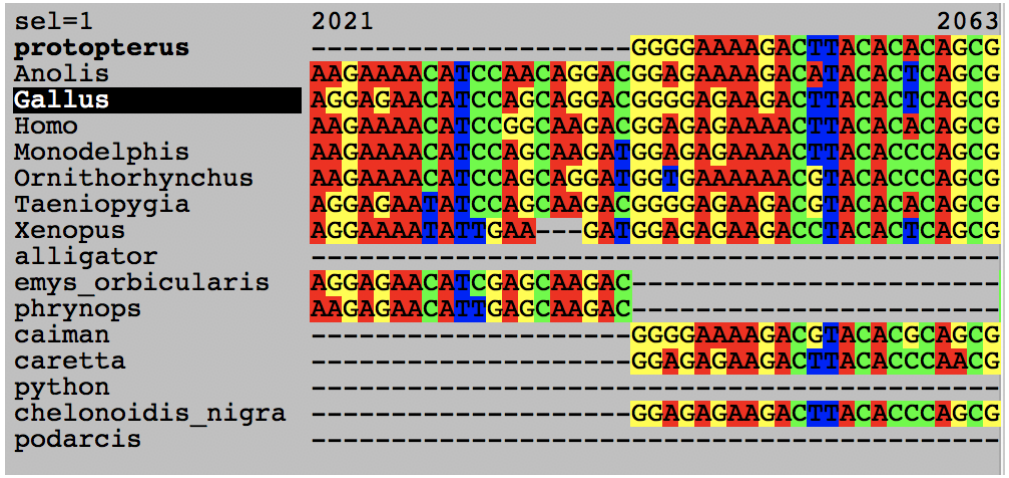
1) Input data



Gene boundaries are obvious within the dataset, as most genes are not present for all species.

The four Turtles (emys, phrynops, caretta, chelonodis) and two Crocodile (caiman, alligator) species have much more missing data than most other species in the alignment. This might make their position in the tree more difficult to resolve.

2) Inferring the first phylogeny

- Chiari et al. (2022) supported (Turtle,(Crocodile,Bird)) topology.
- The best-fit model is **GTR+F+I+R3**.
- BIC score is 232826.2430.
- Meaning of the model:
  - General-time reversible (GTR) model of sequence evolution
  - Base frequencies calculated empirically from the alignment (+F)
  - a proportion of invariable sites (+I)
  - free rate heterogeneity with three categories (+R3), not constrained to the Gamma distribution.
- The inferred tree supported (Bird,(Turtle,Crocodile)).
- Relevant clade (Turtle,Crocodile) has UFBoot support of 82%.
- This tree does NOT agree with Chiari et al. (2022)

2) Inferring the first phylogeny (snippet from .iqtree file)

Model of substitution: GTR+F+I+R3

Rate parameter R:

A-C: 1.8412
A-G: 4.7523
A-T: 1.2184
C-G: 1.4073
C-T: 7.5919
G-T: 1.0000

State frequencies: (empirical counts from alignment)

