Epigenetic Age Estimation

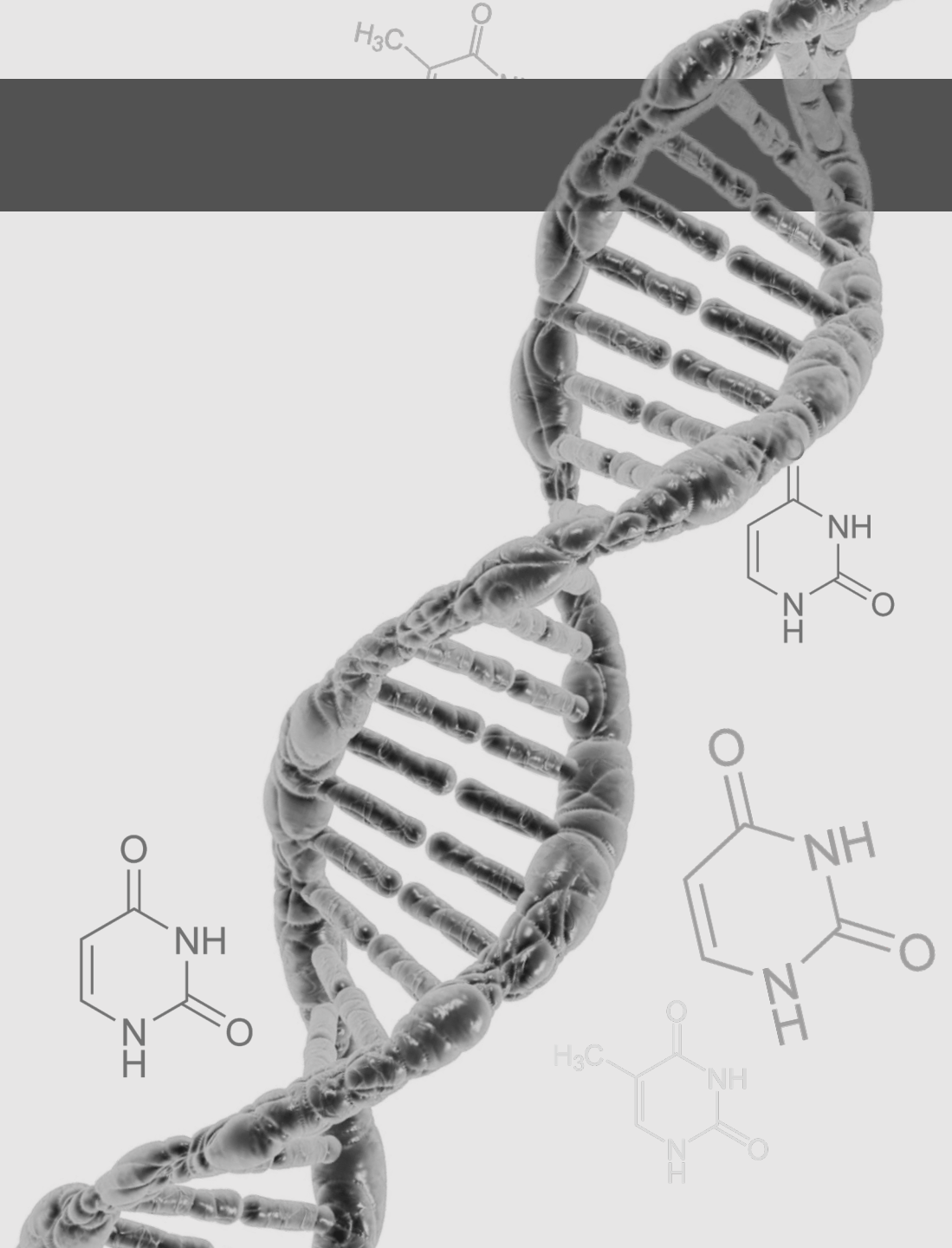
GulfFIN Committee Meeting 2024

D.S. Portnoy[1] , D.N. Weber[1], A.T. Fields[1], B.K. Barnett[2,3], D.W. Chamberlin[3,4], K. Lyons[5], J. Wyffels[6,7] & W.F. Patterson III[3]

[1]Marine Genomics Laboratory, Texas A&M University–Corpus Christi
[2]Southeast Fisheries Science Center, NOAA Fisheries
[3]Fisheries & Aquatic Sciences, University of Florida
[4]Alaska Fisheries Science Center, NOAA Fisheries
[5]Center for Species Survival, Georgia Aquarium
[6]Delaware Biotechnology Institute, University of Delaware
[7]Ripley's Aquariums

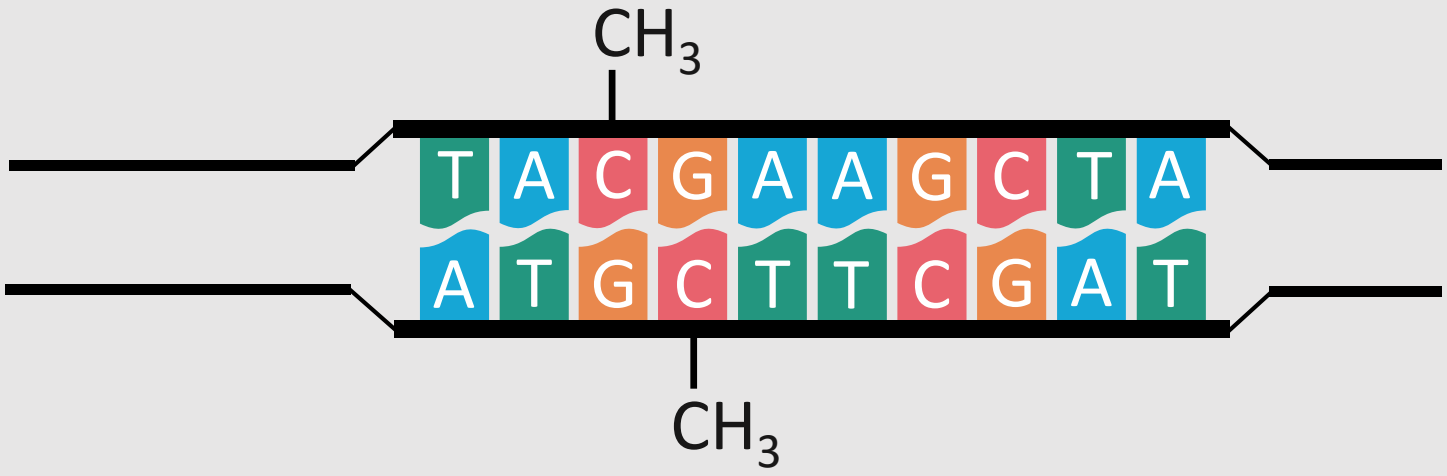
Outline

- Epigenetic ageing
  - Concepts and theory
- Major approaches
  - Cross-amplification
  - Species-specific clocks
- Production ageing
  - Time and cost



What is Epigenetics?

Mechanisms that affect gene expression without altering DNA sequence

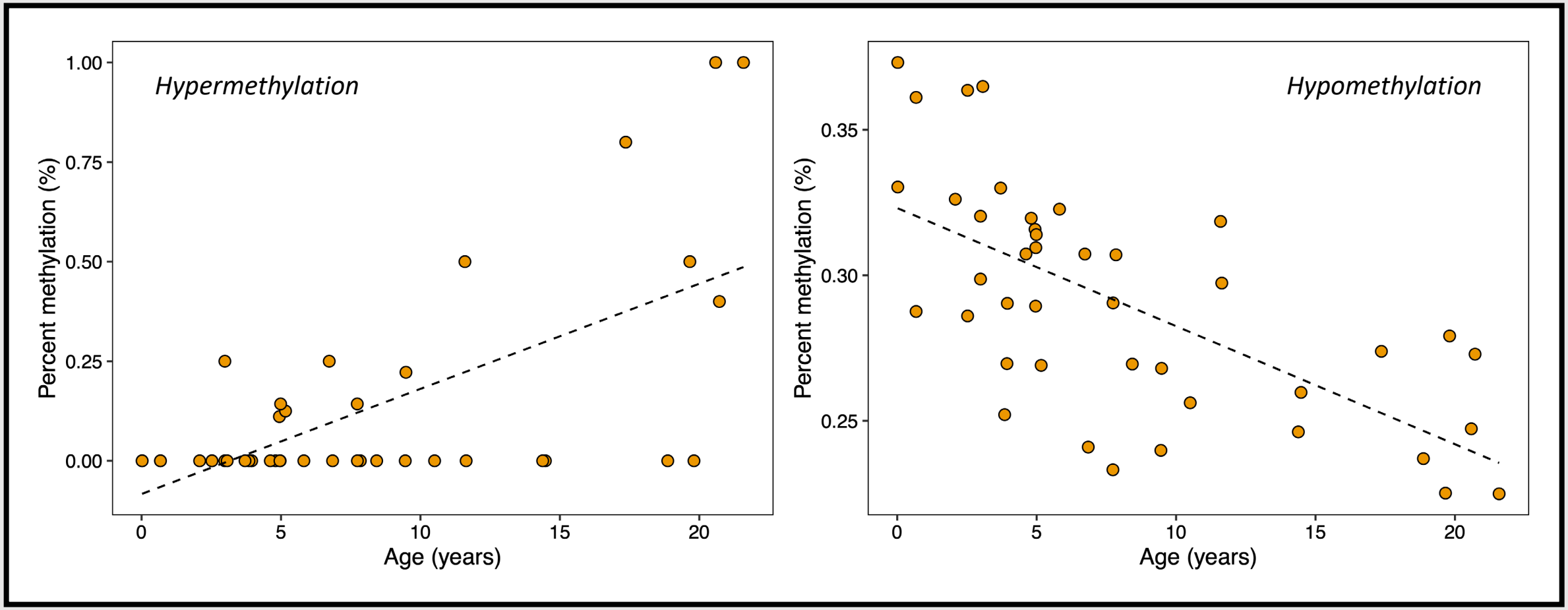


DNA Methylation

- Addition of CH_3 to cytosine, often at CpG sites
- Changes in DNA methylation at select CpG sites correlate with age

→ **Epigenetic clocks**

“… models that summarize age-associated increases or decreases in DNA methylation across specific CpG sites which can be used collectively to estimate age …"



Epigenetic Clocks for Fishes

European seabass (Anastasiadi and Piferrer 2019)

Zebrafish (Mayne et al. 2020)

- Model organisms with well-annotated genomes

Two major approaches:

1) Cross-amplification

   Target CpG sites previously identified in a different species

   - Demonstrated in mammals (e.g., dogs and wolves)

