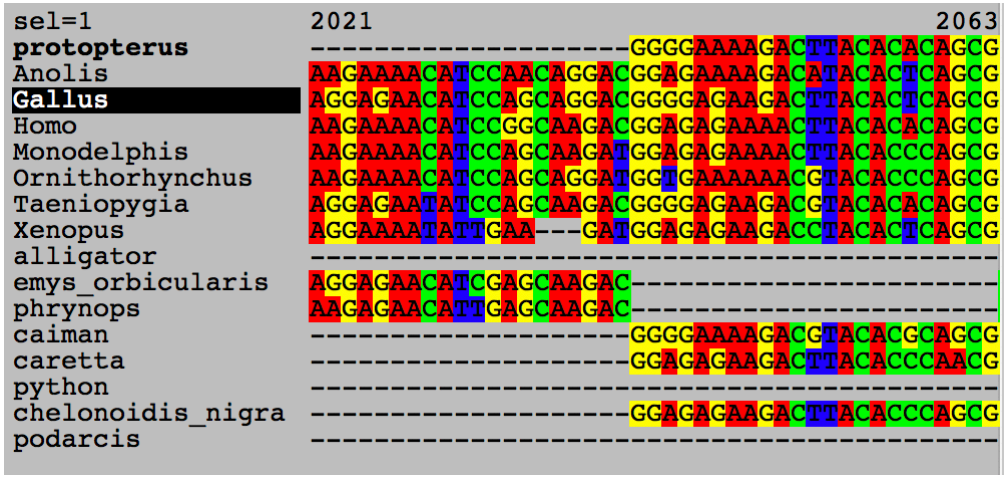
1) Input data



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

The four turtle (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

The best-fit model found by ModelFinder was GTR+F+R3.
This means:

- the GTR model of sequence evolution
- base frequencies calculated empirically from the alignment (as opposed to inferred under ML)
- three categories of rate heterogeneity, with rates and weights inferred by ML, not constrained to the Gamma distribution.

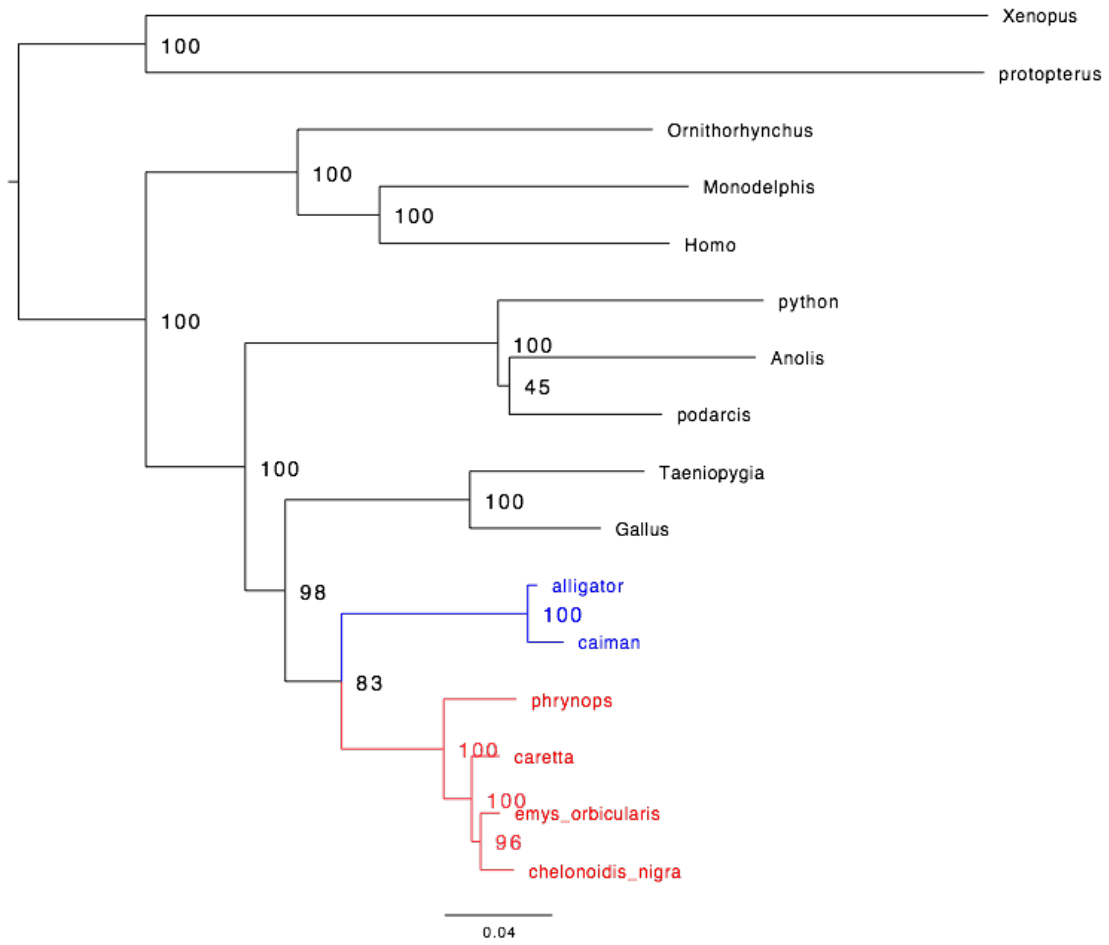
2) Inferring the first phylogeny

```
SUBSTITUTION PROCESS
--------------------


Model of substitution: GTR+F+R3


Rate parameter R:


  A-C: 1.8331
  A-G: 4.7130
  A-T: 1.2195
  C-G: 1.4045
  C-T: 7.5043
  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
```

