<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Perry, George H.; Jr., Edward E. Louis; Ratan, Aakrosh; Bedoya-Reina, Oscar C.; Burhans, Richard C.; Lei, Runhua; Johnson, Steig E.; Schuster, Stephan C.; Miller, Webb</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2013</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/18430">http://hdl.handle.net/10220/18430</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2013 The Authors. This paper was published in Proceedings of the National Academy of Sciences of the United States of America and is made available as an electronic reprint (preprint) with permission of the authors. The paper can be found at the following official DOI: [<a href="http://dx.doi.org/10.1073/pnas.1211990110">http://dx.doi.org/10.1073/pnas.1211990110</a>](<a href="http://dx.doi.org/10.1073/pnas.1211990110">http://dx.doi.org/10.1073/pnas.1211990110</a>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.</td>
</tr>
</tbody>
</table>
Aye-aye population genomic analyses highlight an important center of endemism in northern Madagascar

George H. Perrya,b,c, Edward E. Louis, Jr.b,1, Aakrosh Ratanb, Oscar C. Bedoya-Reinac, Richard C. Burhansc, Runhua Leib, Steig E. Johnsona, Stephan C. Schusterc, and Webb Millerb,c,2

*aDepartments of Anthropology and Biology and bCenter for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802; cCenter for Conservation and Research, Omaha’s Henry Doorly Zoo and Aquarium, Omaha, NE 68107; dDepartment of Anthropology, University of Calgary, Calgary, AB, Canada T2N 1N4; and eSingapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551

Edited by Elwyn L. Simons, Duke University, Durham, NC, and approved February 8, 2013 (received for review July 14, 2012)

We performed a population genomics study of the aye-aye, a highly specialized nocturnal lemur from Madagascar. Aye-ayes have low population densities and extensive range requirements that could make this flagship species particularly susceptible to extinction. Therefore, knowledge of genetic diversity and differentiation among aye-aye populations is critical for conservation planning. Such information may also advance our general understanding of Malagasy biogeography, as aye-ayes have the largest species distribution of any lemur. We generated and analyzed whole-genome sequence data for 12 aye-ayes from three regions of Madagascar (North, West, and East). We found that the North population is genetically distinct, with strong differentiation from other aye-ayes over relatively short geographic distances. For comparison, the average FST value between the North and East aye-aye populations—separated by only 248 km—is over 2.1-times greater than that observed between human Africans and Europeans. This finding is consistent with prior watershed- and climate-based hypotheses of a center of endemism in northern Madagascar. Taken together, these results suggest a strong and long-term biogeographical barrier to gene flow. Thus, the specific attention that should be directed toward preserving large, contiguous aye-aye habitats in northern Madagascar may also benefit the conservation of other distinct taxonomic units. To help facilitate future ecological- and conservation-motivated population genomic analyses by noncomputational biologists, the analytical toolkit used in this study is available on the Galaxy Web site.

conservation genomics | landscape species concept | genomics of non-model species

Madagascar maintains one of the highest levels of unique biodiversity—coupled with imminent extinction risk—in the world (1–4). The endemic primates, lemurs, are among the most diverse faunal groups on Madagascar, with ~100 distinct extant taxa (5). Because of a unique ecological and demographic profile, the lemur species with the largest geographical distribution, the aye-aye (Daubentonia madagascariensis) (6), may also be among the most sensitive to continuing degradation of Madagascar’s forests. Specifically, aye-ayes have very large individual home-range size requirements relative to other lemurs (7–11), population densities that are inferred to be very low (12), a relatively slow life history (13), and the lowest nuclear genetic diversity of any primate yet studied (14). Therefore, their ability to maintain sufficient individual numbers for long-term population viability in remaining forest patches may be at risk.