2) Species-specific clocks

   Develop *de novo* clocks in species of interest

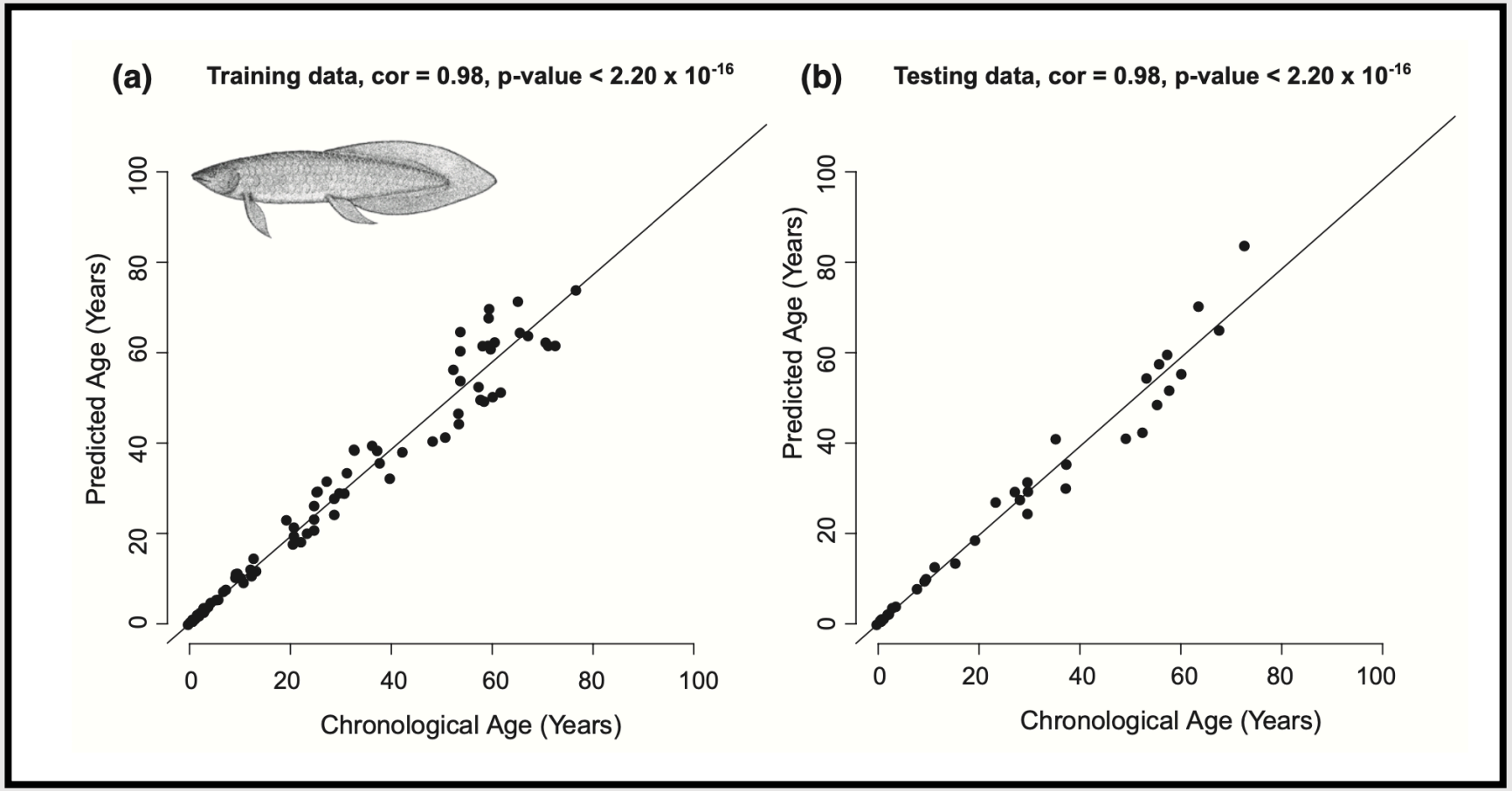
1: Target sites from zebrafish clock

Theory: Search genome of interest for 1,311 age-correlated CpG sites from zebrafish, and design primers to target those sites

1: Target sites from zebrafish clock

Australian Lungfish

- 31 zebrafish sites present
- Median error: 0.86 years
- Median error > 40: 6.10 years

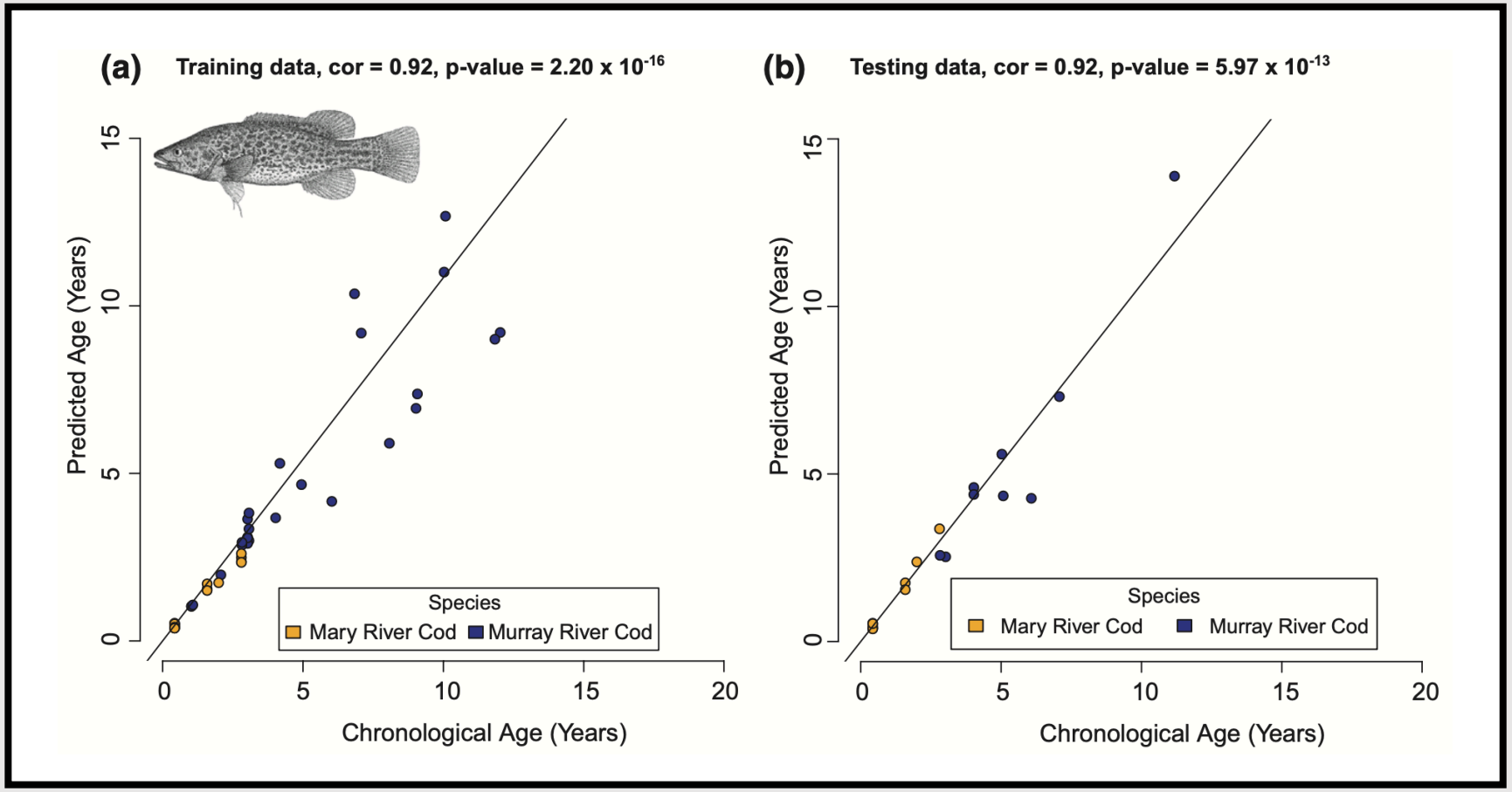


Mayne et al. 2021

1: Target sites from zebrafish clock

Murray & Mary River Cod

- 26 zebrafish sites present
- Median error: 0.35 years
- Median error > 10: 2.86 years



Mayne et al. 2021

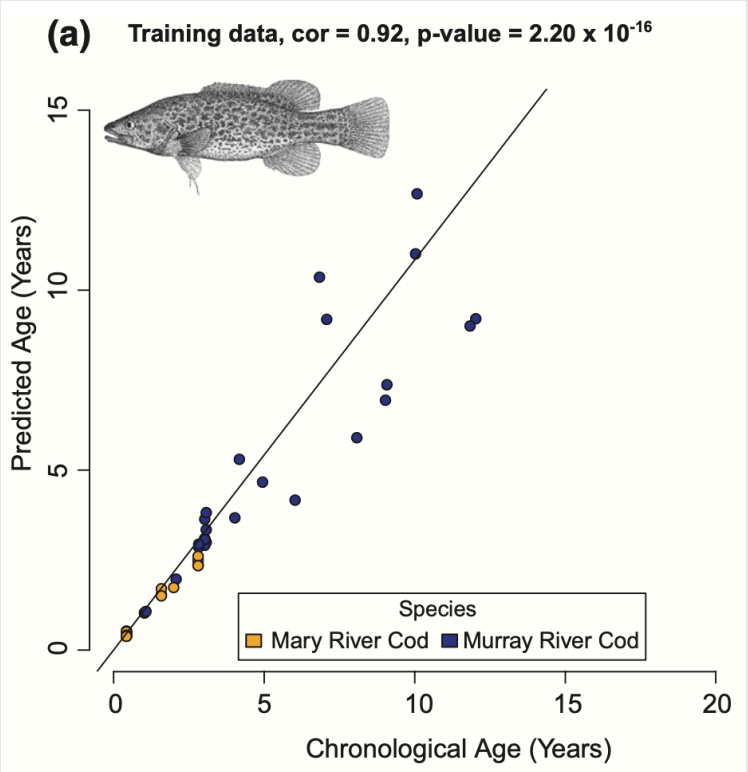
1: Target sites from zebrafish clock

Drawbacks:

- Accuracy dependent on number of conserved sites
- Accuracy decreases as age increases
- Missing species-specific age-informative sites

Benefits:

- Cheaper and quicker than *de novo* clock construction

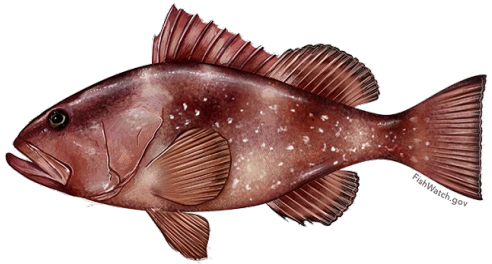


2: *De novo* clock development

Theory: Identify all age-correlated CpG sites in the species of interest and select best subset of CpG sites to predict age

