the Lepilemuridae family, in which folivory evolved independently (28, 103), could also potentially shed greater light on common physiological mechanisms involved in adaptation to a leaf-based diet in lemurs.

The rapid evolution of ribosomal proteins, as well as other gene families involved in translation and proteome maintenance, is of potential interest as these families have been found to show rapid evolution in desert- and/or heat-adapted animal species (104, 105). High temperatures can disrupt transcription and translation and cause the denaturation and degradation of proteins (106). Western sifakas, including P. coquereli, which provides the reference genome reported here, live in seasonally very hot and dry tropical desert/savannah/dry forest biomes (1).

The lemurs of Madagascar represent an adaptive radiation of primates exploiting a diverse array of habitats and ecological niches. Sifakas evolved to exploit a variety of forest types across Madagascar through a series of genetic, morphological, and microbial adaptations to folivory in combination with locomotor adaptations to life in the trees. Although folivory presents many digestive challenges (94), the capacity to meet nutritional needs with a leaf-based diet might have been especially beneficial to sifakas, given the erratic fruit availability in Madagascar caused by unpredictable seasonality and/or the low nitrogen content of Malagasy fruit (5, 6). Our analysis of sifaka genomes reveals a number of genetic characteristics that likely played a role in sifakas’ successful movement into and exploitation of forest niches where leaves were an abundant resource. Moreover, a flexibly folivorous dietary repertoire today continues to contribute to an ability to tolerate a degree of habitat disturbance and partly account for the high levels of genetic diversity that sifakas currently appear to retain. Conservation efforts to prevent further habitat loss and/or fragmentation while these populations appear to largely remain viable are critical. Further study of other lineages...
within the lemuriform clade is likely to be equally revealing of aspects of primate evolutionary adaptation and their underlying molecular mechanisms.

**MATERIALS AND METHODS**

**DNA extraction and sequencing**

For the sifakas housed at the DLC, we extracted DNA using the QIAGEN Blood and Cell Culture Kit. For the wild Verreaux’s sifaka individual, we extracted DNA as described in (107) from banked ear tissue. All samples then underwent quality control via Picogreen quantification and gel imaging for concentration and purity. We used the manufacturer’s recommended procedures to generate Illumina 100-bp paired-end whole-genome sequencing libraries. We sequenced those libraries to 104.7× whole-genome coverage via runs of 12 lanes on an Illumina Hi-Seq 2000 instrument and 1 lane on an Illumina Hi-Seq 2500 instrument for the individual used for the assembly. The additional individuals were sequenced on two lanes of an Illumina Hi-Seq 2000 instrument each. For the Verreaux’s sifaka from Beza Mahafaly Special Reserve, 150-bp paired-end libraries were generated and sequenced on a single Illumina Hi-Seq X lane.

**Genome assembly and annotation**

We assembled the resulting quality-filtered sequence reads using standard approaches and methods (BioProject PRJNA260602). Briefly, we performed the initial assembly steps using ALLPATHS-LG version R43839 (108). We then conducted additional scaffolding using AtlasLink v1.0 (www.hgsc.bcm.edu/software/atlas-link). Where possible, we filled remaining gaps within scaffolds using the original Illumina read data and AtlasGapFill v2.2 (www.hgsc.bcm.edu/software/atlas-gapfill). The NCBI annotation team performed the annotation of Pcoq_1.0 using their Eukaryotic Annotation Pipeline (www.ncbi.nlm.nih.gov/genome/annotation_euk/process).

**P. diadema, P. tattersalli, and P. verreauxi sequences**

We aligned diadem, golden-crowned, and Verreaux’s sifaka reads to the Pcoq_1.0 assembly and called variants in regions of sufficient coverage. To obtain sequences for the comparative genomic and gene family evolution analyses, we applied variants in the various sifakas’ sequences to the Coquerel’s sifaka reference sequence using GATK (Genome Analysis Toolkit) FastaAlternateReferenceMaker (109).

**Diversity analysis**

We divided the number of heterozygous sites by the number of callable sites to estimate genome-wide heterozygosity for each sifaka individual. We used vcftools (110) to count heterozygous sites.

**Demographic reconstruction**

We used PSMC (43) to infer past effective population size ($N_e$). We used several different ages for the generation time parameter and three mutation rate estimates to scale coalescent time to historic time (see section S2). To obtain confidence intervals, we performed 100 bootstrap replicates.

