

Compleasm in VGP Phase I

Tom Brown



**Leibniz Institute for Zoo
and Wildlife Research**
IN THE FORSCHUNGSVERBUND BERLIN E.V.

BeGenDiv



BUSCO -> Compleasm

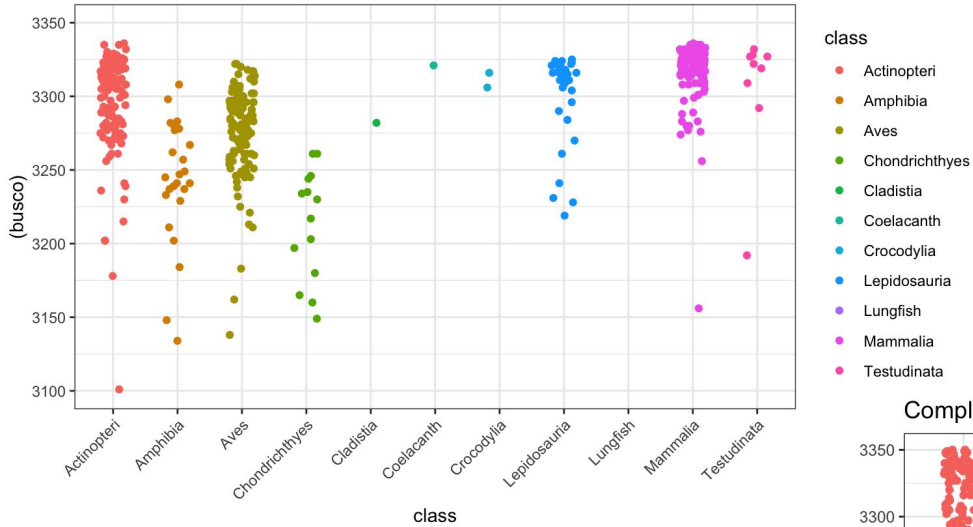


Comparing statistics from running BUSCO in –miniprot mode and Compleasm using the vertebrata_odb10 database

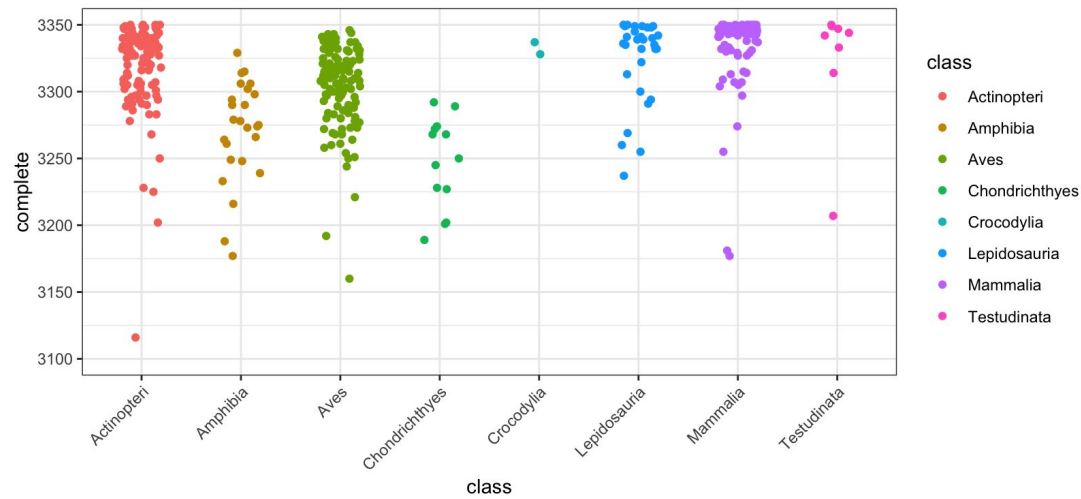
While BUSCO reports number of genes containing stop codons, Compleasm reports the number of frameshift events in complete matches. Here I have counted the number of BUSCO genes found with at least one frameshift event