2) Inferring the first phylogeny

```
Rate matrix Q:

  A    -0.848    0.1821    0.5233    0.1426
  C    0.2528    -1.286    0.1559    0.8776
  G    0.6499    0.1396   -0.9064    0.1169
  T    0.1682    0.7457     0.111    -1.025

Model of rate heterogeneity: FreeRate with 3 categories
Site proportion and rates:  (0.566,0.08039) (0.376,1.639)
(0.05796,5.832)

 Category  Relative_rate  Proportion
  1         0.08039        0.566
  2         1.639          0.376
  3         5.832          0.05796
```

2) Inferring the first phylogeny



Tree inferred under GTR+F+R3 without partitioning the alignment. Turtles (red) are sister to crocodiles (blue), in contradiction with the published tree, in which turtles are sister to archosaurs (crocodiles and birds).

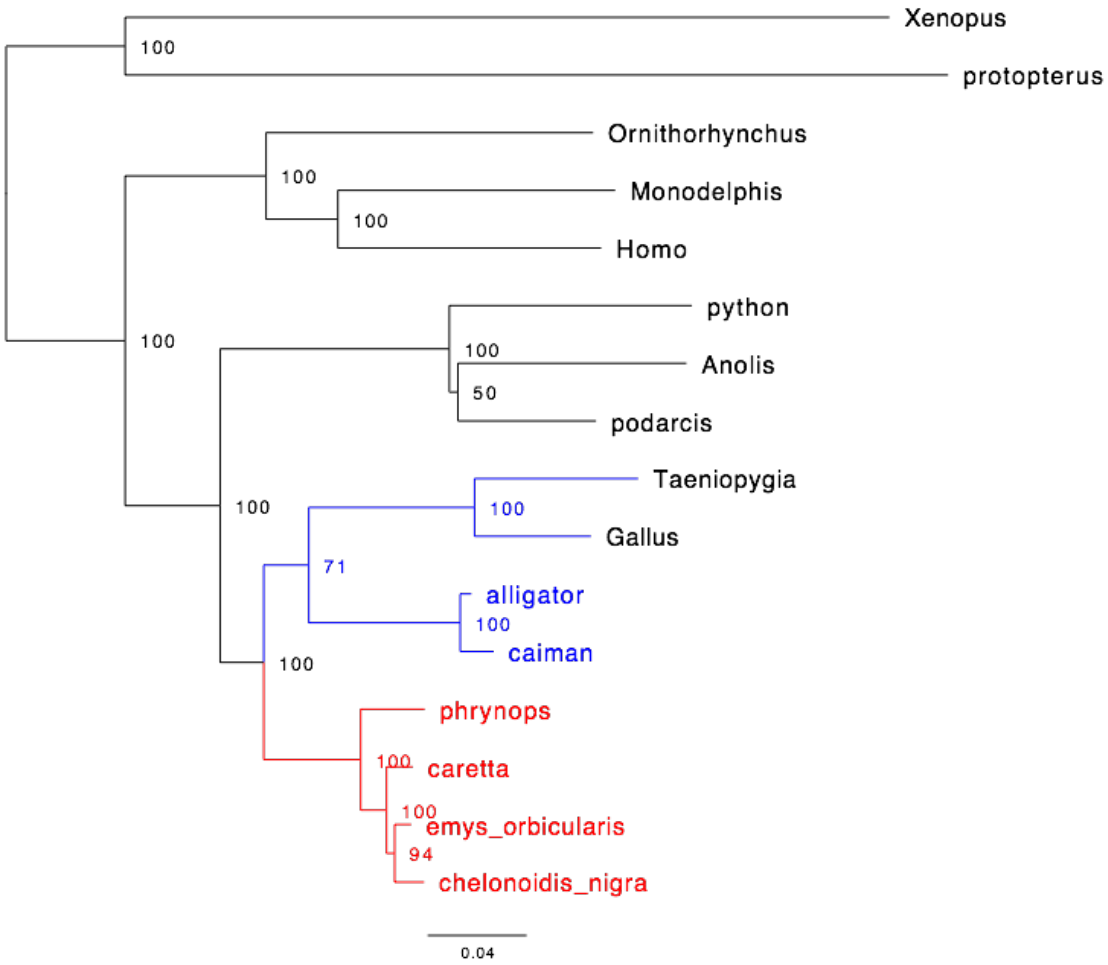