pi(A) = 0.2964
pi(C) = 0.2136
pi(G) = 0.2386
pi(T) = 0.2514

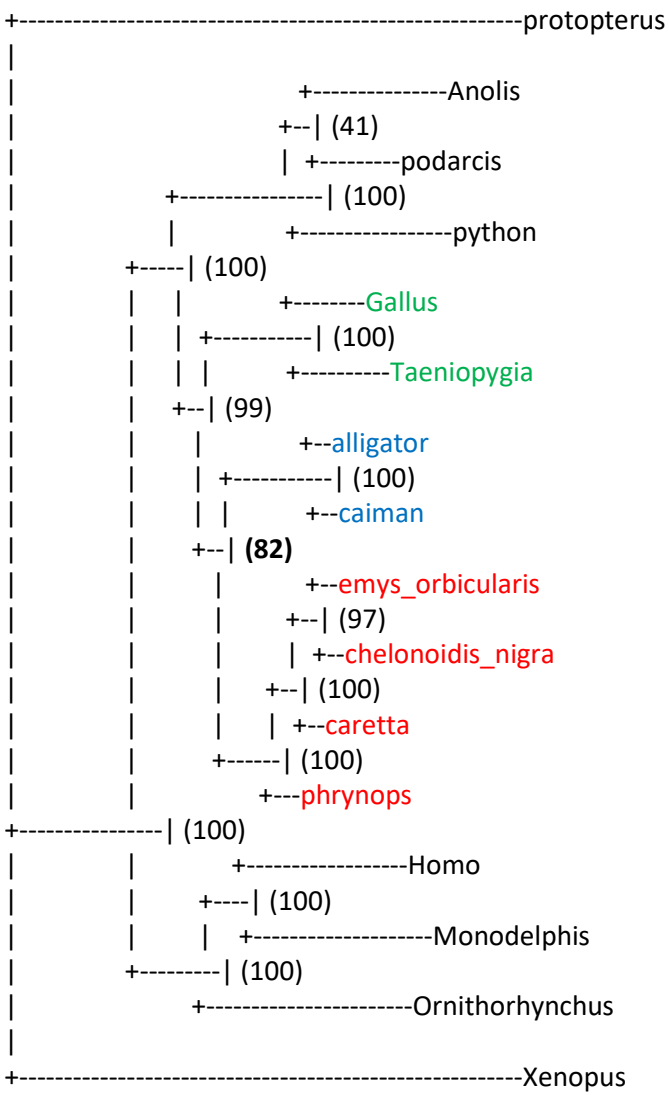
Rate matrix Q:

A	-0.8467	0.1816	0.5237	0.1414
C	0.252	-1.288	0.1551	0.8812
G	0.6504	0.1388	-0.9053	0.1161
T	0.1668	0.7487	0.1102	-1.026

Model of rate heterogeneity: Invar+FreeRate with 3 categories
Proportion of invariable sites: 0.3841
Site proportion and rates: (0.3007,0.4947) (0.2934,2.191) (0.02181,9.563)

Category	Relative_rate	Proportion
0	0	0.3841
1	0.4947	0.3007
2	2.191	0.2934
3	9.563	0.02181

2) Inferring the first phylogeny (snippet from .iqtree file)