Requirement: DNA samples (e.g., fin clips) from known-age individuals

Red Snapper
Lutjanus campechanus



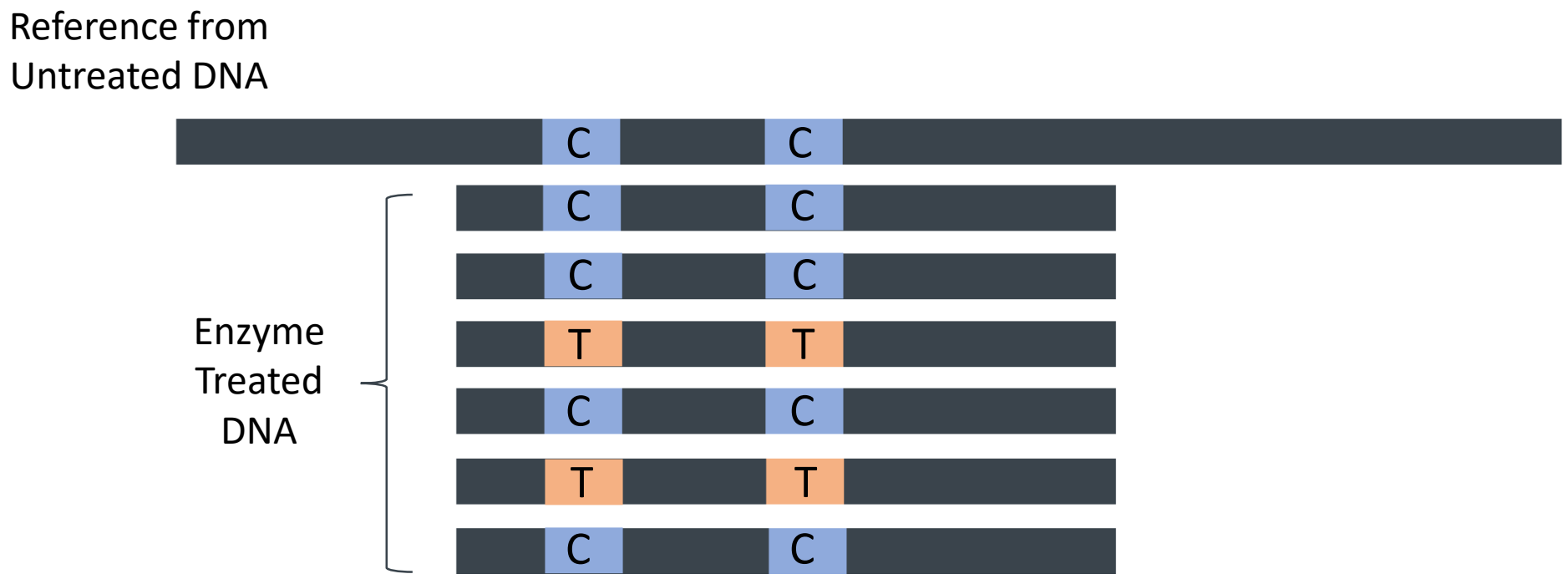
Red Grouper
Epinephelus morio

Blackbelly Rosefish
Helicolenus dactylopterus

Cownose Ray
Rhinoptera bonasus

Genomic Approach

radEM-seq: restriction site-associated enzymatic methyl-sequencing



Data Analysis

1) Identify all CpG sites that exhibit age-correlated methylation

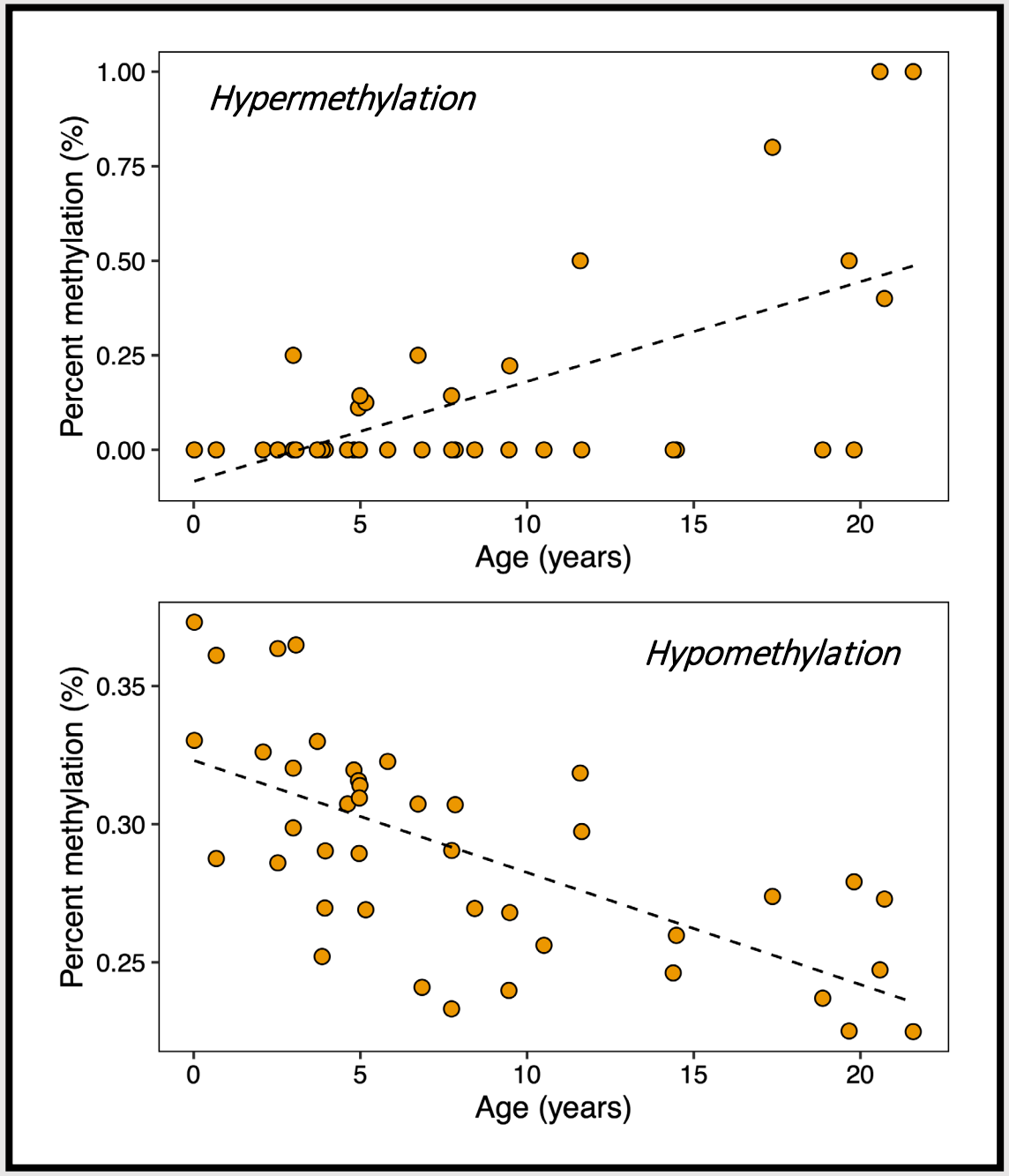
Bayesian GLM

- Age as fixed factor
- Sample as random factor
- $\frac{\#\ \text{methylated reads}}{\#\ \text{total reads}}$ as response

2) Identify subset of CpG sites that best predict age

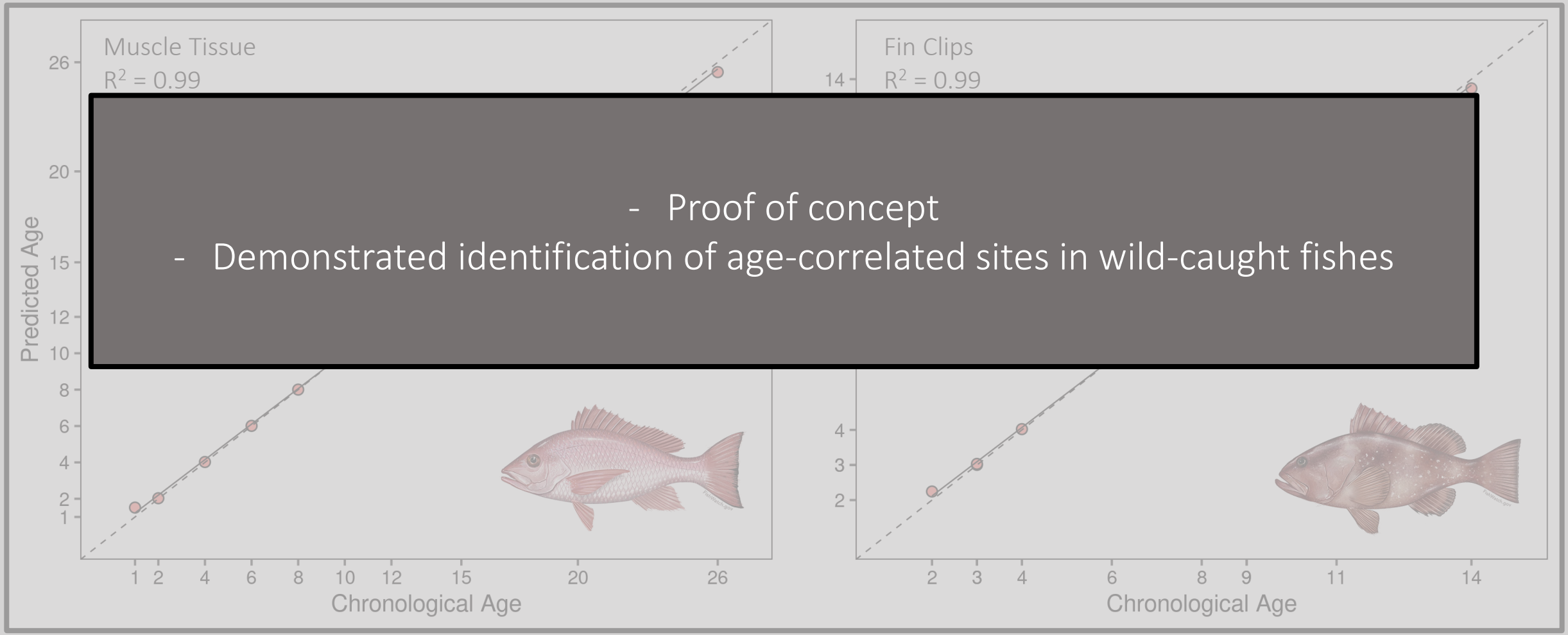
Penalized Regression

- Elastic net version of `glmnet` in R



Red Snapper
- 1,674,121 CpG sites identified
- 3,224 CpG sites age-correlated
- 199 CpG sites in final model

Red Grouper
- 1,238,719 CpG sites identified
- 690 CpG sites age-correlated
- 49 CpG sites in final model



- Proof of concept
- Demonstrated identification of age-correlated sites in wild-caught fishes

