Supplementary Materials

Article

RADseq data suggest occasional hybridization between the *Microcebus murinus* and *M. ravelobensis* in northwestern Madagascar

Helena Teixeira [1,†], Tobias van Elst [1,†], Malcolm S. Ramsay [1,2], Romule Rakotondravony [3,4] , Jordi Salmona [5], Anne D. Yoder [6], Ute Radespiel [1,*]

[1] Institute of Zoology, University of Veterinary Medicine Hannover, Bünteweg 17, Hannover, 30559, Germany
[2] Department of Anthropology, University of Toronto, Toronto, Canada
[3] Ecole Doctorale Ecosystèmes Naturels (EDEN), University of Mahajanga, 5 Rue Georges V— Immeuble KAKAL, Mahajanga Be, B.P. 652, Mahajanga 401, Madagascar
[4] Faculté des Sciences, de Technologies et de l'Environnement, University of Mahajanga, 5 Rue Georges V - Immeuble KAKAL, Mahajanga Be, B.P. 652, Mahajanga 401, Madagascar
[5] CNRS-UPS-IRD, UMR5174, Laboratoire Évolution & Diversité Biologique, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse, France; jordi.salmona@gmail.com
[6] Department of Biology, Duke University, Durham, NC, USA
* Correspondence: UR: ute.radespiel@tiho-hannover.de
† These authors contributed equally to this work.

Text S1 – GATK best practice filtering:

After filtering the genotype calls output by Stacks with VCFtools v0.1.17 [1] to retain only sites present in at least 50% of individuals, we used VariantAnnotator of GATK v3.8.1 [2] to annotate the VCF file with INFO fields for strandedness of reference vs. alternative allele (FS), root mean square mapping quality (MQ), mapping quality of reference vs. alternative allele (MQRankSum), read position of reference vs. alternative allele (ReadPosRankSum) and allele balance (ABHet). Subsequently, sites where one of the following conditions applied were set to FILTER using VariantFiltration of GATK v3.8.1 and removed using VCFtools: FS > 60.0; MQ < 40.0; MQRankSum < −12.5; ReadPosRankSum < −8.0; ABHet < 0.2 or ABHet > 0.8. More information on GATK best practice filtering recommendations can be found at https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants.

Text S2 – Demographic modelling with *fastsimcoal2*:

To infer and date the occurrence of gene flow between *M. murinus* and *M. ravelobensis*, we firstly compared three simple demographic models assuming panmictic populations (model P1 – P3). While the first model assumes that there is no gene flow between *M. murinus* and *M. ravelobensis* (null model, P1), the second and third models consider a constant admixture rate (gene flow model, P2) and changes in admixture rate over time (changing model, P3), respectively. Second, we allowed for population structure between the northern samples (Mariarano) and the ANP samples, as suggested by the clustering results. For this, we compared four demographic models assuming that ancestral *M. murinus* and *M. ravelobensis* populations each split into a northern and southern cluster at time T1 (M1 – M4). While the first of the four models assumes no gene flow between *M. murinus* and *M. ravelobensis* (null model, M1), the second model considers gene flow only between the two ancestral unstructured mouse lemur populations (i.e., before T1; ancient gene flow model, M2). The third model assumes only recent gene flow after the *M. murinus* and *M. ravelobensis* populations became structured (i.e., after T1; recent gene flow model, M3). The fourth model assumes both ancient and recent gene flow (i.e., before and after T1; ancient & recent gene flow model, M4). Since M1 – M4 assume that the ancestral mouse lemur populations split into northern and southern clusters at the same time point T1. Additionally, we reran the best ranking structured model but assuming that *M. murinus* and *M. ravelobensis* became structured at different time

points (i.e., at T1 and T2; model M5). For all models, we considered constant effective population size (Ne) but allowed for different Ne-values for extant populations.

We ran fastsimcoal2 v.2.6 using 200,000 coalescent simulations per set of parameters and 40 ECM cycles during parameter estimation from the SFS (see [3] for methodological details). A total of 100 independent fastsimcoal2 v.2.6 [4] runs were performed for each demographic model to determine the parameter estimates that maximize the composite-likelihoods [4] (see Additional Tables A2 and A3 for details about models parameters). The Akaike Information Criteria (AIC) [5] was used to rank all demographic models, and the models with the lowest ΔAIC within each model category (i.e., population panmixia and structure) were selected as the best demographic models. Finally, Arlecore was used to generate 100 non-parametric bootstrap SFS replicates. A total of 100 independent fastsimcoal2 runs were then performed with each of the bootstrapped SFS. The parameter estimates with the highest likelihood from each independent run were used to estimate the 95% confidence intervals with the Rmisc R package [6].