Gene Completeness per Class

BUSCO completeness

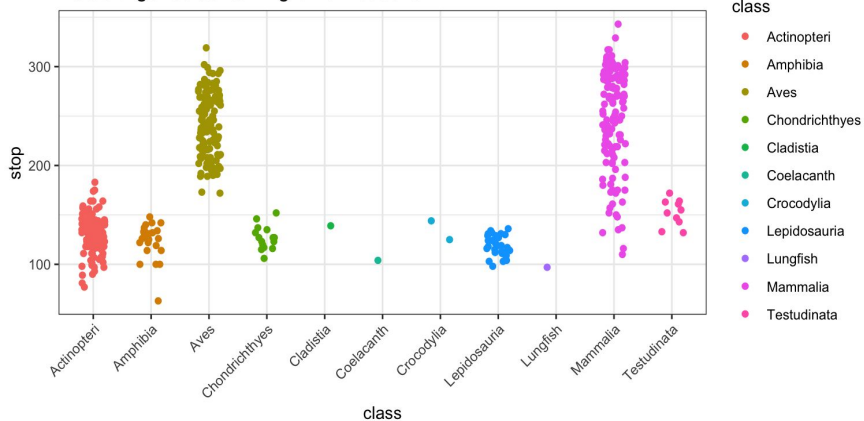


Compleasm complete genes

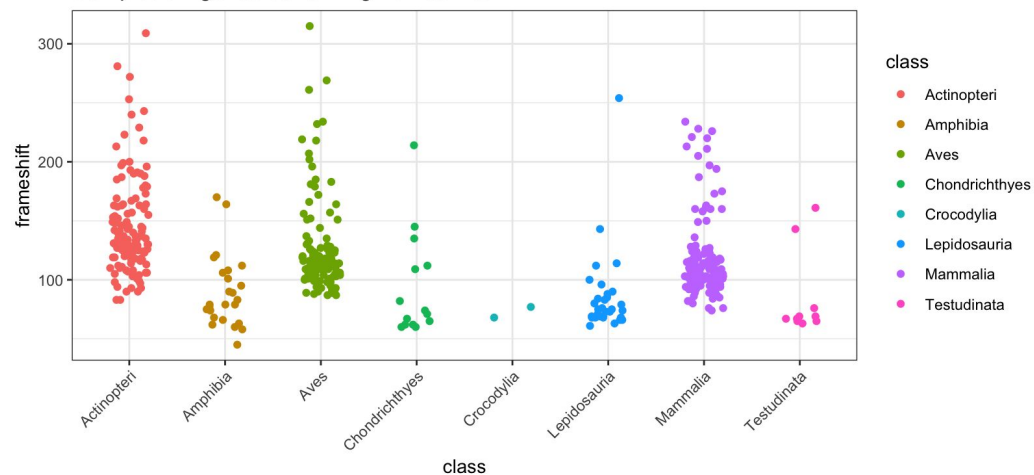


Stop Codons vs Frameshifts