Blackbelly Rosefish

Helicolenus dactylopterus

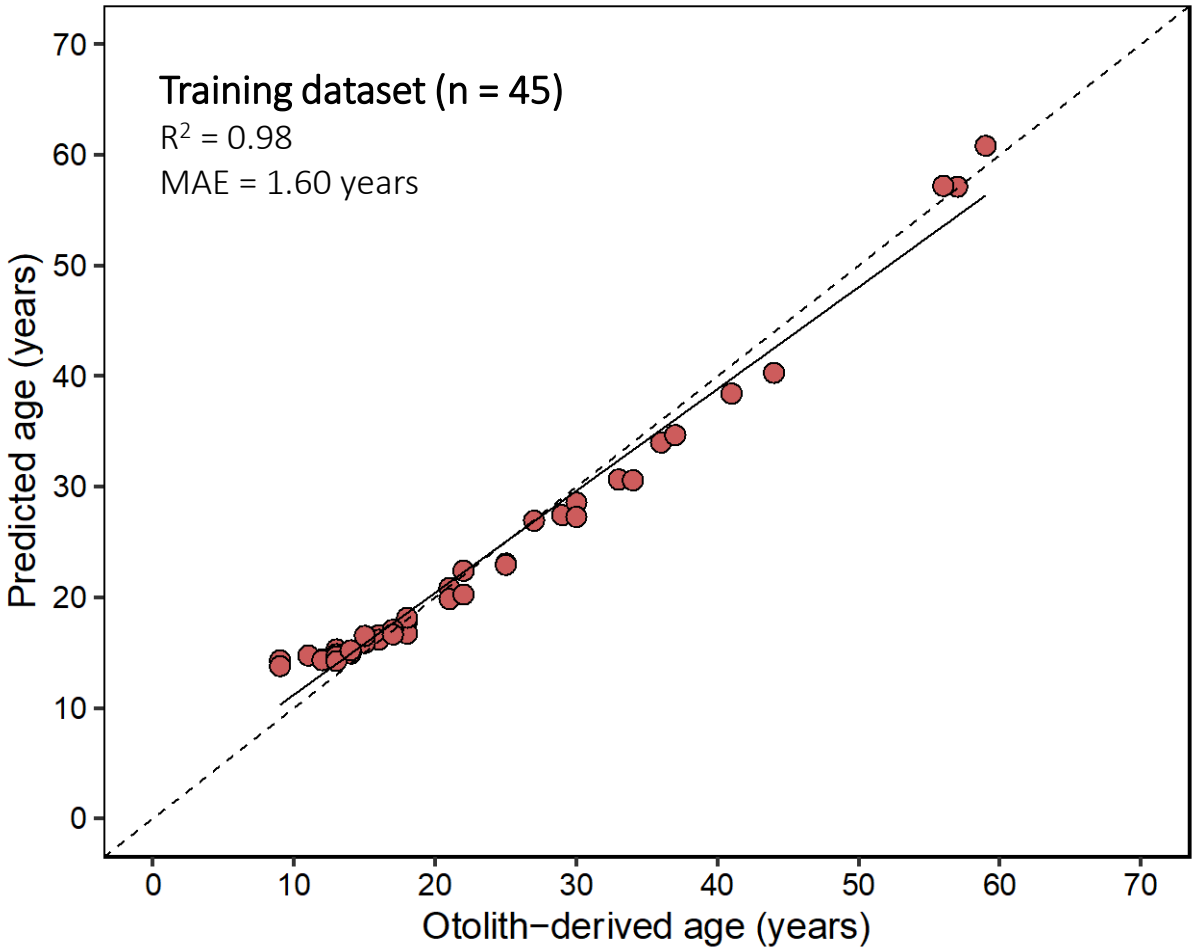
- Long-lived (>90 years)
- Deepwater reef fish (150 to 600 m)
- Difficult to age (age validation)

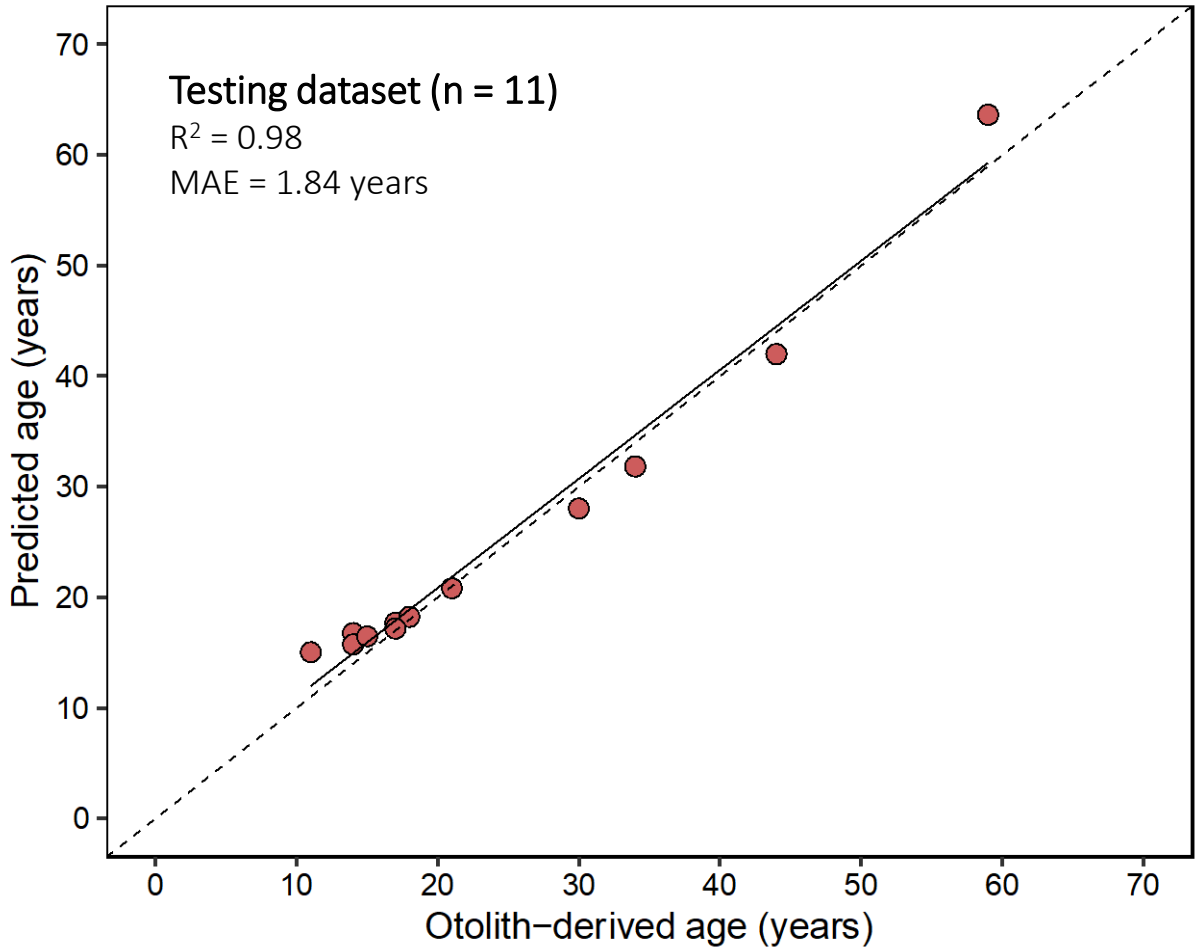
Bayesian GLMs and penalized regression:

- 2,959,164 CpG sites identified
- 10,139 CpG sites age-correlated
- 350-450 CpG sites in final models

Fin Clip Clock

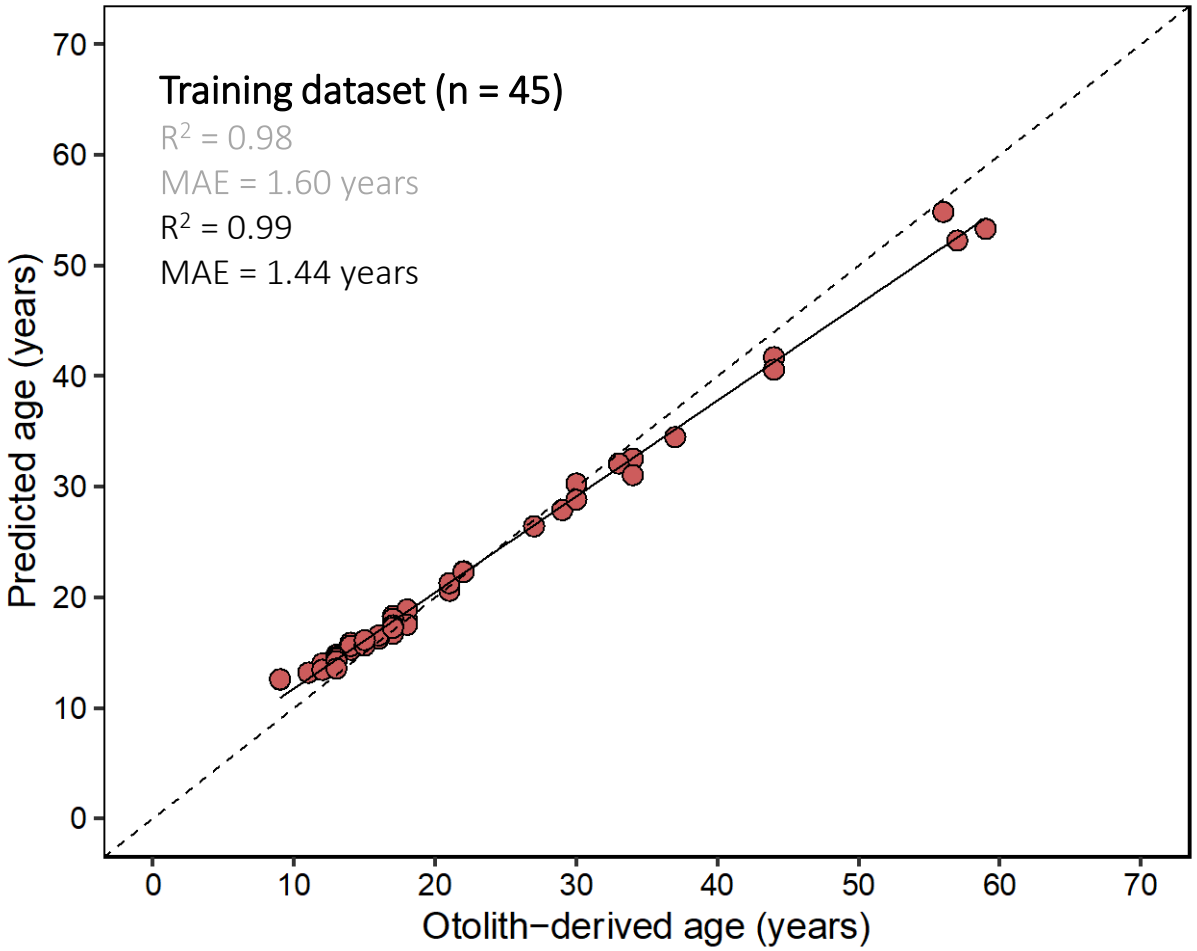
- 56 individuals (9-60 years)
- 316 CpG sites