Aye-ayes are highly specialized extractive foragers, with relatively large, continuously growing incisors that are used to gnaw through decaying tree bark (deadwood) or bamboo to access wood-boring insect larvae and through the endocarp of seeds from the ramy tree (Canarium) to access endosperm (7, 15–17). A slender, flexible, probing third digit is used to extract these foods and bring them to the mouth (18, 19). Limitations in the availability of either deadwood or Canarium resources may explain the large individual home-range requirements, but this has not yet been shown. Aye-ayes are also nocturnal, cryptic, and primarily solitary, making them difficult to study and sample in the wild. As a result, no comparative population studies of this species have been published previously.

Adequate conservation planning requires knowledge of both long-term landscape dynamics and patterns of species distribution in suitable habitats (20). Wider geographic scale assessments are particularly important for species with large-range requirements (e.g., refs. 21 and 22), such as aye-ayes. Thus, it is important to understand the patterns of genetic differentiation that exist among surviving aye-aye populations. Because of their extensive individual home-range sizes and low population densities, conservation efforts relevant to aye-ayes will need to prioritize the preservation of large and contiguous forests. Although such protected areas do exist in Madagascar [albeit many of them currently under stress (23)], prior to this study we have not been able to assess whether current protected areas and conservation strategies maximize the preservation of distinct aye-aye populations and overall genetic diversity, because the genetic relationships among aye-aye populations have been unknown.

The analysis of population-level genome sequence data offers potentially powerful insights into both demographic and evolutionary processes. Although such analyses could thus benefit behavior, conservation, and ecological research across many taxa (24–27), large-scale whole-genome sequencing population studies conducted to date have typically focused on humans and model organisms (e.g., refs. 28–30). With continued increases in sequencing capacity, genomic-scale population studies of nonhuman, nonmodel organisms are increasingly feasible. Indeed, transcriptome sequencing, other reduced representation methods, and whole-genome sequencing approaches have been used in multiple recently published studies (14, 31–35).

In this study, we have generated and analyzed intermediate-coverage whole-genome sequence data for 12 aye-aye individuals from three regions of Madagascar (Fig. 1). We investigated population structure and quantified population differentiation using an analytical toolkit that we have made available through the Galaxy Web site (36, 37) to facilitate similar, future studies


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Sequences have been deposited in the National Center for Biotechnology Information Information short read archive, www.ncbi.nlm.nih.gov/Traces/sra (accession no. SRA066440). Analysis workflows and the full input table of 4,555,737 aye-aye SNPs are available on the Galaxy Web site (ugegalaxy.org/aye-aye).

1G.H.P. and E.E.L. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: ghp3@psu.edu or webb@bx.psu.edu.
of other species. Although the analyses in this study benefitted from our previous assembly of an aye-aye reference genome (38) and high-quality DNA samples isolated from blood and tissue, the absence of such resources would not preclude the application of a similar pipeline to other nonmodel species (SI Text). For example, for this study we collected a total of more than twice as much genome sequence data than we had used to assemble the aye-aye reference genome (38). As an initial step in analyzing other species, a genome assembly could be constructed, benefitting from continually improving genome-assembly algorithms. For analyses like those reported in this article, long-range contiguity of the assembly is immaterial and thus relatively simple assembly methods are adequate. Alternatively, analyses can use the available genome sequence and gene annotations of a related species, such as analyzing bear sequences used to assemble the aye-aye reference genome (38). To limit the incorporation of erroneous genotypes into our analyses, we focused on the genotypes of 666,256 SNPs (of a total of 4,555,737 SNPs) that were covered by a minimum of four sequence reads per individual (see Materials and Methods).

**Aye-Aye Population Structure.** To examine relationships among the aye-aye individuals from the three regions of Madagascar, we constructed a neighbor-joining tree based on genome-wide SNP genotype distances. Individuals from each of the three populations cluster, and the East and West population clusters are more similar to each other than either is to the North (Fig. 2A). In addition, we constructed a rooted neighbor-joining tree by first aligning the aye-aye and human reference sequences (hg19). Given the relatively high levels of sequence divergence in introns and intergenic regions, to ensure orthology it was necessary to focus on gene coding regions. We then analyzed only autosomal synonymous SNPs from the gene-coding regions, because such SNPs do not affect amino acid sequences and are thus presumably neutral. Although the resulting tree is based on a distance matrix constructed from magnitudes of fewer SNPs than that of the genome-wide dataset (857 vs. 666,256), the patterns of aye-aye population structure are the same and the root of the tree separates the North from the East and West populations.