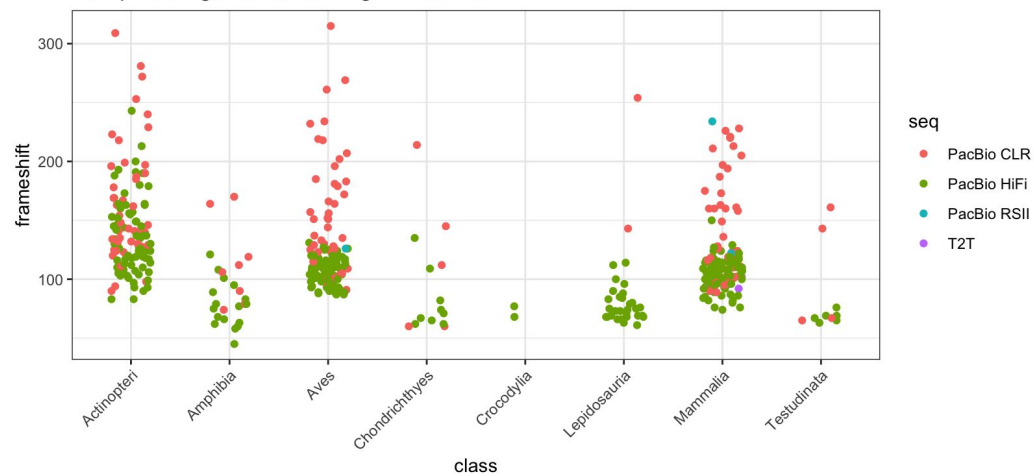
BUSCO genes containing STOP codons



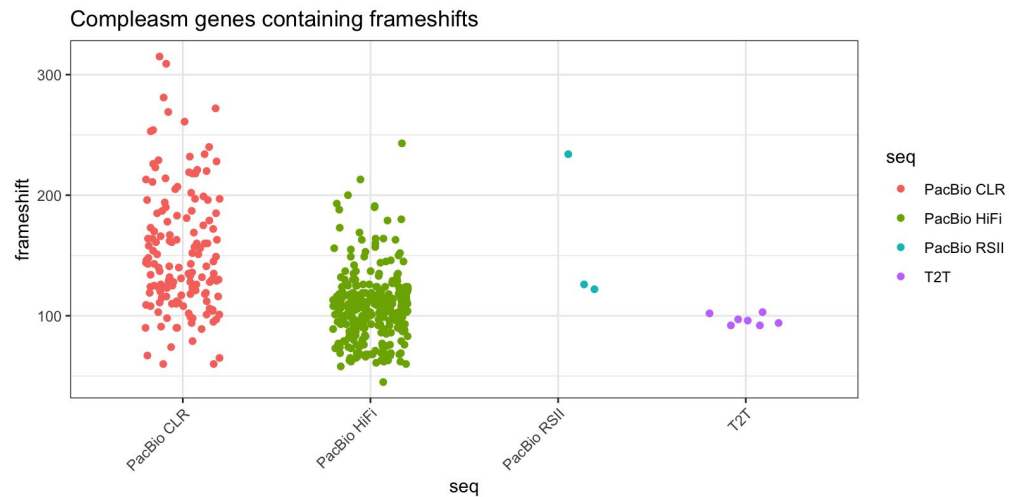
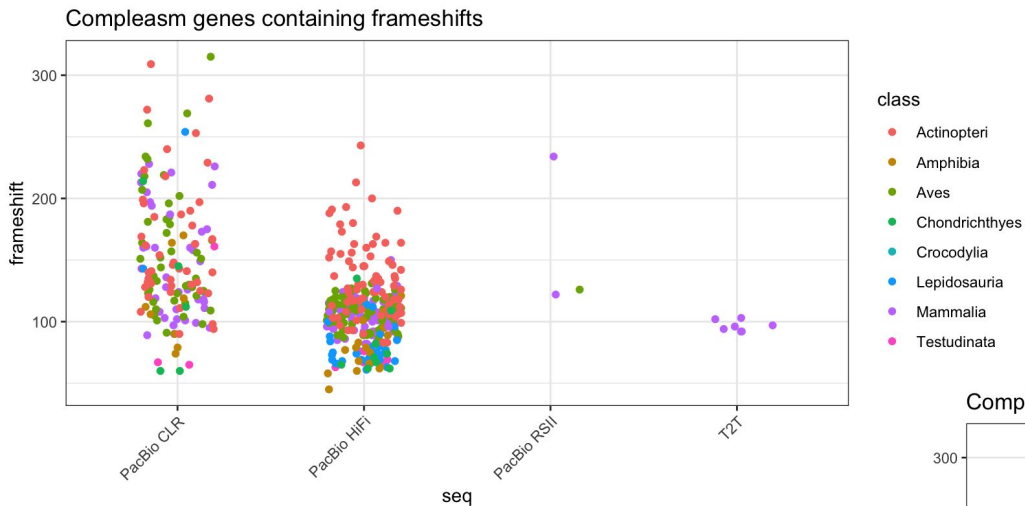
Compleasm genes containing frameshifts



