MOLECULAR ECOLOGY

Supplemental Information for:

Little genetic structure in a Bornean endemic small mammal across a steep ecological gradient

Lillian D. Parker†, Melissa T.R. Hawkins†, Miguel Camacho-Sanchez, Michael G. Campana, Jacob A. West-Roberts, Tammy R. Wilbert, Haw Chuan Lim, Larry L. Rockwood, Jennifer A. Leonard & Jesús E. Maldonado

†these authors should be considered joint first author

Table of Contents:

MOLECULAR ECOLOGY

13
14

MOLECULAR ECOLOGY

I. Methods and Results

Mitogenome assembly from UCE-enriched library sequences

In addition to assembling mitogenomes from amplicon sequence data (methods described in main text), we also assembled mitogenomes from UCE-enriched library sequence reads. We used the Geneious assembler in Geneious v.9.1.2 to map reads that were trimmed with the illumiprocessor program of the Phyluce pipeline v.1.5.0 (Faircloth, 2016) to the appropriate species-specific reference. We used the Medium-Low Sensitivity option and up to 5 iterations for fine-tuning. In iterative assembly, Geneious calls a consensus sequence from the initial mapping results and repeats mapping of reads to the consensus. We then used Geneious to generate a consensus sequence (lowest coverage to call a base 5×, and Highest Total Quality parameters). To compare mitogenome consensus sequences generated by the two methods (UCE-enriched library sequence vs. amplicon sequence data), we aligned sequences with the MAFFT v1.4.0 plugin (Katoh, Misawa, Kuma, & Miyata, 2002) in Geneious. We did not find differences in the mitogenome sequences generated by the two different methods for any individual; however, in a few cases we were able to resolve missing nucleotides in the amplicon-generated mitogenomes. In cases where we did not have amplicon data, we used the UCE-enrichment-derived mitogenome for further analysis.

Phylogenetic analysis

We aligned mountain treeshrew mitogenome sequences with the MAFFT v1.4.0 plugin (Katoh et al., 2002) in Geneious v7.17 (Biomatters Ltd.), including the large treeshrew (*Tupaia tana*), the pygmy treeshrew (*T. minor*), and the ruddy treeshrew (*T. splendidula*) as outgroups. We then used PartitionFinder v2.0 (Lanfear, Calcott, Ho, & Guindon, 2012) to select substitution models and partitions. We designated three user-defined schemes with the alignment separated by gene, by codon position, and with the first and second positions separate from the third in each codon. Genes were identified according to annotated CDS from the northern treeshrew (*T. belangeri*) reference (Genbank accession NC_002521, Schmitz, Ohme, & Zischler, 2000). The final alignment included 92 sequences and was 14,956 bp long, including all coding and non-coding genes and the first 789 bp of D-loop. All tRNAs were removed. The optimal scheme according to the AICc score included 29 partitions, separating the first and second codon position from the third for all coding genes. Substitution models were either HKY or GTR as listed in Table S5. Using the appropriate partitions, we ran RaxML v7.2.8 using the rapid bootstrapping algorithm while searching for the best tree under maximum likelihood (Stamatakis, 2006). We also performed two replicate runs of MrBayes v3.2.6 (Ronquist & Huelsenbeck, 2003) for 1 million generations each sampled every 100 steps across four chains with a temperature of 0.2. MrBayes runs were assessed for convergence and adequate effective sample size (ESS > 200) for all parameters in Tracer v 1.6.0. Consensus trees were compared from each run (sump and sumt commands) for topological congruence. Finally, we calculated the average distance between the two mitochondrial lineages we observed using the Maximum Composite Likelihood model in MEGA7 (Kumar, Stecher, & Tamura, 2016). Rate variation

among sites was modeled with a gamma distribution (shape parameter = 0.7).

Divergence Dating

We used BEAST v1.8.4 (Drummond & Rambaut, 2007) to estimate the date of divergence between the mountain treeshrew mitochondrial lineages identified in our phylogenetic analysis of mitogenomes (see above). We rooted the tree with the ruddy treeshrew, which we confirmed as the sister species to the mountain treeshrew (Roberts, Lanier, Sargis, & Olson, 2011). All other more distantly related treeshrew species were removed in the divergence dating, reducing the number of taxa from 92 to 84 individuals. Both coding and non-coding genes were included, but tRNAs and the first 789 bp of D-loop were included. No codon partitioning was used due to complexity of the model, and the HKY substitution model was selected over GTR due to difficulty in reaching stationarity with GTR. We used a strict clock and a Yule speciation tree prior. We used a secondary calibration point, capping the root height at the estimated date of divergence between the mountain treeshrew and the ruddy treeshrew which was estimated by Roberts et al. (2011). We used a lognormal distribution prior, with a median of 4.5 million years ago (Mya), and 95% quantile of 7.44 Mya. Operators were left at the default classic operator mix. Three hundred million chains were executed across three independent runs. Parameters were logged every 1,000 chains. An empty alignment was run to test if the priors were driving the posteriors for 300 million chains. Run replicates were evaluated for convergence and adequate ESS in Tracer v1.6.0, and log and tree files were combined in LogCombiner v1.8.4. TreeAnnotator v1.8.4 was used to generate maximum clade credibility trees, which were drawn in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/).

Bayesian skyline plot

We used BEAST v2.0 (Bouckaert et al., 2014) to perform a Bayesian coalescent skyline plot analysis. First, we partitioned the alignment of 34 unique mountain treeshrew haplotypes by gene and selected the partitioning scheme and mutation models based on corrected Akaike Information Criterion (AICc) scores using PartitionFinder v2.0 (Lanfear et al., 2012). We applied a strict molecular clock with linked trees and unlinked site and clock models. We used a time to most recent common ancestor (TMRCA) prior of 450,000 years before present (lognormal distribution, $\mu = 0.45$, $\sigma = 0.2$), the estimated date of divergence between the two mitochondrial lineages as determined by the dating analysis performed in BEAST. We ran the MCMC chain for 10 million generations, with a burn-in of one million steps. We checked for convergence and adequate ESS values in Tracer, and plotted the results using R v3.4.1 (R Core Team 2017). The resulting plot showed a flat line, indicating no evidence of a recent expansion or bottleneck. The 95% highest posterior density of the change parameter include zero, meaning we cannot rule out the possibility of zero changes in effective population size in the past 60,000 years.

We performed the analysis as described above for all haplotypes in the 'unrelated dataset' combined ($n = 34$) as well as for only haplotypes in haplogroup 1 ($n = 25$), and results were consistent for both analyses. We also performed the analysis for haplogroup 2, but it contained too few haplotypes ($n = 9$) and the run did not converge.

vcf2aln

vcf2aln is a command-line script that converts a multi-sample, all-site variant-call format (VCF) file to a multiple sequence alignment (MSA) in FASTA format. It can process both previously generated VCFs and uncompressed VCF streams in a pipeline. For each sample in the VCF, vcf2aln simultaneously applies simple indels and SNP calls to the reference base to generate a final consensus sequence. By generating the consensus sequences simultaneously, sequence indexing is maintained so that a separate re-alignment step of individually generated consensus sequences (such as those generated using SAMtools: Li et al., 2009) is obviated. Unlike previously existing tools (e.g. Genome Analysis Toolkit FastaAlternateReferenceMaker: McKenna et al., 2010), vcf2aln does not infer the reference variant in individual-specific regions of zero coverage. Instead, vcf2aln treats these regions as missing data. Missing data are represented by a '?' to distinguish missing data from unresolved, but sequenced, bases (Ns). Additionally, users can apply a variety of variant filters (quality scores, allelic depth, genotype likelihoods, etc.), site filters (depth, missingness, FILTER annotations), and sample filters (depth, missingness) to ensure data quality. Individual filtered variants are replaced by missing data, while filtered sites and samples are removed from the final alignment.

vcf2aln can generate alignments from both haploid and diploid data. In the case of diploid data, it will generate (pseudo)haplotypes or sequences using ambiguity codes depending on user settings and the phasing of the data. vcf2aln utilizes phasing information where present to generate haplotypes, but will randomly select an allele where phasing is unresolved. By default, all alleles are equally likely to be randomly selected. Optionally, the probabilities of pseudo-haplotype allele selection can be weighted by individual allele sequencing depths to account for high-error or low-coverage data (e.g. ancient DNA sequences: Kuhn, Manuel, Jakobsson, & Günther, 2018). For SNP data, it can also output ambiguity codes at these unresolved sites rather than generate pseudo-haplotypes. Depending on user preferences, vcf2aln can also output individual contig/chromosome-level alignments or concatenated alignments and omit sites not present in the VCF.

vcf2aln is written in Ruby (version 2.0 or greater: Matsumoto, 2013) and is compatible with most UNIX and UNIX-like operating systems. vcf2aln has no additional dependencies beyond the Ruby standard library. The program is available under the Smithsonian Institution's terms of use (http://www.si.edu/termsofuse) at GitHub (https://github.com/campanam/vcf2aln).