**Results**

We analyzed intermediate-coverage whole-genome sequence data for 12 aye-aye individuals from three regions of Madagascar: North (n = 4 individuals), West (n = 3), and East (n = 5) (Fig. 1). We identified SNPs following sequence-read alignment to an aye-aye reference genome (38). To limit the incorporation of erroneous genotypes into our analyses, we focused on the genotypes of 666,256 SNPs (of a total of 4,555,737 SNPs) that were covered by a minimum of four sequence reads per individual (see Materials and Methods).

**Fig. 1.** Aye-aye geographical species range and sampling locations. Remaining forests (light green) and presumed current species distribution of aye-ayes (dark green) in Madagascar. Capture locations for the North (blue squares), West (yellow triangles), and East (red circles) aye-aye population samples included in this study are indicated. Species distribution is based on the selection of remaining forest identified from classified satellite imagery, deforestation data from Harper, et al (2007), courtesy of Cambridge University Press (51) using a vector polygon of aye-aye distribution from Andrainarivo et al. (71), which was then further modified according to our field observations of aye-aye feeding traces and occasional sightings. Nonforested areas were not represented in the aye-aye distribution.

**Fig. 2.** Aye-aye population structure. Analyses of estimated genotype SNPs with minimum 4× sequence coverage in each of the 12 individuals studied, and maximum 120× coverage in those individuals combined (Materials and Methods). (A) Neighbor-joining tree estimated from a genotype distance matrix based on all 666,256 genome-wide SNPs. Pairwise distances were calculated as total SNP genotype distance, with distance for an individual SNP the difference between two individuals’ genotypes scored as 0, 0.5, and 1 (e.g., AA, AT, and TT, respectively). (B) Rooted neighbor-joining tree estimated from a distance matrix based on 857 autosomal synonymous SNPs from gene coding regions that could be aligned to the human genome (hg19). Pairwise distances were calculated as described above. The nucleotide of the human reference sequence was different from both aye-aye alleles for 73 of the 857 SNPs; in these cases the human genotype was scored as 0.5. (C) Population structure analyses based on all 666,256 genome-wide SNPs. Cluster membership proportions for each individual are depicted for both k = 2 and k = 3 populations. Each individual is represented as a vertical bar with population origins indicated below the bars.
(Fig. 2B). A population structure analysis produced consistent results, with the North individuals distinguished from all others at $k = 2$ populations, and individuals from each of the three regions completely separated at $k = 3$ (Fig. 2C). Results from a principal component analysis are also similar. The first principal component clearly separates the North individuals from all others, and the second principal component separates West and East individuals (Fig. S1).

**Population Differentiation.** To quantify the level of genetic differentiation between aye-aye populations, or the amount of total genetic variation that can be explained by population structure, we estimated $F_{ST}$ for each SNP that was not fixed for the same allele in each of the two populations being compared. We calculated $F_{ST}$ values using an unbiased estimator from Reich et al. (41) that is not adversely affected by small population sample sizes (42). The average $F_{ST}$ values were 0.169 for the North vs. East populations (596,785 SNPs), 0.194 for the North vs. West populations (517,323 SNPs), and 0.129 for the East vs. West populations (536,734 SNPs).