Supplementary references:

1. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158, doi:10.1093/BIOINFORMATICS/BTR330.

2. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303, doi:10.1101/gr.107524.110.

3. Teixeira, H.; Montade, V.; Salmona, J.; Metzger, J.; Bremond, L.; Kasper, T.; Daut, G.; Rouland, S.; Ranarilalatiana, S.; Rakotondravony, R.; et al. Past Environmental Changes Affected Lemur Population Dynamics Prior to Human Impact in Madagascar. *Commun. Biol.* **2021**, *4*, 1084, doi:10.1038/s42003-021-02620-1.

4. Excoffier, L.; Dupanloup, I.; Huerta-Sánchez, E.; Sousa, V.C.; Foll, M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet.* **2013**, *9*, doi:10.1371/journal.pgen.1003905.

5. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716–723.

6. Harrell Jr, F.E.; Harrell Jr, M.F.E. Package 'Hmisc.' *CRAN2018* **2019**, *2019*, 235–236.

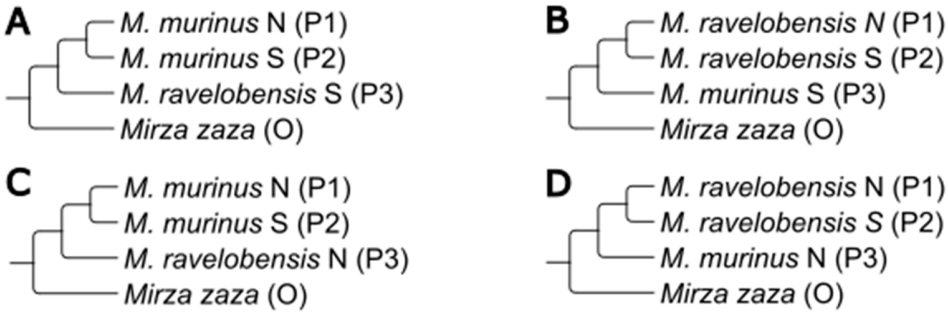

Figure S1. Configuration of the four introgression tests given the tree topology (((P1, P2), P3), O). We tested for introgression between *M. murinus* and *M. ravelobensis* southern clusters (A and B) and northern clusters (C and D). *Mirza zaza* was used as the outgroup. N = northern cluster; S = southern cluster.

