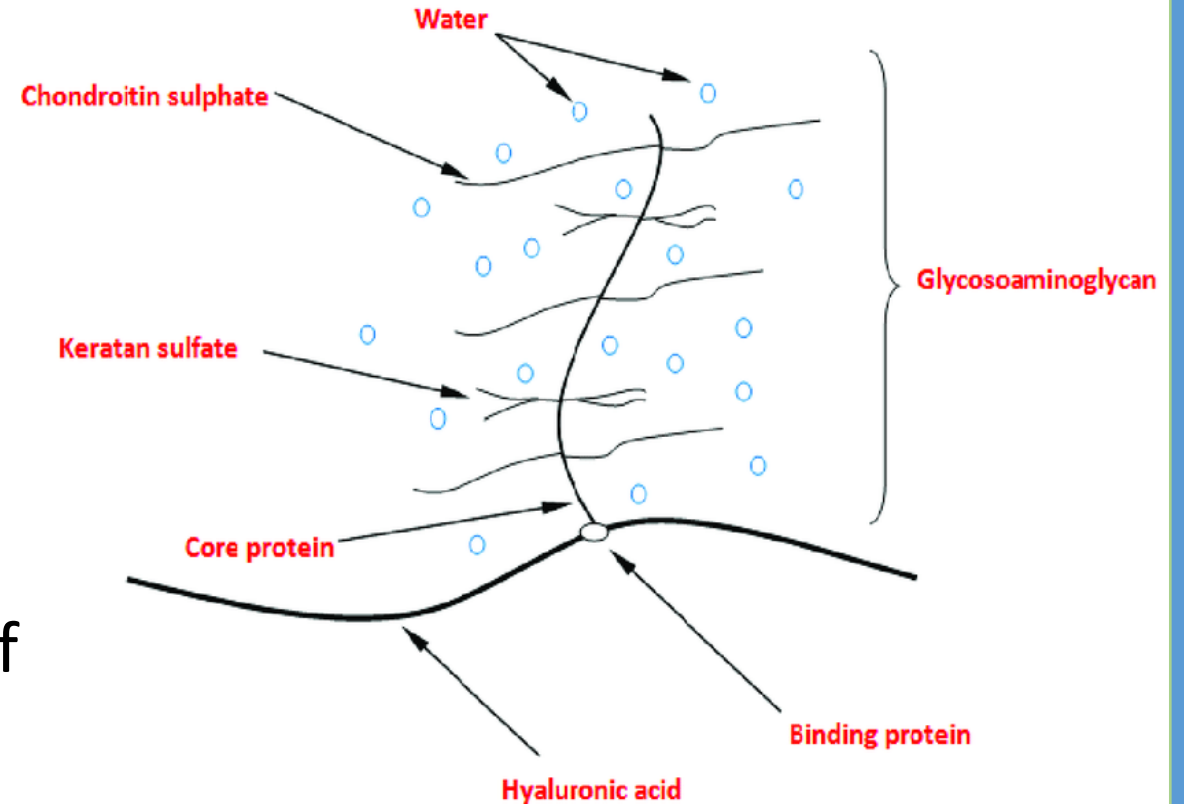


Mutation ratios of proteoglycan synthesis genes in organisms with different bone classifications

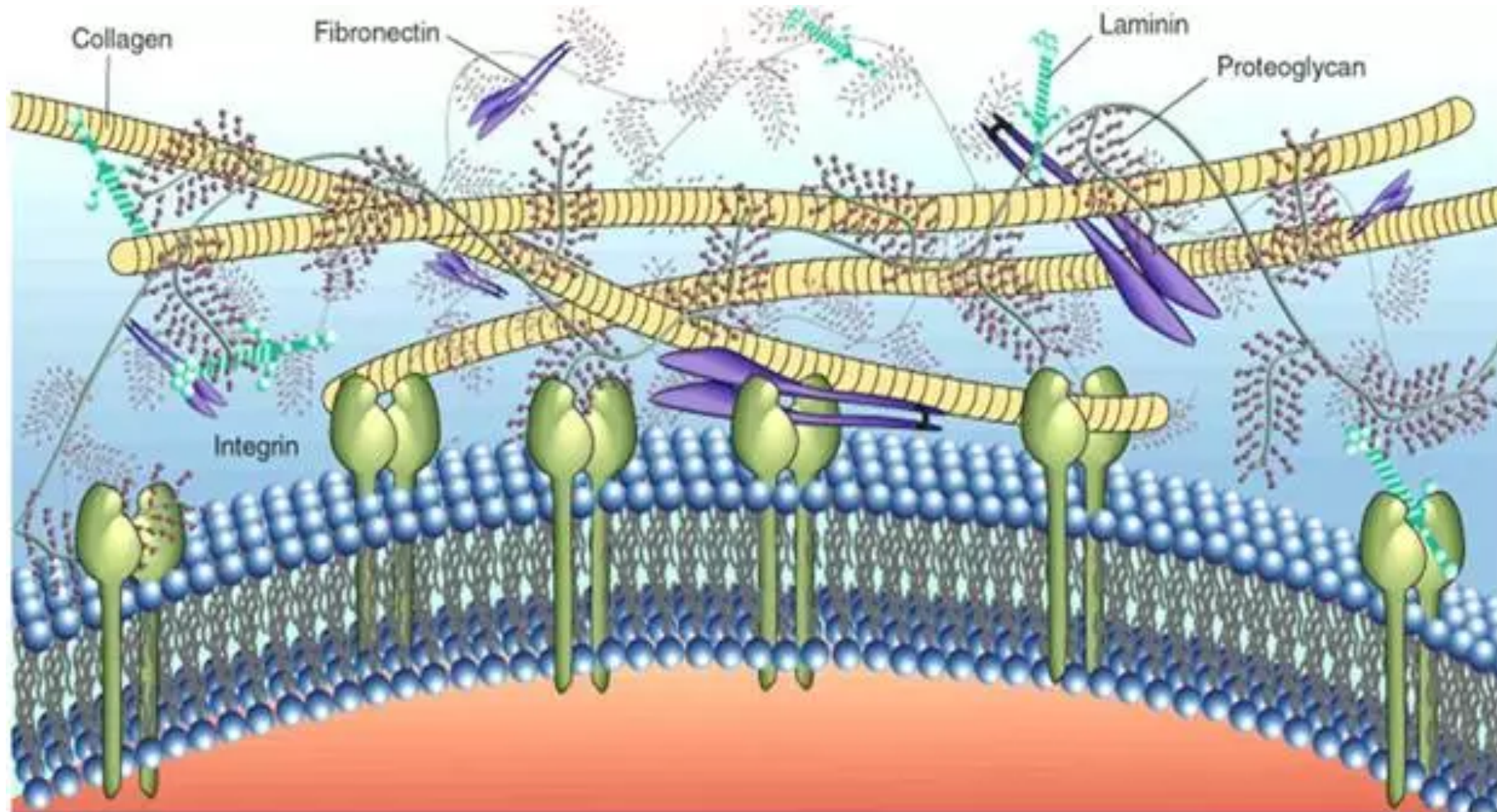
Flaviu Vadan, Ian McQuillan, Brian Eames