Spatial PCA

We performed a spatial PCA (sPCA) analysis with the *sPCA* function in the R package Adegenet v2.1.1 (Jombart, 2008). We used the Delaunay triangulation method to calculate a connection network between sampling locations. Based on the variance and autocorrelation of eigenvalues visualized through a screeplot (Figure S4b), we retained eigenvalue 1 which is well distinguished from the rest of the eigenvalues. We then interpolated principal components using the *interp* function from the R package Akima v2.1 (Akima, 1978) and plotted them to visualize the genetic cline (Figure S4a).

Effective population size

We explored the effect of including close relatives and subdividing population clusters on the estimation of N_e with the linkage disequilibrium (LD) method implemented in NeEstimator v2 (Do et al., 2014). We found evidence of the Wahlund effect (Wahlund, 1928), i.e., combining population clusters led to a decrease in the estimated effective population size. Including relatives in the dataset also resulted in lower N_e estimates. With all individuals (n = 80), N_e = 38 (95% CI 37–38); after removing first-degree relatives (n = 58), N_e = 96 (95% CI 92–100). Using only unrelated individuals and dividing them into two populations as assigned by the STRUCTURE analysis, removing individuals that could not be assigned to a cluster ($q_k < 0.6$), the N_e of MK + low elevation MT (n = 36) is 180 (95% CI 160–205), and high elevation MT (n = 18) is 57 (95% CI 52–63) (Table 2b). Assigning individuals to 3 populations as defined by the STRUCTURE output, again removing unassigned individuals, N_e of MK (n = 22) is 125 (95% CI 105–152), low elevation MT (n = 19) is 202 (95% CI 157–282), and high elevation MT (n = 11) is 48 (95% CI 40–59) (Table 2a). Because the LD method assumes discrete, non-overlapping generations, an assumption that we know is violated in mountain treeshrews, N_e estimates should be considered as the number of breeding individuals that gave rise to the cohort from which our samples were taken. The LD method also assumes that only genetic drift is responsible for the signal in the data (Waples & Do, 2010). It is possible that the strong purifying selection that UCE loci are subject to resulted in nonrandom association of UCE-associated SNPs and artificially decreased our N_e estimates. However, the selective forces are unlikely to vary across population clusters; therefore, our estimates are most useful to compare the effective population sizes across clusters.

Estimated effective migration surface modeling

We used estimated effective migration surface modeling (EEMS) to calculate and visualize the decay of genetic similarity across Kinabalu National Park (KNP) (Petkova, Novembre & Stephens, 2016). EEMS identifies areas where pairwise genetic distance decreases more than expected due to isolation-by-distance, revealing regions or landscape features that may represent historical barriers to migration. We estimated the migration surface using KNP boundaries for the habitat polygon. We used the *bed2diffs* script included in the EEMS package to calculate a matrix of pairwise genetic distances. We ran the analysis for several different deme sizes (200, 300, 400, and 500). For each deme size, we ran three independent chains for 20 million MCMC iterations with a burn-in of 10 million steps. Convergence was assessed, runs were combined, and data was visualized by running the Reemsplots R package (R Core Team 2017).

We found that the peaks of both MT and MK show significantly faster decay in pairwise genetic distances than average in Kinabalu Park (Fig S3a). This could be because the steeper slopes closer to the peaks result in shorter dispersal distances; however, this pattern could also be due to the 2-dimensional projection of our samples which underestimates the true straight-line distance between points due to elevational changes. The projection would have the effect of

187 making it appear as if pairwise genetic distances on the peaks increase faster than expected due
188 to distance alone. We also find higher-than-average genetic diversity among high elevation MT
189 individuals (>2000 meters above sea level, masl), and lower-than-average genetic diversity
190 among high elevation MK individuals (≥1600 masl) (Figure S3b).
191

MOLECULAR ECOLOGY

II. Tables

Table S1. Individuals sequenced in this study. EBD, Estación Biológica de Doñana, Seville, Spain

Specimen Field ID (BOR#)*	Specimen Accession Number	Transect Elevation (meters above sea level)	Mountain	Taxon	Marker Sequenced (UCEs, Mitogenome, or Both)	Collected Material
010	EBD30330M	500	Tambuyukon	*T. tana*	Mitogenome	Voucher
016	Pending**	500	Tambuyukon	*T. tana*	Mitogenome	Voucher
038	Pending	500	Tambuyukon	*T. tana*	Mitogenome	Voucher
050	EBD30336M	900	Tambuyukon	*T. tana*	Mitogenome	Voucher
056	EBD30341M	900	Tambuyukon	*T. tana*	Mitogenome	Voucher
059	Pending	900	Tambuyukon	*T. montana*	Both	Voucher
060	Pending	900	Tambuyukon	*T. montana*	Both	Voucher
062	Pending	900	Tambuyukon	*T. montana*	Both	Voucher
063	Pending	900	Tambuyukon	*T. montana*	Both	Voucher
066	Pending	1600	Tambuyukon	*T. montana*	Both	Voucher
067	Pending	1600	Tambuyukon	*T. montana*	Both	Voucher
068	Pending	1400	Tambuyukon	*T. montana*	UCEs	Voucher
072	EBD30344M	1300	Tambuyukon	*T. montana*	Both	Voucher
076	NA	1300	Tambuyukon	*T. montana*	Both	Ear Punch
081	EBD31352M	1300	Tambuyukon	*T. montana*	Both	Voucher
082	EBD31353M	1300	Tambuyukon	*T. montana*	Both	Voucher
083	EBD31354M	1300	Tambuyukon	*T. montana*	Mitogenome	Voucher
084	EBD31355M	1300	Tambuyukon	*T. montana*	Both	Voucher
087	Pending	1600	Tambuyukon	*T. montana*	Mitogenome	Voucher
088	Pending	1600	Tambuyukon	*T. montana*	Both	Voucher
090	Pending	1600	Tambuyukon	*T. montana*	Both	Voucher
096	EBD30348M	1600	Tambuyukon	*T. montana*	Mitogenome	Voucher
136	EBD31357M	2000	Tambuyukon	*T. montana*	Both	Voucher
137	EBD30351M	2000	Tambuyukon	*T. montana*	Both	Voucher
138	EBD31358M	2000	Tambuyukon	*T. montana*	Both	Voucher
139	EBD30352M	2000	Tambuyukon	*T. montana*	Both	Voucher
146	EBD31359M	2000	Tambuyukon	*T. montana*	Both	Voucher
147	EBD31360M	2000	Tambuyukon	*T. montana*	Both	Voucher
149	NA	2000	Tambuyukon	*T. montana*	Both	Ear Punch
202	EBD31361M	2400	Tambuyukon	*T. montana*	Both	Voucher
203	EBD31362M	2400	Tambuyukon	*T. montana*	Both	Voucher
204	EBD31363M	2400	Tambuyukon	*T. montana*	Both	Voucher
205	EBD31364M	2400	Tambuyukon	*T. montana*	Both	Voucher
206	EBD31365M	2400	Tambuyukon	*T. montana*	Both	Voucher
211	EBD31366M	2400	Tambuyukon	*T. montana*	Both	Voucher