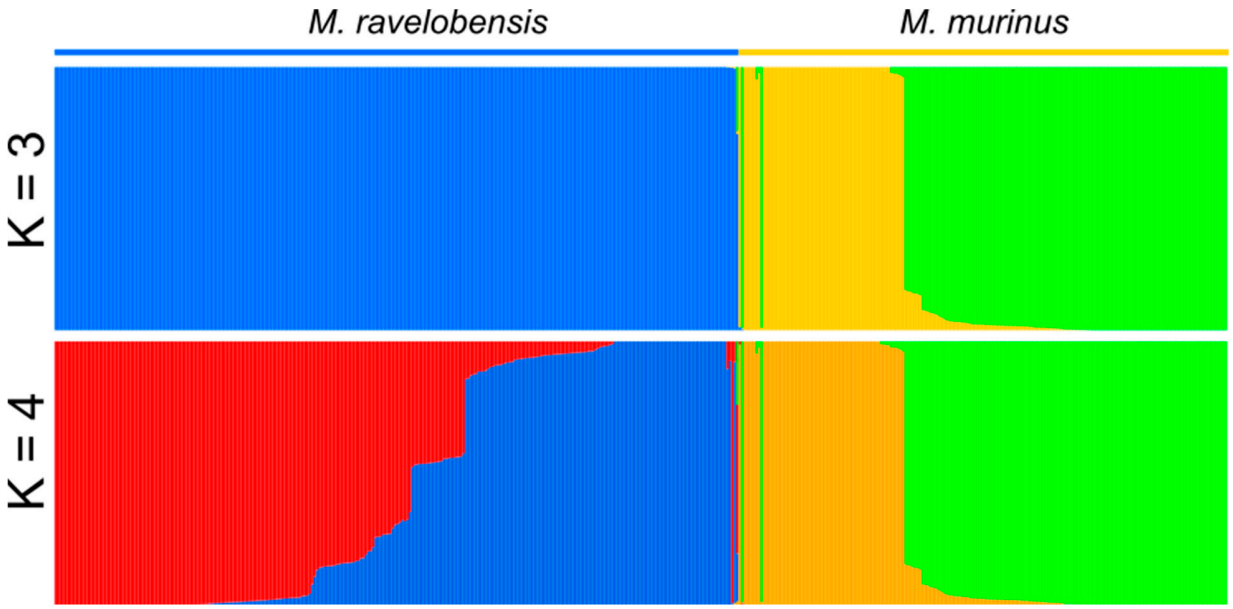

Figure S2. Clustering assignment of 480 mouse lemur individuals to three (K = 3) and four (K = 4) genetic clusters. Each vertical bar represents an individual and each colour a distinct genetic cluster. The analyses suggested a stronger population structure signature for *M. murinus* than for *M. ravelobensis*, when considering K = 3. For K = 4, the *M. murinus* and *M. ravelobensis* individuals sampled at the northern and southern sites were assigned to two geographic clusters with some levels of admixture between them. Samples are sorted according to sampling site, with the individuals with admixed ancestry

located in between the two species.

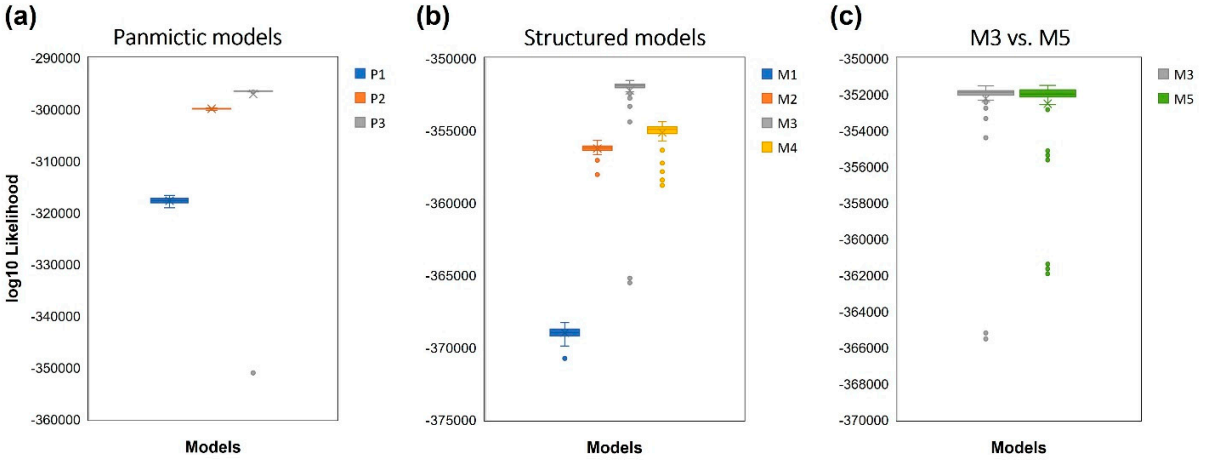

Figure S3. Boxplots showing the log10 likelihood from 100 expected SFS simulations under the parameters that maximize the likelihood of each model. **(a)** Assuming population panmixia, the changing model (P3) exhibited the highest log10 likelihood. **(b)** Considering population structure, the model with the highest log10 likelihood was the recent gene flow model (M3). **(c)** The model assuming that ancestral *M. murinus* and *M. ravelobensis* became structured at the same period (M3) or at different times (M5) yielded a similar log10 likelihood distribution suggesting a similar model fit. P1 = null model (panmixia); P2 = gene flow model; P3 = changing model; M1 = null model (structure); M2 = ancient gene flow model; M3 = recent gene flow model; M4 = ancient and recent gene flow model; M5 = recent gene flow & asymmetric structure.