3) Applying partition model

- The slowest evolving gene is the 10th gene, with a rate of 0.4683.
- The fastest evolving gene is the 18th gene, with a rate of 1.8421.

- The BIC of the non-partitioned model is 232837.7889.
- The BIC of the partitioned model is 233126.4205.

Even though the partitioned model has a higher likelihood than the non-partitioned model, the non-partitioned model has a smaller (better) BIC, and on that basis should be preferred. This is because the partitioned model has 221 free parameters, compared to just 41 for the non-partitioned model.

3) Applying partition model



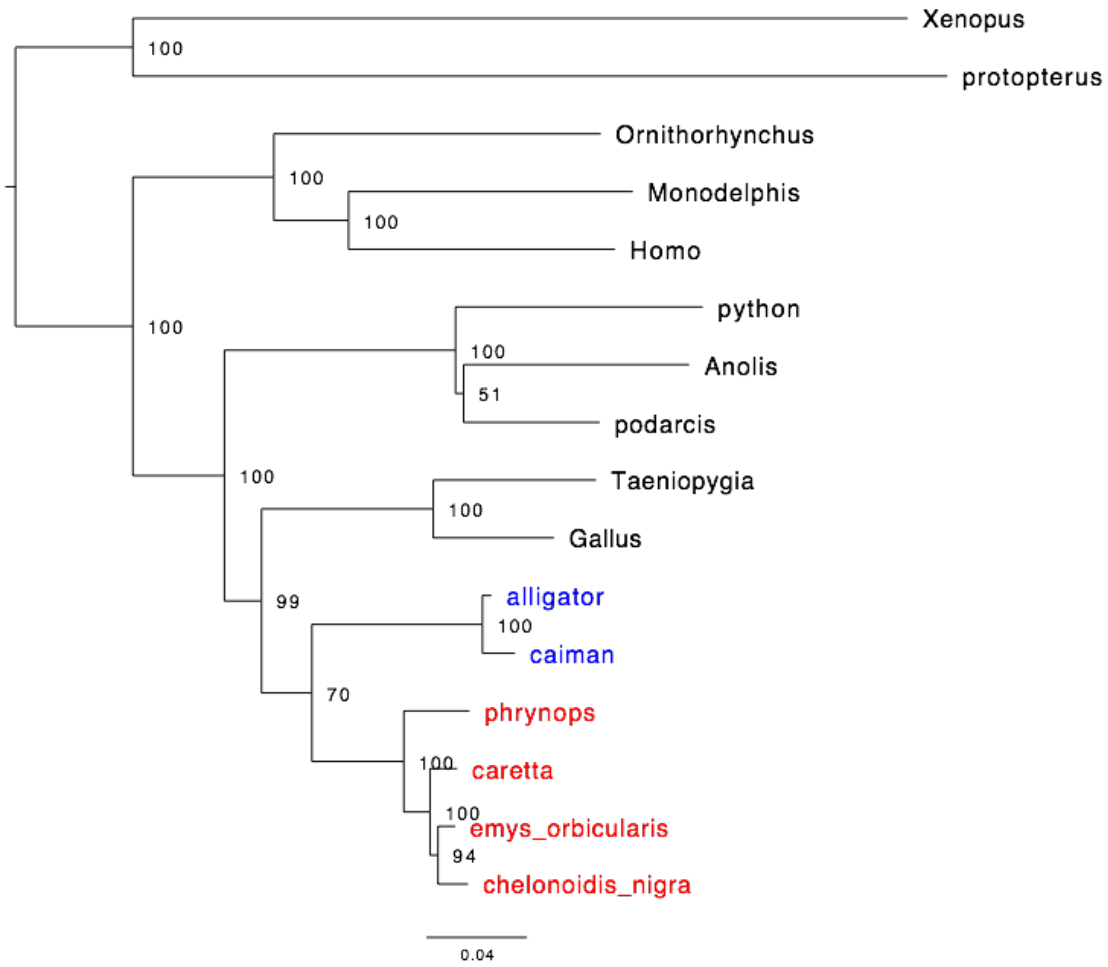
Tree inferred under the partition model. Turtles (red) are sister to archosaurs (blue), concurring with the published tree.

