**Electronic Supplement Material 1**

**The genetics of morphological and behavioural island traits in deer mice**

Felix Baier & Hopi E. Hoekstra

Howard Hughes Medical Institute, Department of Molecular & Cellular Biology, Department of Organismic & Evolutionary Biology, Museum of Comparative Zoology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138

email: hoekstra@oeb.harvard.edu

**Body weight of museum specimens.** To minimize confounding effects of reproductive status and age, we excluded females and specimens of unknown sex, juveniles, subadults, and embryos. We then removed specimens with a body length < 70 mm and/or a tail length < 60 mm, which represents minimum values in a sample of adult *P. maniculatus* we collected in British Columbia (see below). To remove specimens with out-dated taxonomy that no longer are classified as *P. maniculatus*, notably *P. keeni* [1], we filtered out records labelled as *P. keeni* or any of its synonyms or subspecies. We also excluded specimens with both a tail/body ratio of > 1.1 and a tail length > 95mm, which are morphological criteria to identify *P. keeni* [2].

**Admixture analysis and divergence dating.** Genomic DNA was extracted using an AutoGenPrep 965 (AutoGen), and the quality of extractions was verified with Quant-iT (Thermo Fisher Scientific). We digested genomic DNA samples with NlaIII and MluCI enzymes (New England Biolabs) and ligated fragments to biotinylated barcoded adapters. We size-selected for 216-276bp fragments using a PippinPrep (Sage Science), cleaned fragments with Streptavidin-coupled M-280 Dynabeads (Thermo Fisher Scientific), and PCR amplified fragments using 12 uniquely indexed PCR primers with Phusion DNA polymerase (New England Biolabs). After a bead clean-up and evaluation of the library quality on a 2200 TapeStation (Agilent Technologies), we generated 125bp paired-end reads with a standard v4 run on an Illumina HiSeq 2500 sequencer.

We used a custom-made pipeline to demultiplex reads and combine them by sample into one R1 and R2 file. We then used BWA 0.7.15 [3] to map reads to an in-house *de novo* assembly (v2.1) of the *P. maniculatus bairdii* (strain BW) reference genome. After calling variants for each sample separately with GATK 3.5 [4] using "HaplotypeCaller" (including options -minPruning 1, -minDanglingBranchLength 1, and -het 0.005), we produced one vcf file containing all samples with GATK's "GenotypeGVCFs" (including option het -0.005). After quality filtering (QualByDepth (QD) < 2; FisherStrand (FS) > 60 for SNPs, > 200 for indels; RMSMappingQuality (MQ) < 40; MappingQualityRankSumTest (MQRankSum) < 12.5; ReadPosRankSumTest (ReadPosRankSumTest) < -8 for SNPs, < -20 for indels; StrandOddsRatio (SOR) > 3 for SNPs, > 10 for indels), we used vcftools [5] to explore depth and missingness. We then removed samples of *P. keeni*, and filtered sites by minimum and maximum depth across individuals (-min-meanDP 1 and -max-meanDP 264).

To prepare the input file for the principal component analysis (PCA) and admixture analysis, we next filtered by minimum and maximum depth within individuals (-minDP 4 and -maxDP 25) and by missingness across individuals (-max-missing 0.6), and then removed samples with <25% of these sites. To account for the genotype uncertainty inherent in low-coverage data, we generated genotype probabilities using ANGSD 0.911 [6] with the -doPost 1 and -doGeno 32 options. This dataset contained 80,248 variants across 54 *P. maniculatus* samples. For the PCA, we used the ngsCovar function in ngsTools 0.615 [7] to compute a covariance matrix. For the admixture analysis, we generated a beagle file with the -doGlf function in ANGSD, and used the ngsAdmix function in ngsTools to calculate admixture proportions. We plotted PCs and admixture proportions following this tutorial: https://github.com/mfumagalli/ngsTools/blob/master/TUTORIAL.md.