Specimen Field ID (BOR#)	Specimen Accession Number	Transect Elevation (meters above sea level)	Mountain	Taxon	Marker Sequenced (UCEs, Mitogenome, or Both)	Collected Material
250	EBD31368M	1600	Kinabalu	*T. montana*	Both	Voucher
256	EBD31369M	1600	Kinabalu	*T. montana*	Both	Voucher
257	EBD31370M	1600	Kinabalu	*T. montana*	Both	Voucher
261	EBD31371M	1600	Kinabalu	*T. montana*	Both	Voucher
262	EBD31372M	1600	Kinabalu	*T. montana*	Both	Voucher
263	EBD31373M	1600	Kinabalu	*T. montana*	Both	Voucher
287	EBD31374M	2200	Kinabalu	*T. montana*	Both	Voucher
288	EBD31375M	2200	Kinabalu	*T. montana*	Both	Voucher
290	EBD31376M	2400	Kinabalu	*T. montana*	Both	Voucher
291	EBD31377M	2200	Kinabalu	*T. montana*	Both	Voucher
292	EBD31378M	2200	Kinabalu	*T. montana*	Both	Voucher
293	EBD31379M	2200	Kinabalu	*T. montana*	Both	Voucher
322	EBD31380M	2700	Kinabalu	*T. montana*	Both	Voucher
323	EBD31381M	2700	Kinabalu	*T. montana*	Both	Voucher
324	EBD31382M	2700	Kinabalu	*T. montana*	Both	Voucher
327	EBD31383M	2700	Kinabalu	*T. montana*	Both	Voucher
329	EBD31384M	2700	Kinabalu	*T. montana*	Both	Voucher
330	EBD31385M	2700	Kinabalu	*T. montana*	Both	Voucher
386	Pending	3200	Kinabalu	*T. montana*	Both	Voucher
387	EBD31386M	3200	Kinabalu	*T. montana*	Both	Voucher
389	EBD31387M	3200	Kinabalu	*T. montana*	Both	Voucher
390	EBD31388M	3200	Kinabalu	*T. montana*	Both	Voucher
395	EBD31389M	3200	Kinabalu	*T. montana*	Both	Voucher
396	EBD31390M	3200	Kinabalu	*T. montana*	Both	Voucher
423	Pending	500	Kinabalu	*T. minor*	Mitogenome	Voucher
429	EBD31391M	900	Kinabalu	*T. montana*	Both	Voucher
441	EBD31392M	900	Kinabalu	*T. montana*	Both	Voucher
443	Pending	500	Kinabalu	*T. minor*	Mitogenome	Voucher
450	EBD31393M	900	Kinabalu	*T. montana*	Both	Voucher
451	EBD31394M	900	Kinabalu	*T. montana*	Both	Voucher
461	EBD31395M	900	Kinabalu	*T. montana*	Both	Voucher
462	EBD31396M	900	Kinabalu	*T. montana*	Both	Voucher
501	EBD31397M	2000	Tambuyukon	*T. montana*	Both	Voucher
503	NA	2000	Tambuyukon	*T. montana*	Both	Ear Punch
506	EBD31398M	2000	Tambuyukon	*T. montana*	Both	Ear Punch
511	EBD31399M	2000	Tambuyukon	*T. montana*	Both	Ear Punch
513	EBD31400M	2000	Tambuyukon	*T. montana*	Both	Voucher
517	Pending	2000	Tambuyukon	*T. montana*	Both	Ear Punch
Specimen	**Specimen**	**Transect**	**Mountain**	**Taxon**	**Marker**	**Collected**

Field ID (BOR#)	Accession Number	Elevation (meters above sea level)			Sequenced (UCEs, Mitogenome, or Both)	Material
518	NA	2000	Tambuyukon	*T. montana*	Both	Ear Punch
521	NA	2000	Tambuyukon	*T. montana*	Both	Ear Punch
525	NA	2000	Tambuyukon	*T. montana*	Both	Ear Punch
527	EBD31401M	2400	Tambuyukon	*T. montana*	Both	Voucher
530	EBD31402M	2400	Tambuyukon	*T. montana*	Mitogenome	Voucher
536	EBD31403M	2400	Tambuyukon	*T. montana*	Both	Voucher
537	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
538	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
539	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
542	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
545	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
546	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
549	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
550	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
551	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
552	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
553	NA	2400	Tambuyukon	*T. montana*	Both	Ear Punch
555	EBD31404M	2400	Tambuyukon	*T. montana*	Both	Voucher
NA	UMMZ174429	NA	Mount Palung National Park	*T. splendidula*	Mitogenome	Voucher

*****Additional information for each BOR sample can be found in DatasetSF1 in Camacho-Sanchez, Hawkins, Tuh Yit Yu, Maldonado, & Leonard (2019).

******Voucher specimens with pending accession numbers will be accessioned in the Kinabalu Park Museum.

MOLECULAR ECOLOGY

Table S2. Mitochondrial genome partitions and models with the lowest AICc as determined by PartitionFinder.

Subset	Best Model	# sites	Partition
1	GTR+I+G+X	952	12S
2	GTR+G+X	1580	16S
3	GTR+I+X	638	ND1_pos1, pos2 ND1_pos2
4	HKY+I+X	319	ND1_pos3
5	HKY+G+X	684	ND2_pos1, pos2 ND2_pos2
6	HKY+X	342	ND2_pos3
7	TRN+I+X	1028	CO1_pos1, pos2 CO1_pos2
8	TRN+I+X	514	CO1_pos3
9	HKY+I+X	456	CO2_pos1, pos2 CO2_pos2
10	HKY+I+X	228	CO2_pos3
11	TRN+I+X	136	ATP8_pos1, pos2 ATP8_pos2
12	HKY+I+X	68	ATP8_pos3
13	HKY+I+X	425	ATP6_pos1, pos2 ATP6_pos2
14	HKY+X	212	ATP6_pos3
15	HKY+I	523	CO3_pos1, pos2 CO3_pos2
16	HKY+I+X	261	CO3_pos3
17	HKY+I+X	231	ND3_pos1, pos2 ND3_pos2
18	HKY+X	115	ND3_pos3
19	HKY+G+X	198	ND4L_pos1, pos2 ND4L_pos2
20	HKY+G+X	99	ND4L_pos3
21	GTR+I+X	914	ND4_pos1, pos2 ND4_pos2
22	HKY+I+G+X	456	ND4_pos3
23	GTR+I+G+X	1208	ND5_pos1, pos2 ND5_pos2
24	TRN+X	604	ND5_pos3
25	TRN+G+X	349	ND6_pos1 ND6_pos2
26	HKY+G+X	174	ND6_pos3
27	HKY+I	759	CB_pos1 CB_pos2
28	GTR+I+X	379	CB_pos3
29	HKY+I+G+X	800	DL

MOLECULAR ECOLOGY

205 **Table S3.** Mountain treeshrew mitogenome haplotypes per elevation.
206

Specimen Field ID (BOR#)	Transect Elevation (masl)	Mountain	Haplotype Number	Haplogroup
462	900	Kinabalu	Hap_5	1
441	900	Kinabalu	Hap_27	1
429	900	Kinabalu	Hap_30	2
450	900	Kinabalu	Hap_6	1
451	900	Kinabalu	Hap_34	1
461	900	Kinabalu	Hap_28	2
250	1600	Kinabalu	Hap_33	1
263	1600	Kinabalu	Hap_22	1
256	1600	Kinabalu	Hap_28	2
257	1600	Kinabalu	Hap_29	2
261	1600	Kinabalu	Hap_28	2
262	1600	Kinabalu	Hap_33	1
288	2200	Kinabalu	Hap_23	1
292	2200	Kinabalu	Hap_32	2
287	2200	Kinabalu	Hap_32	2
290	2200	Kinabalu	Hap_32	2
291	2200	Kinabalu	Hap_35	1
293	2200	Kinabalu	Hap_35	1
324	2700	Kinabalu	Hap_32	2
327	2700	Kinabalu	Hap_32	2
322	2700	Kinabalu	Hap_20	2
323	2700	Kinabalu	Hap_20	2
329	2700	Kinabalu	Hap_31	2
330	2700	Kinabalu	Hap_36	2
386	3200	Kinabalu	Hap_35	1
390	3200	Kinabalu	Hap_21	1
387	3200	Kinabalu	Hap_26	1
389	3200	Kinabalu	Hap_35	1
395	3200	Kinabalu	Hap_24	1
396	3200	Kinabalu	Hap_25	1
Specimen Field ID	**Transect Elevation**	**Mountain**	**Haplotype Number**	**Haplogroup**

(BOR#)	(masl)			
060	900	Tambuyukon	Hap_12	1
062	900	Tambuyukon	Hap_20	2
063	900	Tambuyukon	Hap_1	1
059	900	Tambuyukon	Hap_5	1
084	1300	Tambuyukon	Hap_3	1
072	1300	Tambuyukon	Hap_3	1
076	1300	Tambuyukon	Hap_3	1
081	1300	Tambuyukon	Hap_15	1
082	1300	Tambuyukon	Hap_17	1
083	1300	Tambuyukon	Hap_16	1
066	1600	Tambuyukon	Hap_2	1
096	1600	Tambuyukon	Hap_4	1
067	1600	Tambuyukon	Hap_11	1
087	1600	Tambuyukon	Hap_11	1
088	1600	Tambuyukon	Hap_10	1
090	1600	Tambuyukon	Hap_11	1
136	2000	Tambuyukon	Hap_13	1
147	2000	Tambuyukon	Hap_13	1
149	2000	Tambuyukon	Hap_5	1
137	2000	Tambuyukon	Hap_7	1
138	2000	Tambuyukon	Hap_5	1
139	2000	Tambuyukon	Hap_10	1
146	2000	Tambuyukon	Hap_14	1
506	2000	Tambuyukon	Hap_10	1
511	2000	Tambuyukon	Hap_13	1
513	2000	Tambuyukon	Hap_5	1
501	2000	Tambuyukon	Hap_5	1
503	2000	Tambuyukon	Hap_5	1
517	2000	Tambuyukon	Hap_19	2
518	2000	Tambuyukon	Hap_9	1
521	2000	Tambuyukon	Hap_9	1
525	2000	Tambuyukon	Hap_9	1
527	2400	Tambuyukon	Hap_9	1
Specimen Field ID	**Transect Elevation**	**Mountain**	**Haplotype Number**	**Haplogroup**