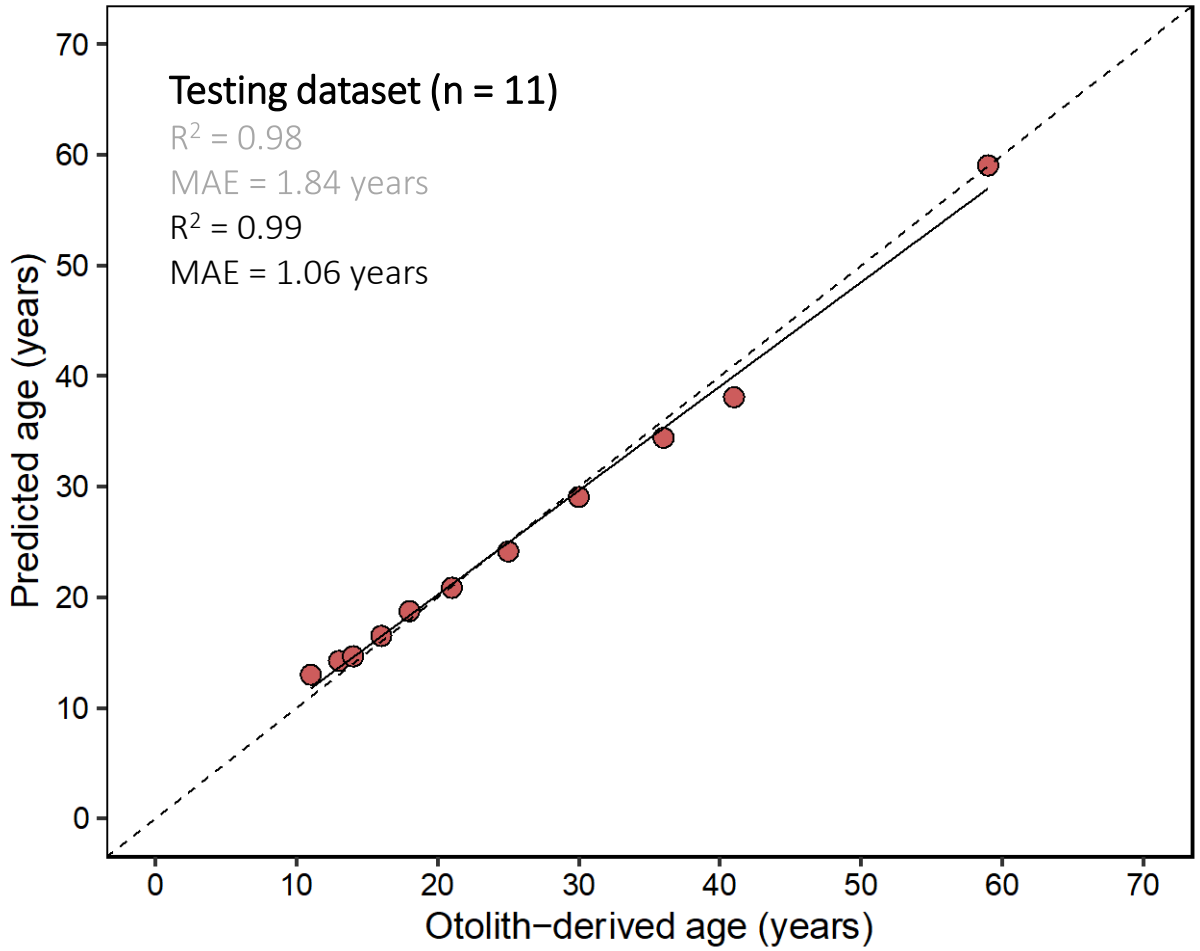




Fin Clip Clock + Length Data

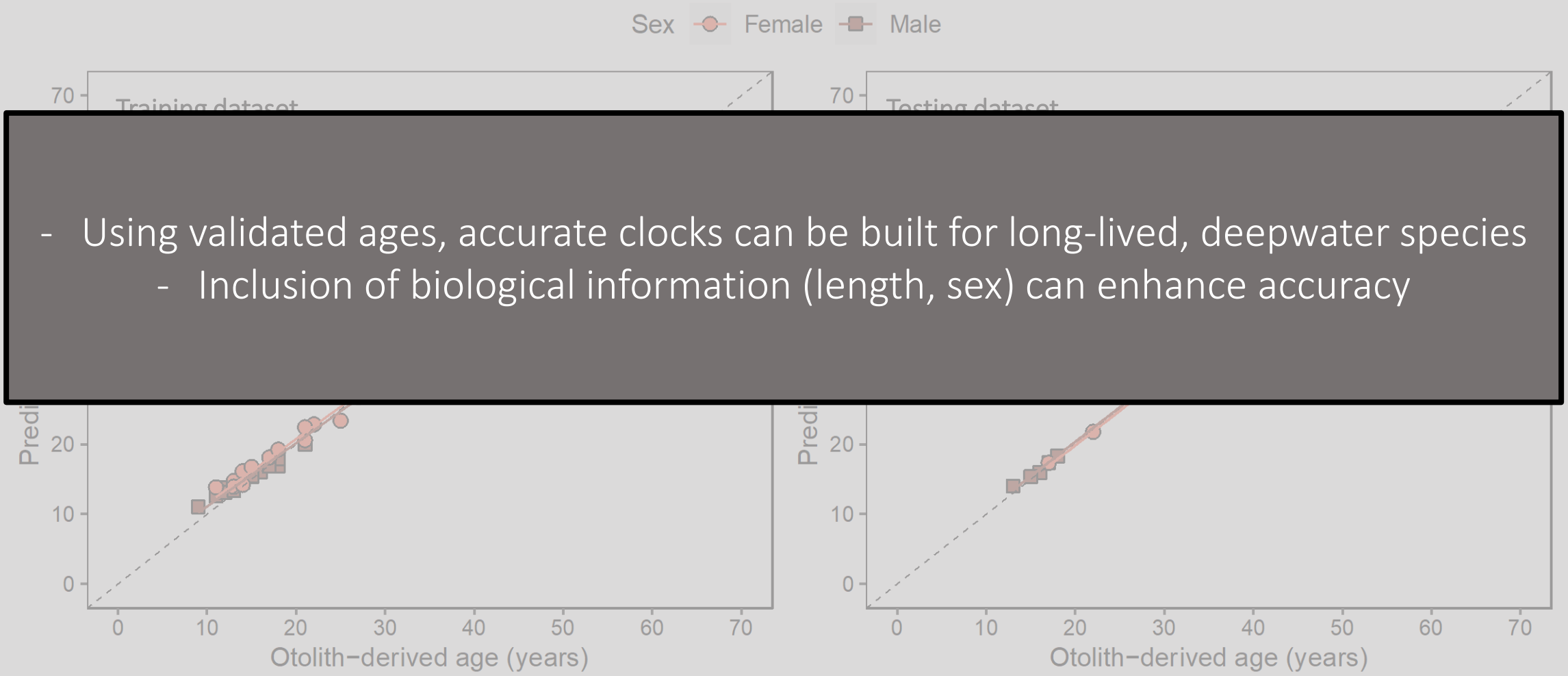
- 56 individuals (9-60 years)
- 315 CpG sites





Fin Clip Clock + Length Data by Sex

- Females: 308 CpG sites
- Males: 450 CpG sites



- Using validated ages, accurate clocks can be built for long-lived, deepwater species
  - Inclusion of biological information (length, sex) can enhance accuracy

Cownose Ray

Rhinoptera bonasus

- Benthopelagic batoid
- Commonly displayed in aquariums worldwide
- Known dates of birth

Bayesian GLMs and penalized regression:

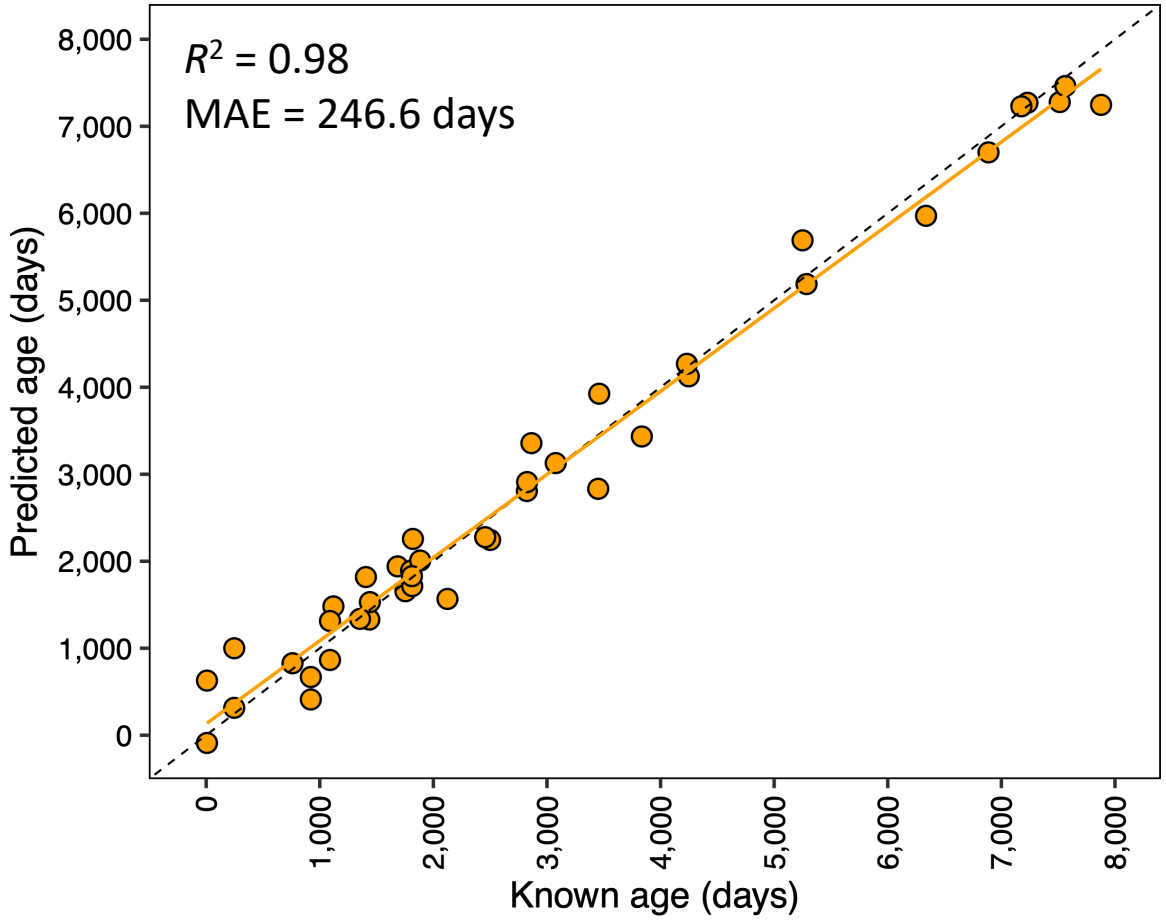
- 8,042,910 CpG sites identified
- 7,813 CpG sites age-correlated
- 30-62 CpG sites in final models