Proteoglycans (PGs)

- A class of macromolecules
 - Two components: core protein, glycosaminoglycans (GAGs)
- Specific function dictated by type of attached carbohydrate
- PGs are found in connective tissue
- Serve important functions extracellularly
 - Bind other PGs or collagen

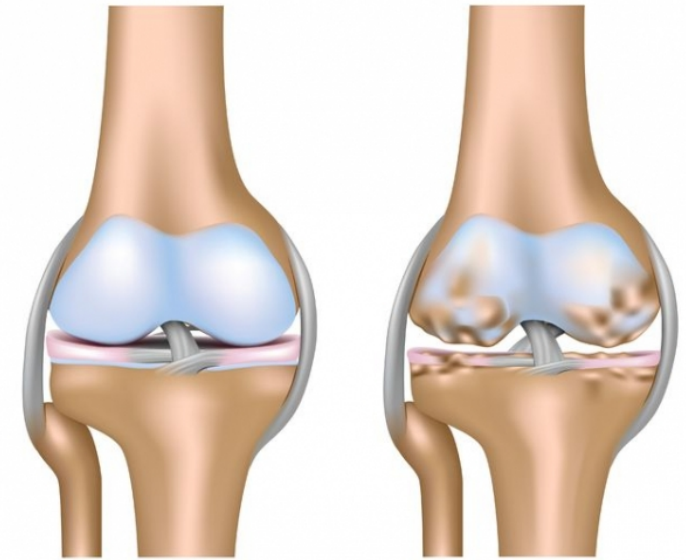


Proteoglycans



Proteoglycans

- Known to bind collagen to facilitate cartilage synthesis
 - Collagen type I, II, and III
- Loss of extracellular PGs associated with osteoarthritis
 - Wreak havoc in the ECM!
- Loss of specific PGs, such as B4GALT7, associated with Ehlers-Danlos syndrome
 - Loss of ability to find and replace different disaccharides on core proteins



Healthy knee joint

Osteoarthritis

Proteoglycans

- Current problem in skeletal evolution is the lack of understanding of evolution of PG genes
- Understanding may guide efforts towards studying genes in relation to their contribution towards osteoarthritis
- Can be studied using comparative genomics approaches

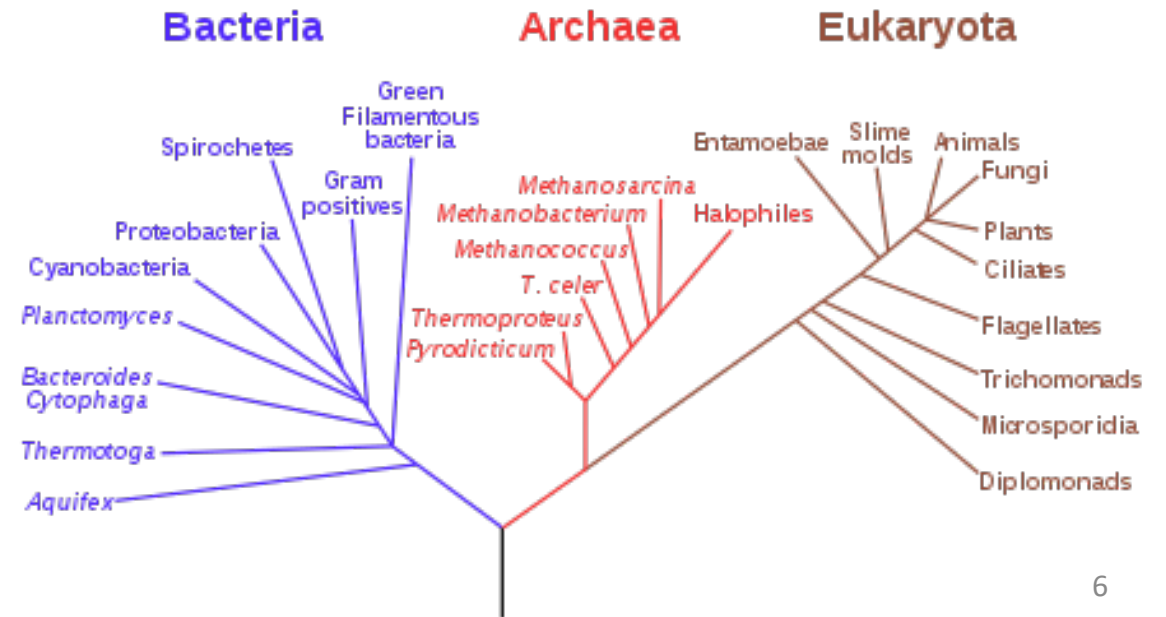
Approaches

- Sequence alignments
 - Essential for analyzing homology between sequences
- Phylogenetic profiling
 - Tree topologies suggestive of coevolution

Sequence alignment

```
A C T C G C A A T A T G C
A C T _ _ _ _ T T A T G C
A C T T G T C T T A T G C
A C T _ G _ _ T T A _ _ C
```