birds

crocodiles

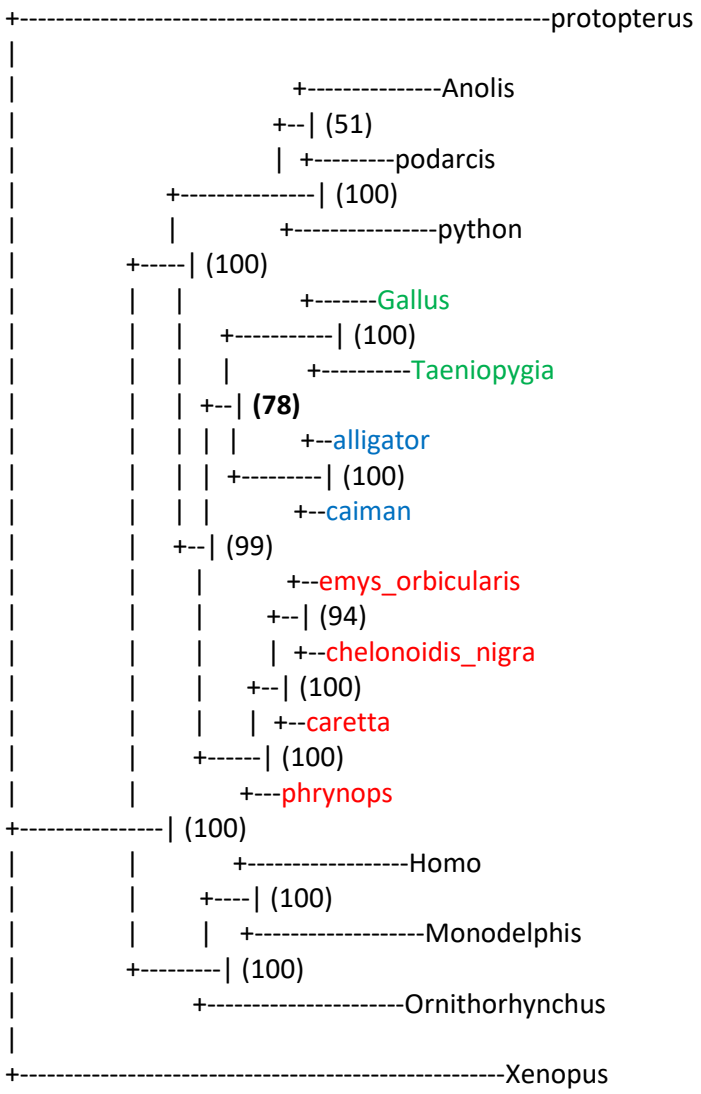
turtles

Turtles (red) are sister to crocodiles (blue) with 82% UFBoot support.
But the published tree has crocodiles and birds (green) as sister group.

3) Applying partition model

- BIC of partition model is: 233078.3986.
- This is **worse** than the single model (232826.2430).
- It supports (Turtle,(Crocodile,Bird)), agreeing with Chiari et al. (2012).
- The relevant clade “(Crocodile,Bird)” has UFBoot support of **78%.**

3) Applying partition model (snippet from .iqtree file)



birds

crocodiles

turtles

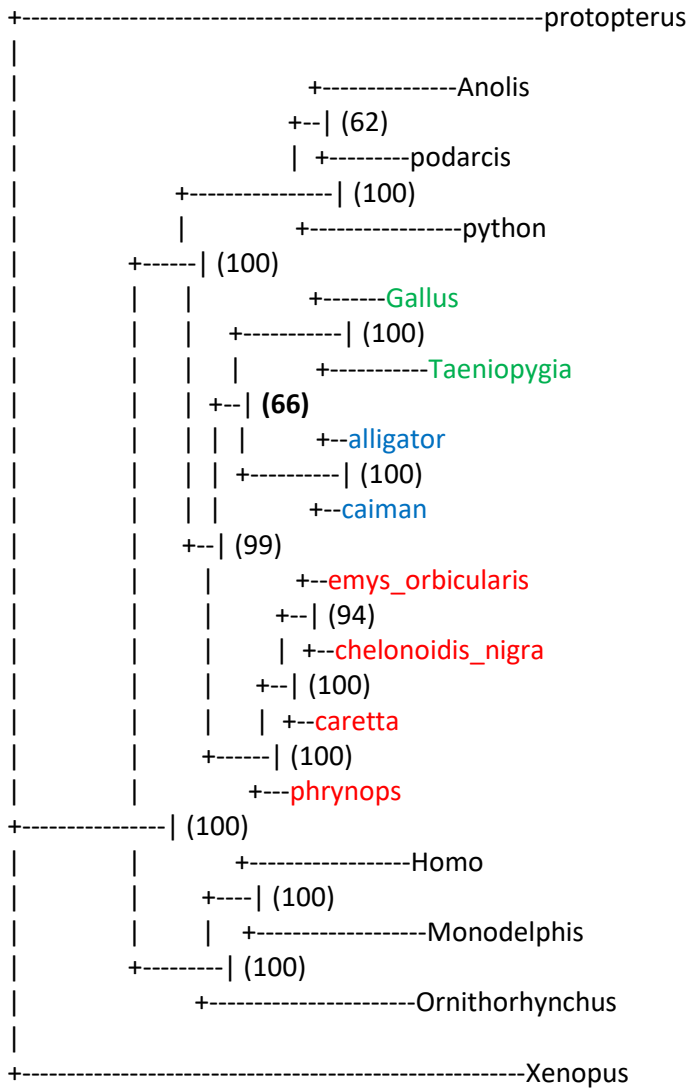
Birds (green) and crocodiles (blue) form a monophyletic group with 78% UFBoot support, which is low but concurs with the published tree.