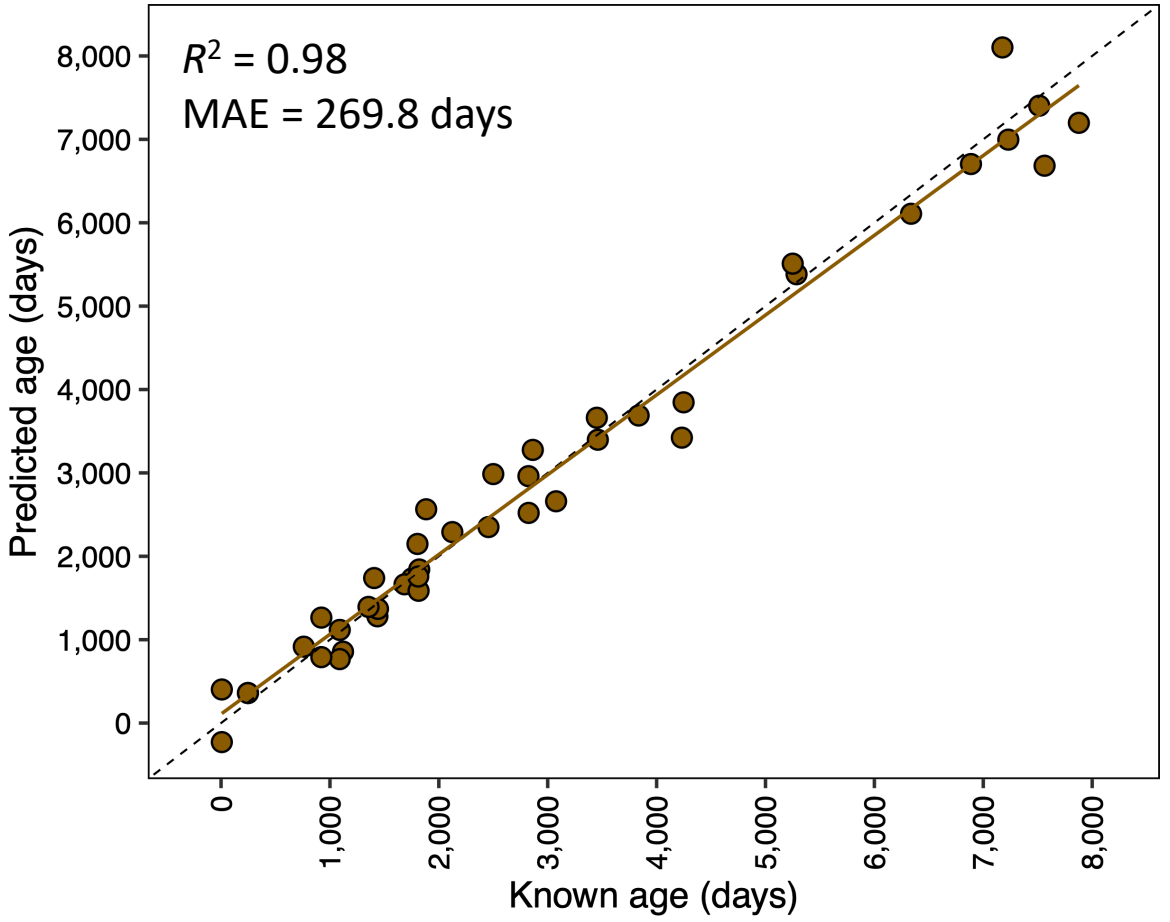
Fin Clip Clock

- 42 individuals (0-21 years)
- 32 CpG sites

Whole Blood Clock

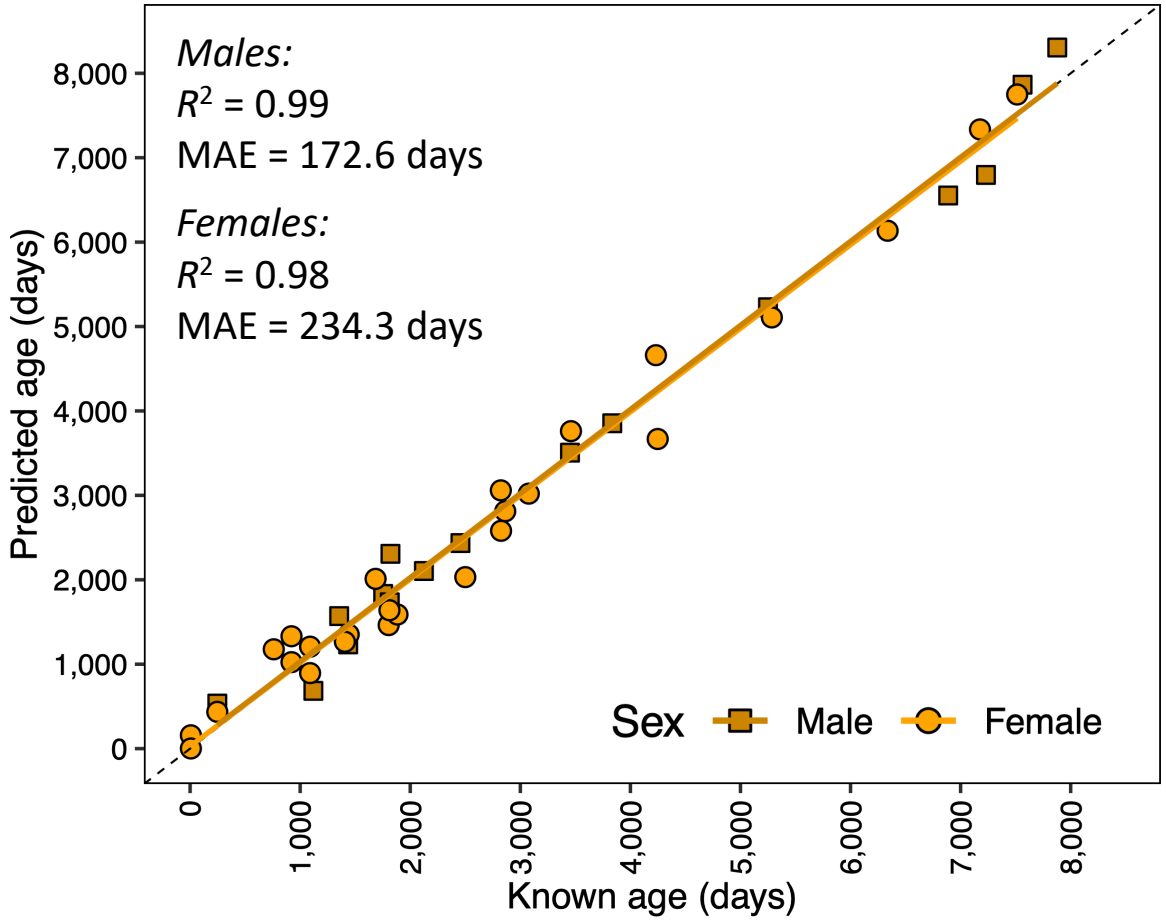
- 42 individuals (0-21 years)
- 30 CpG sites





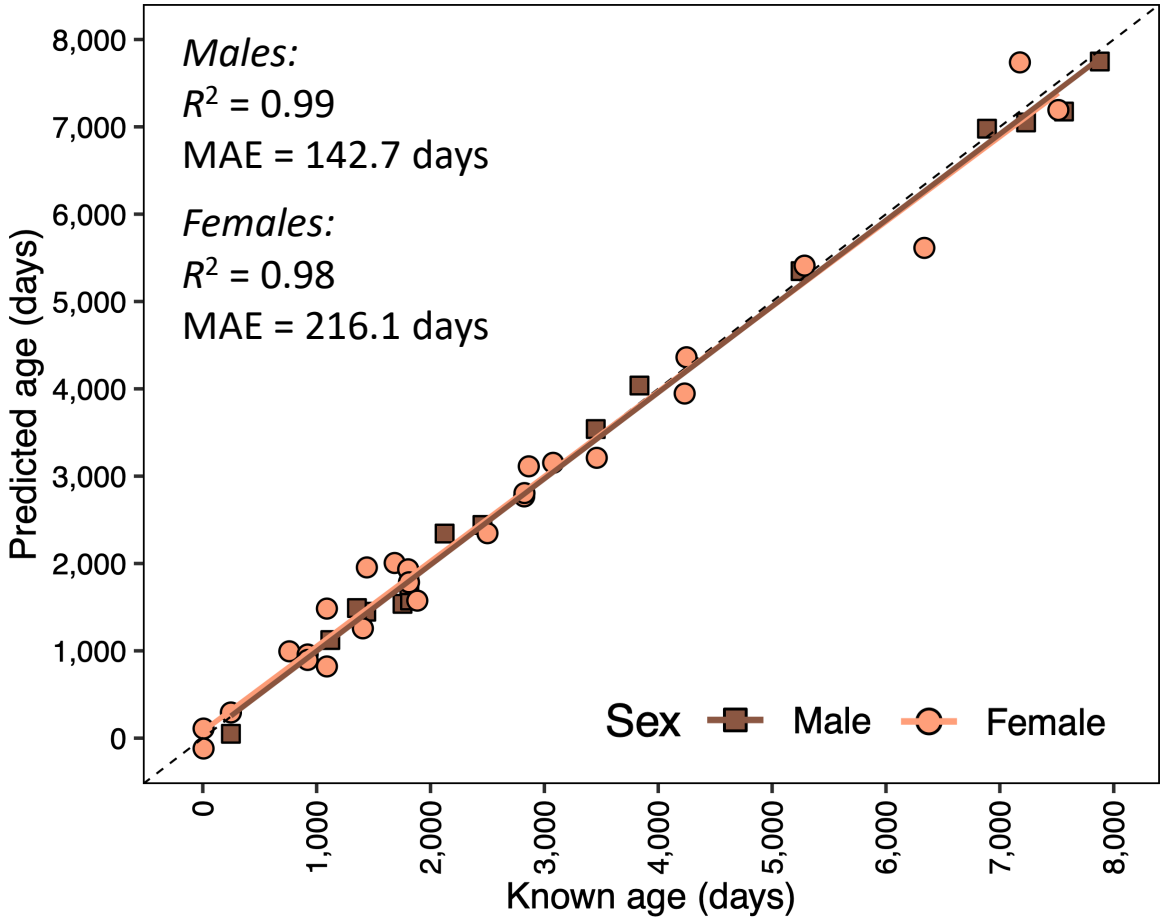
Fin Clip Clock by Sex

- 42 individuals (0-21 years)