(BOR#)	(masl)			
203	2400	Tambuyukon	Hap_18	2
206	2400	Tambuyukon	Hap_18	2
211	2400	Tambuyukon	Hap_18	2
202	2400	Tambuyukon	Hap_18	2
204	2400	Tambuyukon	Hap_6	1
205	2400	Tambuyukon	Hap_19	2
530	2400	Tambuyukon	Hap_6	1
536	2400	Tambuyukon	Hap_10	1
555	2400	Tambuyukon	Hap_8	1
537	2400	Tambuyukon	Hap_9	1
538	2400	Tambuyukon	Hap_8	1
539	2400	Tambuyukon	Hap_9	1
542	2400	Tambuyukon	Hap_9	1
545	2400	Tambuyukon	Hap_8	1
546	2400	Tambuyukon	Hap_8	1
549	2400	Tambuyukon	Hap_9	1
550	2400	Tambuyukon	Hap_8	1
551	2400	Tambuyukon	Hap_9	1
552	2400	Tambuyukon	Hap_10	1
553	2400	Tambuyukon	Hap_9	1

207
208

MOLECULAR ECOLOGY

Table S4. STRUCTURE Harvester output **a.** Full dataset, *n* = 58 **b.** MK, *n* = 25 **c.** MT, *n* = 33

a.

K	Reps	MeanLnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-30542.24	1.5131	NA	NA	NA
2	10	**-29778.24**	0.8168	764	539.33	**660.321581**
3	10	**-29553.57**	3.491	224.67	312.24	**89.442533**
4	10	-29641.14	77.3951	-87.57	17.6	0.227405
5	10	-29746.31	33.3188	-105.17	70.77	2.124027
6	10	-29922.25	106.0196	-175.94	127.7	1.204494
7	10	-29970.49	191.0689	-48.24	NA	NA

b.

K	Reps	MeanLnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	**-12450.78**	1.7015	NA	NA	NA
2	10	-12513.04	21.459	-62.26	1385.53	64.566337
3	10	-13960.83	1957.4401	-1447.79	2531.28	1.293158
4	10	-12877.34	122.6713	1083.49	1235.28	10.069834
5	10	-13029.13	418.7107	-151.79	NA	NA

c.

K	Reps	MeanLnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-17401.11	1.0826	NA	NA	NA
2	10	**-17065.31**	2.5164	335.8	523.07	**207.867**
3	10	-17252.58	31.7557	-187.27	128.55	4.048098
4	10	-17311.3	76.32	-58.72	101.97	1.336085
5	10	-17268.05	16.256	43.25	NA	NA

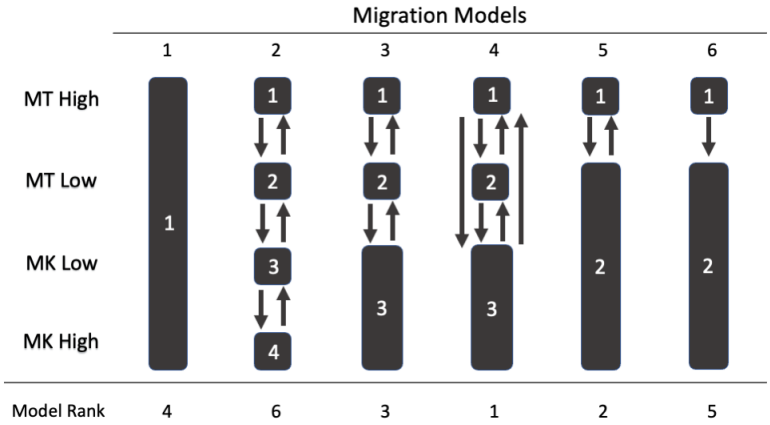
MOLECULAR ECOLOGY

Table S5. Log marginal likelihood values used to rank models evaluated using MIGRATE-N. Models are described in Methods and Figure 2.

Model	Parameters	Log(ML)	LBF	Rank
4	**9**	**-982795.38**	**0**	**1**
5	4	-984710.56	-1915.18	2
3	7	-985966.85	-3171.47	3
1	1	-986106.13	-3310.75	4
6	3	-986285.77	-3490.39	5
2	10	-988709.71	-5914.33	6

MOLECULAR ECOLOGY

Table S6. Parameter estimates from MIGRATE-N analysis. Figure shows population cluster labels for each model. Theta values are the mutation-scaled effective population sizes ($\Theta_i = 4N_e^{(i}\mu$) and M is the mutation-scaled migration rate ($M_i = m_i/\mu$).



Model	Parameter	Mean	2.5%	97.50%
1	Θ1	0.00403	0.0029	0.0052
2	Θ1	0.00191	0.0007	0.0030
2	Θ2	0.00178	0.0006	0.0029
2	Θ3	0.00203	0.0008	0.0032
2	Θ4	0.00229	0.0011	0.0035
2	M1->2	11377.2	11020.0	11720.0
2	M2->1	585.0	330.0	830.0
2	M2->3	11259.6	10880.0	11265.0
2	M3->2	9543.5	9140.0	9960.0
2	M3->4	11247.4	10930.0	11520.0
2	M4->3	9736.8	9190.0	10180.0
3	Θ1	0.00187	0.0007	0.003
3	Θ2	0.00174	0.0006	0.0029
3	Θ3	0.00277	0.0016	0.0039
3	M1->2	12165	11780.0	12540.0
3	M2->1	1926	1670.0	2180.0
3	M2->3	11387.1	11120.0	11395.0
3	M3->2	6770.8	6370.00	6775.00
4	Θ1	0.00138	0.0002	0.0025
4	Θ2	0.00159	0.0004	0.0028
4	Θ3	0.00182	0.0006	0.0029
4	M1->2	10569.9	10170.0	10575.0

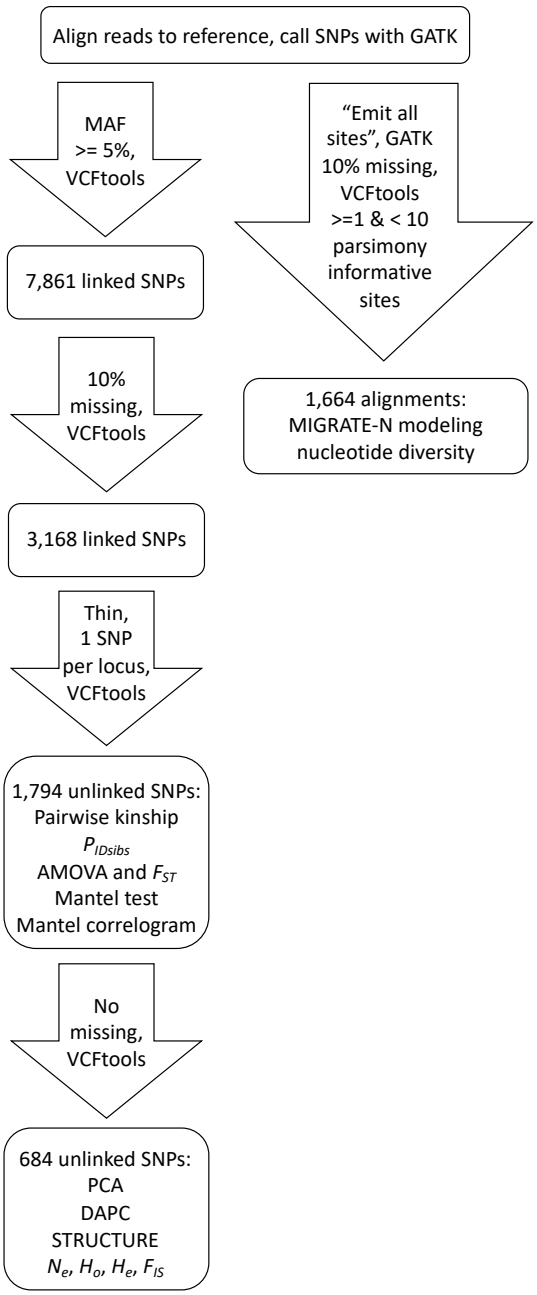
Model	Parameter	Mean	2.5%	97.50%
4	M2->1	8360.8	7840.0	8930.0
4	M1->3	10697.7	10380.0	10990.0
4	M3->1	6605.7	6080.0	6605.7
4	M2->3	10460.2	10090.0	10820.0
4	M3->2	8729.3	8340.0	9120.0
5	Θ1	0.002	0.0008	0.0031
5	Θ2	0.00259	0.0014	0.0037
5	M1->2	11252.9	10990.0	11500.0
5	M2->1	2639.2	2350.0	2920.0
6	Θ1	0.00302	0.0018	0.0042
6	Θ2	0.02304	0.0209	0.0250
6	M1->2	8469.2	8210.0	8720.0