Phylogenetic tree



Approaches

- Quantifying changes in codon changes
 - Non-synonymous (dN) vs. synonymous mutations (dS)
 - Non-synonymous mutations change the amino acid sequence, synonymous mutations do not
- Statistical methods to quantify mutation ratios
 - $dN/dS < 1$ = conservation, $dN/dS \sim 1$ = neutral, $1 < dN/dS$ = mutation
 - Helps infer direction and magnitude of natural selection acting on protein-coding genes

Objectives

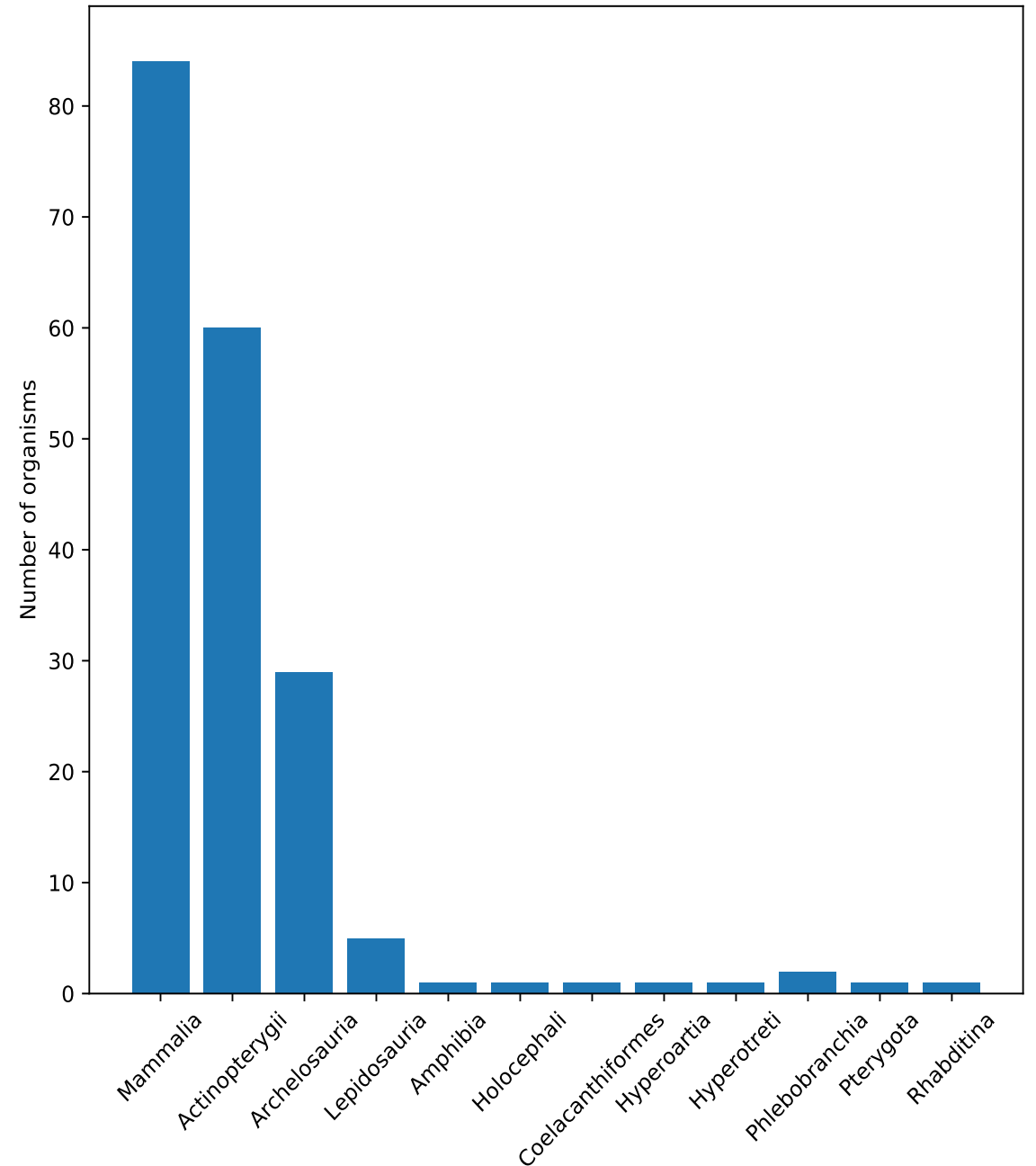
- Hypothesis: genes of the PG pathway showcase higher dN/dS in organisms that synthesize bone/cartilage
 - Transition from water to land = gravity vs. drag
- Study homology of PG genes
- Understand gene distribution across organisms
- Quantify mutation ratios

Data collection

- Collected 51 gene orthologs' sequences from Ensembl
 - Ensembl genomes are annotated and curated
- Organisms with automatically annotated genomes were removed
 - *Projection builds* are not manually curated

Taxa information

- 199 organisms collected, 188 used
- 12 organisms' genomes were *projection builds*
- Verified using NCBI's BLAST
 - Identified some organisms' genomes to include “missing” genes
- Bias towards vertebrates
- Selected a set of 21 representatives

