

Comparing genome-wide measures of phylogenetic support in squamates

Genevieve G. Mount

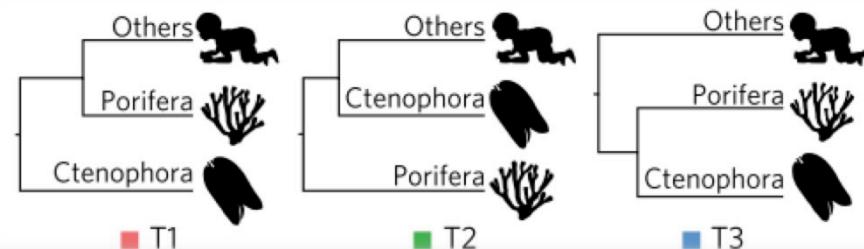
Jeremy M. Brown

LSU, Dept. of Biological Sciences &
Museum of Natural Science



Discordant topologies with strong support values

- Birds tree of life
- Turtle placement
- Metazoan origins



Methods to quantify variation in support across multiple topologies

- Locus by locus approach

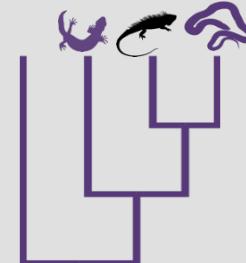
A sequence alignment diagram showing 10 DNA sequences. The sequences are color-coded to highlight matching bases: A (red), T (blue), C (green), and G (yellow). The sequences are as follows:

```
TTCCACACAAACATACGAGCCGGAAAGCATAAAGCTAAA  
CGAACGGGGGGAGAGGGGGTTCCTTAATGGGGCG  
TGACTCCACACAAACATACGACCCGGGAAGCATAAAGCTAAA  
TCGCCCAACACGGGGGGAGGGGGTTCCTTAATGGGGCG  
CCAAATGACCTCCACACAAACATACGACCCGGGAAGCATAAAGCTAAA  
TGACTCCACACGGGGGGAGGGGGTTCCTTAATGGGGCG  
TCGCCCAACAGCTCCACACAAACATACGACCCGGGAAGCATAAAGCTAAA  
TGACTCCACACGGGGGGAGGGGGTTCCTTAATGGGGCG  
TCGCCCAACAGCTCCACACAAACATACGACCCGGGAAGCATAAAGCTAAA  
TGACTCCACACAAACATACGACCCGGGAAGCATAAAGCTAAA  
TCGCCCAACAGCTCCACACAAACATACGACCCGGGAAGCATAAAGCTAAA
```

Methods to quantify variation in support across multiple topologies

- Locus by locus approach
- Compare relative support for each topology

```
TGCCACACAACTACGAGCCGGAAAGCATAAAGTGAAA  
CCAACGGCGGGAGAGCGGGTTCCTTAATGGGGCG  
TGACTCCACACAACTACGAGCCGGAGCAAAAGGAGAAA  
TCCCCAACTACGAGCCGGAGCGGGTTCCTTAATGGGGCG  
CCAATGACTCCACACAACTACGAGCCGGAGCAAAAGGAGAAA  
TGACTCCACACAACTACGAGCCGGAGCGGGTTCCTTAATGGGGCG  
TCCCCAACTACGACTCCACACAACTACGAGCCGGAGCAAAAGGAGAAA  
TGACTCCACACAACTACGAGCCGGAGCGGGTTCCTTAATGGGGCG  
TCCCCAACTACGACTCCACACAACTACGAGCCGGAGCAAAAGGAGCAA  
TGACTCCACACAACTACGAGCCGGAGCGGGTTCCTTAATGGGGCG  
TCCCCAACTACGACTCCACACAACTACGAGCCGGAGCAAAAGGAGCAA  
TGACTCCACACAACTACGAGCCGGAGCGGGTTCCTTAATGGGGCG  
TCCCCAACTACGACTCCACACAACTACGAGCCGGAGCAAAAGGAGCAA  
TGACTCCACACAACTACGAGCCGGAGCGGGTTCCTTAATGGGGCG  
TCCCCAACTACGACTCCACACAACTACGAGCCGGAGCAAAAGGAGCAA
```



Methods to quantify variation in support across multiple topologies

- Bayes Factors (BF)
- Difference in gene log-likelihood scores (dGLS)
- Gene genealogy interrogation (GGI)

Methods for quantifying support

Bayes Factors (BF)

- Marginal likelihood ratio
- Relative fit of model to data

Methods for quantifying support

Bayes Factors (BF)

- Marginal likelihood ratio

Difference in gene log likelihood (dGLS)

- Sum of site specific likelihoods across locus
- Maximum likelihood

Methods for quantifying support

Bayes Factors (BF)

- Marginal likelihood ratio

Difference in gene log likelihood

- Maximum likelihood difference

Gene genealogy interrogation (GGI)

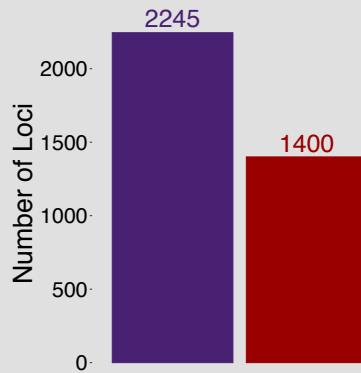
- Approximately unbiased test
- Rank topologies with p-value

Do all these methods give us the same answer??

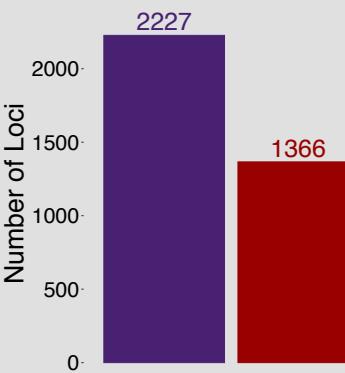
- Bayes Factors (BF)
- Difference in gene log-likelihood scores (dGLS)
- Gene genealogy interrogation (GGI)

Do the majority of loci support the same topology across methods?

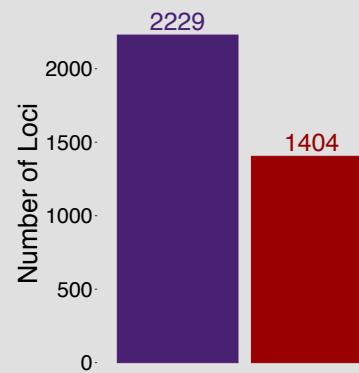
Do the majority of loci support the same topology across methods?



vs

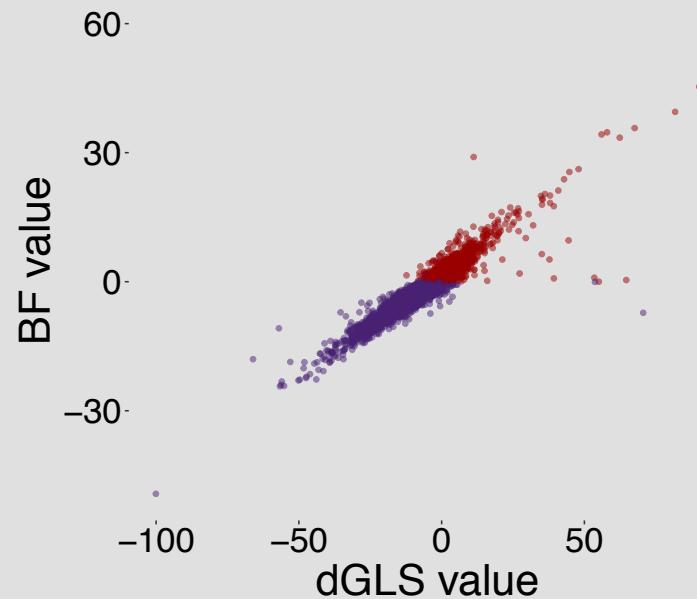


vs

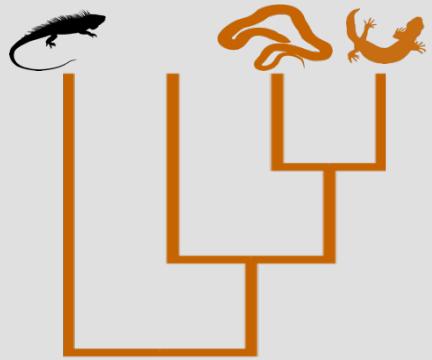


Are locus support values correlated across methods?

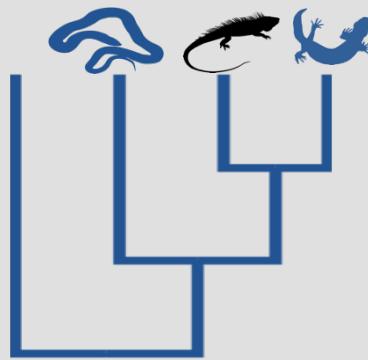
Large BF value =
large dGLS value =
GGI Rank 1?