4) Choosing the best partitioning scheme

- The BIC of the non-partitioned model is 232837.7889.
- The BIC of the partitioned model is 233126.4205.
- The BIC of the best partition model is 232401.3940

By merging similar genes, we have reduced the number of partitions from 29 to 10. This has reduced the number of parameters in the model from 221 to 106 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

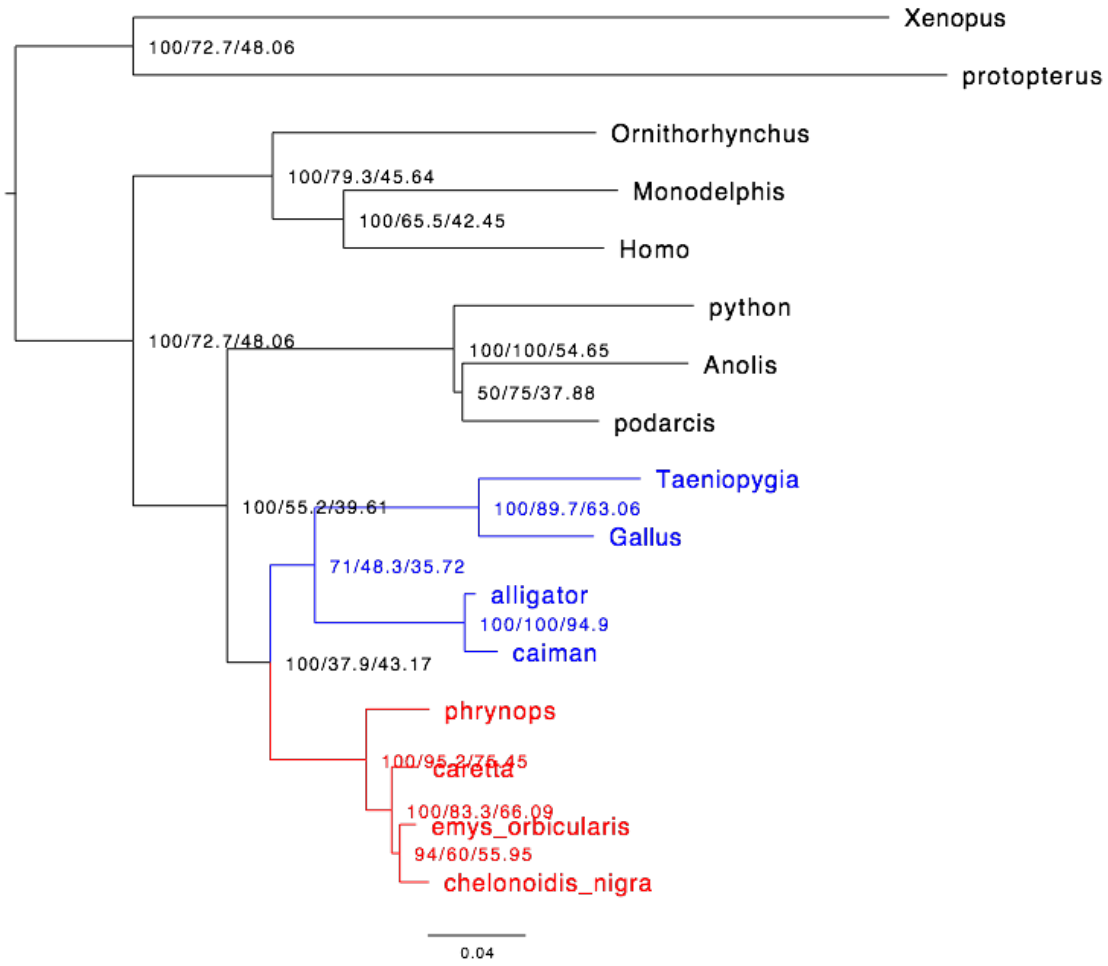


Tree inferred under best partition scheme. This topology agrees with that inferred by the non-partitioned model, and conflicts with the published tree. The bootstrap support for the conflicting branch has fallen from 83 to 70.