227
228

MOLECULAR ECOLOGY

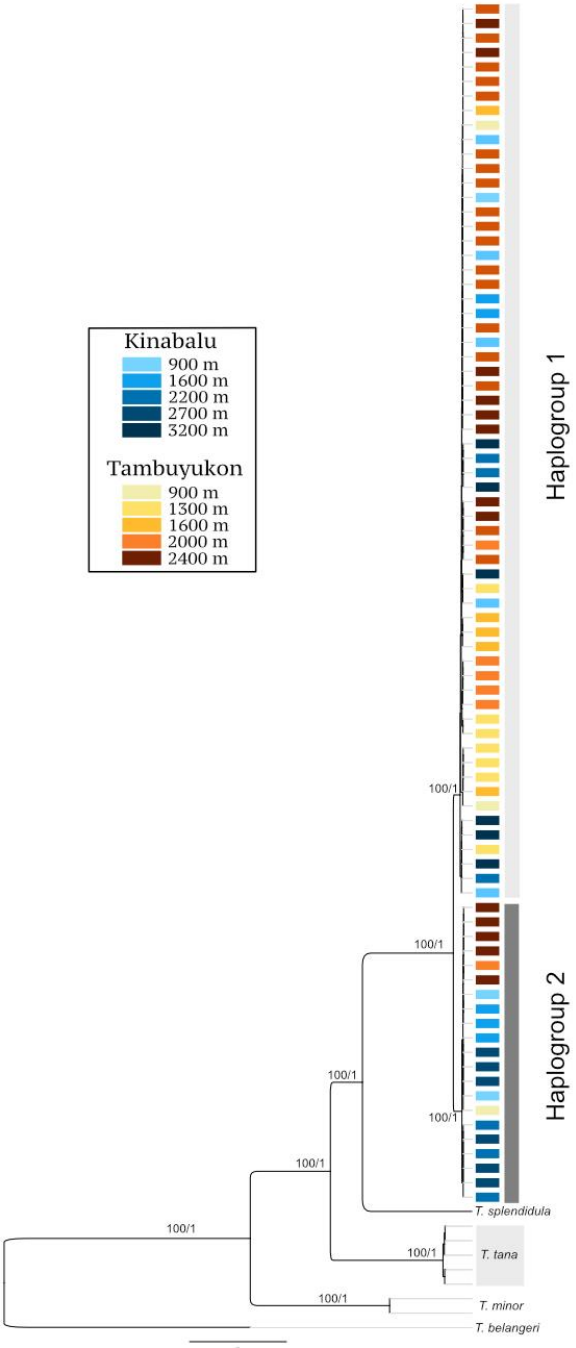
III. Figures

Figure S1. SNP filtering. Diagram of SNP filtering process and analyses performed on each dataset.



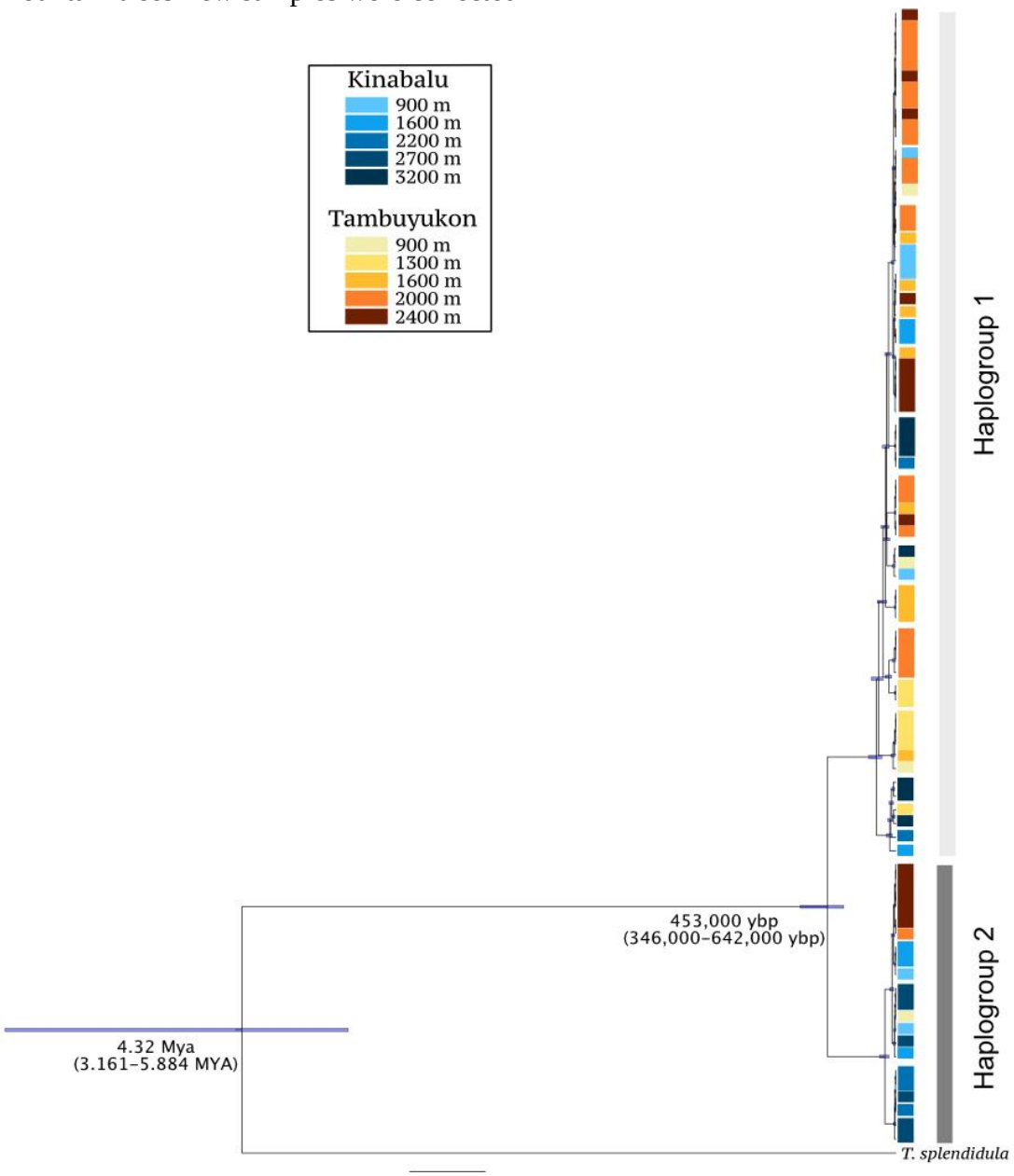
MOLECULAR ECOLOGY

Figure S2. Mitogenome phylogeny. Phylogenetic tree of all treeshrew mitogenomes sequenced in this study. Support values are from RAxML and MrBayes (bootstrap/posterior probability). Tree topology is concordant with previously identified relationships (Roberts et al., 2011). All nodes have strong support (100/1). Colored tips depict the elevation at which each mountain treeshrew was collected. The two major haplogroups are labelled.