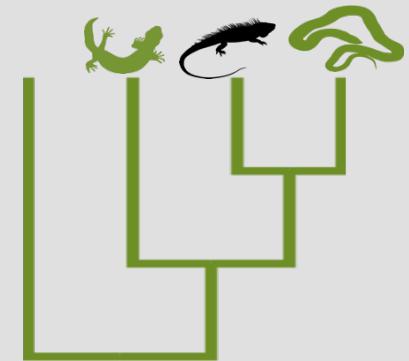
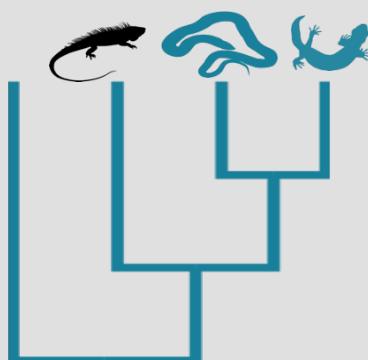
Squamate phylogeny



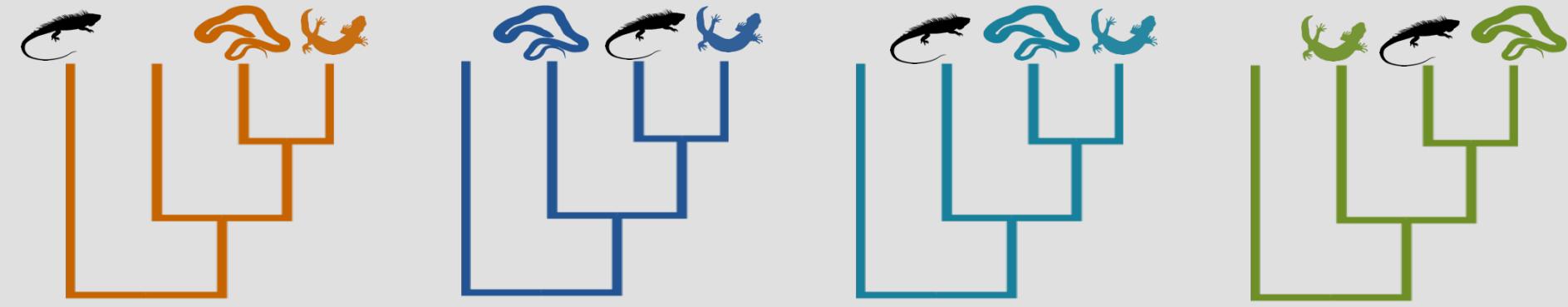
Scleroglossa



Toxicofera



Squamate phylogeny



Scleroglossa

Anguimorph

Snake

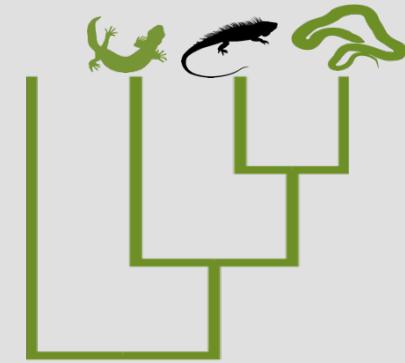
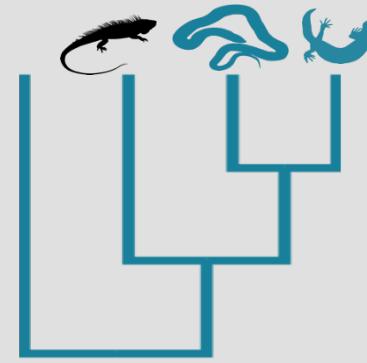
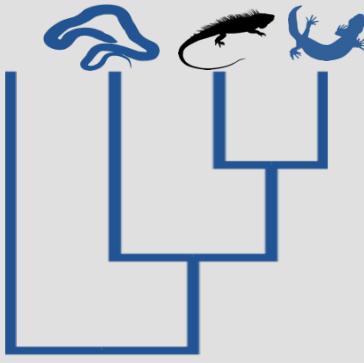
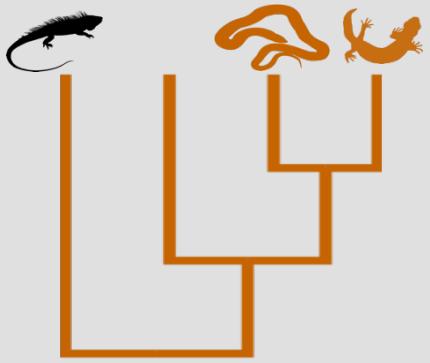
Iguania

Iguania

Anguimorph

Snake

Squamate data



Singhal et al 2017 – ~3800 loci, 18 taxa

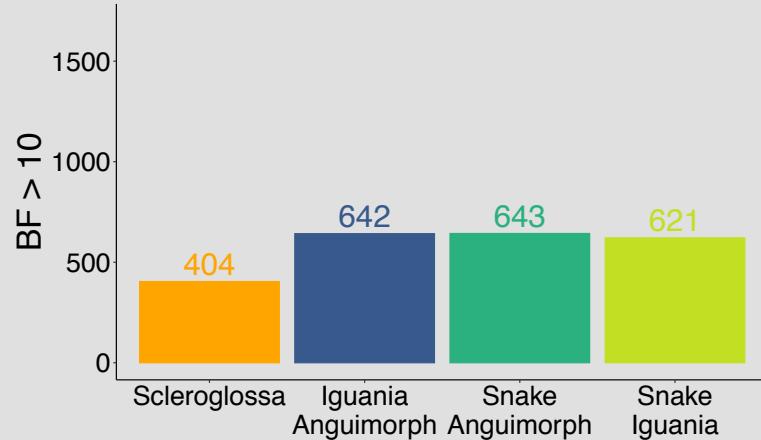
Streicher & Weins 2017 - ~ 4100 loci, 32 taxa

Quantifying “strong” support

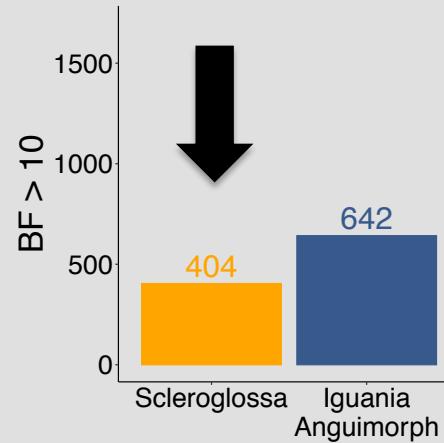
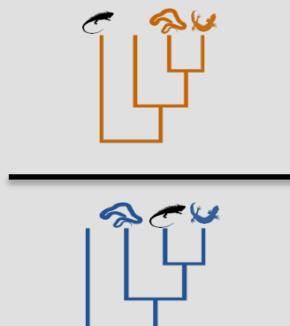
- “Strong” support
 - $\text{BF} > 10$
 - $\text{dGLS} > 0.5$
 - GGI, Rank 1 $p > 0.95$

Quantifying “strong” support

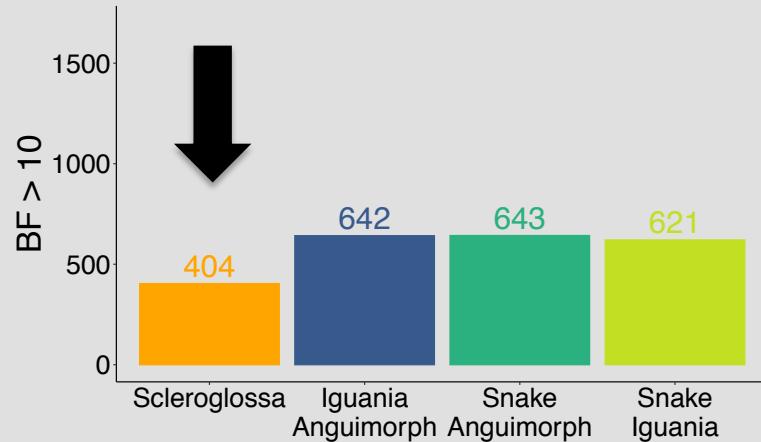
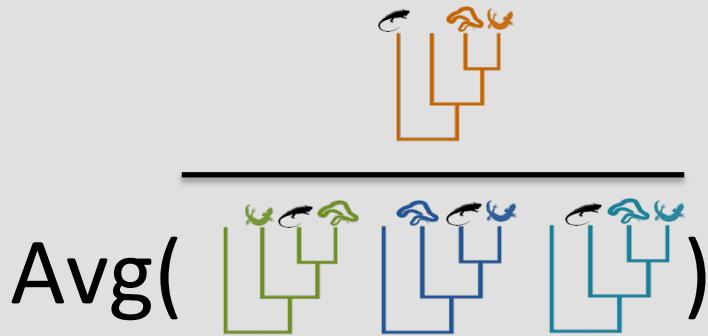
- “Strong” support
 - $BF > 10$
 - $dGLS > 0.5$
 - GGI, Rank 1 $p > 0.95$



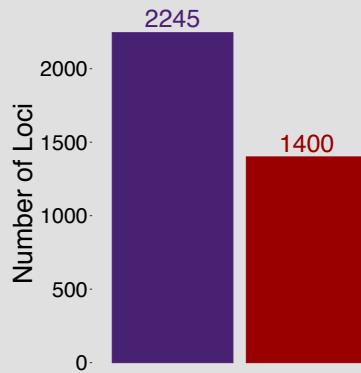
Pairwise Comparisons



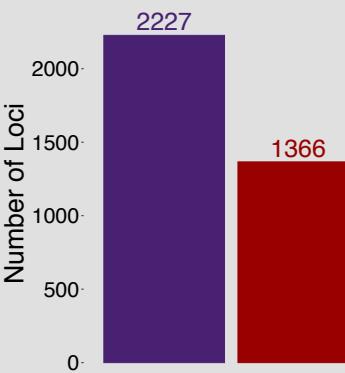
Compare all



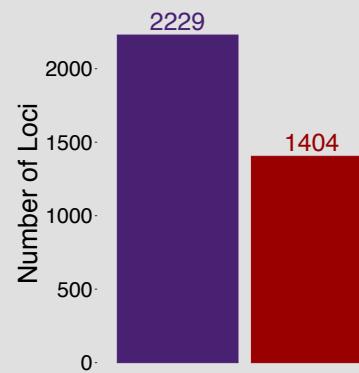
Do the majority of loci support the same topology across methods?