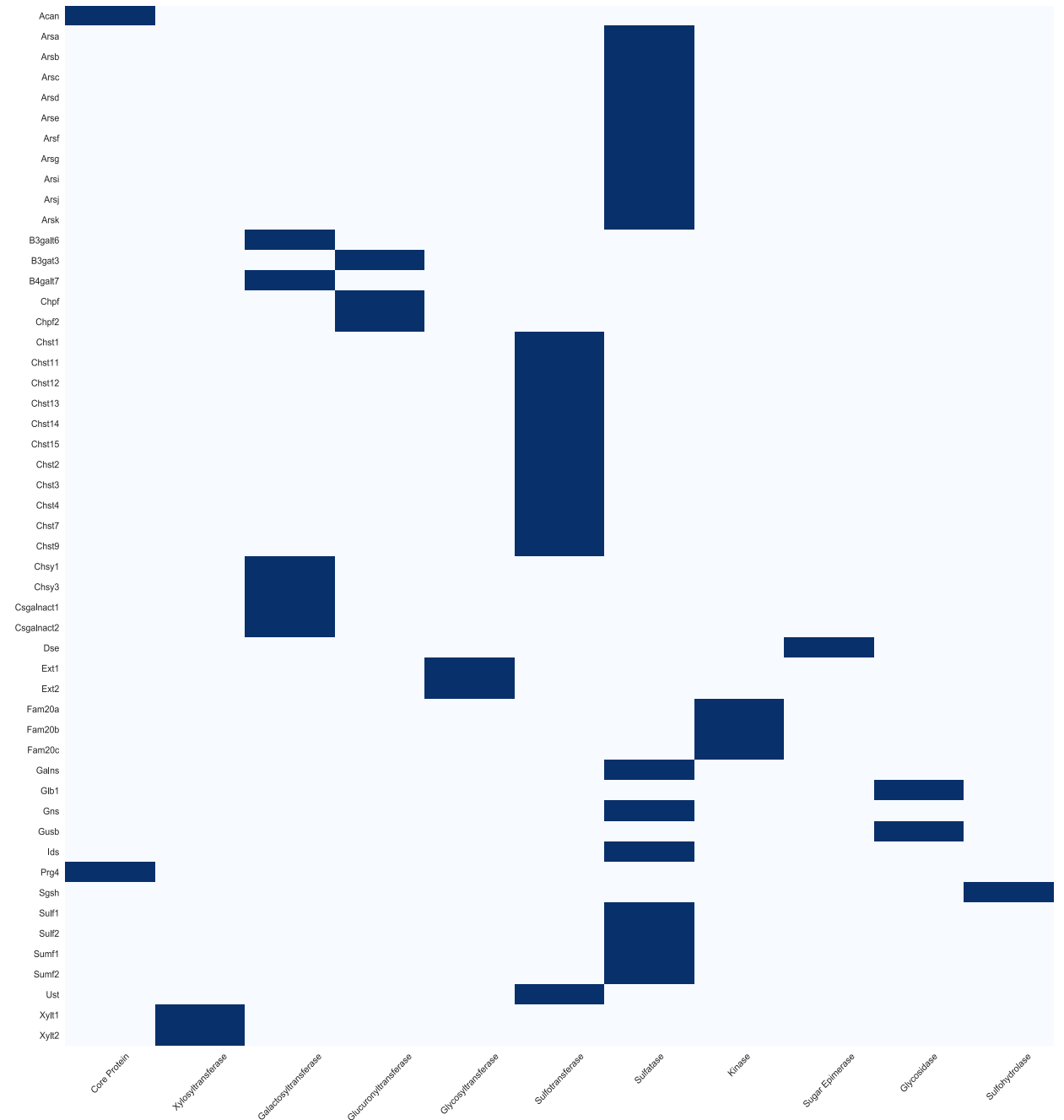


Organism representatives

Class	Organism	Cartilage	Bone	Reason
Mammalia	<i>Homo sapiens</i>	✓	✓	All 51 genes
	<i>Mus musculus</i>	✓	✓	Close to <i>Homo sapiens</i>
	<i>Ornithorhynchus anatinus</i>	✓	✓	Egg-laying mammal
Actinopterygii	<i>Danio rerio</i>	✓	✓	Model organism in Eames Lab
	<i>Oryzias latipes</i>	✓	✓	Cartilage only
	<i>Takifugu rubripes</i>	✓	✓	Cartilage only
	<i>Lepisosteus oculatus</i>	✓	✓	Cartilage only
Archelosauria	<i>Anas platyrhynchos</i>	✓	✓	Lives on land, flies
	<i>Gallus gallus</i>	✓	✓	Good genome, economic interest
	<i>Chrysemys picta bellii</i>	✓	✓	Lives on land and water
	<i>Crocodylus porosus</i>	✓	✓	Lives on land and water
Lepidosauria	<i>Notechis scutatus</i>	✓	✓	Smaller in size but still have bone
	<i>Anolis carolinensis</i>	✓	✓	Smaller in size but still have bone
Amphibia	<i>Xenopus tropicalis</i>	✓	✓	Xenopus with diploid genome
Holocephali	<i>Callorhynchus milii</i>	✓		Cannot turn cartilage into bone
Coelacanthiformes	<i>Latimeria chalumnae</i>	✓		Closely related to amphibians
Hyperoartia	<i>Petromyzon marinus</i>	✓		Jawless fish
Hyperotreti	<i>Eptatretus burgeri</i>	✓		Jawless fish
Phlebobranchia	<i>Ciona intestinalis</i>			Single org. available
Pterygota	<i>Drosophila melanogaster</i>			Important model org.
Rhabditina	<i>Caenorhabditis elegans</i>			Important model org.

Gene functions

- Diverse gene functions
 - Sulfatases, Kinases, Epimerases, etc.
- Multiple genes perform the same function
 - Suggests possibility of redundancy



