Goal 1: Create a methylation aging clock to use to reliably age mada bat species from wing punches instead of dentition

Goal 2: Use this clock to create updated PVA’s for each species

We are FIRST going to try building this clock for myotis species, whose methylation levels were generated using a simpler lab methodology. This is going to be a proof of concept to then be able to use the same lab methodology for our mada bat samples.

Background

Epigenetic clocks allow us to measure the age of tissue by looking at changes in DNA

This works using DNA methylation.

Methyl groups attach to cytosines over time. You can study changes and patterns in DNA methylation over time to determine age.

Wilkinson et al 2021 – paper that uses bat dna methylation to age bats by building epigenetic clock

Input: methylation, confirmed age

Output: model

Methods: Creation of epigenetic clocks using penalized regression

We developed epigenetic clocks for bat wing tissue by regressing chronological age on all CpGs that map to at least one of the ten bat genomes. To improve linear fit we transformed chronological age to sqrt(age + 1). Penalized regression models were created in the R package glmnet[61](https://www.nature.com/articles/s41467-021-21900-2#ref-CR61). We investigated models produced by elastic net regression (alpha = 0.5). The optimal penalty parameters in all cases were determined automatically by using a tenfold internal cross-validation (cv.glmnet) on the training set. By definition, the alpha value for the elastic net regression was set to 0.5 (midpoint between Ridge and Lasso-type regression) and was not optimized for model performance. We performed two cross-validation schemes for arriving at unbiased estimates of the accuracy of the different DNAm based age estimators. One type consisted of leaving out a single sample (LOO) from the regression, predicting an age for that sample by regressing an elastic net on the methylation profiles of all other samples and iterating over all samples. We conducted LOO analyses using all samples from all species, using all samples from each species and using all samples from several species in the same genus. The second type consisted of leaving out a single species (LOSO) from the regression, thereby predicting the age of each sample using the data for all other species.

Creating epigenetic clock for myotis spp

THE DATA

Manny has run the pipeline on the data coming out of the lab procedure to determine the methylation levels.  
https://github.com/docmanny/methylationAging

Quick facts

5 species

* Myotis californicus
* Myotis evotis
* Myotis lucifugus
* Myotis thysanodes
* Myotis Volans

File structure

Output/methylation\_extracted

Table

Description automatically generated

Output/bismark\_bedgraph

According to samples.tsv

M\_californicus

B1S1-4

B1S1-4-sub

M\_evotis

B1S1-2

B1S1-3

B1S1-3-sub

B1S1-2-sub

M\_lucifugus

ML3

ML5

ML7

ML3-sub

ML5-sub

ML7-sub

M\_colans

MV2

MV2-sub

M\_thysanodes

MT

MT-sub

METHODS

General elastic net regression ML theory

Target variable: age

Independent variable: dna methylation levels

\*use methylation to predict age

1. Split into X and Y
   1. X = independent
   2. Y = target only
2. Split into train and test
   1. X\_train =
   2. X\_test =
   3. Y\_train =
   4. Y\_test =
3. Set up hyperparameters in elastic net
   1. Elastic net(alpha = 0.5, L1\_ratio = 0.5)
4. Fit model
   1. Enet.fit(X\_train, Y\_train)
5. Make predictions
   1. Y\_pred = Enet.pred(X\_test)
6. Evaluate
   1. MSE (y\_test, y\_pred)