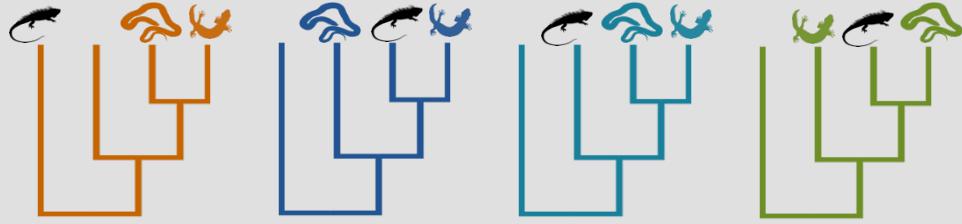
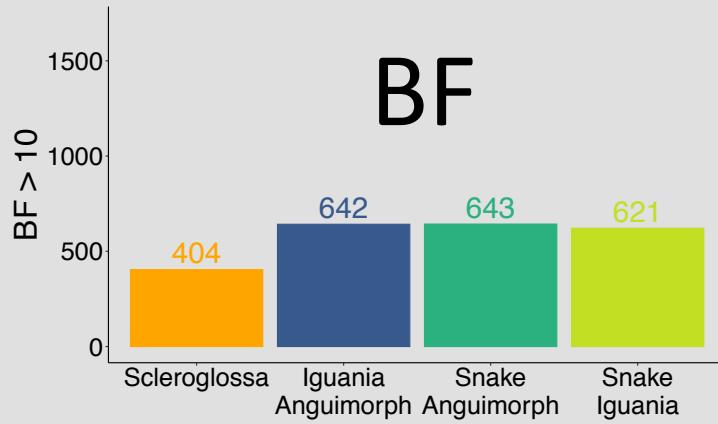
vs



vs



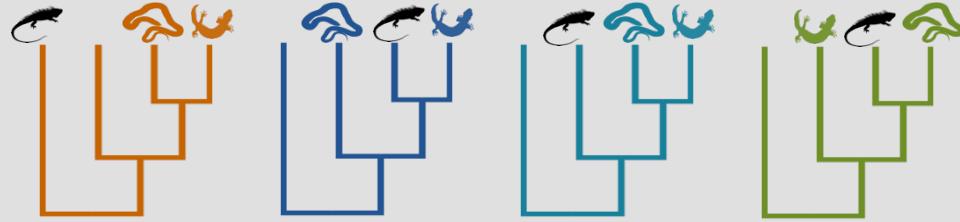
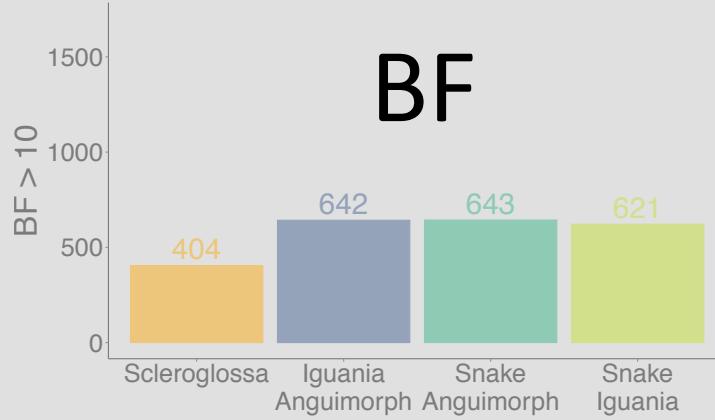
Do the majority of loci support the same topology across methods?



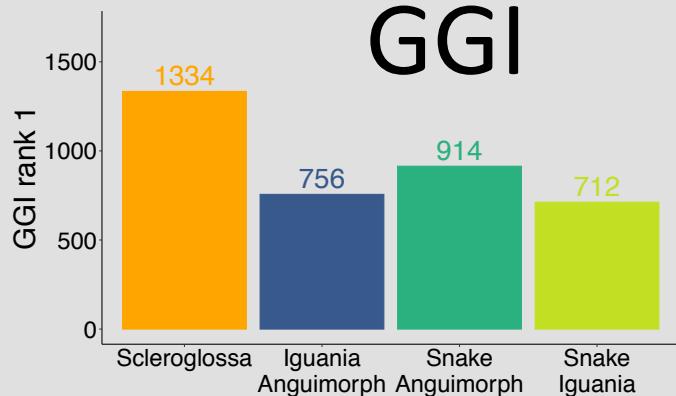
Singhal

Do the majority of loci support the same topology across methods?

BF



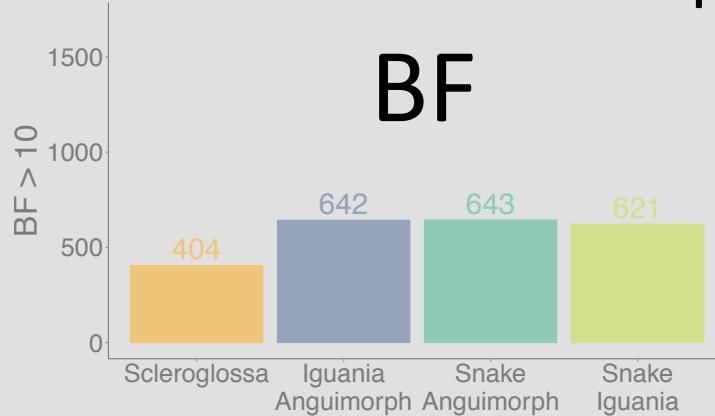
GGI



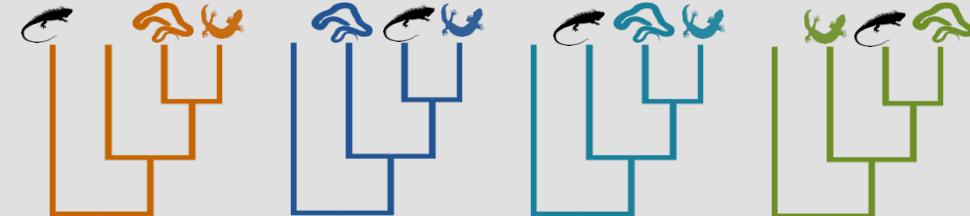
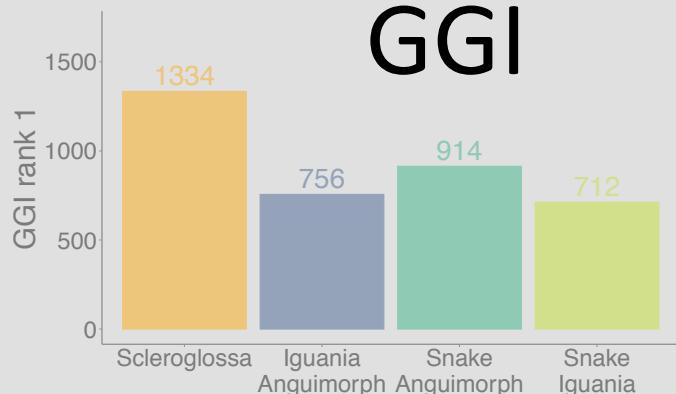
Singhal