To prepare the input file for the divergence dating analysis, we selected a subset of high coverage samples (3 Saturna Island *P. maniculatus*, 5 Pender Island *P. maniculatus*,3 mainland *P. maniculatus*, 3 *P. keeni*), and filtered by minimum and maximum depth within individuals (-minDP 5 and -maxDP 25) and missingness across individuals (-max-missing 0.8). This resulted in a dataset of 7,092 variants across 14 samples. We then converted the vcf file to nexus format with custom Python code, and used SNAPP 1.3.0 [8] to create a phylogenetic tree with divergence time estimates. Briefly, we loaded the nexus file into a SNAPP template in BEAUti 2.4.8 [9], assigned species IDs to samples, calculated mutation rates, logged every 500 trees, and used a chain length of 106. We ran 10 separate iterations of the resulting XML file in BEAST 2.4.8 [9], and combined log and tree files across iterations in LogCombiner with a by-iteration burn-in of 10%. We checked for convergence and high ESS values in Tracer 1.6 [10], and used TreeAnnotator to calculate a maximum clade credibility tree with median heights. We produced the final tree using FigTree 1.4.2.

**Establishment of laboratory colonies and animal husbandry.** We housed animals on Bed-o'Cobs 1/4" bedding (The Andersons, Maumee, Ohio) in ventilated standard rodent cages (Allentown Inc., Allentown, NJ) on a 16h light: 8h dark cycle at 23°C. Animals were provided with a red translucent polycarbonate hut, Enviro-Dri nesting material, and a cotton nestlet. Animals were given *ad libitum* access to irradiated Prolab Isopro RMH 3000 5P74 (LabDiet) and water. We weaned litters at 23 days of age and kept animals in groups of up to five individuals of the same sex and strain.

**Behavioural experiments.** Approximately 3-4 hours before the trial, the female was removed from the resident's cage, resident and intruder were weighed, and the intruder was marked for easy identification with permanent marker on the tail. At the start of the experiment, we placed the resident's cage into black blinders, removed the lid and the red hut, and placed a custom-built divider with a gate and a transparent lid onto the cage. The divider set apart a small portion of the cage for the intruder, but was perforated to enable contact. The resident was habituated to the divider for 10 mins. Then we introduced the intruder into the area separated by the divider, and both mice were habituated for another 10 mins. Next, we opened the gate and the intruder was allowed to enter the resident's area. We video-recorded behaviour from above the cage at 15 fps; trials lasted for 15 mins.

**Statistical analysis.** We tested for differences in weight, body composition, and skeletal traits with linear fixed effects models with a strain (island/mainland) by origin (field/lab) interaction. We tested for differences in birth weight of parental strains (or F1 hybrids) using linear fixed effects models with the natural logarithm of weight as the dependent variable and the natural logarithm of litter size and strain (or maternal strain, respectively) as explanatory variables. We tested for differences in growth rate between parental strains (or F1 hybrids) using linear fixed effects models with the natural logarithm of growth rate as the dependent variable and the natural logarithm of litter size and an interaction of strain (or maternal strain, respectively) and day as explanatory variables, and adjusted p-values for multiple testing with the Holm method. Growth rates were obtained by taking the first derivative of cubic smoothing splines fit to the weight data using smoothPspline {pspline}. Our sample size for the F1 hybrids was moderate, therefore we ran additional tests that controlled for family effects. We did not include parent ID (as a proxy for family effects) into the model to avoid model estimation problems introduced by the perfect collinearity of parent ID and strain. Instead, we accounted for family effects by adjusting the standard errors of the model using wild cluster bootstrapping as implemented in cluster.wild.glm {clusterSEs}. To compare litter weight before and after the first milk meal, we used a repeated measures linear mixed effects model with the natural logarithm of weight as dependent variable, the natural logarithm of litter size and an interaction of feeding status (pre- vs. post-feeding) and strain as explanatory variables and litter ID as random effect. We tested for differences in weight by maternal strain in adult hybrid mice with a linear fixed effects model with the natural logarithm of weight as the dependent variable and strain as the explanatory variable. In the cross-fostering experiment, we tested for differences in the natural logarithm of growth rate of conspecific non-fostered and fostered litters with a linear fixed effects model, with the natural logarithm of litter size and an interaction of state (fostered vs. non-fostered) and day as explanatory variables, and adjusted p-values for multiple testing with the Holm method.