4) Choosing the best partitioning scheme

- The best partitioning scheme has 6 partitions.
- BIC score is 232341.9089.
- It's **the best** because it's lower than single model (232826.2430) and fully partition model (233078.3986).
- It supports (Turtle,(Crocodile,Bird)), agreeing with Chiari et al. (2012).
- The relevant clade (Crocodile,Bird) has UFBoot support of **66%**, which is quite low.

By merging similar genes, we have reduced the number of partitions from 29 to 7. This has reduced the number of parameters in the model from 218 to 85 and consequently, the best partition scheme now has the lowest BIC score of the three models considered so far.

4) Choosing the best partitioning scheme



birds

crocodiles

turtles

This tree agrees with that inferred by the fully partitioned model and the published tree. But the UFBoot support 66% for birds and crocodiles together is relatively low.

5) Tree topology tests

```
USER TREES
----------

See turtle.test.trees for trees with branch lengths.

Tree      logL    deltaL  bp-RELL    p-KH     p-SH      c-ELW       p-AU
-------------------------------------------------------------------------
  1 -115770.1154  4.2647   0.432 +   0.44 +   0.44 +     0.431 +    0.456 +
  2 -115765.8508       0   0.568 +   0.56 +      1 +     0.569 +    0.544 +

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 10000 resamplings using the RELL method.
```

- Tree by single model has worse log-likelihood.
- But you can't reject this tree by SH test, the p-value is 0.44, way higher than 5%.
- You also can't reject this tree by AU test, the p-value is 0.456, way high than 5%.

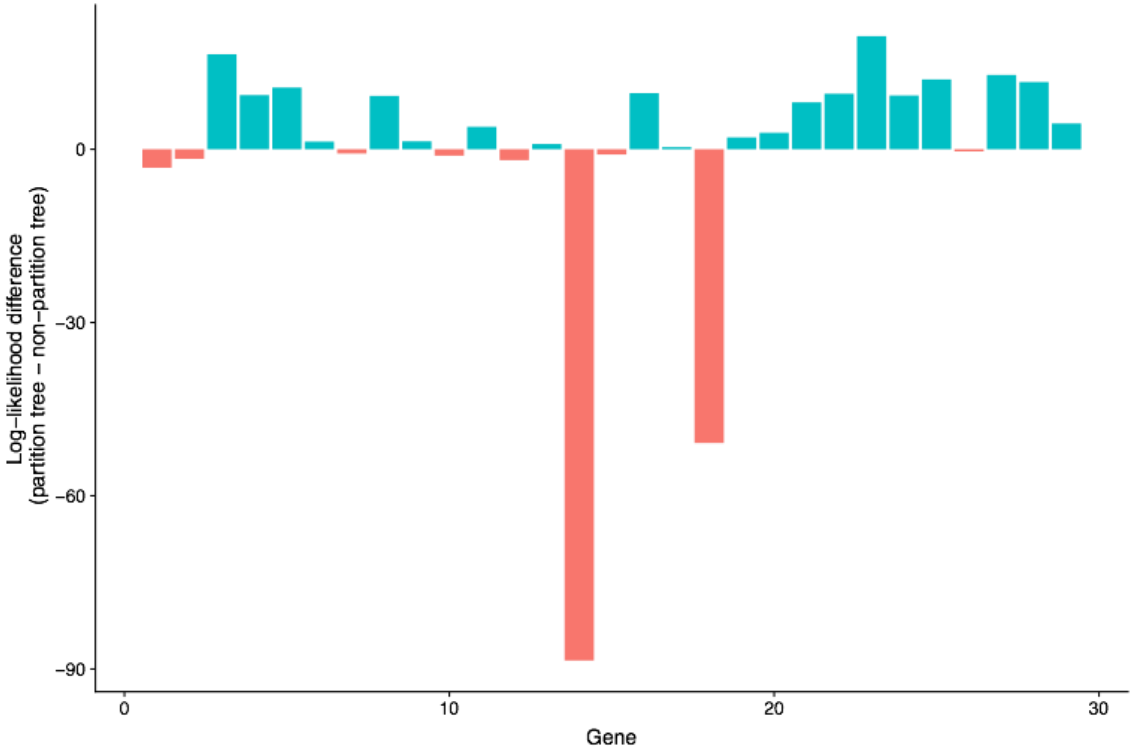
6) Tree mixture model (MAST)