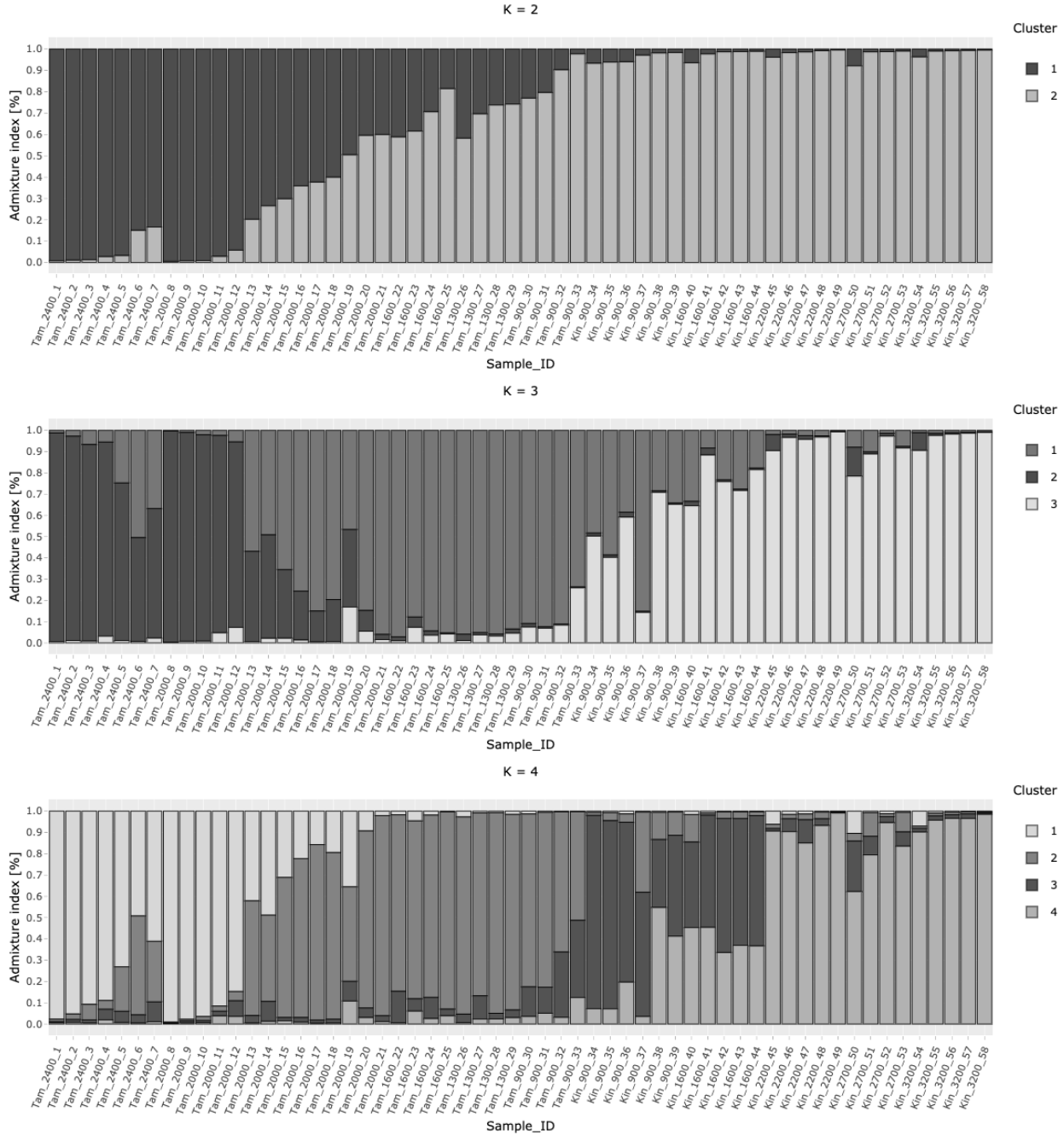
MOLECULAR ECOLOGY

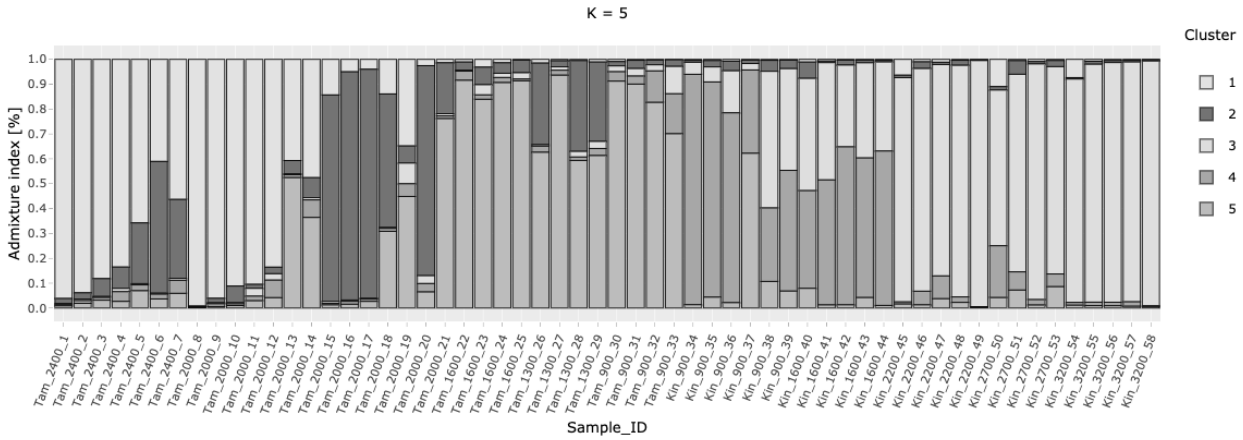
Figure S3. Dated mitogenome phylogeny. Dated phylogenetic tree from BEAST including data from the mountain treeshrew and its sister species, the ruddy treeshrew (*T. splendidula*). The date of divergence between the two species (Roberts et al., 2011) was used as a calibration point to estimate the divergence between the two mountain treeshrew haplogroups. The text below the nodes shows the mean and 95% highest posterior density in parentheses. All nodes are strongly supported (posterior probability=1). Colored tree tips correspond to the elevation at which mountain treeshrew samples were collected.

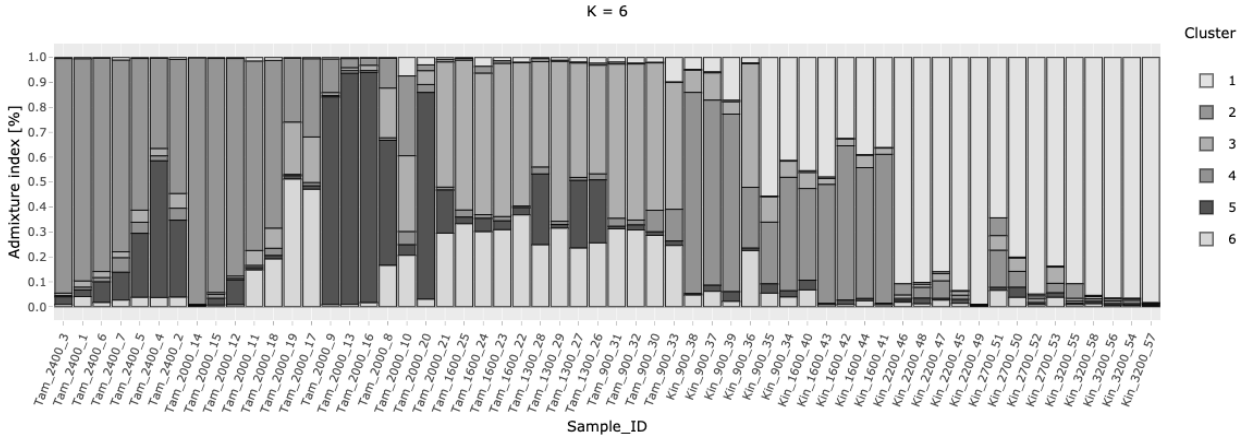


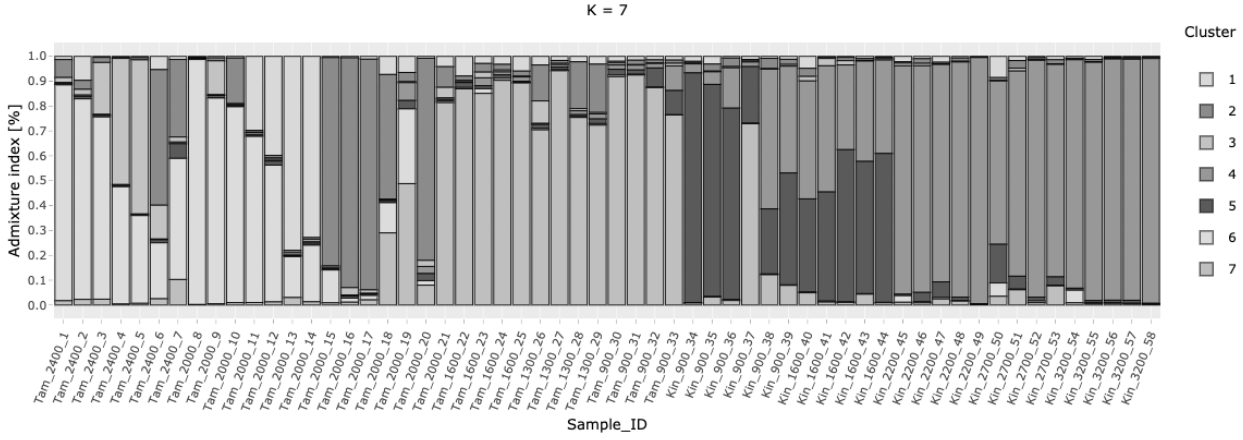
MOLECULAR ECOLOGY

Figure S4. STRUCTURE plots. Cluster membership assigned by STRUCTURE analyses for *K* = 2–7. Each vertical line represents a single individual, with shading indicating how much of each's ancestry can be attributed to each cluster. Individuals are arranged left to right from high elevation MT to low elevation MT followed by low elevation MK to high elevation MK. Using the Δ*K* method (Evanno, Regnaut, & Goutdet, 2005), *K* = 2 is the best fit to the data, but *K* = 3 has the highest likelihood.