To reduce the dimensionality in the behaviour of the wild-caught resident animals, we performed a PCA on behaviours averaged across trials using prcomp {stats}, with variables scaled to unit variance and shifted to be zero centred. We next selected variables that contributed most to the first PC and focused on these behaviours for subsequent analyses. Specifically, we compared levels in these three behaviours (wrestling, chasing, and pindown) in wild-caught residents, captive-born residents one week after they were paired with a female, and breeding pairs after the onset of mating attempts. For this, we used a two-model hurdle approach to test for differences between strains. This included a binary model to test whether different proportions of individuals express a behaviour, and count and duration models to test whether differences exist in the number of bouts and duration once animals express a given behaviour. We initially evaluated the impact of trial number and weight difference between resident and intruder on aggression by comparing repeated measures linear mixed effects models with and without these factors using likelihood ratio tests. Because these factors did not improve models for most behaviours, we averaged data across trials (average number of bouts were rounded) and ran logistic regression models using glm {stats} with binomial (binary) and quasipoisson (count) error distributions and linear fixed effects models (duration). To compare wrestling duration before and after reproduction, we fit a repeated measures mixed effects linear model on the averaged dataset with reproductive status and strain as explanatory variables and resident ID as random effect. To evaluate other potential impacts on the summed duration of wrestling, chasing, and pindown in wild-caught mice, we fit a repeated measures mixed effects linear model with days since capture, days since last litter was sired, weight difference, and strain as explanatory variables and resident ID as random effect. Finally, we tested for differences in aggression behaviors (wrestling, chasing, pindown) between wild-caught and captive-born mice (Suppl. Fig. 14) with linear fixed effects models with a strain (island/mainland) by origin (field/lab) interaction.

**LIST OF SUPPLEMENTAL FIGURES**

**Suppl. Fig. 1.** Haplotype network based on 381bp of the mitochondrial cytochrome *b* gene. Numbers on edges indicate the number of DNA substitutions separating adjacent haplotypes. Diameter of nodes is proportional to the number of specimens that carry the respective haplotype (see schematic at the top of the graph).

**Suppl. Fig. 2.** Histograms of body weight of mainland (top) and island (bottom panel) mice. The median weights for all populations are indicated by dashed lines.

**Suppl. Fig. 3.** Body weight in adult female wild-caught island (red) and mainland (blue) mice. Statistical significance evaluated by linear fixed effects model (see Methods for details). \*\* *P* < 0.01.

**Suppl. Fig. 4.** Size-corrected (**A**) skull length, (**B**) skull width, (**C**) zygomatic width, (**D**) right zygomatic length, (**E**) right humerus length, (**F**) right femur length, (**G**) right ulna length, (**H**) right tibia length, and (**I**) right metatarsal calcaneal length in male wild-caught (left) and captive-born (right) island (red) and mainland (blue) mice. Statistical significance evaluated by linear fixed effects model (see Methods for details). NS=not significant, \* *P* < 0.05, \*\* *P* < 0.01, \*\*\* *P* < 0.001, \*\*\*\* *P* < 0.0001.

**Suppl. Fig. 5.** Fat mass in adult male wild-caught (left) and captive-born (right) island (red) and mainland (blue) mice. Statistical significance evaluated by linear fixed effects model (see Methods for details). NS=not significant, \*\* *P* < 0.01, \*\*\*\* *P* < 0.0001.

**Suppl. Fig. 6.** Litter size of island (red) and mainland (blue) mice. Arrows indicate the median litter size for each strain. Statistical significance evaluated by Kruskal-Wallis test (see Methods for details). \* *P* < 0.05.

**Suppl. Fig. 7.** (**A**) Mean weight of island (red) and mainland (blue) litters before and after the first milk meal. Statistical significance evaluated by repeated measures linear mixed effects model (see Methods for details). NS=not significant, \*\* *P* < 0.01, \*\*\*\* *P* < 0.0001. (**B**) Example of pup with milk present in stomach (outlined).

**Suppl. Fig. 8.** **(A-B)** Growth curves of (**A**) island (red) and mainland (blue) litters and (**B**) F1 hybrid litters by maternal strain. Points and error bars represent mean and SEM of growth of litters. **(C-D)** Weight of adult hybrid mice plotted by maternal strain identity. The natural logarithm of weights of (**C**) males and (**D**) females were separately compared by maternal strain with linear fixed effects models (see Methods for details). \* *P* < 0.05, \*\* *P* < 0.01, \*\*\* *P* < 0.001.