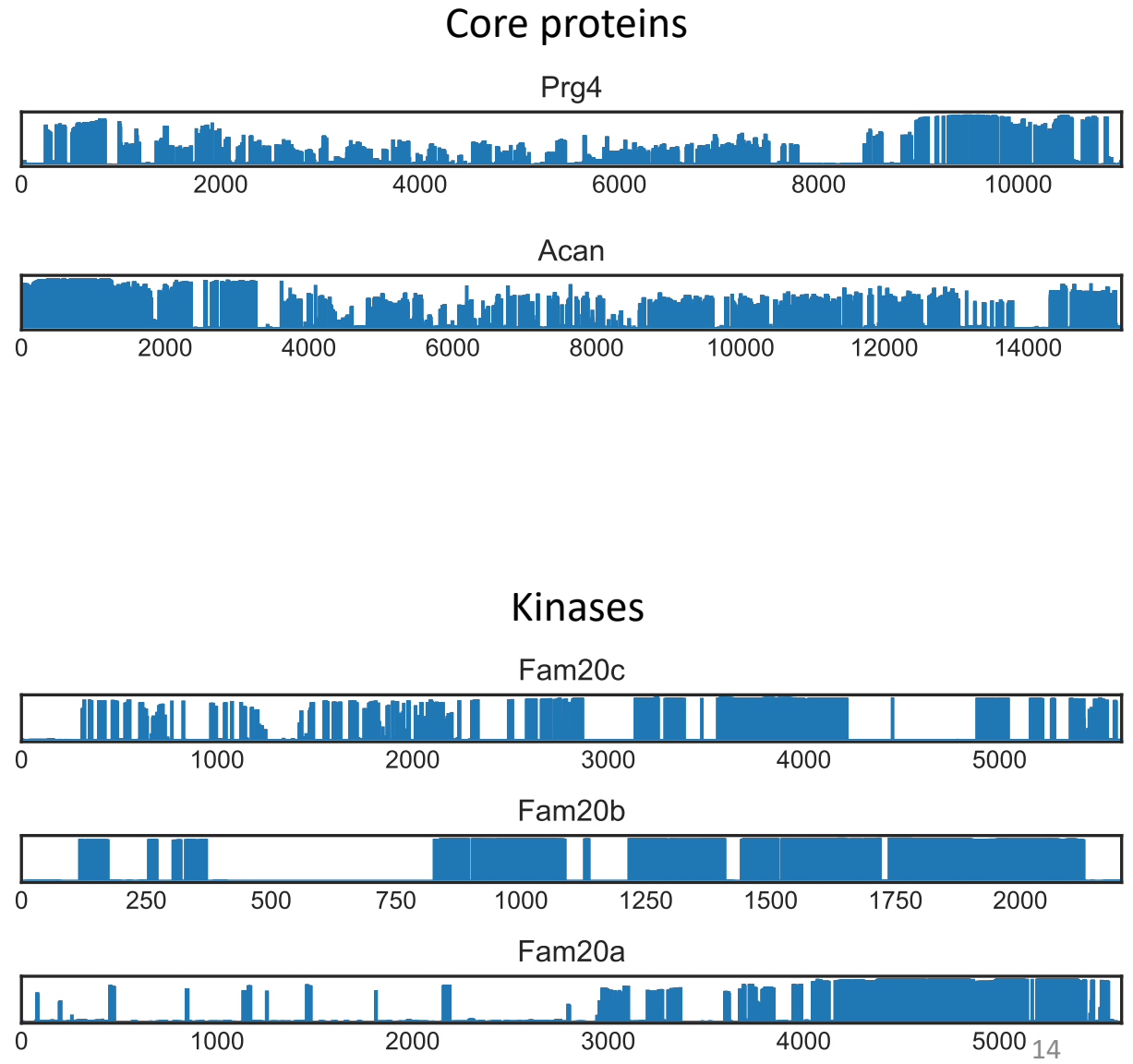
Multiple sequence alignments

- Multiple sequence alignments (MSAs) were created using EBI's *kalign*
- Defined a column similarity metric
 - Ratio between the number of non-gap characters and total number of characters in an MSA column
- Similarity used to understand conservation and visualize MSAs

```
A C T C G C A A T A T G C
A C T _ _ _ T T A T G C
A C T T G T C T T A T G C
A C T _ G _ _ T T A _ _ C
```

Multiple sequence alignments

- Grouped genes showcase some similarities
- Core proteins showcase conservation across MSA
- Kinases showcase sparsity
 - Important regions are conserved (R304)



Phylogenetic profiling

- Phylogenetic relationships between organisms were identified using NCBI's Common Tree
 - Allowed the creation of a central phylogenetic tree
- 51 trees were generated, one for each gene
 - Organisms that do not have the gene represented by the tree were marked
- Trees were paired to reduce dimensionality
 - Used Jaccard's index (intersection/union)
 - $S1 = \{1, 2, 3\}$, $S2 = \{2, 3, 4\}$
 - $S1 \cap S2 = \{2, 3\}$, $S1 \cup S2 = \{1, 2, 3, 4\}$
 - Index = $2/4$