5) Tree topology tests

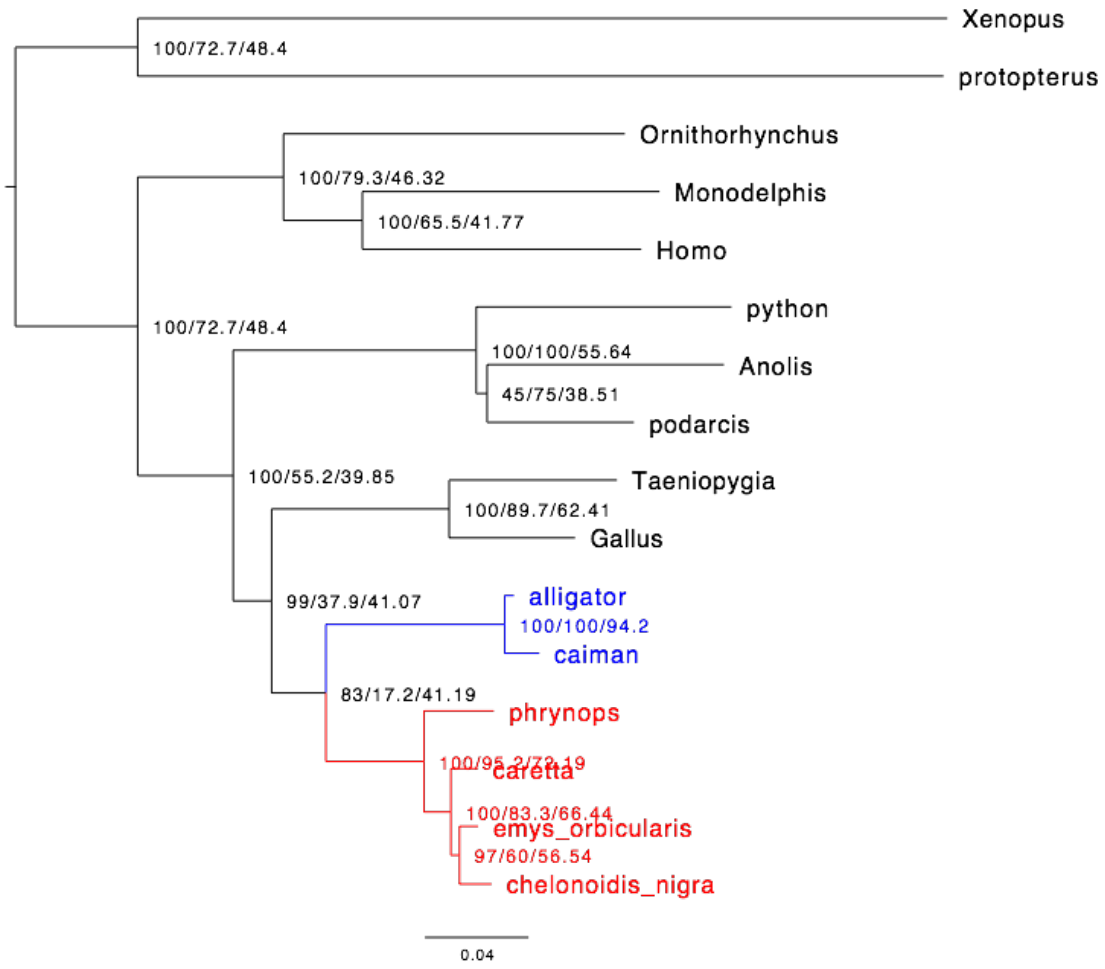
```
USER TREES
----------

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

Tree      logL    deltaL  bp-RELL    p-KH     p-SH       c-ELW
------------------------------------------------------------------
  1 -115476.8396  6.7446   0.399 +  0.394 +  0.394 +     0.401 +
  2  -115470.095       0   0.601 +  0.606 +      1 +     0.599 +

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).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.
```