Snippet from turle.mix.iqtree:

Tree weights: 0.437, 0.563

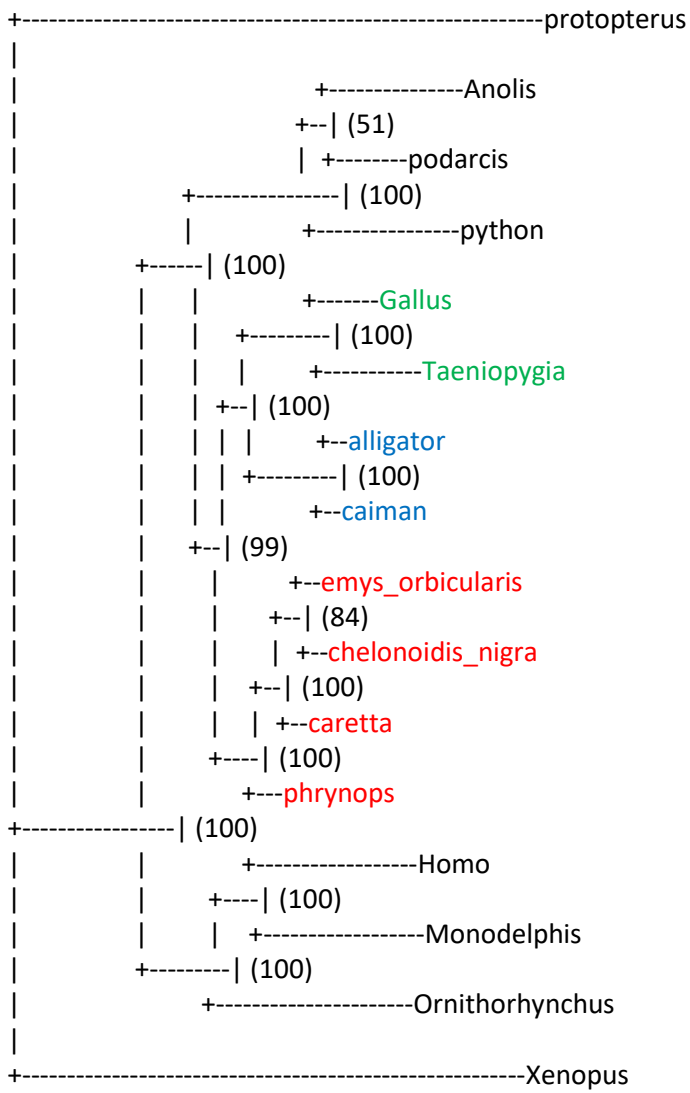
- Highest weight is 0.563.
- It is the tree found by partition model: (Turtle,(Crocodile,Bird)), and also by Chiari et al. (2012).
- It's telling us that the majority of sites in the alignment supports this tree, although the signal is weak, just slightly >50%.

7) Identifying most influential genes



- Examining differences in log-likelihoods between the two trees on a per-gene basis, the 14th and 18th genes (red bars) strongly support non-partition tree: ENSGALG00000008916 and ENSGALG00000011434.
- These 2 genes are also found by Brown and Thompson (2017).
- These two genes happen to contain paralogous sequences! That may distort tree topology (https://doi.org/10.1093/sysbio/syw101).