Whole Blood Clock by Sex

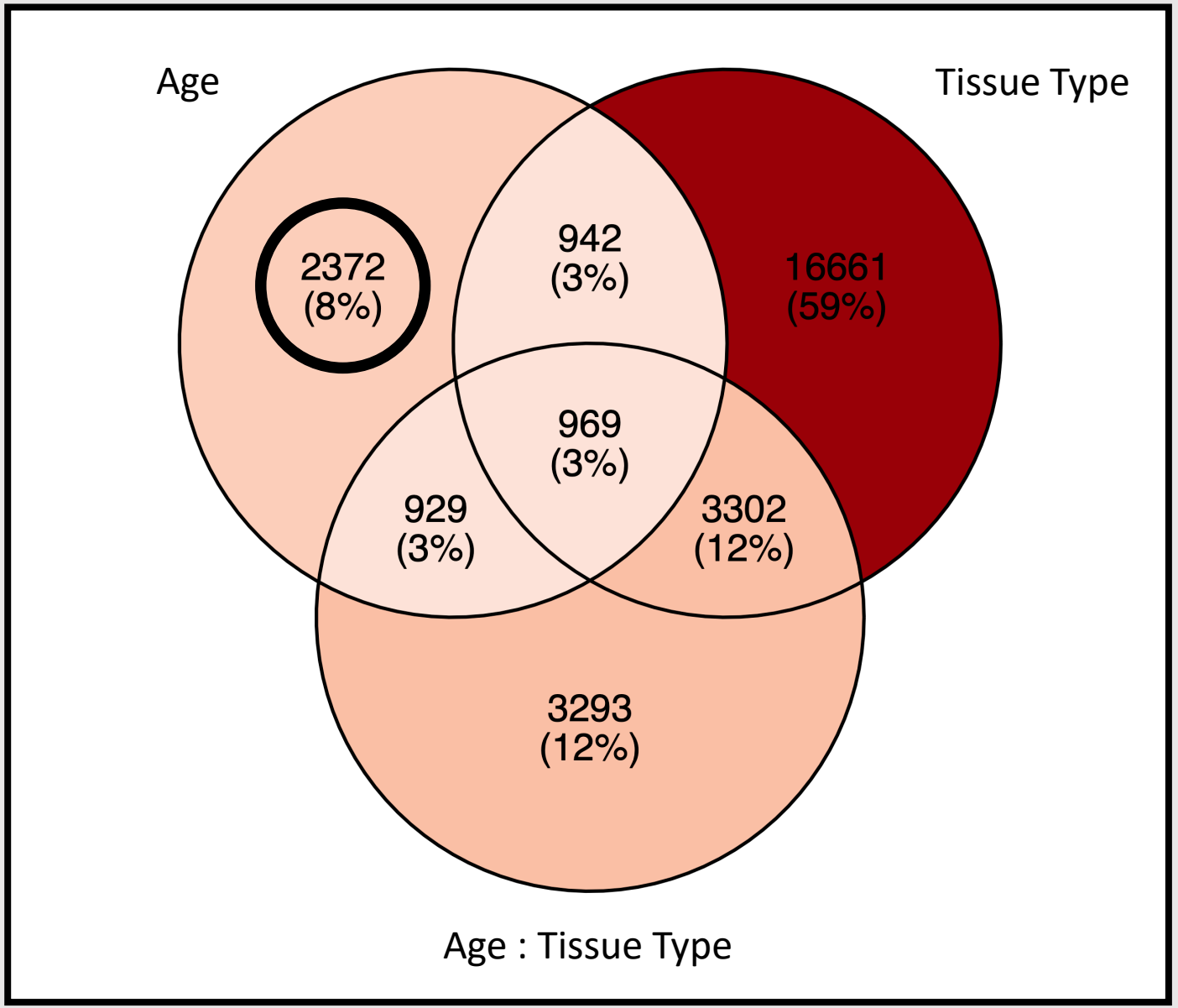
- 42 individuals (0-21 years)



Combined Tissue Clock

Bayesian GLM

- Age as fixed factor
- *Tissue type as fixed factor*
- Sample as random factor
- $\frac{\#\ methylated\ reads}{\#\ total\ reads}$ as response
- 95% HPDI's



Combined Tissue Clock

- Leave-one-out
- 62 CpG sites

Fin Clip:

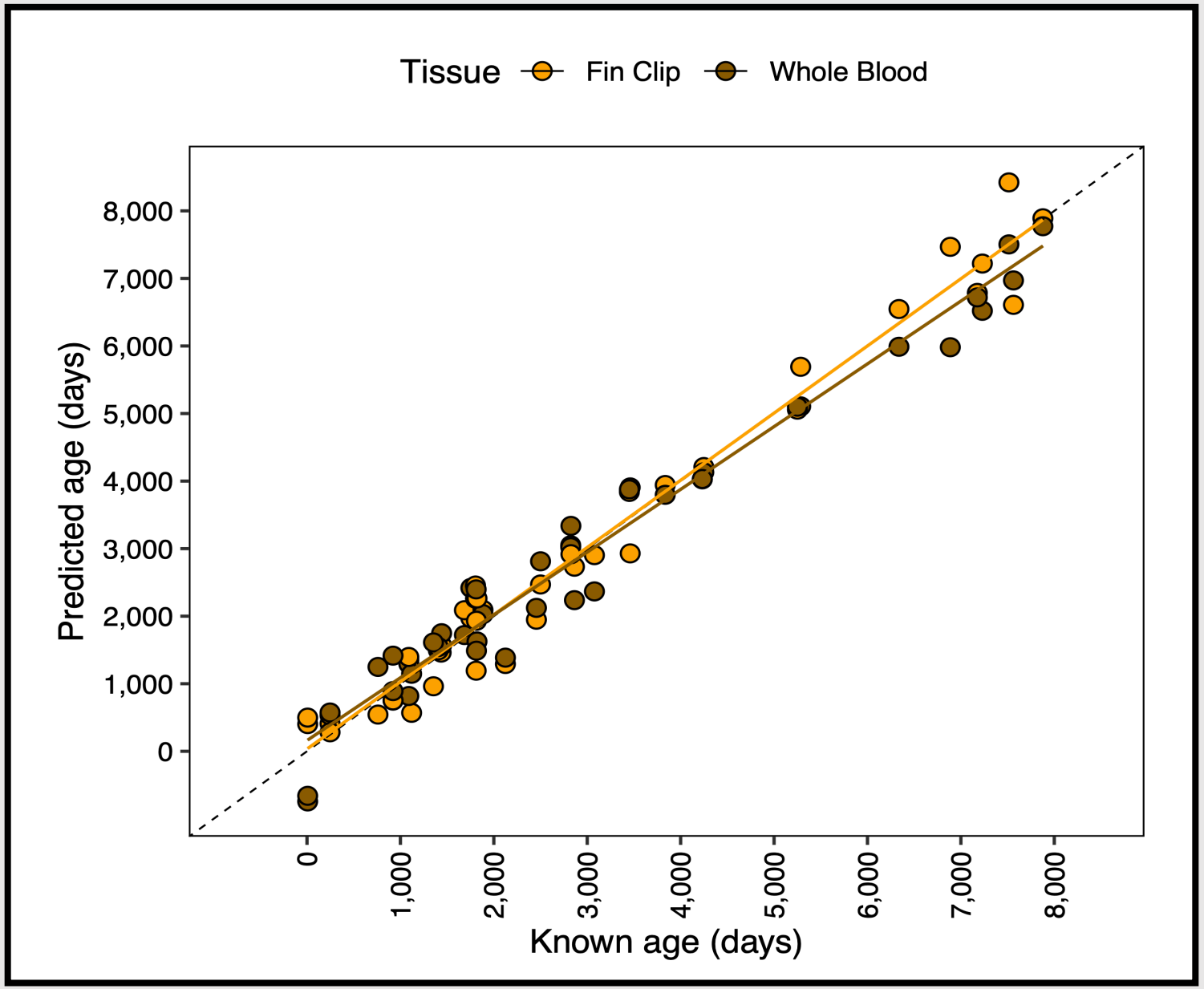
R^2 = 0.97

MAE = 309.0 days

Whole Blood:

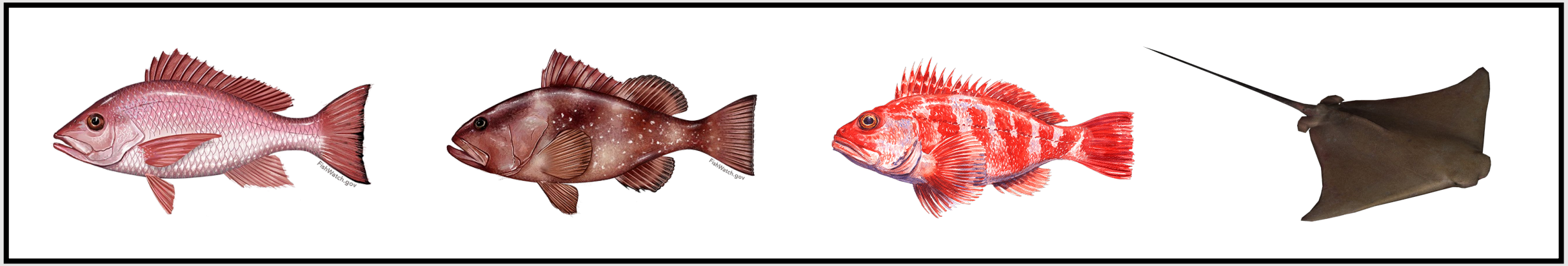
R^2 = 0.97

MAE = 354.4 days



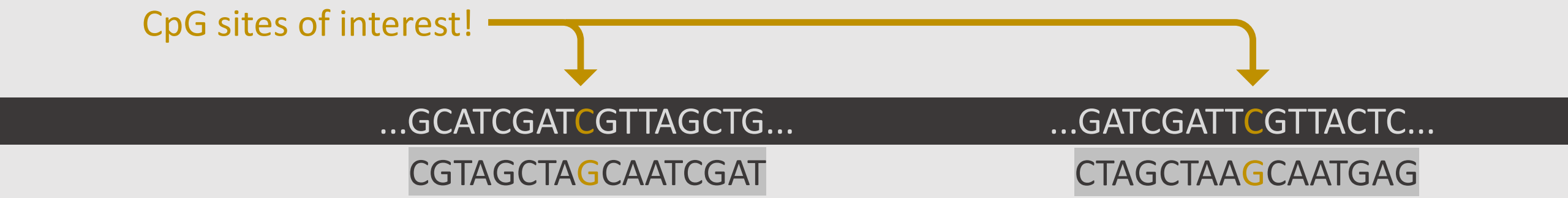
What we've learned so far…

- Accurate epigenetic clocks can be developed for wild-caught fishes
- Inclusion of biological info (length, sex) can enhance accuracy and precision
- Bayesian GLMs are flexible (removing unwanted variation, multi-tissue clocks)
- Multi-tissue clocks can be developed, though likely not as accurate



Production Ageing

Once epigenetic clocks are developed, design panels of primers (25 bp in length) to target age-correlated CpG sites



→ Genotyping-in-thousands by sequencing (GTseq)

- Low cost, high-throughput

Production Ageing

Timeline

- *Labwork*: 1,000 samples every 2 weeks per technician
- *DNA Sequencing*: ~2 weeks of “waiting”
- *Analysis*: 1 day to generate age estimates for all 1,000 samples

Cost

- $14 per sample for labwork (start to finish)

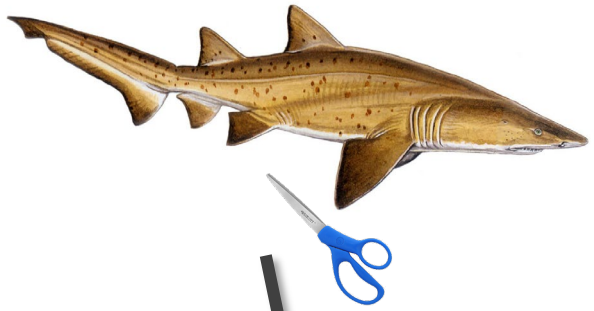
Requirements

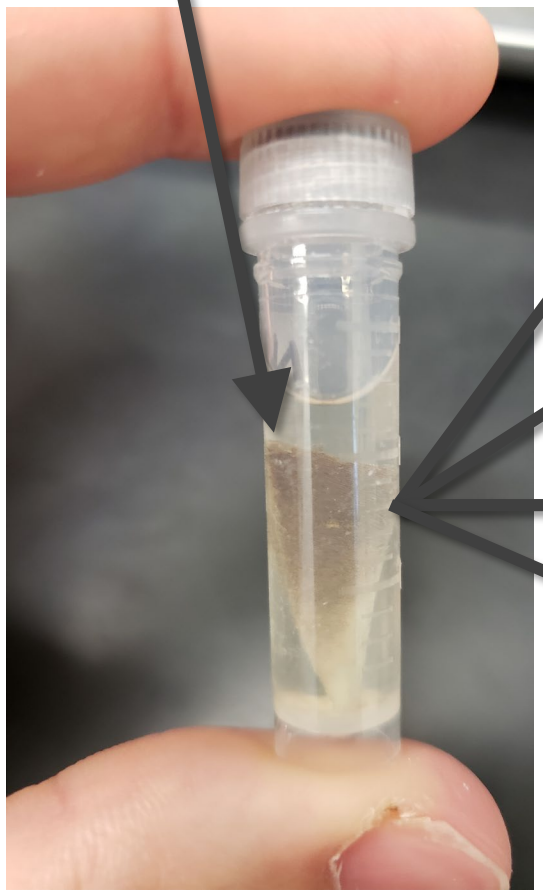
- Typical genetics lab (extract DNA, run/image gels, perform PCR)
- No “clean” room necessary

