READ ME FILE FOR DATA USED FOR “GENETIC ANCESTRY PREDICTS MALE-FEMALE AFFILIATION IN A NATURAL BABOON HYBRID ZONE”

ARIELLE S. FOGEL, EMILY M. MCLEAN, JACOB B. GORDON, ELIZABETH A. ARCHIE, JENNY TUNG, SUSAN C. ALBERTS

The CSV files **groom_anonymized.csv** and **prox_anonymized.csv** contain all data necessary for running and reproducing the results of the multivariate logistic regression models predicting grooming and proximity behavior, respectively, presented in the main text. In both files, each row of data corresponds to a unique, co-resident female-male dyad in a given two-month interval. These files contain the following columns:

- ***female_id***: the identifier of the female baboon represented by that row
- ***male_id***: the identifier of the male baboon represented by that row
- ***groom_two_month* or *prox_two_month***: the grooming or proximity behavior of the female-male dyad in a given two-month interval represented by that row; assigned a value of “1” if the dyad was observed grooming or in proximity at least once during the two-month interval and a “0” if they were not
 - ⇒ This is the only column name that differs between the **groom_anonymized.csv** and **prox_anonymized.csv** files.
- ***genetic_ancestry_female***: the genetic ancestry estimate of the female represented by that row (i.e., the proportion of the female’s genome estimated to be from anubis ancestry); this ranges continuously from 0 to 1, where 0 corresponds to unadmixed yellow baboon ancestry and 1 corresponds to unadmixed anubis baboon ancestry
- ***genetic_ancestry_male***: the genetic ancestry estimate of the male represented by that row (i.e., the proportion of the male’s genome estimated to be from anubis ancestry); this ranges continuously from 0 to 1, where 0 corresponds to unadmixed yellow baboon ancestry and 1 corresponds to unadmixed anubis baboon ancestry
- ***assortative_genetic_ancestry_index***: the pairwise assortative genetic ancestry index of the female-male dyad represented by that row
- ***heterozygosity_female***: the genetic diversity estimate of the female represented by that row based on the genotype data in **geno_anonymized.csv**
- ***heterozygosity_male***: the genetic diversity estimate of the male represented by that row based on the genotype data in **geno_anonymized.csv**
- ***genetic_relatedness***: the genetic relatedness estimate of the male-female dyad represented by that row using the method of Queller and Goodnight (1989) based on the genotype data in **geno_anonymized.csv**
- ***rank_female***: the ordinal dominance rank (i.e., the highest-ranking individual holds rank 1 and lower-ranking individuals occupy ranks of successively higher numbers) of the female in the two-month interval represented by that row; this is the average of the female’s rank across both months since rank is assigned on a monthly basis
- ***rank_male***: the ordinal dominance rank (i.e., the highest-ranking individual holds rank 1 and lower-ranking individuals occupy ranks of successively higher numbers) of the male in the two-month interval represented by that row; this is the average of the male’s rank across both months since rank is assigned on a monthly basis
- ***female_age***: the age of the female baboon represented by that row, averaged across each two-month analysis window (i.e., her age at the start of the second month)

- ***female_age_transformed***: a measure of female age (*female_age*) transformed to reflect the relationship between female age and conception probability in the study population
- ***females_in_group***: the number of adult females in the social group (*group_id*) of the male-female dyad averaged across the two-month period represented by that row
- ***males_in_group***: the number of adult males in the social group (*group_id*) of the male-female dyad averaged across the two-month period represented by that row
- ***reproductive_state***: the reproductive state of the female baboon in a given two-month interval represented by that row; assigned a value of “-1” if the female was pregnant and a “1” if she was lactating
- ***pair_coresidency***: the total number of days in each two-month interval that the male and female in the male-female dyad represented by that row were censused in the same social group
- ***observer_effort***: the average number of minutes of focal sample data collected per adult female per social group (*group_id*) in the two-month interval represented by that row
- ***group_id***: the identifier of the female-male dyad’s social group
⇒ Not included in the multivariate logistic regression models but necessary for recreating Table S1.
- ***interval_id***: the identifier of the female’s two-month interval represented by that row
⇒ Not included in the multivariate logistic regression models but necessary for running permutation tests to assess statistical significance.

The CSV file **geno_anonymized.csv** contains all microsatellite data for study subjects used for calculating genetic ancestry (*genetic_ancestry_female* and *genetic_ancestry_male* above), genetic diversity (*heterozygosity_female* and *heterozygosity_male* above), and genetic relatedness (*genetic_relatedness* above). The first column, ***id***, is the identifier of the individual baboon represented by that row. This column also corresponds to either the *female_id* or *male_id* in *groom_anonymized.csv* and *prox_anonymized.csv*. The remaining 28 columns correspond to the 14 microsatellite markers. Each microsatellite locus (e.g., “agat006”) has two columns which are distinguished by a “_1” and “_2” (e.g., “agat006_1” and “agat006_2”). NAs correspond to missing genotype data.

The CSV file **unique_social_group_months_FigS1.csv** contains all data necessary for recreating Figure S1. In this file, each row of data corresponds to a unique, social group-month combination from the initial monthly-based data set. It contains the following columns:

- ***group_id***: the identifier of the social group (corresponds to *group_id* above in *groom_anonymized.csv* and *prox_anonymized.csv*)
- ***females_in_group***: the number of adult females in the social group for the social group-month combination represented by that row
- ***total_adult_female_point_samples_in_month***: the total number of point samples recorded for adult females (10 point samples are collected per focal sample, if the sample is complete) for the social group-month combination represented by that row