Do the majority of loci support the same topology across methods? NO

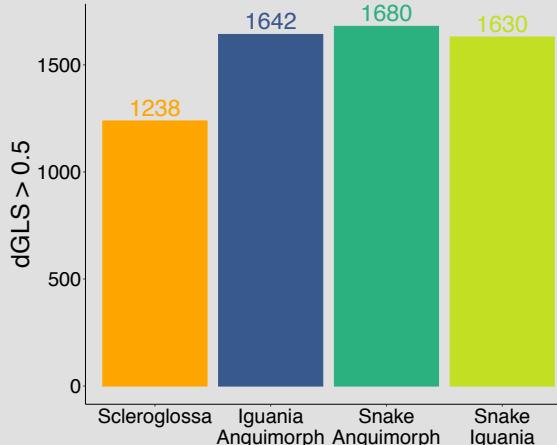
BF



GGI

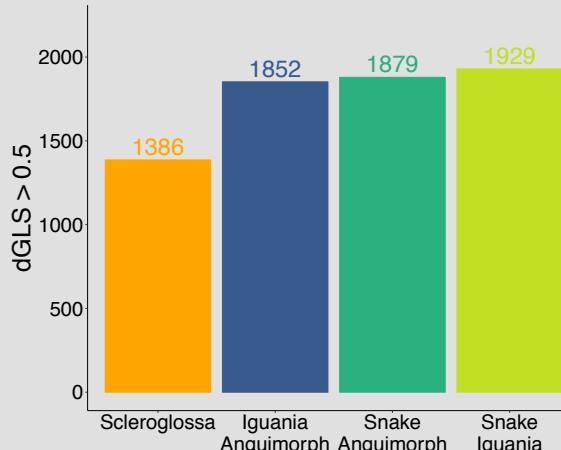
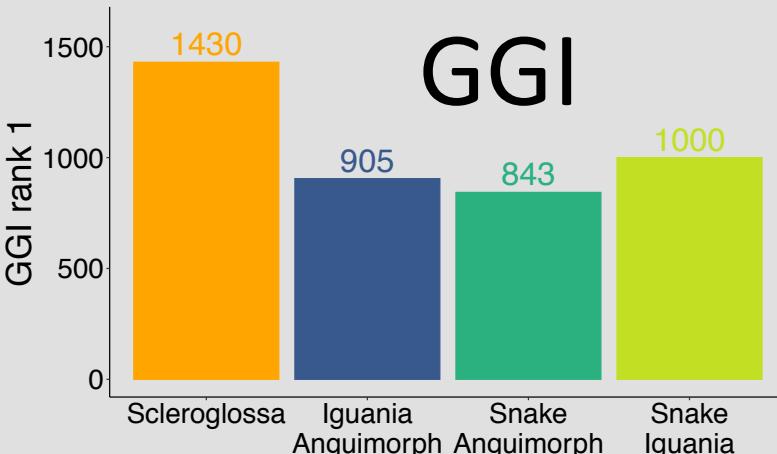
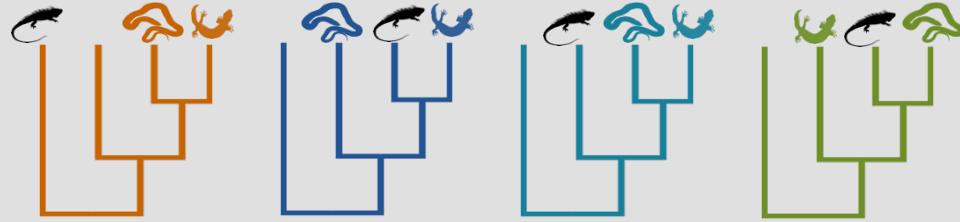
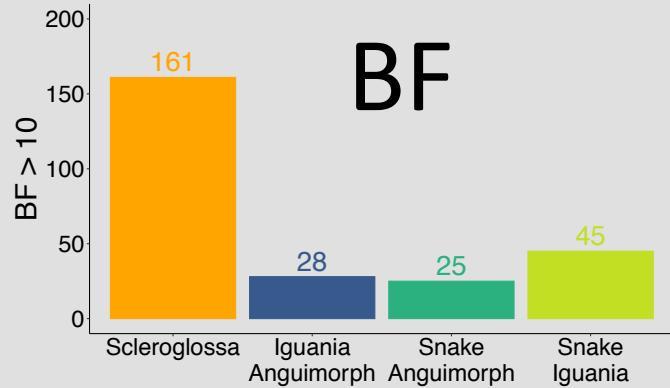


dGLS



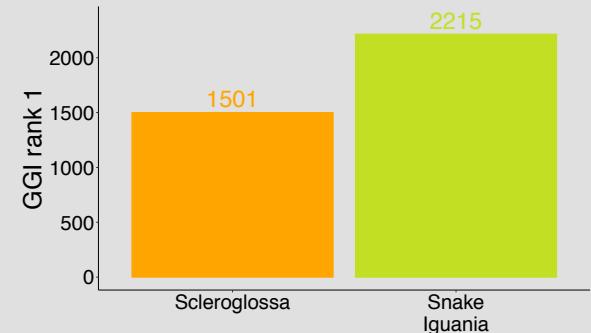
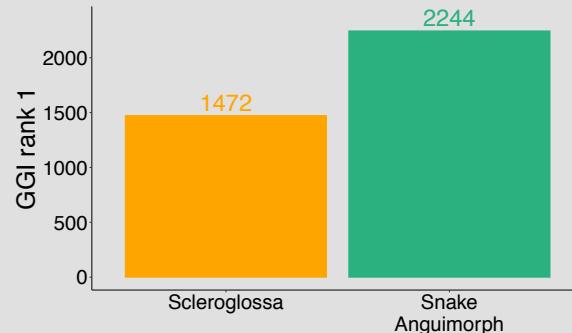
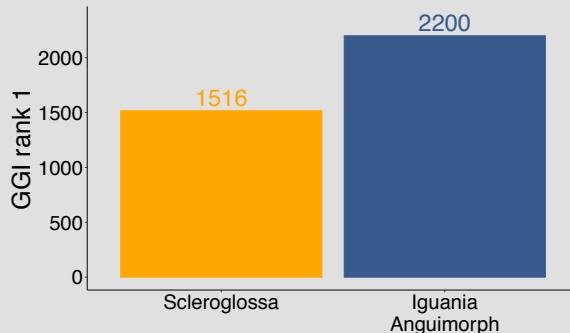
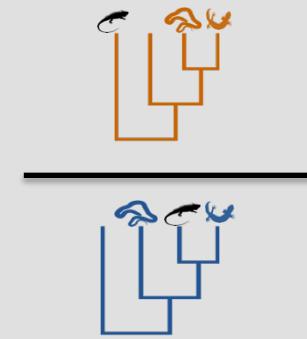
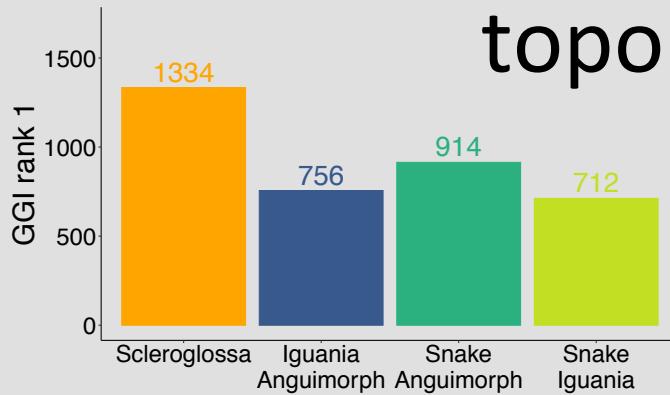
Singhal

Do the majority of loci support the same topology across methods? NO



dGLS
Streicher

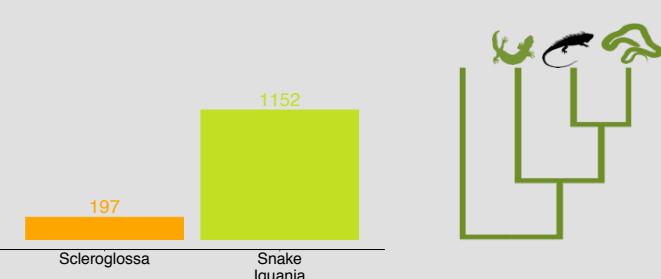
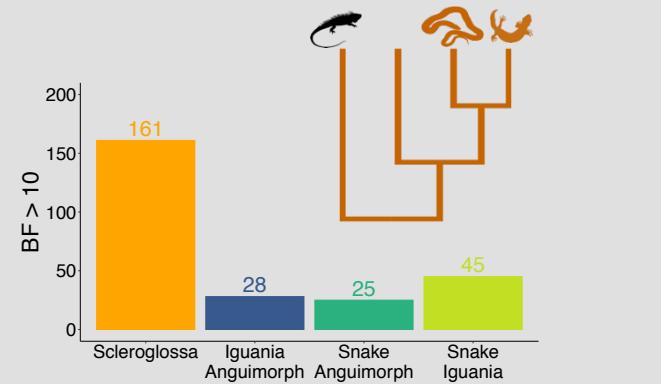
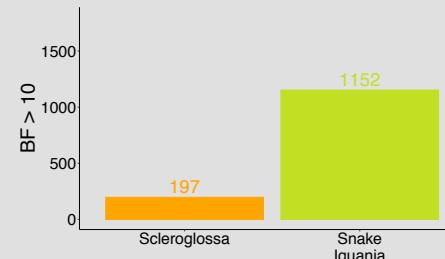
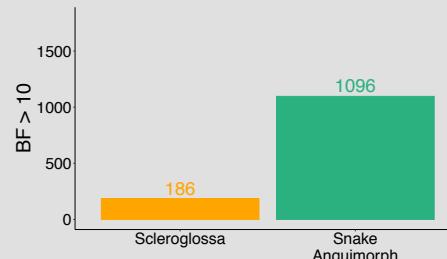
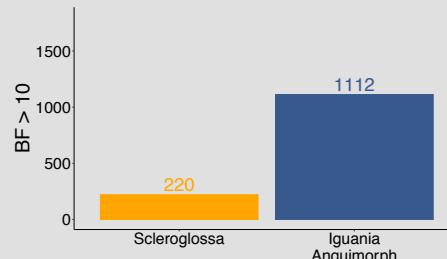
Pairwise comparisons give different than all topology comparison