We next assessed the level of observed genetic differentiation between aye-aye populations in a comparative context. To do so, we created an equivalent dataset for humans based on publicly available data. Specifically, we obtained genome sequence data for a total of 12 human individuals from three populations that were, as were our aye-aye data, generated using Illumina sequencing technology. The sampled human populations were Sub-Saharan African agriculturalists ($n = 4$ individuals), European ($n = 5$ individuals), and Southeast Asian ($n = 5$ individuals). We matched sequence coverage levels to the aye-aye data at both the individual and population levels (Fig. S2), and used the same pipeline for sequence alignment, SNP genotype estimation, data filtering (e.g., minimum fourfold sequence coverage per individual), and $F_{ST}$ analysis.

The average $F_{ST}$ values were 0.078 for the African vs. European populations (1,061,671 SNPs), 0.091 for the African vs. Asian populations (988,646 SNPs), and 0.069 for the European vs. Asian populations (748,232 SNPs). Thus, the level of genetic differentiation between aye-aye populations from the North and East regions of Madagascar is more than 2.1-times greater than that between human Africans and Europeans based on an equivalent dataset (Fig. 3). The relative level of aye-aye versus human population differentiation was similar for each of the other equivalent comparisons. Furthermore, the two least-differentiated aye-aye populations, East and West (average $F_{ST} = 0.129$), are likely more genetically differentiated than Africans and Asians (average $F_{ST} = 0.091$), the two most differentiated human populations in our analysis. We obtained consistent results from SNP subsets generated using sequence coverage cutoffs of five-, six- and sevenfold per individual, and when using Weir and Cockerham’s unbiased estimator for $F_{ST}$ (43) or Wright’s original $F_{ST}$ definition (44) instead of Reich et al.’s unbiased $F_{ST}$ estimator (41) (Fig. S3).

**Neutral Genetic Diversity.** Aye-ayes have the lowest level of genetic diversity of any studied primate species (14, 38, 45), which is likely a function of large individual range requirements and low population densities throughout the aye-aye species distribution. Alternatively, because the previous genetic diversity estimates were based primarily on individuals with ancestry from only one region of Madagascar (equivalent to our East population), they could reflect population-specific rather than species-wide demographic processes. We sought to address this issue by estimating neutral genetic diversity separately for each of the three aye-aye populations in our study. Although we expect to underestimate true genetic diversity with the intermediate coverage sequence data (because of the undercalling of rare SNPs and heterozygous sites), these estimates should be generally comparable both across aye-aye populations and against humans (using our sample size and sequence coverage-matched human dataset).

We identified the total number of autosomal synonymous sites covered by a minimum of four sequence reads per individual for each species (aye-aye = 368,675 synonymous sites; human = 915,245 sites) (SI Materials and Methods) and computed average pairwise genetic diversity ($\pi$) for each population: aye-aye North $\pi = 0.054\%$, aye-aye East $\pi = 0.057\%$, aye-aye West $\pi = 0.049\%$, human African $\pi = 0.093\%$, human European $\pi = 0.070\%$, human Asian $\pi = 0.066\%$. As expected, these estimates are lower than those from previous studies that used higher coverage sequence data (14), but the magnitude of the between-species differences are similar. Moreover, in contrast to the variability observed among human populations [as expected, with higher genetic diversity in Africa (46)], genetic diversity levels are similar among aye-aye populations suggesting that relatively low genetic diversity is a species-wide characteristic.

**Discussion**

We conducted this study to characterize patterns and levels of genetic differentiation among aye-aye populations for conservation planning purposes as well as to contribute to the general understanding of biogeographical processes in Madagascar, as this species has the widest geographical distribution of any lemur (6, 12). The aye-aye’s demographic profile suggests a particular sensitivity of this species to the rapid degradation and fragmentation of Madagascar’s forests. Specifically, aye-aye home-range sizes of 120–215 ha for males and 30–40 ha for females [with travel distances up to 4.4 km/night (47)] are very large for a solitary animal only ~2.5 kg in size and considerably larger than those of other lemurs across a diversity of activity patterns, social systems, and body sizes [for example, 1–2 ha for woolly lemurs (8), 5 ha for fork-marked lemurs (9), 5.3 ± 5.2 ha for blue-eyed black lemurs (10), and 5.7–10.1 ha for Verreaux’s sifakas (11)]. Because there is minimal same-sex overlap in female home ranges (47), aye-aye population densities are inferred to be relatively very low (12). Perhaps unsurprisingly, the level of...
estimated nuclear genomic diversity in aye-ayes is the lowest of any primate yet studied (14). Thus, the ability of aye-ayes to maintain sufficient individual numbers for long-term population viability in remaining forest patches may be at risk.