**Suppl. Fig. 9.** Growth rates of (**A**) mainland litters cross-fostered by island parents (purple) compared to mainland litters that were not cross-fostered (blue); (**B**) island litters cross-fostered by mainland parents (orange) compared to island litters that were not cross-fostered (red). Points and error bars represent mean and SEM of growth of litters. Statistical significance evaluated by linear fixed effects model (see Methods for details). \* *P* < 0.05.

**Suppl. Fig. 10.** Count data of wrestling, chasing, and pindown behaviours in (**A**) wild-caught founder animals, (**B**) captive-born males paired with a female for one week before testing, and (**C**) captive-born breeding males tested after verified copulation. Statistical significance evaluated by generalized linear models (see Methods for details). NS=not significant, \* *P* < 0.05.

**Suppl. Fig. 11.** Effect of (**A**) time since capture at testing, (**B**) time since a litter was last sired, and (**C**) weight difference between resident and intruder on the aggressive behaviour of wild-caught deer mice (N=7 island [red], N=6 mainland [blue]). Replicate trials are plotted separately. Statistical significance evaluated with a repeated measures mixed effects linear model (see Methods for details). NS=not significant.

**Suppl. Fig. 12.** Effect of reproductive experience on territorial aggression. (**A**) Time of mating after pairing in a subset of mice from the experiment in Fig. 5A-B. We backdated for females that went on to give birth when the litter was sired. The horizontal line marks the time when males were tested in the resident-intruder assay (7 days after pairing); at this time, few males in this subset had sired litters. (**B**) We re-tested a subset of these males after they had sired a litter, and compared wrestling duration before and after siring a litter with a repeated measures linear mixed effects model (see Methods for details). Only mice were included that had not sired a litter by the time of first testing. Mean and SEM across strains is shown in black. \* *P* < 0.05.

**Suppl. Fig. 13.** Schematic showing birth and mating events in breeding pairs. Each row corresponds to one pair. The grey boxes indicate the maximum duration of behaviours (it was not possible to observe females continuously, e.g. when they were inside the nest at the time of birth). The light cycle (16h:8h, light:dark) is shown at the bottom.

**Suppl. Fig. 14.** Direct comparison of aggressive behavior of mice exposed to comparable reproductive experience and experimental conditions in captivity (i.e., data from captive-born mice in Fig. S12B [right column] and from wild-caught mice in Fig. 4). Statistical significance evaluated with a fixed effects linear model with a strain (island/mainland) by origin (captive/wild) interaction. NS=not significant, \* *P* < 0.05.

**Supplemental References**

[1] Hogan, K.M., Hedin, M.C., Koh, H.S., Davis, S.K. & Greenbaum, I.F. 1993 Systematic and taxonomic implications of karyotypic, electrophoretic, and mitochondrial-DNA variation in *Peromyscus* from the Pacific Northwest. *Journal of Mammalogy* **74**, 819-831. (doi:10.2307/1382420).

[2] Zheng, X., Arbogast, B.S. & Kenagy, G.J. 2003 Historical demography and genetic structure of sister species: deermice (*Peromyscus*) in the North American temperate rain forest. *Molecular Ecology* **12**, 711-724. (doi:10.1046/j.1365-294X.2003.01770.x).

[3] Li, H. & Durbin, R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760. (doi:10.1093/bioinformatics/btp324).

[4] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303. (doi:10.1101/gr.107524.110).

[5] 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. 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158. (doi:10.1093/bioinformatics/btr330).

[6] Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. 2014 ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356. (doi:10.1186/s12859-014-0356-4).

[7] Fumagalli, M., Vieira, F.G., Linderoth, T. & Nielsen, R. 2014 ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486-1487. (doi:10.1093/bioinformatics/btu041).

[8] Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. 2012 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**, 1917-1932. (doi:10.1093/molbev/mss086).

[9] Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. 2014 BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537. (doi:10.1371/journal.pcbi.1003537).

[10] Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. 2018 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**, 901-904. (doi:10.1093/sysbio/syy032).