Table S1. Metadata file containing information about all individuals considered in the present study and their respective sampling locations (n = 480). This file also includes a detailed list of the number of raw reads obtained from RADseq sequencing for each individual, the number of reads that passed all quality filters and were used for the genomic analyses, and the mean depth for each individual **(see Additional Excel file**).

Table S2. List of all demographic parameters used in each model during the *fastsimcoal2* analyses, and their respective search ranges assuming population panmixia (P1 – P3). All population size parameters are given in haploid numbers and time estimates in number of generations. NPOP0 = effective population size for *M. murinus* at present time; NPOP1 = effective population size for *M. ravelobensis* at present time; 2Nm = average number of haploid immigrants entering the population per generation. For the changing model, 2Nm0 denotes recent gene flow between the two species and 2Nm1 ancient gene flow. T1 = time when gene flow rate changed; T2 = time to the most recent common ancestor of *M. murinus* and *M. ravelobensis*.

				Search Range		
Parameter Name	**Model (P)**	**Value Type**	**Distribution Type**	**Minimum**	**Maximum**	**Bounded**
NPOP0	1, 2	Integer	Uniform	50	1.20E+05	no
NPOP1	1, 2	Integer	Uniform	50	1.20E+05	no
T1	3	Integer	Uniform	1	1.00E+04	no
T2	1, 2	Integer	Uniform	1	4.00E+06	no
T2	3	Integer	Uniform	1.00E+04	4.00E+06	no
2Nm0	2, 3	Float	Log-Uniform	0.0001	20	yes
2Nm1	3	Float	Log-Uniform	0.0001	20	yes

Table S3. List of all demographic parameters used in each model during the *fastsimcoal2* analyses and their respective search ranges assuming population structure (M1 – M4). All population size parameters are given in haploid numbers and time estimates in number of generations. NPOP0 = effective population size for *M. murinus* southern cluster at present time; NPOP1 = effective population size for *M. murinus* northern cluster at present time; NPOP2 = effective population size for *M. ravelobensis* southern cluster at present time; NPOP3 = effective population size for *M. ravelobensis* northern cluster at present time; 2Nm0 = average number of haploid immigrants entering the population per generation. T1 = time when *M. murinus* and *M. ravelobensis* ancestral populations became structured into a northern and southern cluster, respectively; T2 = time to the most recent common ancestor of *M. murinus* and *M. ravelobensis*.

				Search Range		
Parameter Name	**Model (M)**	**Value Type**	**Distribution Type**	**Minimum**	**Maximum**	**Bounded**
NPOP0	1, 2, 3, 4	Integer	Uniform	50	1.20E+05	no
NPOP1	1, 2, 3, 4	Integer	Uniform	50	1.20E+05	no
NPOP2	1, 2, 3, 4	Integer	Uniform	50	1.20E+05	no
NPOP3	1, 2, 3, 4	Integer	Uniform	50	1.20E+05	no
T1	1, 2, 3, 4	Integer	Uniform	1	1.20E+04	no
T2	1, 2, 3, 4	Integer	Uniform	1.20E+04	4.00E+06	no
2Nm0	2, 3, 4	Float	Log-Uniform	0.0001	20	yes

Table S4. Results of the four introgression tests conducted on filtered genotype calls in Dsuite Dtrios (see Additional Figure S1). Shown are Patterson's *D*, the results of the significance test via block-jackknifing (*Z* score and *p* value), and the amount of BBAA, ABBA and BABA sites given the topology (((P1, P2), P3), O). *Mirza zaza* was used as the outgroup in all the introgression tests. N = northern cluster; S = southern cluster.

Panel in Fig. A1	P1	P2	P3	Patterson's *D*	*Z* score	*P* value	BBAA sites	ABBA sites	BABA sites
A	*M. murinus* S	*M. murinus* N	*M. ravelobensis* S	0.0139	1.3514	0.0883	55205.7	876.21	852.24
B	*M. ravelobensis* N	*M. ravelobensis* S	*M. murinus* S	0.028	5.7071	0	59485.6	943.8	892.36
C	*M. murinus* S	*M. murinus* N	*M. ravelobensis* N	0.0144	1.3932	0.0818	55255.1	874.97	850.21
D	*M. ravelobensis* N	*M. ravelobensis* S	*M. murinus* N	0.0275	5.7645	0	59463	946.08	895.46

Table S5. Demographic parameter estimates that maximized the likelihood for each demographic model after 100 independent simulations per model, assuming population panmixia (P1 – P3). All population size parameters are given in haploid numbers and time estimates in number of generations. NPOP0 = effective population size for *M. murinus* at present time; NPOP1 = effective population size for *M. ravelobensis* at present time; 2Nm = average number of haploid immigrants entering the population per generation. For the changing model, 2Nm0 denotes recent gene flow between the two species and 2Nm1 ancient gene flow. T1 = time when gene flow rate changed; T2 = time to the most recent common ancestor of *M. murinus* and *M. ravelobensis*.