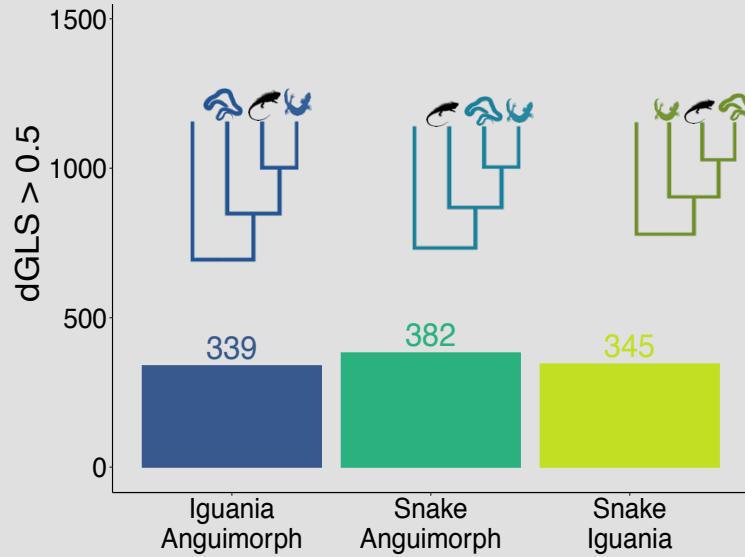
Singhal

Preferred topology may **change** when all topologies compared – BF, GGI

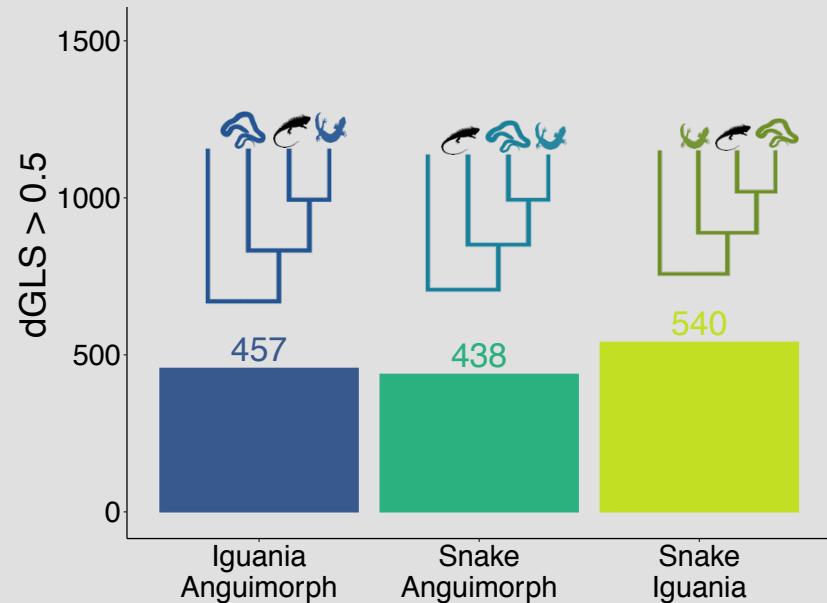
- Majority of loci support toxicofera for pairwise comparisons for all methods



Where does Iguania go?!?!

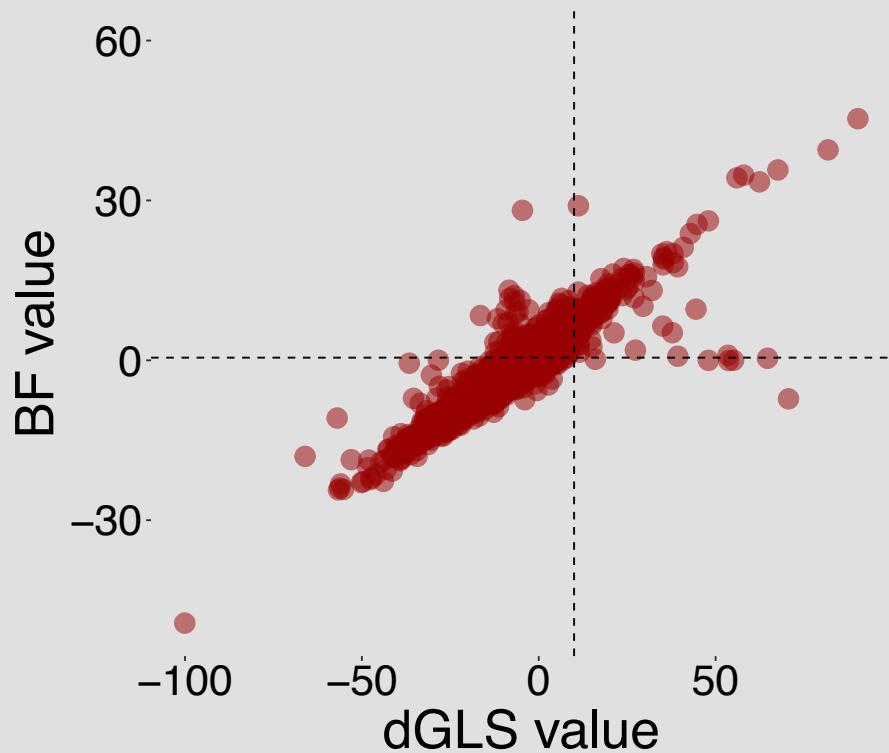


Singhal



Streicher

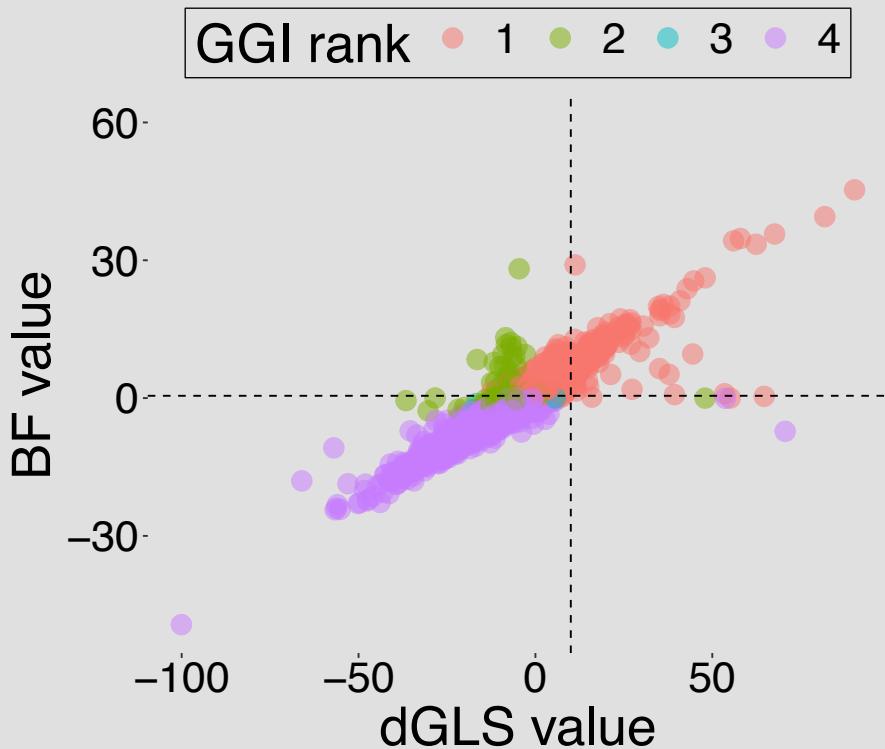
Are locus support values correlated across methods?



If correlated:

- Diagonal line

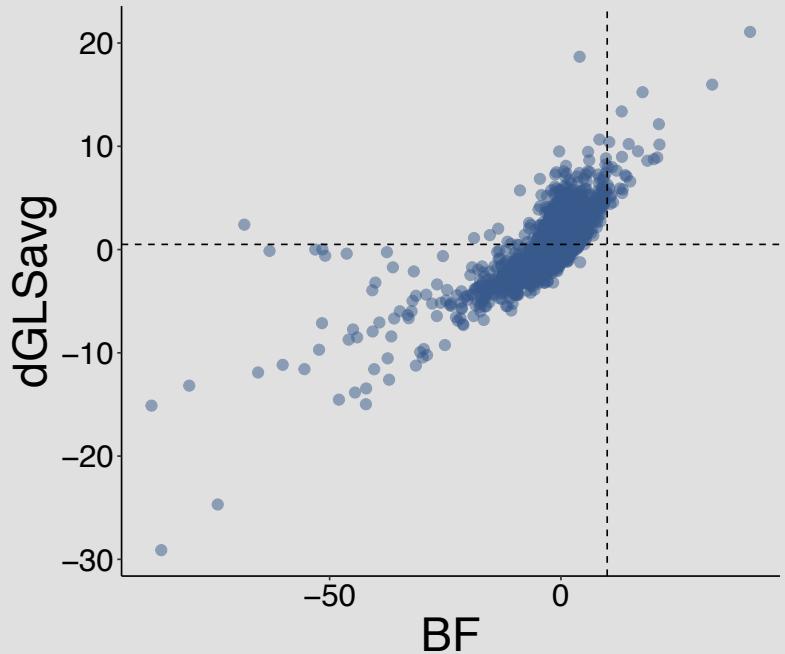
Are locus support values correlated across methods?