- # Arsf
-
- 16

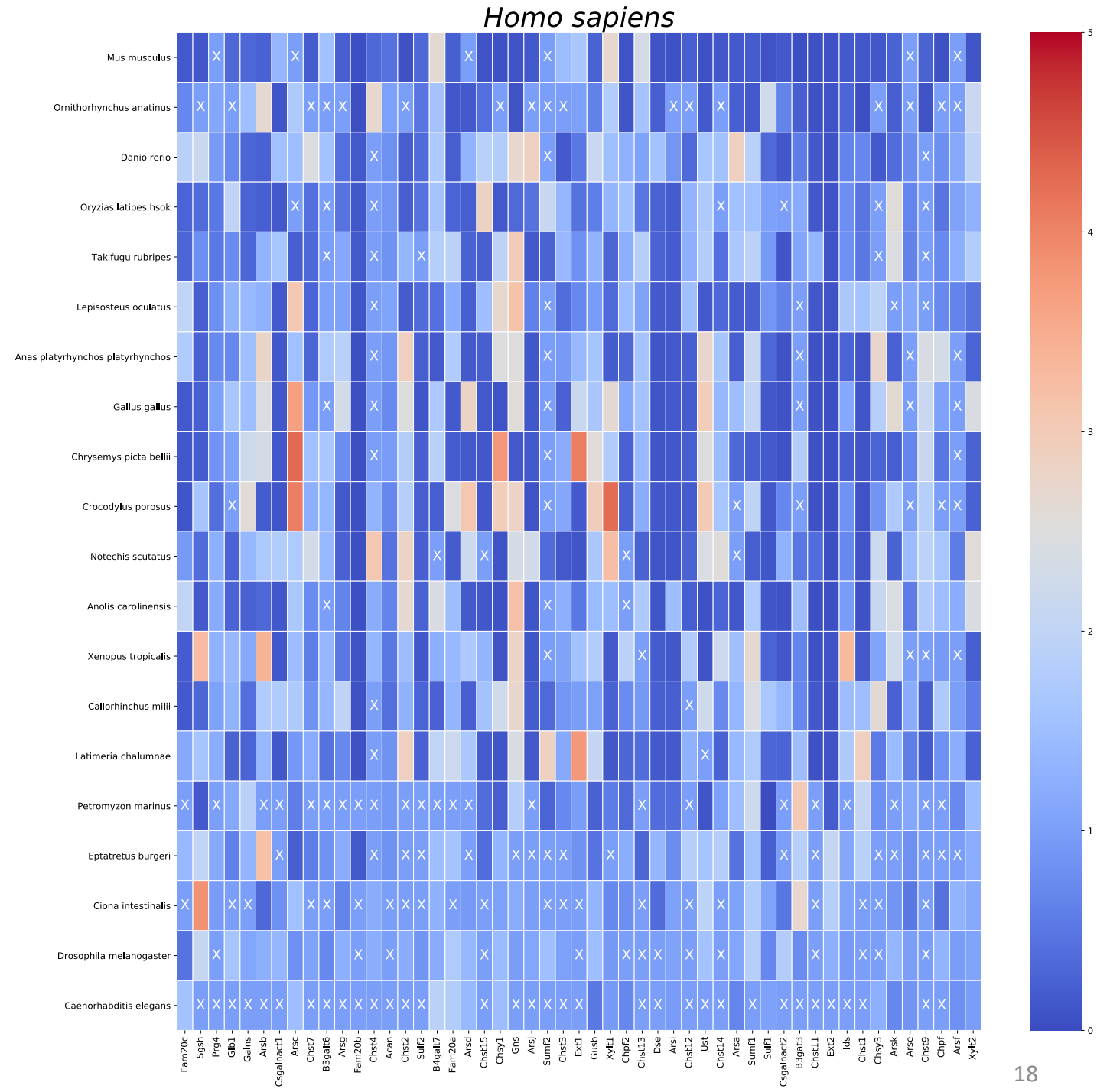
dN/dS

- Advantages of dN/dS
 - Popular method, easy to understand
 - Compared to other population genetic methods, only assumes equal mutation rates of nucleotide loci
- Disadvantages of dN/dS
 - Potential systematic bias because of assumption
 - Many reasons for sequences to be different

$$d_N = -\frac{3}{4} \ln \left(1 - \frac{4p_N}{3} \right) \quad d_S = -\frac{3}{4} \ln \left(1 - \frac{4p_S}{3} \right) \quad dN/dS = \frac{d_N}{d_S}$$

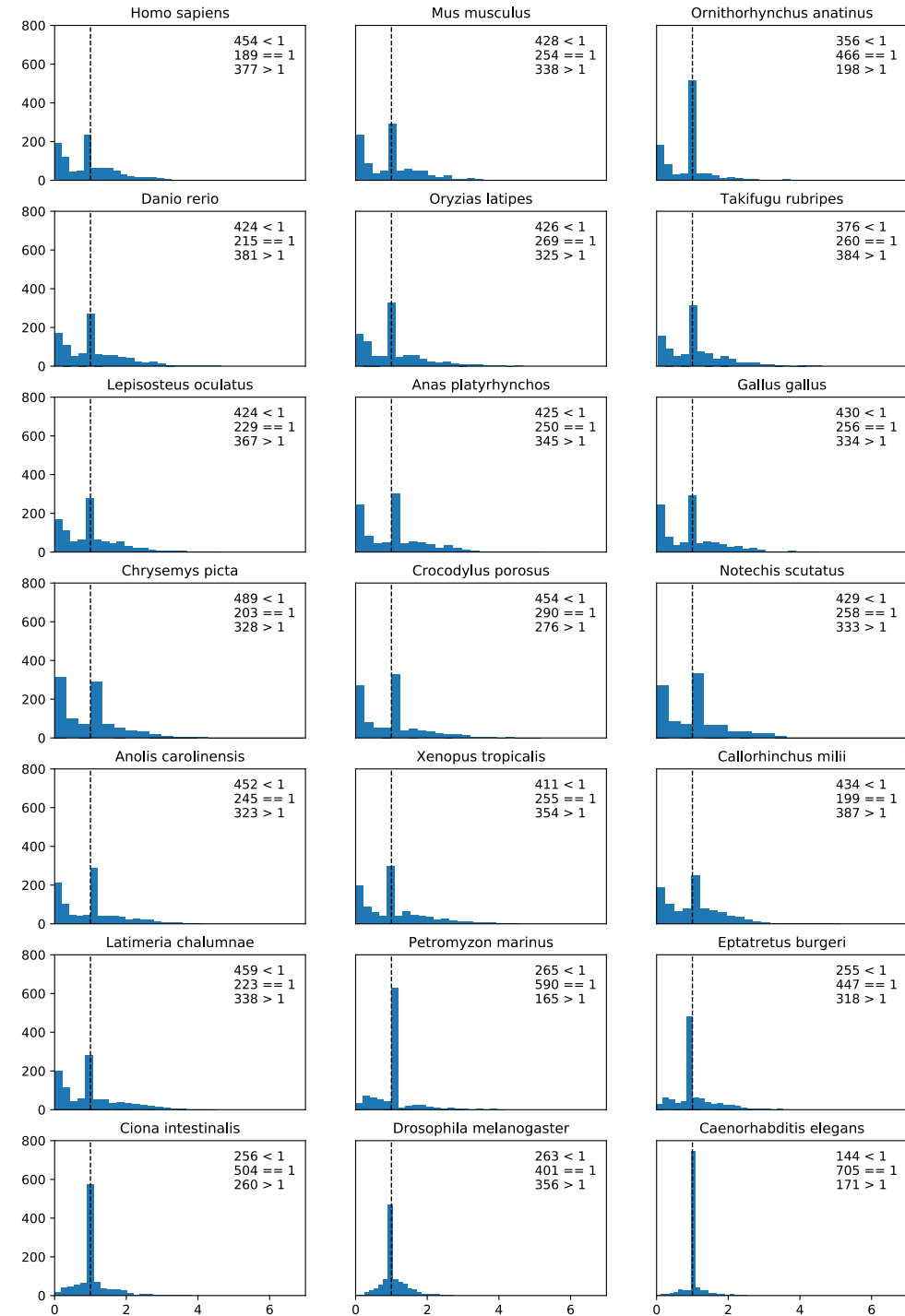
Independent dN/dS

- Independent gene differences consistent with organisms' complexity
 - Compare human and mouse
 - Compare human and nematode