8) Removing influential genes



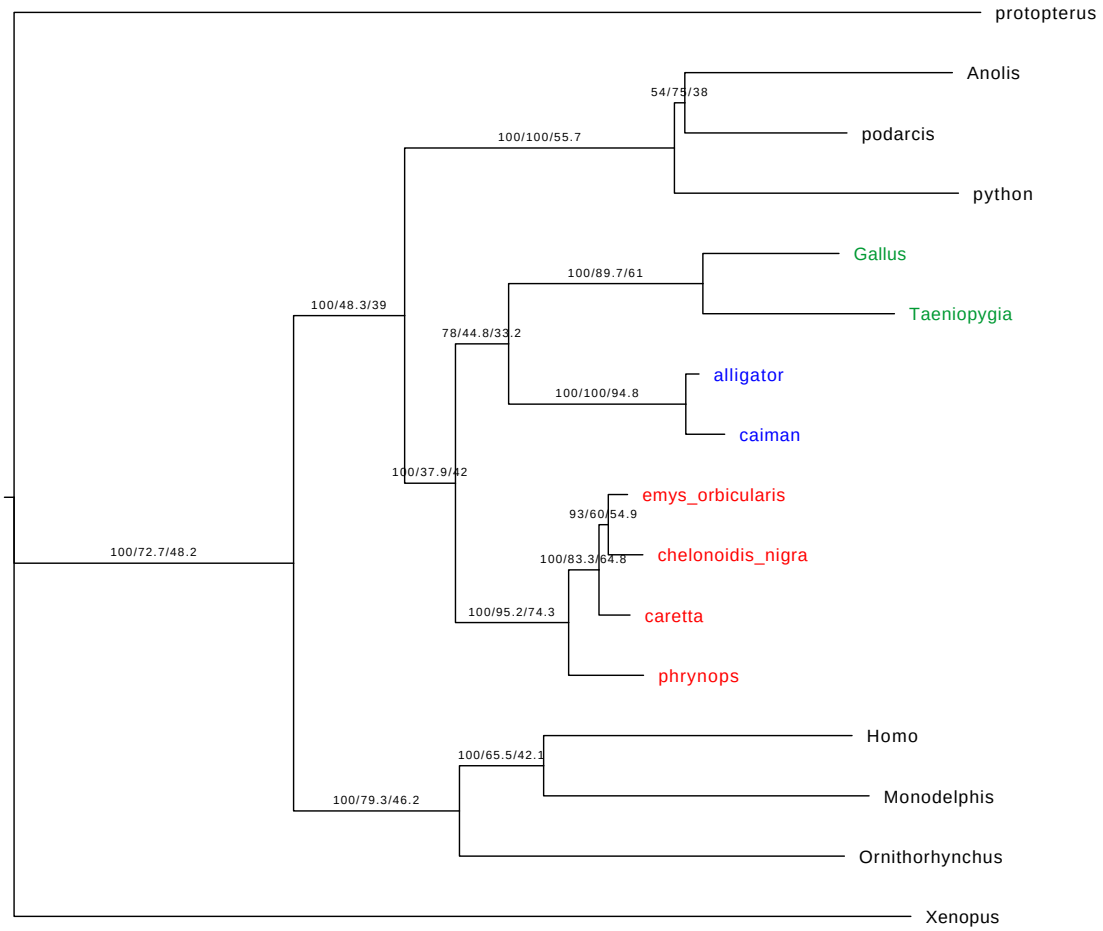
birds

crocodiles

turtles

After removing 2 genes, the tree now agrees with the fully partitioned model and the published tree. And the UFBoot support 100% for birds and crocodiles together.

9) Concordance factors - gCF



- 44.8% of genes (13) support (Bird,Crocodile).
- 13.8% of genes (4) support (Turtle,Crocodile).
- 6.9% of genes (2) support (Turtle,Bird).

9) Concordance factors - sCF

- 33.2% of sites support (Bird, Crocodile).
- 40.1% of sites support (Turtle, Crocodile).
- 100 – 33.2 – 40.1 = 26.7% of sites support (Bird,Turtle).