Our analyses revealed that although aye-ayes from the East and West coasts of Madagascar are distinguished readily by their genome sequences, divergence between either of these populations and the northern Madagascar population is greater. Although connected forests between the habitats of the North and East sampled populations no longer exist (Fig. 1), the level of genetic differentiation between the populations in these regions implies a longer-term reproductive barrier than that which could be attributed to the human-mediated habitat loss that began only within the past 2,300 y (48, 49) before accelerating rapidly over the past century (50, 51). We found that the level of genetic differentiation between the North and East populations is substantially greater than that between, for example, human African and European populations, based on the analysis of an equivalently curated human SNP database. Although the relative level of aye-aye genetic differentiation—across only a relatively small geographic distance (248 km), representing a small part of the total aye-aye range (Fig. 1)—is intriguing, what relevance does this result have for conservation planning? In particular, the level of genetic differentiation among human populations is not typically considered high among primates (52). In addition, the observed level of aye-aye population differentiation may not be unusual for a species with low population sizes and geographical barriers across its range. Future expectations of an expanded population genomic database that will include many additional endangered taxon will help us to better contextualize this result.

As a starting point, we also computed the average $F_{ST}$ between populations of Alaskan and Norwegian polar bears using data from a recent genomics study (34). Although the polar bear sequence coverage levels could not be matched precisely to our aye-aye and human datasets, the average North vs. East aye-aye $F_{ST}$ was more than five times greater than that for polar bears (SI Materials and Methods), strengthening the belief that the observed aye-aye $F_{ST}$ values may be unusual for a wide-ranging animal over such a small geographic distance. For now, we are not suggesting that the North aye-aye population should necessarily be considered a distinct taxonomic unit. However, if general goals of conservation include preserving distinct population and maximizing overall species-level genetic diversity, then this population merits particular protection, especially for a species with a demographic profile that suggests high extinction risk.

Because aye-ayes have a geographical species distribution that only excludes the central highlands and the southwest of Madagascar [the largest distribution of any lemur (6, 12)], these results are also particularly valuable for our broader understanding of the island’s biogeography. Indeed, the distinctiveness of the North aye-aye population accords well with 1 of 12 proposed cross-taxa centers of endemism, based on an analysis of elevation, the location of watersheds, and Quaternary climatic shifts (53). Multiple river systems, such as the Manambato, Bemarivo, and Ankavandra, have the potential to limit the dispersal of aye-ayes and other species between northern Madagascar and adjacent regions. Furthermore, the Tsaratanana Massif, which includes the highest peak in Madagascar, rises above the known elevational limits of aye-ayes (54), and thus may form at least a partial barrier to gene flow. We have sampled from only a portion of the total aye-aye species distribution. The cryptic, nocturnal nature of this species and low population densities constrain sampling efforts, but future population genomic studies that include populations from further south along the west and east coasts of Madagascar would likely contribute further to biogeographic knowledge and to aye-aye conservation efforts.