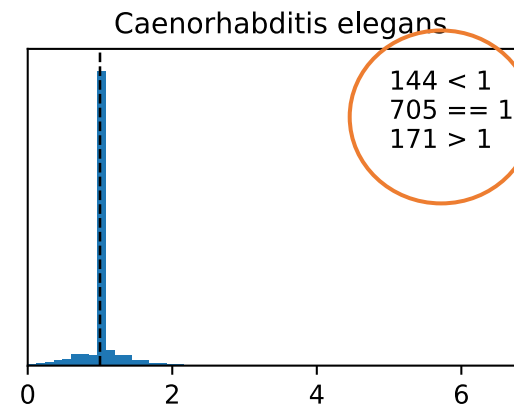
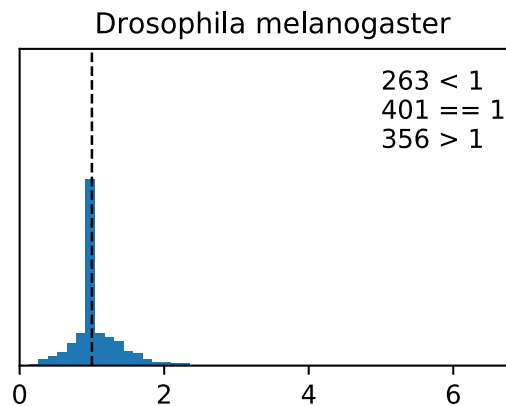
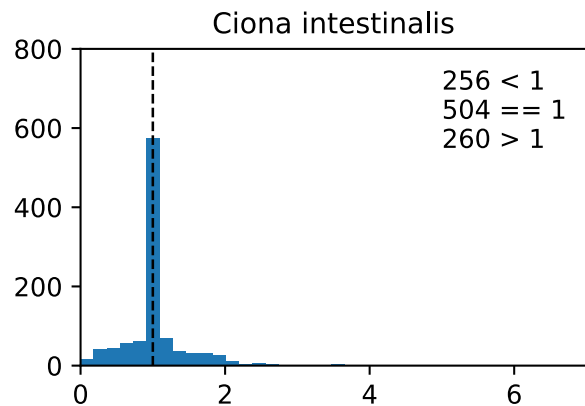
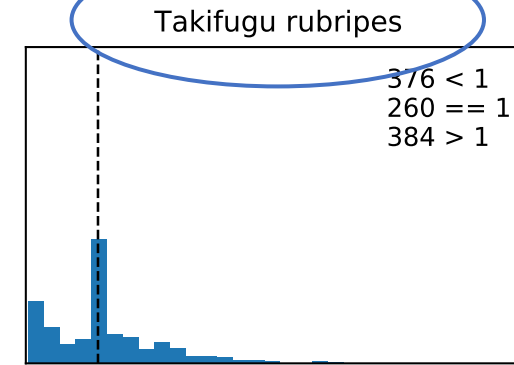
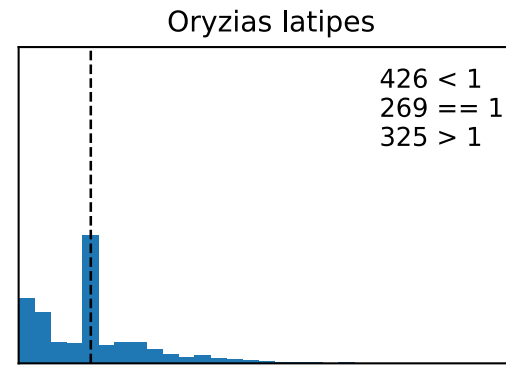
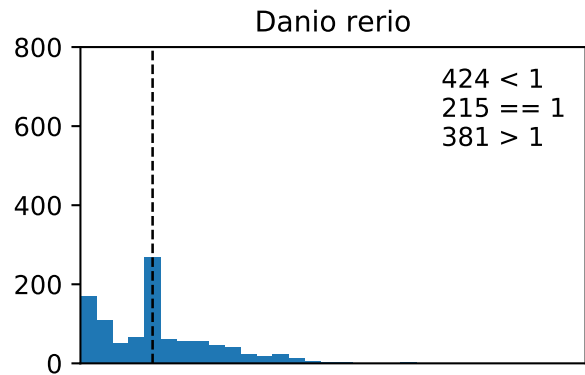
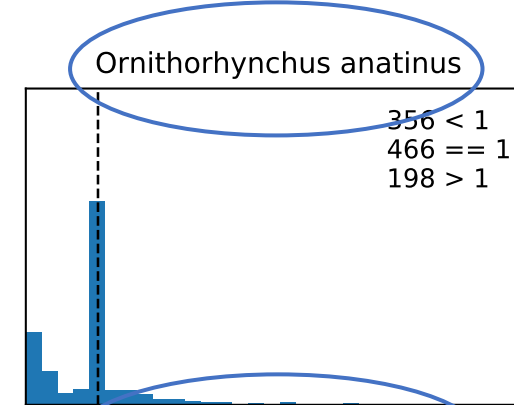
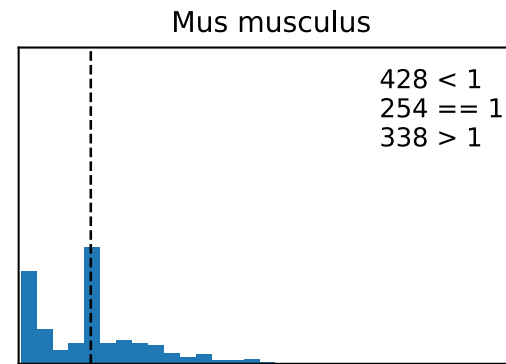
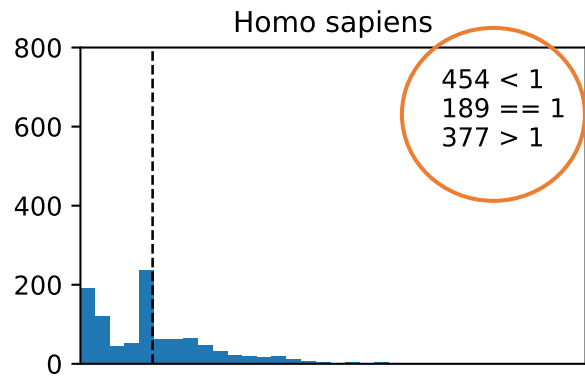


dN/dS summary

- Hypothesis predicted higher mutation rates in organisms with bone, opposite trend identified
- Bimodal distribution caused by comparing two types of organism classes
 - Higher complexity vs. lower complexity