If correlated:

- Diagonal line
- GGI rainbow along line

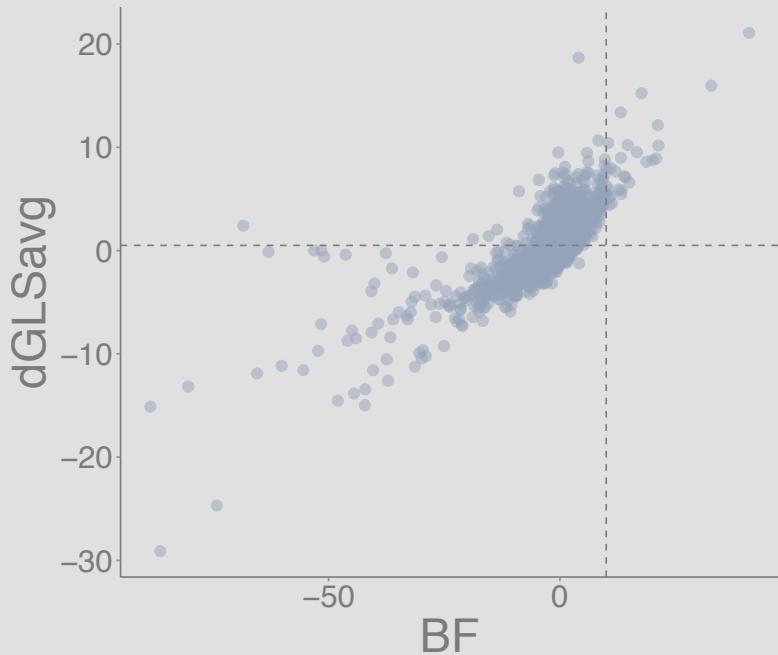
Correlation - BF vs dGLS



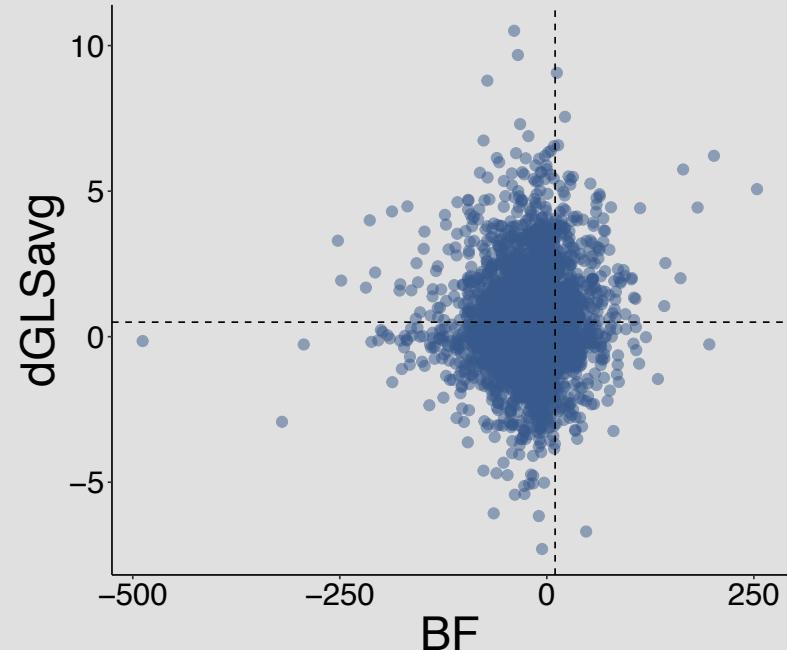
Streicher

Singhal

Correlation - BF vs dGLS

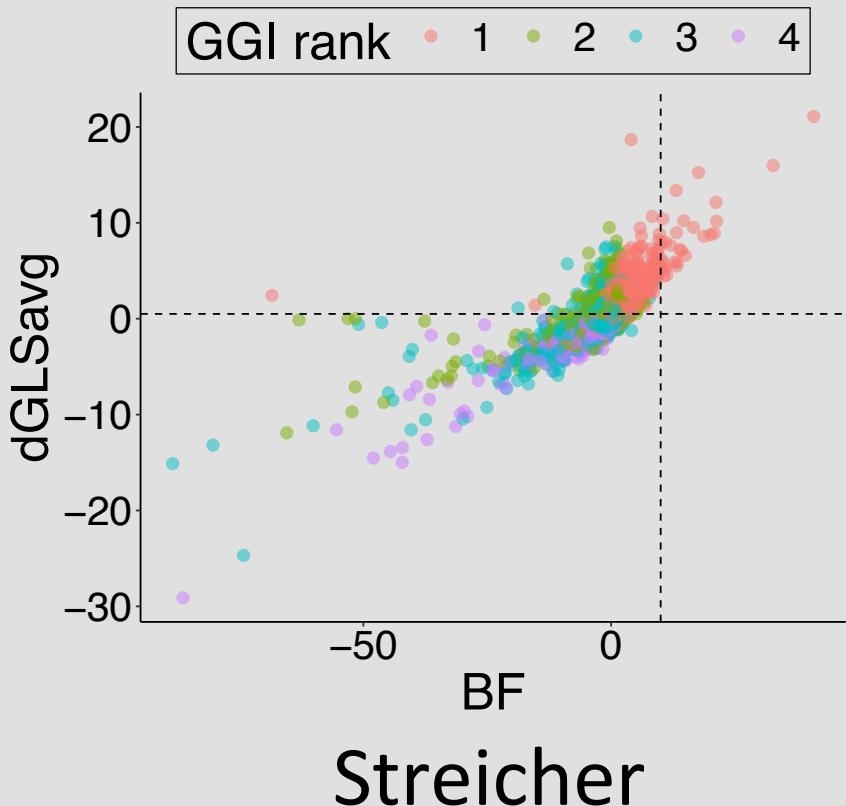


Streicher

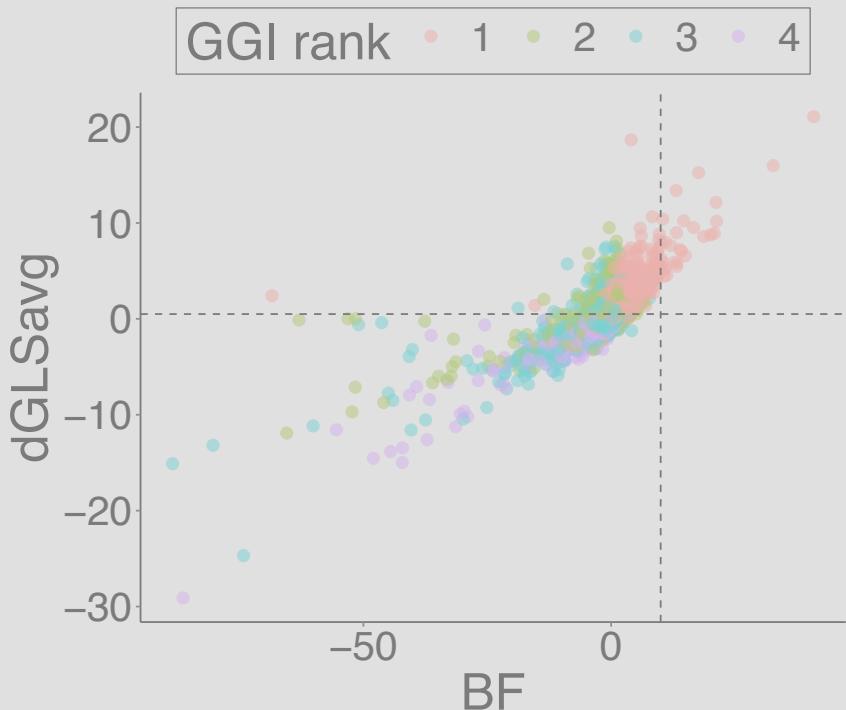


Singhal

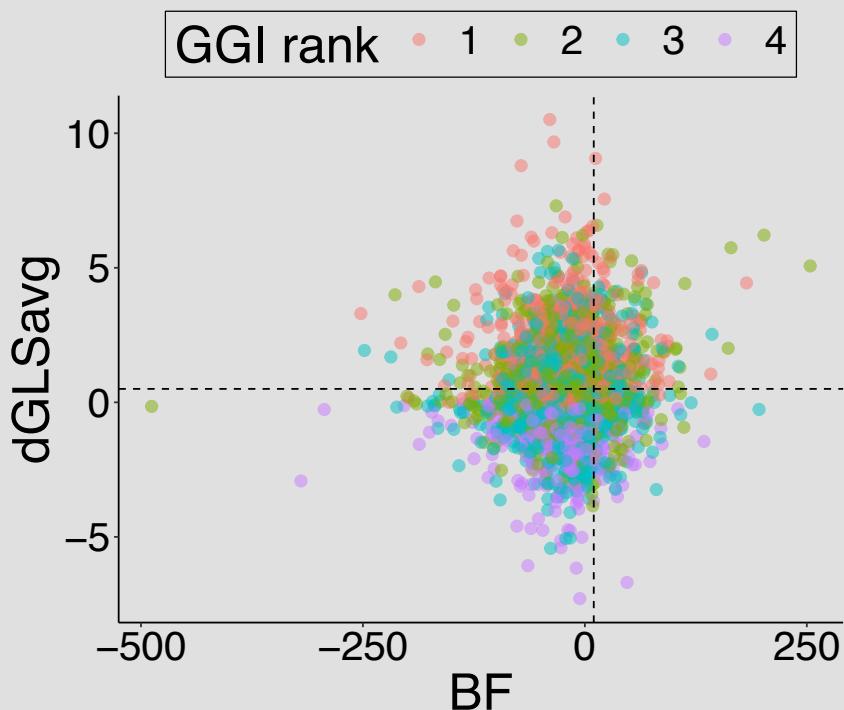
Correlation! BF vs dGLS vs GGI



Correlation? BF vs dGLS vs GGI



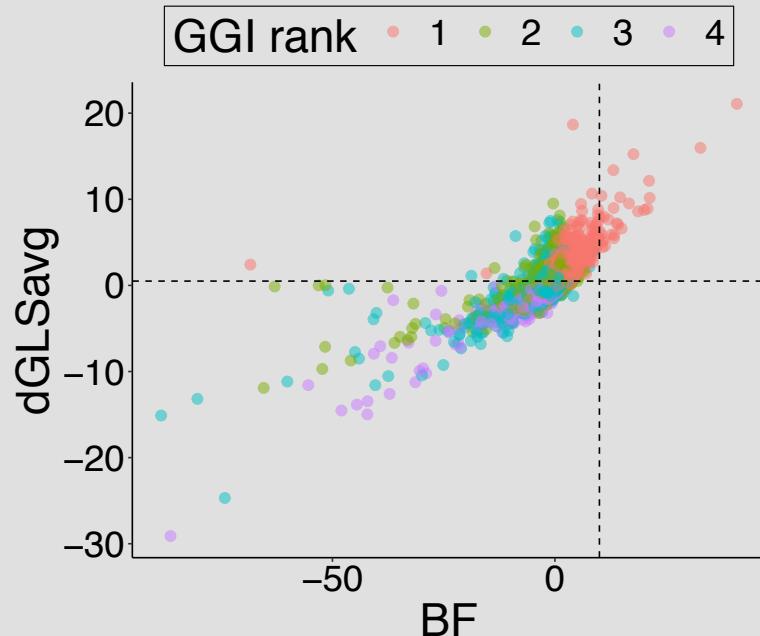
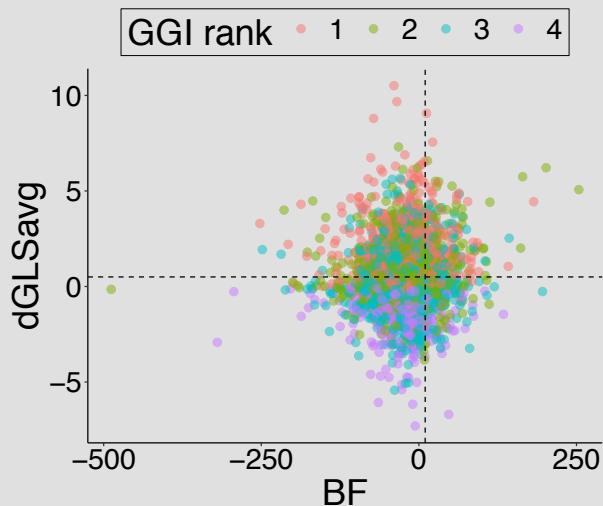