K = 5
Admixture index [%]
Sample_ID
Cluster
1
2
3
4
5

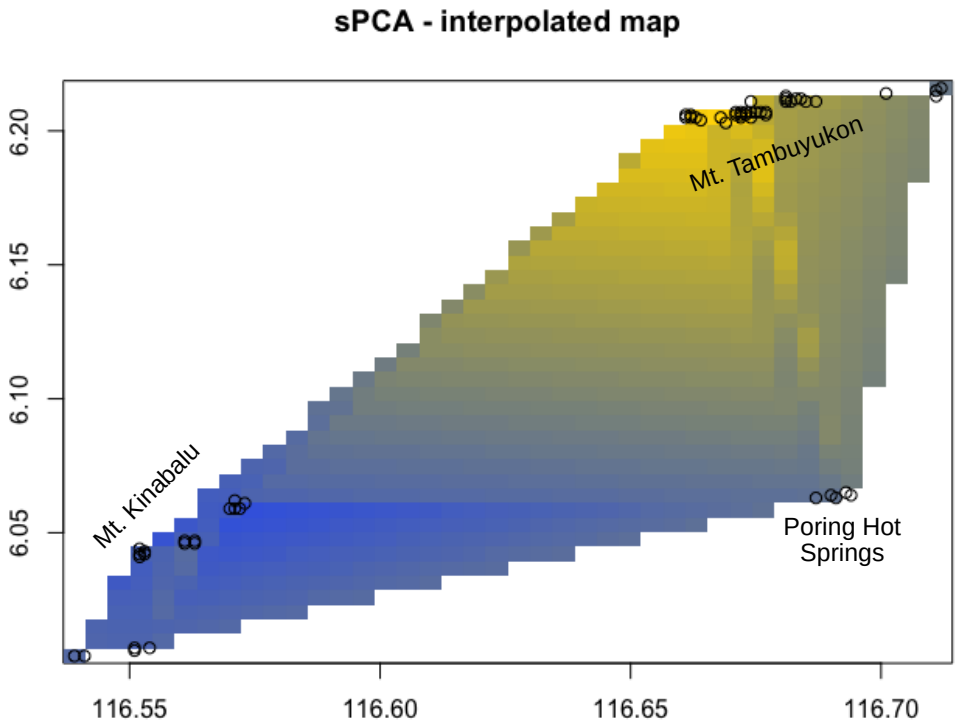

K = 6
Admixture index [%]
Sample_ID
Cluster
1
2
3
4
5
6


K = 7
Admixture index [%]
Sample_ID
Cluster
1
2
3
4
5
6
7

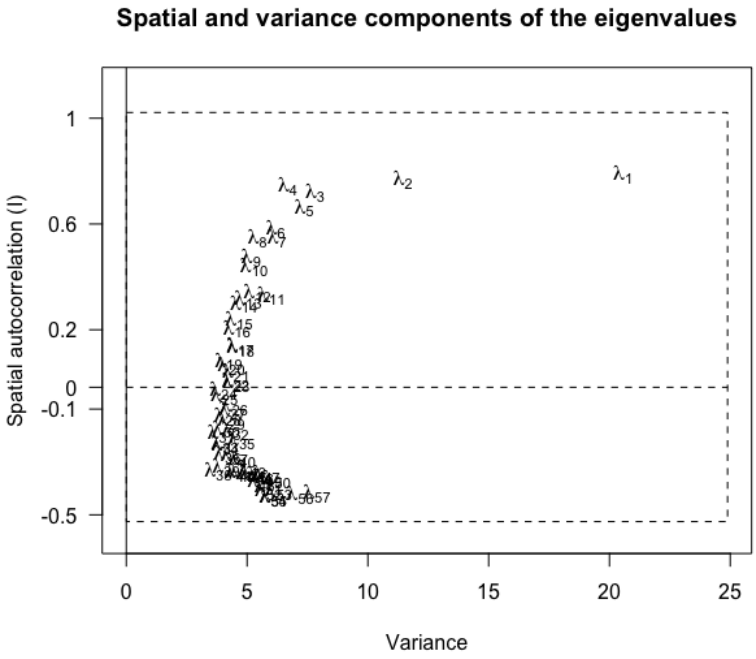
MOLECULAR ECOLOGY

Figure S5. Spatial PCA. a (top). Interpolated sPCA plot with individuals plotted as circles with X and Y coordinates corresponding to decimal longitude and latitude of collection localities, respectively. Poring Hot Springs, on the eastern slope of MK at 900 masl, is shown between the two peaks of MT and MK. Shading indicates interpolated principal components, with yellow representing negative values, blue positive values, and grey values in between. The greatest differentiation is between high elevation MT and MK, with intermediate individuals at low elevation MT and Poring Hot Springs. **b (bottom).** Screeplot of sPCA eigenvalues.

a)

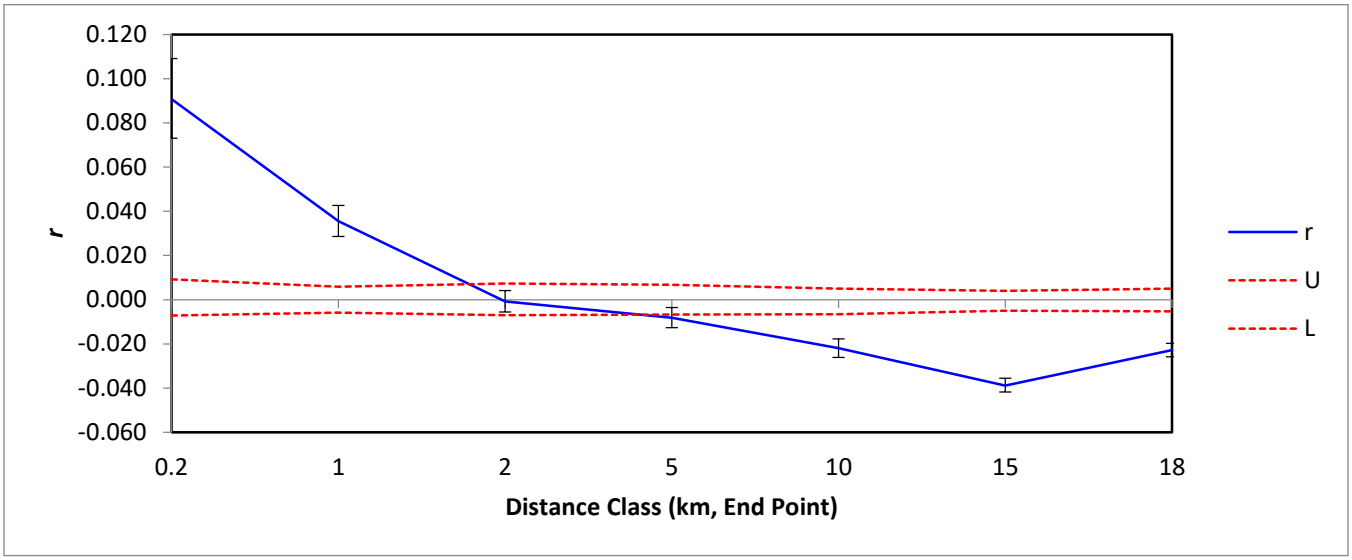


b)


Spatial and variance components of the eigenvalues
Spatial autocorrelation (I)
1
0.6
0.2
0
-0.1
-0.5
0
5
10
15
20
25
Variance
λ1
λ2
λ3
λ4
λ5

MOLECULAR ECOLOGY

Figure S5. Spatial correlogram. Mantel correlogram plotting the correlation coefficient (*r*) between genetic distance and geographic distance in 7 distance classes. There is significant, positive spatial autocorrelation between individuals caught within 1 km of each other (p = 0.0001); by 2 km apart, there is no significant autocorrelation. Dashed lines indicate the upper (U) and lower (L) bounds of the 95% confidence interval built from 1,000 random permutations; confidence intervals around the *r* value at each distance class are based on 10,000 bootstraps.