6) Concordance factors



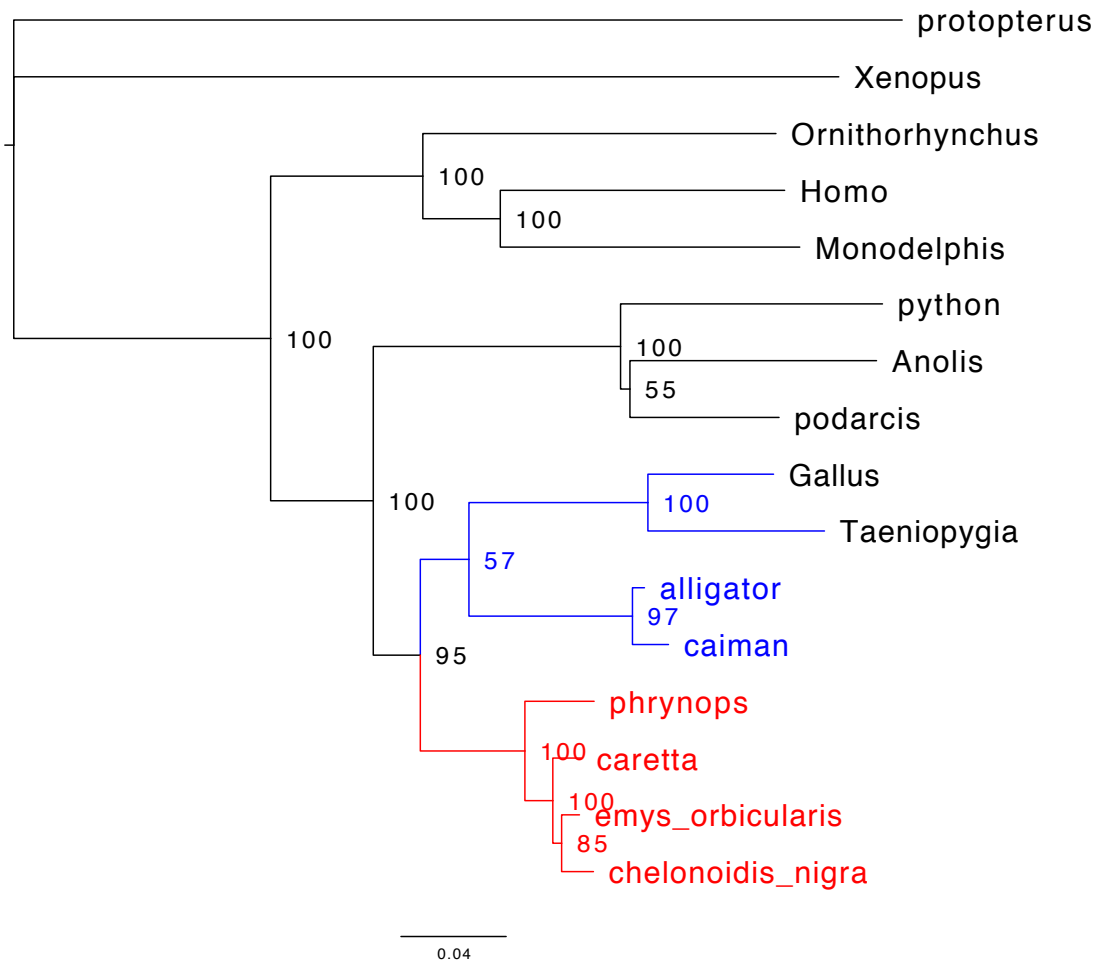
Inferred tree for the full partition model. The node annotation indicates BS/gCF/sCF scores. The contentious branch splitting turtles and archosaurs has gCF of 48.3%, which equates to 14 of the 29 genes supporting this topology.