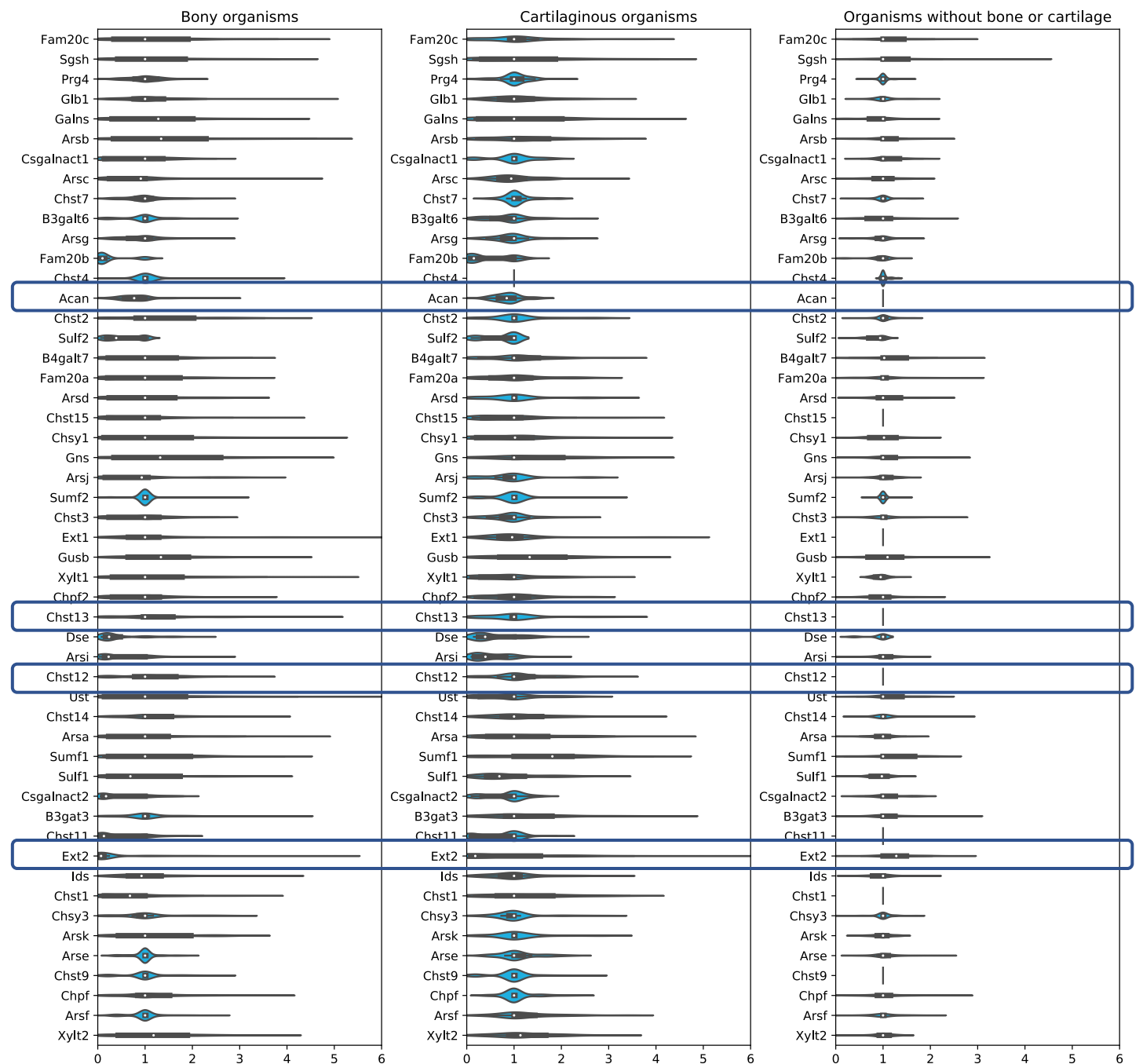


dN/dS summary



dN/dS grouped

- Grouped based on bone presence
- No significant overall differences between distributions
 - Loss of genes in organisms without bone/cartilage
- Range of ratios in organisms without bone/cartilage generally smaller



Summary

- Approx. 2000 lines of code*
 - Easy to replicate study or apply to other data
- 51 MSAs, 59 phylogenetic trees, 21 dN/dS plots, 3 summary plots, various visualizations to understand data

<https://github.com/flaviuvadan/proteoglycan-pathway-evolution>

Conclusion

- Initial hypothesis was orgs. with bone/cartilage have higher dN/dS compared to org. without; opposite trend was found
- Inconclusive evidence to suggest significant differences between dN/dS distributions
 - Some localized differences point towards independent gene loss, or gain
- Cannot use this, yet, to suggest clear directions for osteoarthritis research

Future Work

- Focus on genes independently
 - One study focused on *Fam20*, no other work pursued in the area
- Potentially low signal/noise ratio
 - Patterns are unclear, conclusion may be different
- Focus on genes with similar localization
 - Different groupings may yield different conclusions

Credits

- Brian Eames (supervisor)
- Ian McQuillan (supervisor)
- Jason Bernard (reviewer)
- Everyone who contributes to open-source

Code, data, and paper available:

<https://github.com/flaviuvadan/proteoglycan-pathway-evolution>