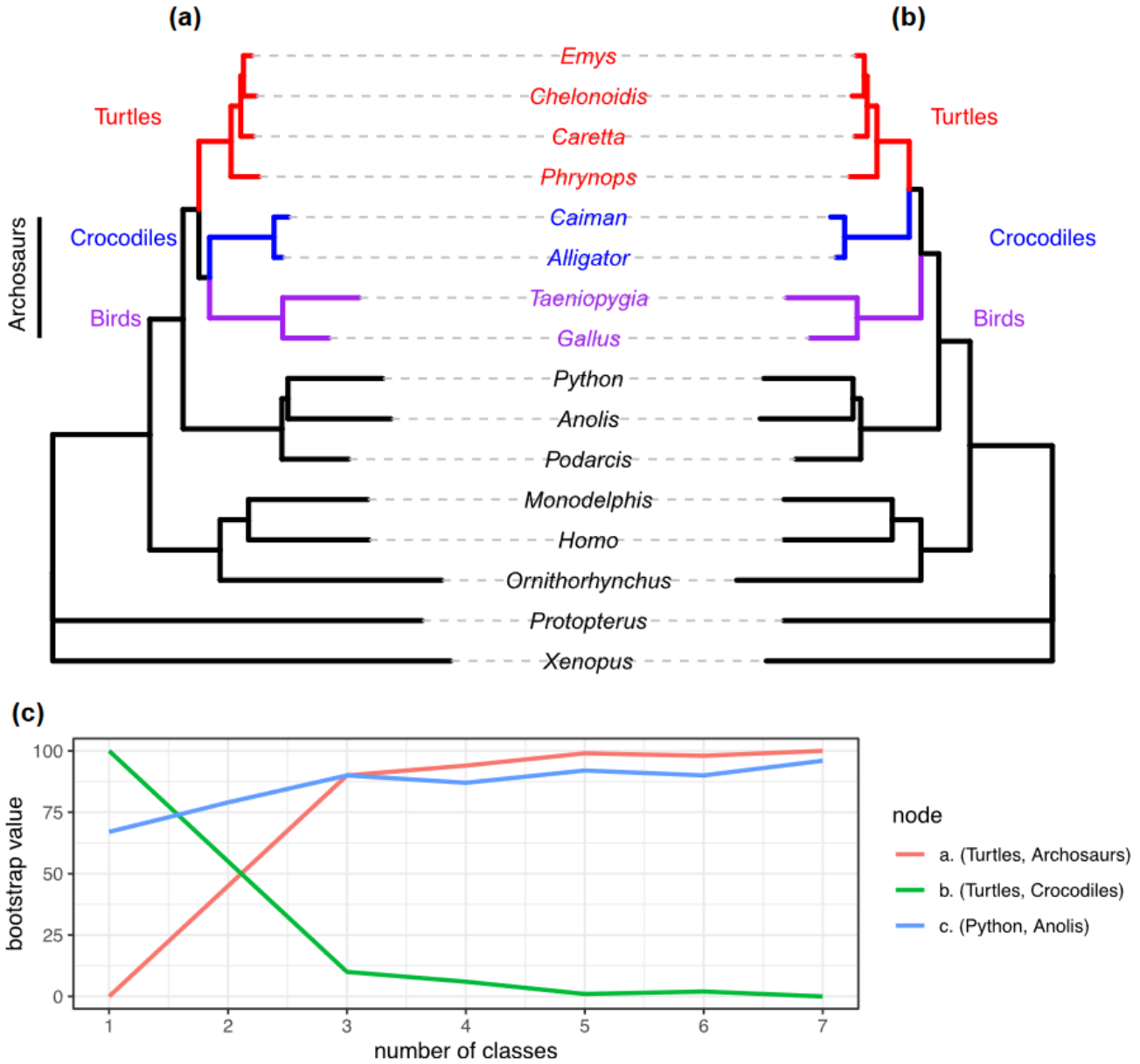
9) Concordance factors - sCF

- 33.2% of sites support (Bird, Crocodile).
- 40.1% of sites support (Turtle, Crocodile).
- 100 – 33.2 – 40.1 = 26.7% of sites support (Bird,Turtle).

10) Mixture models

Classes	BIC	Relation ship	Support
1	232818	Turtle-croc	83
2	231744	Turtle-croc	82
4	231298	Turtle-croc	63
6	231381	Bird-croc	93

- The 6-class model agrees with the results of Chiari et al.
- Yes – support for turtle-croc decreases with number of model classes
- Ren et al. find turtle-croc under a 1-class model, but bird-croc under a 6-class model. Our results agree.



Ren et al. *MBE* 2025