Preservation of the distinctive aye-aye populations in northern Madagascar would likely have indirect benefits for the conservation of other taxa within this center of endemism, the genetic diversity and structure of which are not yet fully characterized. The geographical correspondence of this northern region with patterns of species turnover in other lemur taxa (53) suggests similar underlying biogeographic processes and responses to landscape variation. Although the use of surrogate species for conservation management is strongly debated (55, 56), carefully selected individual species or groups may serve as appropriate indicators of overall biodiversity, although these applications are often highly context-dependent (57). Similarly, protecting species at the highest trophic levels and with large home ranges (e.g., apex predators) can serve ecosystem-level conservation goals and preserve disproportionate amounts of diversity in other taxa (58). Although not strictly predatory, aye-ayes maintain an ecological niche that necessitates large individual ranges, which can incorporate entire communities and populations of other endangered fauna, including other lemurs. Thus, any efforts to preserve contiguous forests large and diverse enough to support a viable population of aye-ayes would be very likely to also meet or exceed the space requirements for other taxa. Additionally, because of their relatively slow life history, aye-ayes may be particularly sensitive to rapid habitat changes. The combination of these factors would make aye-ayes a strong candidate for a focal species in a “landscape species” approach to conservation planning (59, 60).

Conclusion

In this study, we generated and analyzed complete genome sequences from 12 aye-aye individuals to characterize levels and patterns of genetic differentiation and to highlight a distinct population in northern Madagascar. This work serves as a potential model for future conservation- and ecology-motivated population genomic studies of nonmodel species; such research is expected to become more feasible with continuing advances in sequencing technology and capacity. Thus, we have made the analytical toolkit used for our analyses available on the Galaxy Web site. We argue that conservation attention should be directed toward an important center of endemism in northern Madagascar. Such efforts would preserve distinct populations of a species that is one of the world’s most unusual and highly specialized mammals, as well as other potentially distinct populations and taxa in the region.

Materials and Methods

DNA Samples. DNA samples from 13 wild-caught aye-aye individuals were initially included in this study. Genomic DNA was extracted from liver-tissue samples collected from free-ranging individuals in Madagascar. In Madagascar, aye-ayes were immobilized with a CO$_2$—protein mixture, or blowgun with 10 mg/kg of Telazol (Fort Dodge Animal Health), and 1.0 cc/kg of 2% (vol/vol) EDTA, 2% (vol/vol) SDS at room temperature until transferred to the laboratory for storage at −80 °C. All collection and export permits were obtained from Madagascar National Parks, formerly Association Nationale pour la Gestion des Aires Protégées, the Ministère des Eaux et Forêts of Madagascar. All rules and regulations were followed according to Malagasy law. Samples were imported to the United States under Convention for International Trade in Endangered Species permits 08US121039/9, 08US121040/9, and 08US121041/9 from the US Fish and Wildlife Service. Capture and sampling procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Omaha’s Henry Doorly Zoo and Aquarium under IACUC #120104. Genomic DNA was isolated from the samples using a standard Phenol-Chloroform extraction protocol (61).

Sequencing, Sequence Alignment, and SNP Identification. For detailed descriptions of the library preparation, sequencing, alignment, and SNP identification methods see SI Materials and Methods. Each sample was pair-end sequenced for 101 bp from each end using one lane of the Illumina HiSeq 2000 sequencing system. We obtained an average of 204,202,246 total reads

per lane (SD = 40,714,120) (Dataset S1), or ~20 Gb of raw sequence data per individual. Sequence data have been deposited in the National Center for Biotechnology Institute short read archive under accession no. SRA06644A. Sequence reads were aligned to the aye-aye reference genome sequence (38) using the Burrows-Wheeler Aligner (62). On average, we mapped 16 Gb of sequence data per individual (SD 4.1 Gb) (Dataset S1), corresponding to an average of ~5.6-fold coverage of the 2.9-Gb aye-aye reference genome sequence. We used SAMTools (63) to identify the locations of SNPs and estimate genotypes at all SNPs for each individual, regardless of sequence coverage for that SNP and individual.

Quality Control and SNP Filtering. We selected an intermediate-coverage population genomics study design, using one HiSeq lane per individual (resulting in an average of ~5.6× per-individual mapped sequence coverage). For analyses of these data, our approach was to focus on the subset of identified SNPs with a minimum of 4× sequence coverage for each individual in the study. In doing so, we ignored the majority of our data, but we have more confidence in the accuracy of the estimated genotypes of the SNPs that we do analyze. The number of SNPs remaining after this filtering step is still large (see below), sufficient for accurate population genetics inference (e.g., ref. 64). In future population genomics studies, it may become possible to extend the analysis to a larger proportion of variable sites in the genome by developing and implementing more cost-effective methods to sequence individual genomes. However, such a design would have been economically and computationally inefficient for our study, given current output of the Illumina HiSeq. 2000 sequencing system and data storage and analysis computing resource needs.