Model	Topology	NPOP0	NPOP1	T1	T2	2Nm0	2Nm1
P1	Null model	833,895	1,179,793	_	3,067,200	_	_
P2	Gene flow	694,299	1,228,675	_	7,813,479	0.022	_
P3	Changing	750,240	1,264,245	7467	6,602,411	0.432	0.003

Table S6. Demographic parameter estimates that maximized the likelihood for each demographic model after 100 independent simulations per model, assuming population structure (M1 – M4). All population size parameters are given in haploid numbers and time estimates in number of generations. NPOP0 = effective population size for *M. murinus* southern cluster at present time; NPOP1 = effective population size for *M. murinus* northern cluster at present time; NPOP2 = effective population size for *M. ravelobensis* southern cluster at present time; NPOP3 = effective population size for *M. ravelobensis* northern cluster at present time; 2Nm0 = average number of haploid immigrants entering the population per generation. T1 = time when *M. murinus* and *M. ravelobensis* ancestral populations became structured into the northern and southern clusters, respectively; T2 = time to the most recent common ancestor of *M. murinus* and *M. ravelobensis*.

Model	Topology	NPOP0	NPOP1	NPOP2	NPOP3	T1	T2	2Nm0
M1	Null model	736,626	196,699	1,215,156	308,342	31,163	5,051,674	_

M2	Ancient gene flow	660,986	137,494	1,194,398	240,679	21,739	8,498,021	0.025
M3	Recent gene flow	514,230	637,726	1,199,363	841,257	56,915	6,387,850	0.063
M4	Ancient & recent gene flow	551,276	329,651	1,152,654	475,004	43,340	7,945,249	0.017

Table S7. Demographic parameter estimates that maximized the likelihood of model M5 (recent gene flow & asymmetric structure) after 100 independent simulations. All population size parameters are given in haploid numbers and time estimates in number of generations. NPOP0 = effective population size for *M. murinus* southern cluster at present time; NPOP1 = effective population size for *M. murinus* northern cluster at present time; NPOP2 = effective population size for *M. ravelobensis* southern cluster at present time; NPOP3 = effective population size for *M. ravelobensis* northern cluster at present time; T1 = time when *M. murinus* ancestral population became structured into a northern and southern cluster; T2 = time when *M. ravelobensis* ancestral population became structured into a northern and southern cluster; T3 = time to the most recent common ancestor of *M. murinus* and *M. ravelobensis*; 2Nm0 = average number of haploid immigrants entering the population per generation. This model represented a similar model fit to the recent structured model (M3; see Table 1 and Additional Figure A3c).

Model	NPOP0	NPOP1	NPOP2	NPOP3	T1	T2	T3	2Nm0
M5	483103	626850	1247119	780048	58193	44493	6466667	0.072

Table S8. Demographic parameters inferred under the best demographic model (M3). Maximum-likelihood (ML) estimates were obtained from the run with the highest composite likelihood. The 95% confidence intervals (CI) were generated from 100 bootstrapped data sets. Population size estimates are given in number of diploid copies. Time changes were scaled considering GT = 2.5 years. See Figure 4 for a schematic illustration of M3. In bold: ML estimates that lie within the 95% CI. Note that we are mainly interested in dating the occurrence of gene flow and the T1 estimate falls inside the 95% confidence interval range.

		95% CI	
Parameter	**ML estimate**	**Lower bound**	**Upper bound**
NPOP0	257,115	241,123	244,269
NPOP1	**318,863**	313,120	319,878
NPOP2	599,682	607,629	611,606
NPOP3	420,629	398,102	404,558
T1	**142,288**	141,293	143,558
T2	15,969,625	16,437,440	16,577,038
2Nm0	**0.06**	0.06	0.06