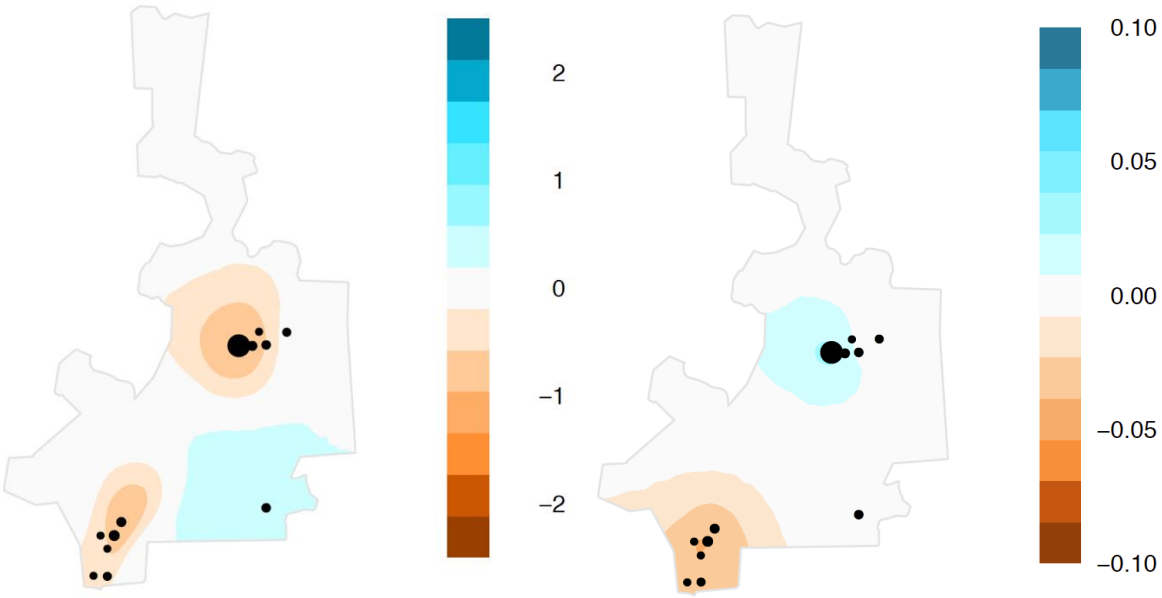


MOLECULAR ECOLOGY

Figure S6. a (left). Estimated effective migration surface modeling (EEMS) plot showing posterior mean migration rate deviation from average on a log scale. **b (right).** Posterior mean genetic diversity rates on a log scale. Outline corresponds to Kinabalu National Park boundary, and black circles are sampling localities. Circle area is proportional to the number of samples.

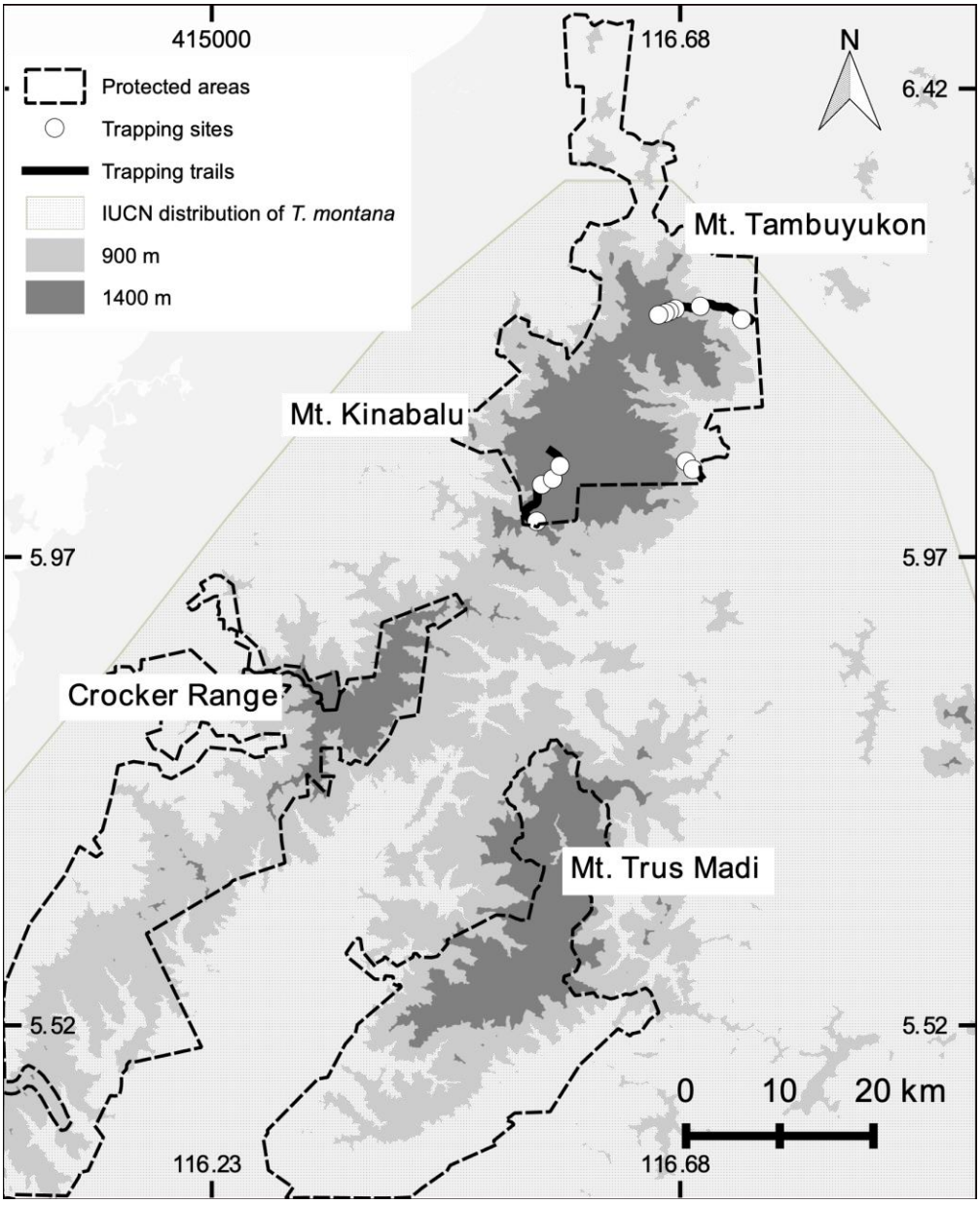
The EEMS analysis reveals two regions where the estimated migration rate is significantly lower than the average across the park (posterior probability > 95%). These areas correspond to the tops of both MK (lower left) and MT (upper right) (S5a, left). The EEMS plot of posterior mean genetic diversity rates (S5b, right) shows greater genetic diversity on high elevation MT (>2000 masl) than MK (≥1600 masl) .

a) **b)**



MOLECULAR ECOLOGY

Figure S7. Estimated mountain treeshrew habitat in the year 2100 CE assuming mild IPCC scenarios. Light grey shows elevations ≥ 900 masl, which is current mountain treeshrew habitat; dark grey indicates ≥ 1400 masl, which is potential mountain treeshrew habitat in 2100 assuming mild climate change as projected by the IPCC. Protected areas are demarcated by dashed lines. Transects sampled in this study are shown in black, with sampling locations indicated by white circles.



IV. References cited

Akima, H. (1978). A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. ACM Transactions on Mathematical Software, 4, 148-164.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., …, Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Camacho-Sanchez, M., & Hawkins, M. T. R., Tuh Yit Yu, F., Maldonado, J. E., & Leonard, J. A. (2019). Endemism and diversity of small mammals along two neighboring Bornean mountains. *PeerJ*, 7,e7858. https://doi.org/10.7717/peerj.7858

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evolutionary Biology*, 7(1), 214. https://doi.org/10.1186/1471-2148-7-214

Evanno, G., Regnaut, S., & Goutdet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5), 786–88. https://doi.org/10.1093/bioinformatics/btv646

Jombart, T. (2008). Adegenet: An R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kuhn, M., Manuel, J., Jakobsson, M., & Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PLOS ONE*, 13(4), e0195491 https://doi.org/10.1371/journal.pone.0195491

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29(6), 1695–1701. https://doi.org/10.1093/molbev/mss020

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., …, 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Matsumoto, Y. 2013. Ruby Programming Language. http://www.ruby-lang.org

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., …, DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing

next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48(1), 94–100. https://doi.org/10.1038/ng.3464

Roberts, T. E., Lanier, H. C., Sargis, E. J., & Olson, L. E. (2011). Molecular phylogeny of treeshrews (Mammalia: Scandentia) and the timescale of diversification in Southeast Asia. *Molecular Phylogenetics and Evolution*, 60(3), 358–372. https://doi.org/10.1016/j.ympev.2011.04.021

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Schmitz, J., Ohme, M., & Zischler, H. (2000). The complete mitochondrial genome of *Tupaia belangeri* and the phylogenetic affiliation of Scandentia to other Eutherian orders. *Molecular Biology and Evolution*, 17(9), 1334–1343. https://doi.org/10.1093/oxfordjournals.molbev.a026417

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), 2688–2690. doi: 10.1093/bioinformatics/btl446.

Wahlund, S. (2010). Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas*, 11(1), 65–106.

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, *3*(3), 244–262. https://doi.org/10.1111/j.1752-4571.2009.00104.x