With only 4× coverage, some error in the estimated genotypes is expected (e.g., the probability of observing both SNP alleles among four reads for a heterozygous individual = 0.875). However, the 4× threshold is the minimum coverage limit for all SNPs detected in any individual; thus for any given SNP, the coverage levels for most individuals are higher. At this coverage threshold, overall genotype estimate qualities are expected to be high.

Before applying the 4× minimum coverage filter, we considered the coverage distributions (for the 4,555,737 identified SNPs) for each individual (Fig. S2A). One individual, North5, was a coverage outlier, with a mode of only 1.0-fold coverage and 20% of the SNPs for covered by fewer than 2× reads. North5 also had the fewest mapped reads of any individual in the study (Dataset S1). Accordingly, we excluded North5 from further analyses. We also examined the sum total coverage distribution for the remaining 12 individuals (North1 to 4, West1 to -3, East1 to -5) (Fig. S2B). SNPs with relatively low sum total coverage will be filtered because they would not meet the 4× per-individual threshold. Some proportion of the SNPs with relatively high sum total coverage could be located in duplicated regions of the genome, and thus potential false-positives. Therefore, we chose to filter all SNPs with >120× sum total coverage. With this step, we likely excluded many true SNPs, but the remaining number of SNPs was large. Specifically, there were 866,256 SNPs with 4× coverage in each of the 12 individuals, and not more than 120× coverage in those individuals combined. These SNPs were the focus of our population genomic analyses.

Galaxy Tools. We created tools on the Galaxy Web site usegalaxy.org (36, 37) to facilitate ecological- and conservation-motivated analyses of population genomics datasets such as ours, by noncomputational biologists. The user uploads SNP genotype calls with coverage and genotype quality information (alternatively, SAMtools functionality is also available through Galaxy). From the uploaded SNP table, the user may specify populations, compute and display coverage distributions at the individual and population levels, filter SNPs based on individual and population minimum and maximum sequence coverage levels, as well as minimum genotype quality, examine population structure, and perform analyses based on FST. Several popular population genetics and genomics analysis programs, including SMARTPCA (65) and ADMIXTURE (66), have been integrated into the Galaxy functionality.

ACKNOWLEDGMENTS. We thank Sarah Zehr and the Duke Lemur Center for providing captive aye-aye samples; Shannon Engberg and Carolyn Bailey for sample preparation; Lynn Tomsho and John McGraw for library preparation and sequencing; Tracy Wyman for assistance with Fig. 1; Emily Davenport and Kate Thompson for comments on an earlier draft of the manuscript; the Madagascar Biodiversity Partnership for assistance in sample collection and field logistics in Madagascar; and the Madagascar National Parks, formerly Association Nationale pour la Gestion des Aires Protégées, and the Ministère des Eaux et Forêts of Madagascar for sampling permission. Funding for aye-aye sample collection was provided by Conservation International, the Margot Marsh Biodiversity Fund and the Margot Marsh Biodiversity Foundation (E.E.L.), the European Biodiversity Action Fund and the Margot Marsh Biodiversity Foundation (G.H.P.). The development of the Galaxy tools to analyze intermediate-coverage sequence data from multiple individuals was supported by the National Institutes of Health Grant UL1 RR03184-01 (to the Pennsylvania State Clinical and Translational Science Institute), with additional funding from National Science Foundation Award DEB 0733029 and a grant from the Pennsylvania Department of Health using Tobacco Comprehensive Universal Research Enhancement Funds. This is Duke Lemur Center publication #1241.