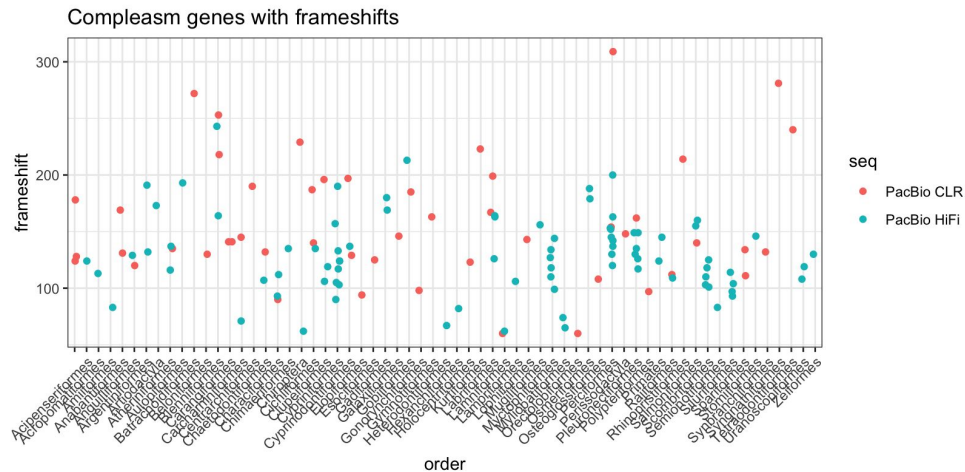
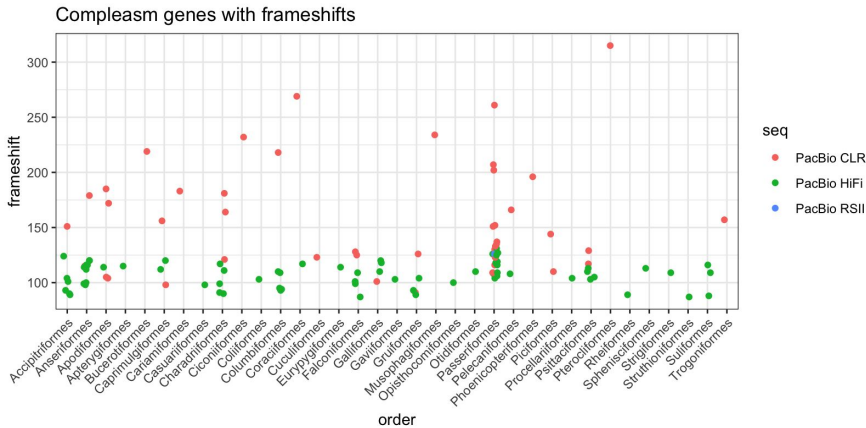
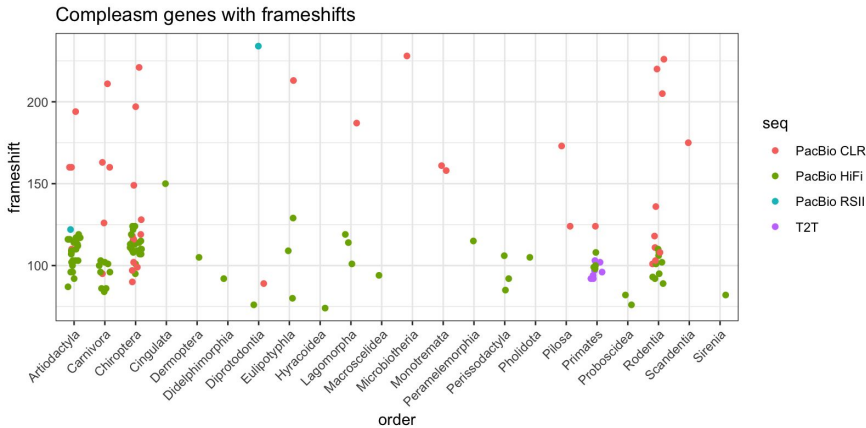
Compleasm genes containing frameshifts