**Comparative genomics**

We used branch and branch site models in codeml (111) within the python wrapper environment for Tree Exploration (ETE) Toolkit [v3; (111)] using orthologous gene alignments for 30 mammals, including 27 primates for 8862 genes and a species tree (28, 112) to model protein-coding evolution along the ancestral sifaka branch. We used likelihood ratio tests to determine whether a model of accelerated evolution on the branch of interest relative to the rest of the tree was better than a null model. To identify genes potentially involved in adaptation to folivory, we compared REGs in the sifaka lineage with genes previously identified as REGs in the genomes of several snub-nosed monkey species (Rhinopithecus roxellana, R. bieti, R. brelichi, and R. slykeri), which also showed rapid evolution in bovine ruminants (35).

**Gene family evolution**

We assessed gene family expansion and contraction on the ancestral sifaka branch using CAFE (Computational Analysis of gene Family Evolution) (113) and a reduced dataset of 12 species, including 11 primates and the Northern treeshrew (Tupaia belangeri; section S5 and fig. S4).

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/17/eabd2274/DC1

View/request a protocol for this paper from Bio-protocol.

**REFERENCES AND NOTES**


63. historical and lemur (A. habrarrnon in Madagascar.

66. M. V. Irwin, Feeding ecology of African primate communities: A


99. B. Deza, C. Gutta microbiome and gut microbiome in non-colonial primates, the howler monkeys (M. fuliginosa).


153. M. D. Dearing, K. D. Kohl, Beyond fertilization: Other important services provided to endothermic herbivores by their gut microbiota. Integr. Comp. Biol. 57, 723–731 (2013).


Acknowledgments: We would like to thank Madagascar National Parks for allowing us to conduct research at Beza Mahafaly Special Reserve; the DLC and the Beza Mahafaly Special Reserve monitoring team for samples; K. Glander, D. Haring, and K. Gerard for helpful information; A. Richard, E. Sargsi, and D. Watts for helpful feedback on earlier drafts of this work; P. Cingolani, the creator of SnpEff, S. Langdon of the Duke DNA Analysis Facility for support; and E. Tapanes and G. Tiley for helpful discussions. We thank the Yale Center for Research Computing and Duke Computer Cluster team for guidance and use of the research computing infrastructure. We also thank the library and sequence production teams and the project managers’ group within the Human Genome Sequencing Center, Baylor College of Medicine for essential assistance in generating the data reported here. This is DLC publication #1474. Funding: We thank the Center for the Advanced Study of Human Paleobiology at The George Washington University, Duke University, and the Wenner-Gren Foundation for funding. Genome sequencing and assembly were funded by National Human Genome Research Institute grant U54 HG003273 to R. Gibbs (HGSC, Baylor College of Medicine). The research reported here complied with governmental and institutional regulations and ethical guidelines. Author contributions: Designed the study and supervised the analysis: J. Rog, K.C.W., and A.D.Y.; managed or supervised the sequence production: M.R. and D.M.M.; managed or supervised the preparation of sequencing libraries: M.R. and D.M.M.; produced the assembly: K.C.W., Y.L., S.M., R.A.H., and D.S.T.H.; project and data management: D.M.M., M.R., and K.C.W.; coding region assembly from resequencing data: E.E.G. and T.W.H.; demographic reconstruction, selection analysis, and gene family analysis: E.E.G.; RNase1 analysis: E.E.G. and L.K.G.; provided essential biomaterials: R.R.L., B.B., L.K.G., J.R., and A.D.Y.; wrote the paper: E.E.G., J.Rog., T.H.W., R.R.L., B.B., L.K.G., and A.D.Y.; All authors read and approved the final manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All the raw sequence data, as well as sample metadata and additional information for this genome assembly and analysis project, are available under BioProject PRJNA251429 at www.ncbi.nlm.nih.gov. This includes the Pcoq_1.0 genome assembly (RefSeq accession GCF_000956105.1) and the supporting read data (SRX accessions SRR701287, SRR701294 to SRR701300, SRR763489 to SRR763493, and SRR1025581 to SRR1025582). The data used for the diversity analyses across several Propithecus species can be accessed under umbrella project PRJNA26602, which covers the data for P. diadema (PRJNA265045), P. tattersalli (PRJNA265041), and P. verreauxi (PRJNA265040). Additional data related to this paper may be requested from the authors.

Submitted 13 June 2020
Accepted 8 March 2021
Published 23 April 2021
10.1126/sciadv.abd2274

Comparative genomic analysis of sifakas (*Propithecus*) reveals selection for folivory and high heterozygosity despite endangered status

Elaine E. Guevara, Timothy H. Webster, Richard R. Lawler, Brenda J. Bradley, Lydia K. Greene, Jeannin Ranaivonasy, Joelisa Ratsirarson, R. Alan Harris, Yue Liu, Shwetha Murali, Muthuswamy Raveendran, Daniel S. T. Hughes, Donna M. Muzny, Anne D. Yoder, Kim C. Worley and Jeffrey Rogers

*Sci Adv* 7 (17), eabd2274.
DOI: 10.1126/sciadv.abd2274