Streicher



Singhal

Some correlation across methods

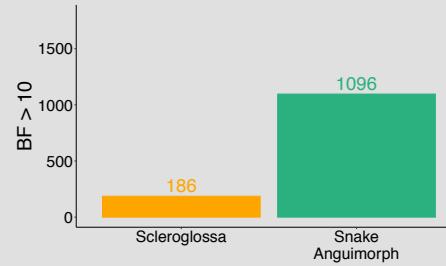
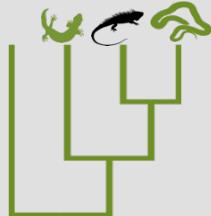
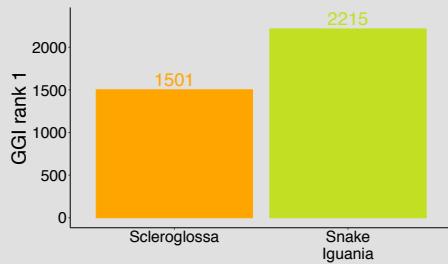
- **Some** correlation across methods for **some** data



Majority of loci support same
topology for pairwise
comparisons

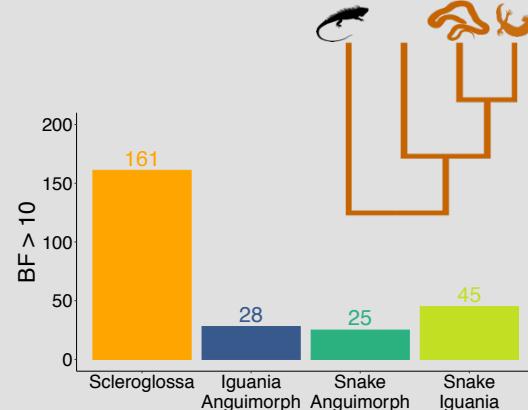
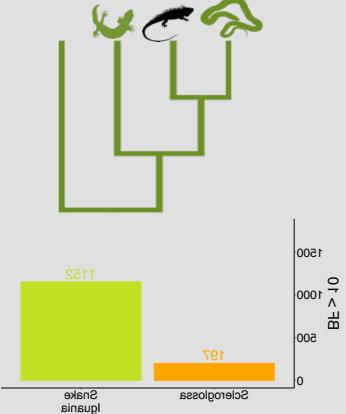
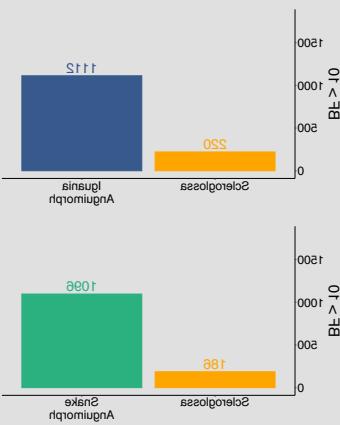
Majority of loci support same topology for pairwise comparisons