Potential Benefits

- More time- and cost-efficient generation of age estimates
  - $14 per sample
  - Age thousands of individuals per month per technician
- Accurate/precise for difficult to age species
- Non-destructive sampling
- Field sampling fast and easy

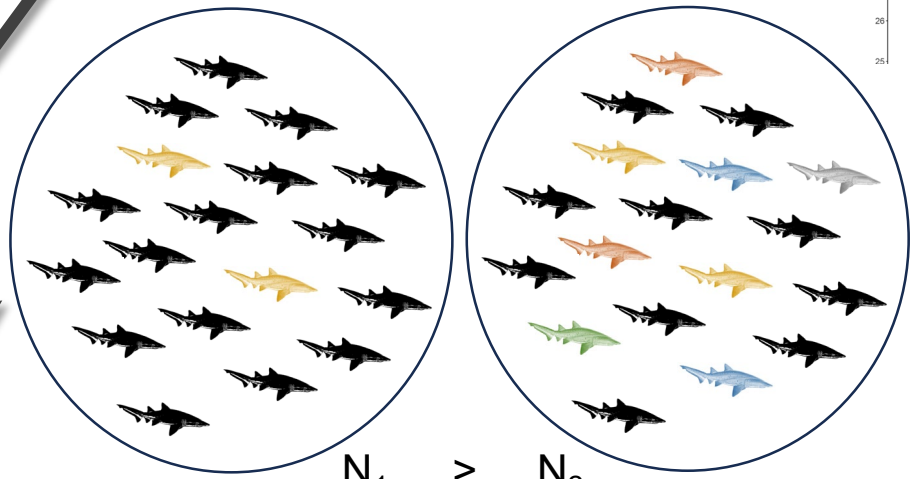
The Genomic Toolbox





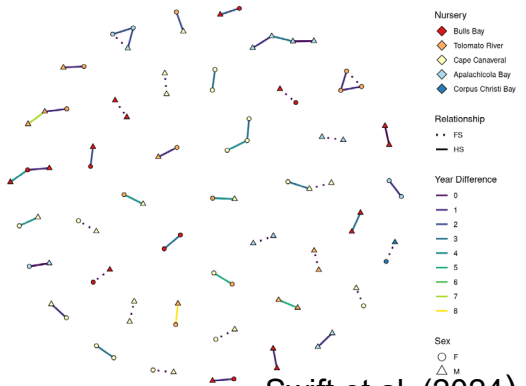
Population Structure

Abundance (CKMR)



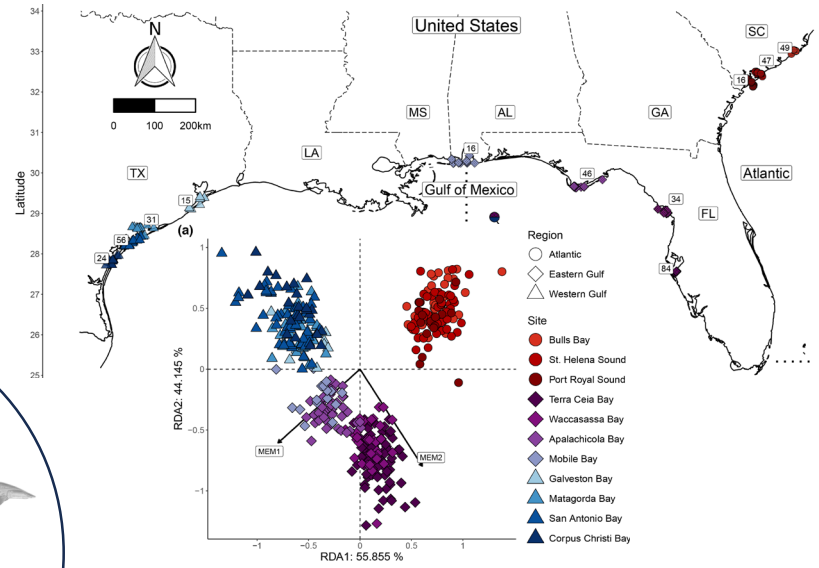
$N_1 > N_2$

Behavior

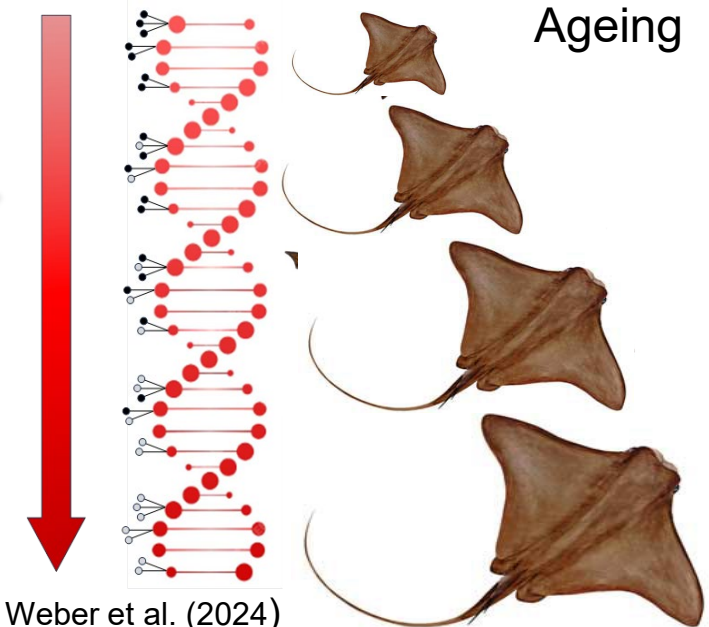


Swift et al. (2024)

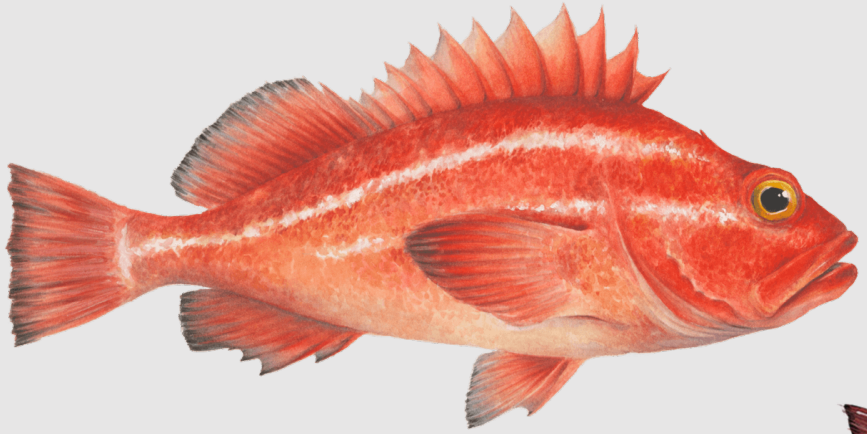
Swift et al. (2022)

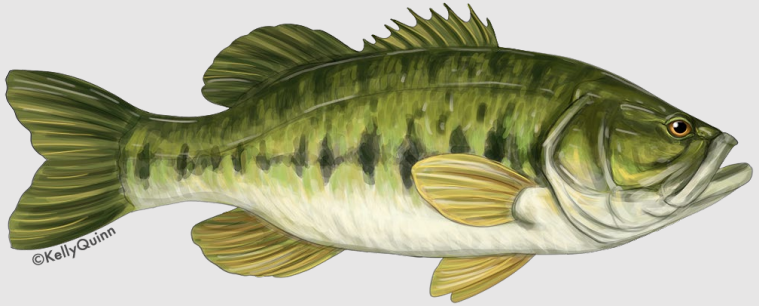


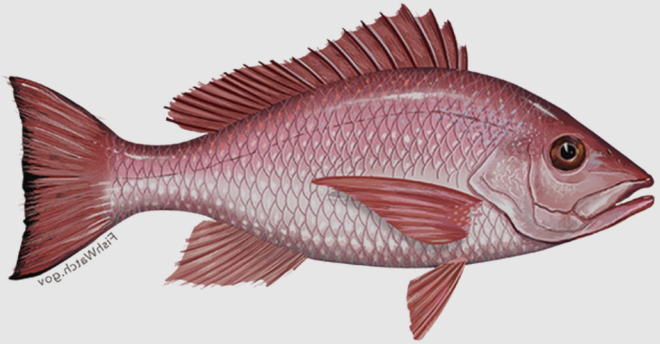
Ageing

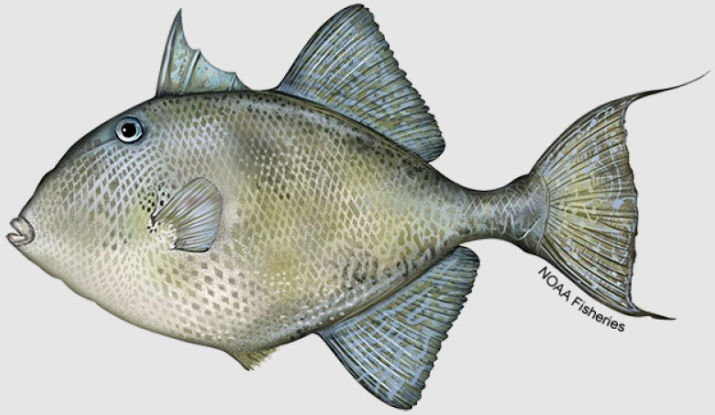


Weber et al. (2024)

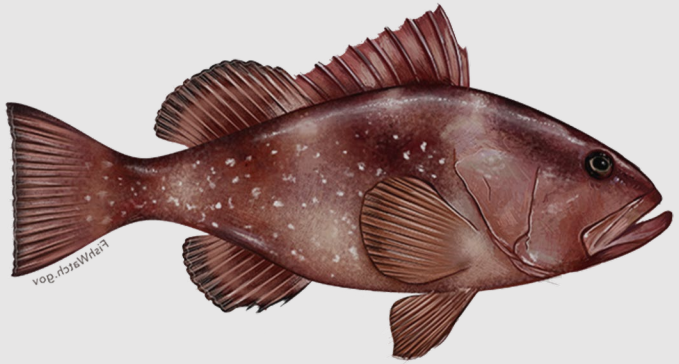








Thank you!



Things to Consider…

- Epigenetic clocks can only be as accurate as age estimates used to construct them
  - Age validation is important
- Epigenetic clocks may require re-calibrating over time
  - Subsample otoliths?

Removing Unwanted Variation

using the Bayesian GLM

- Removing CpG sites with tissue type relationship removes tissue-specific signal