6) Concordance factors



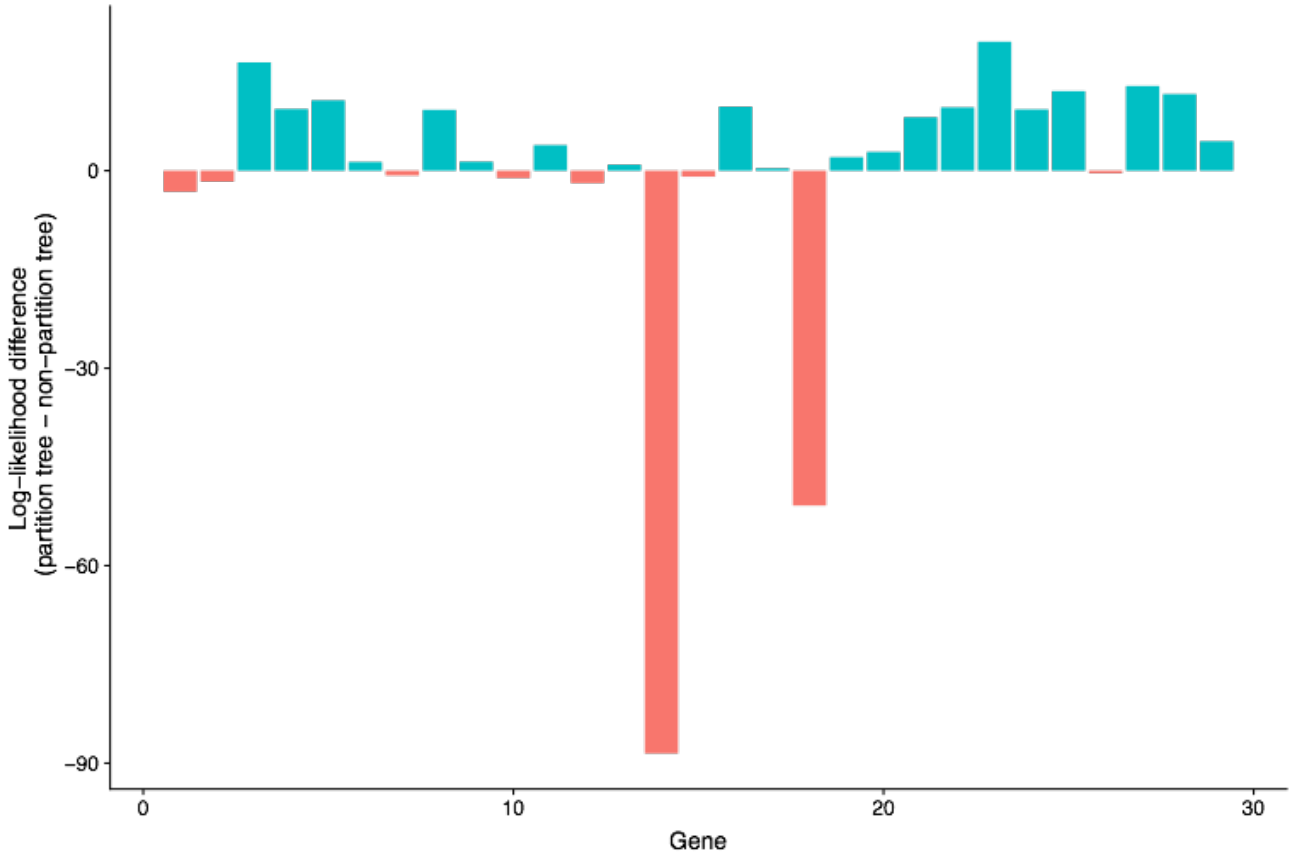
Inferred tree for the non-partition model. The node annotation indicates BS/gCF/sCF scores. The contentious branch splitting turtles and archosaurs has gCF of 17.2%, which equates to 5 of the 29 genes supporting this topology.

7) Resampling partitions and sites



Turtles (red) are sister to archosaurs (blue), concurring with the published tree.

8) Identifying most influential genes



When we examine the difference in log-likelihoods between the two trees on a per-gene basis, we notice that two particular genes strongly support turtles as sister to crocodiles, whereas most other genes are either neutral, or support turtles as sister to archosaurs. These two genes happen to contain paralogous sequences! That may distort tree topology.