Frameshift events - per sequencing technology

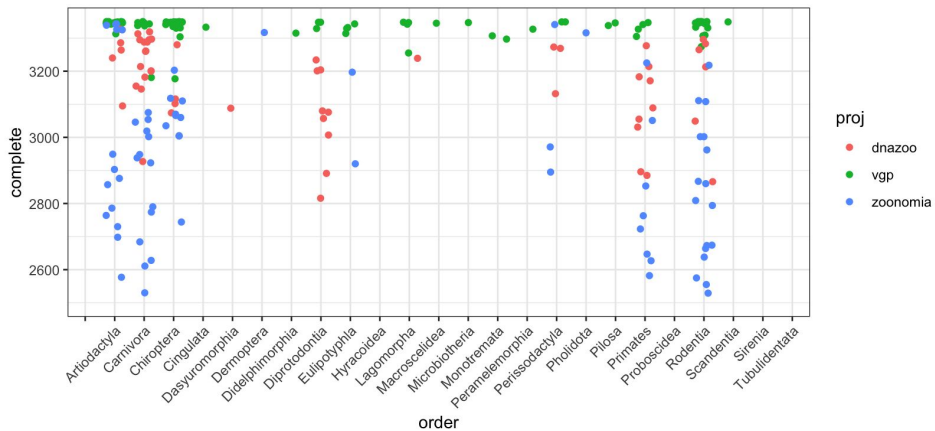


Frameshift events in mammals, birds and fish

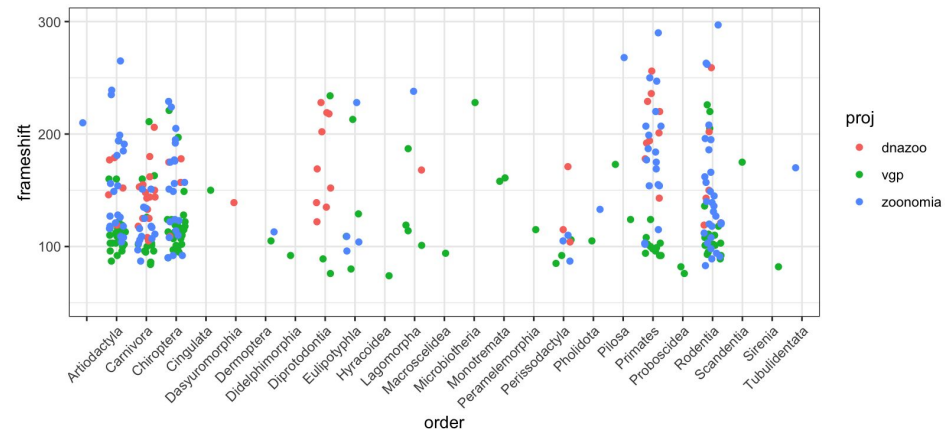


Compleasm - VGP vs DNAZoo vs Zoonomia

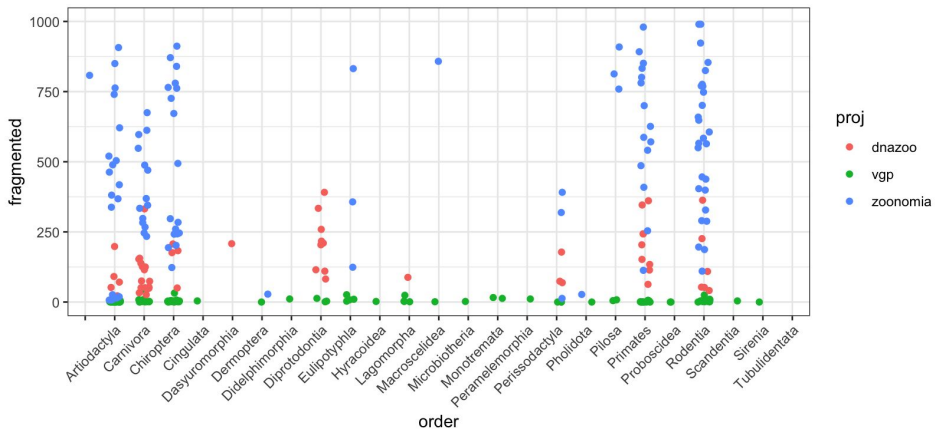
Compleasm complete genes



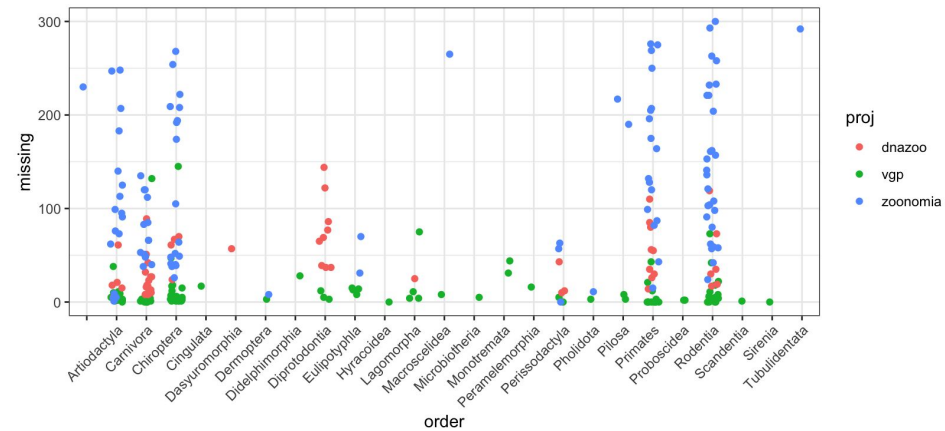
Compleasm genes containing frameshifts



Compleasm fragmented genes



Compleasm missing genes

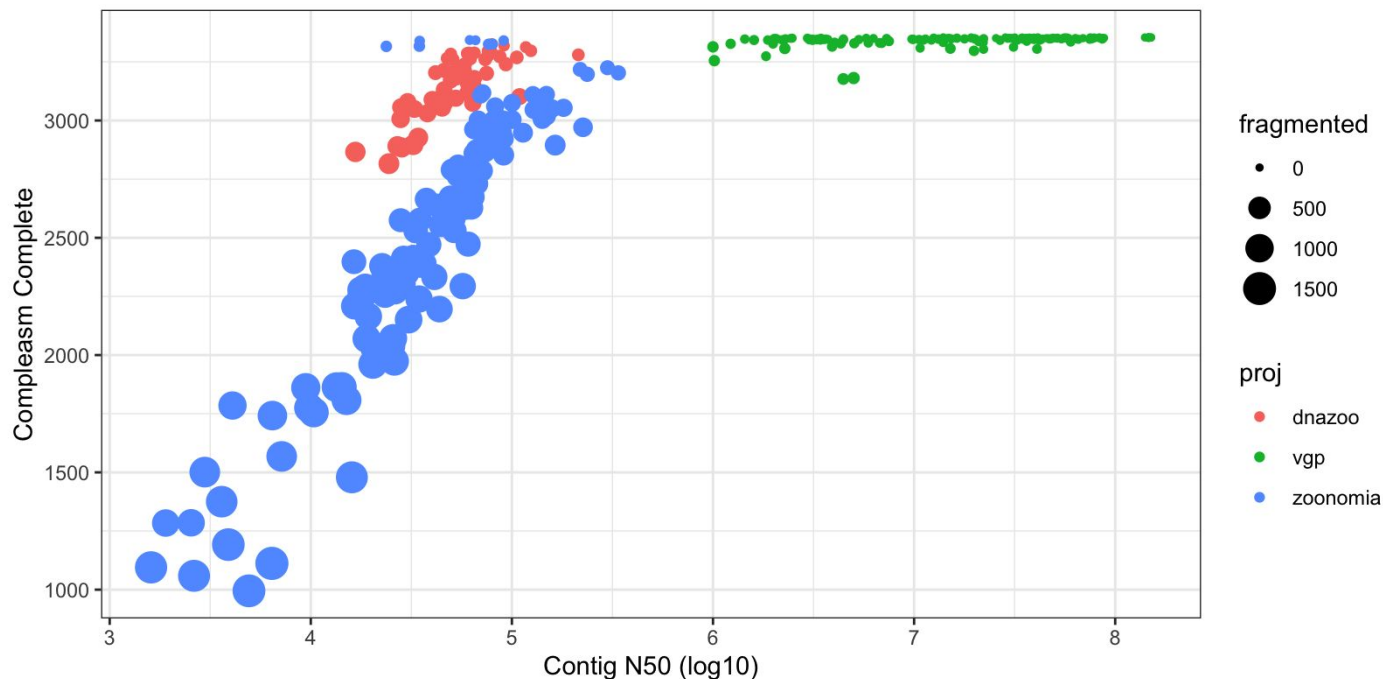


Compleasm & Contiguity - VGP vs DNAZoo vs Zoonomia

Mammalian genomes

Compleasm completeness vs Contig N50

Coloured by project, size is number of Fragmented genes from Compleasm



A scatter plot showing the relationship between the class of an animal and its log10(Length). The x-axis is labeled 'class' and includes categories: Actinopteri, Amphibia, Aves, Chondrichthyes, Crocodylia, Lepidosauria, Mammalia, and Testudinata. The y-axis represents log10(Length) and ranges from 0 to 10. Data points are colored by class: Actinopteri (red), Amphibia (brown), Aves (green), Chondrichthyes (teal), Crocodylia (blue), Lepidosauria (cyan), Mammalia (purple), and Testudinata (pink). Most classes show a positive correlation between log10(Length) and an unobserved variable, with Mammalia and Testudinata having the highest values (around 8-9) and Chondrichthyes having the lowest (around 4-5). There are several outliers, particularly in the Actinopteri and Mammalia classes.