- Across all methods
- Not across data sets

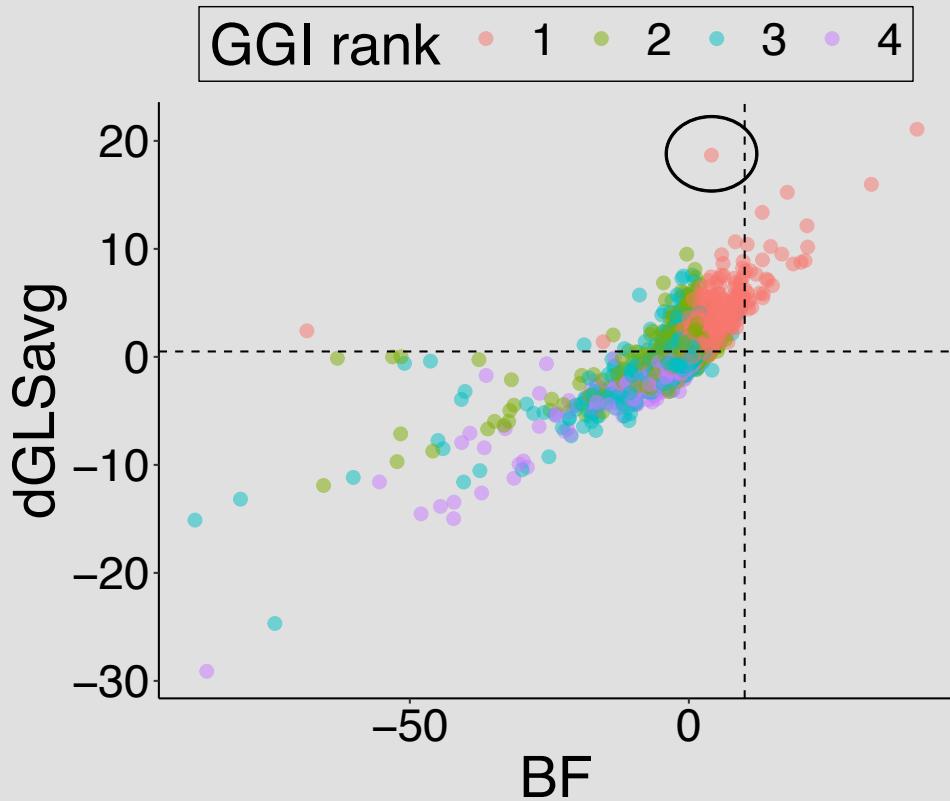


Input topologies matter

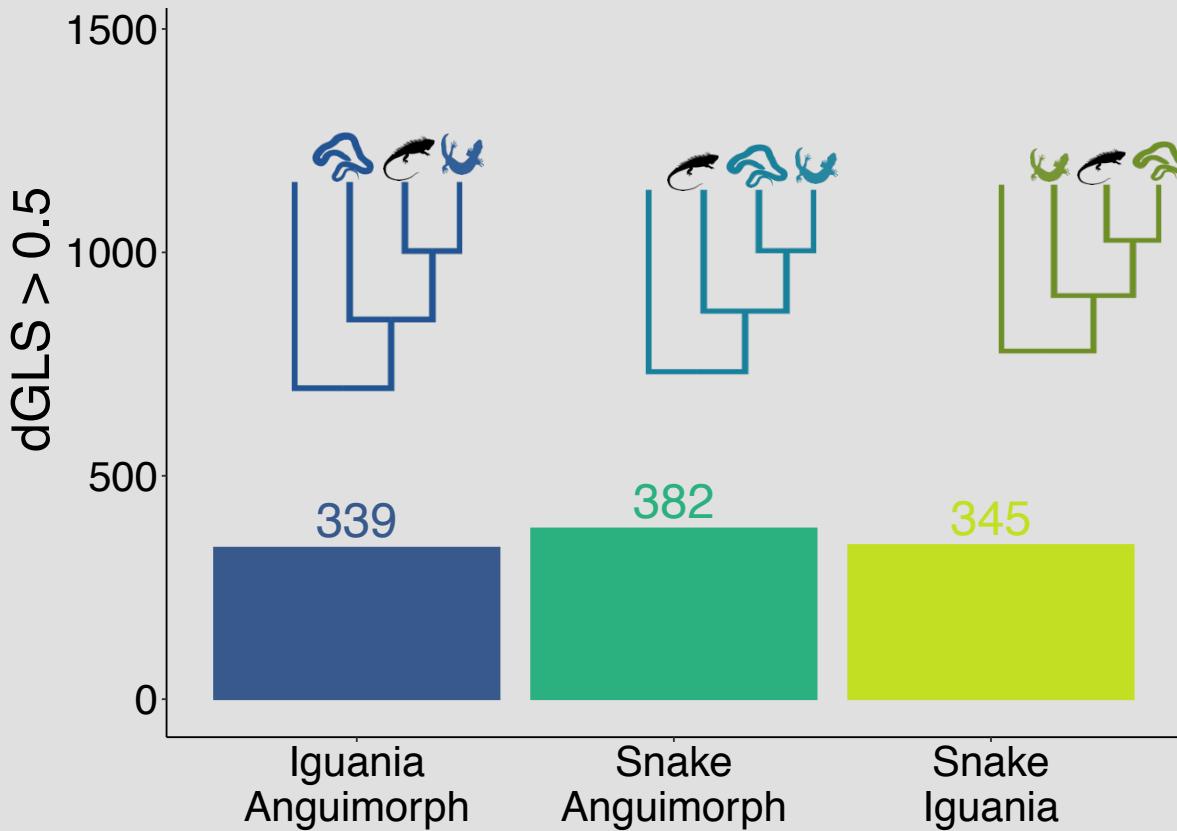
- Preferred topology may **change** when all topologies are compared (BF & GGI)



**Some
correlation
across methods
for some data**



Iguania – Too close to call



Take home

Use multiple methods

Acknowledgements



Dr. Jeremy Brown
Dr. Bob Thomson
Dr. Chris Austin
Dr. Lyndon Coghill
Dr. Vinson Doyle
Dr. Robb Brumfield
Dr. Laura Lagomarsino
Zach Rodriguez
Ben Toups

Acknowledgements

Graduating Summer 2020 (ish)
Looking for post-docs



Dr. Jeremy Brown
Dr. Bob Thomson
Dr. Chris Austin
Dr. Lyndon Coghill
Dr. Vinson Doyle
Dr. Robb Brumfield
Dr. Laura Lagomarsino
Zach Rodriguez
Ben Toups

 @Lizardeve

 gmount1@lsu.edu

Acknowledgements

Graduating Summer 2020 (ish)
Looking for post-docs

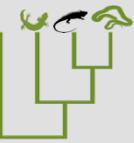
Questions?

 @Lizardeve

 gmount1@lsu.edu



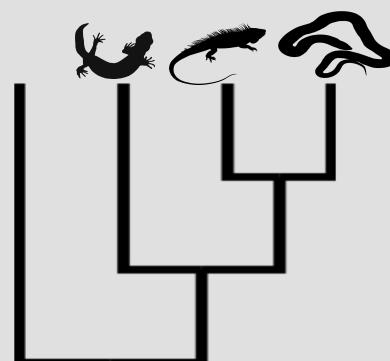
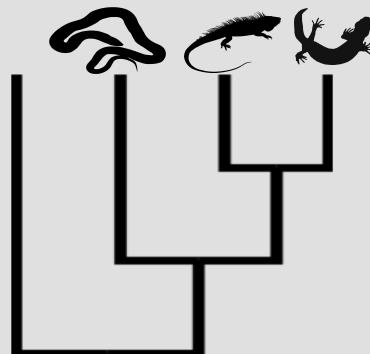
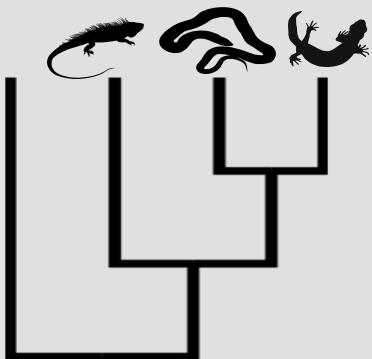
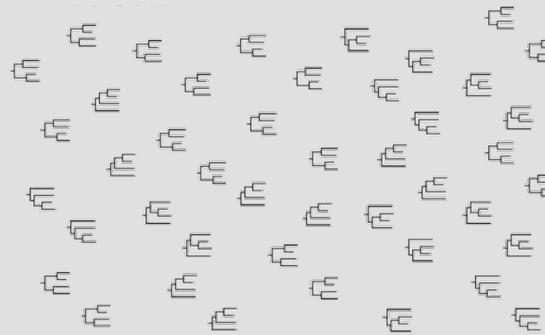
Dr. Jeremy Brown
Dr. Bob Thomson
Dr. Chris Austin
Dr. Lyndon Coghill
Dr. Vinson Doyle
Dr. Robb Brumfield
Dr. Laura Lagomarsino
Zach Rodriguez
Ben Toups

 — Avg( 



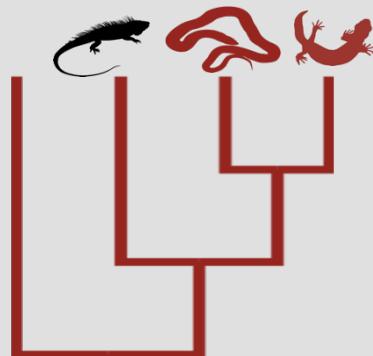
Quantify support for known topological conflict

- Genome wide data
- Information content and magnitude varies across loci



How much does each locus support each topology?

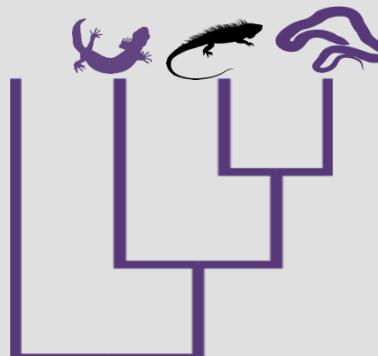
- Relative support for each topology



Locus 1

CCCCACACAACAACTACGGAGCCGGAAACCAAAACCGCAAA
CCACACGCCGGGGAGGGCGGTTTCCGATGGGGCGC
TAGACGCCAACTAGACCTGGGGGAATCATGGCGAACG
TCCACACAACAACTACGGAGCCGGAAACCAAAACCGCAAA

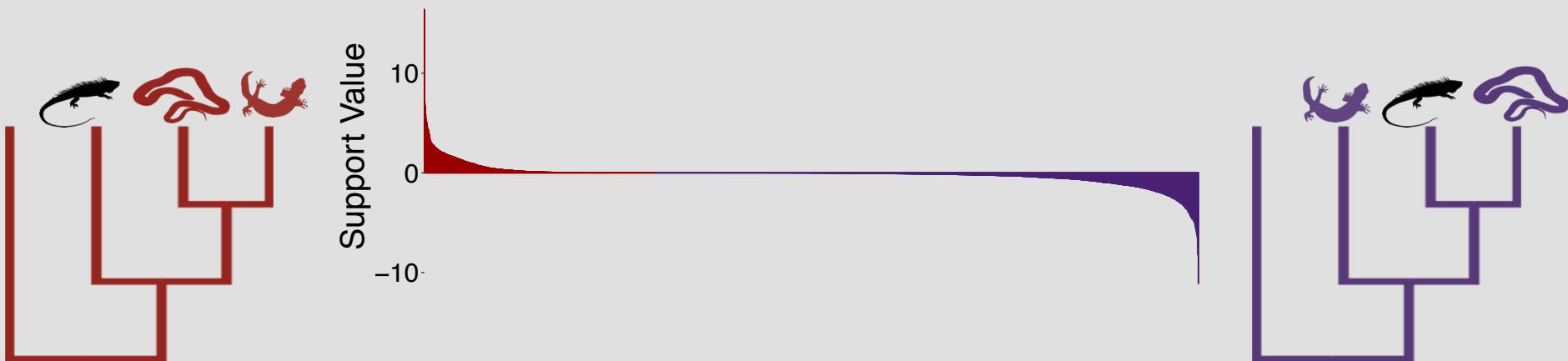
+300



CCCCACACAACAACTACGGAGCCGGAAACCAAAACCGCAAA
CCACACGCCGGGGAGGGCGGTTTCCGATGGGGCGC
TAGACGCCAACTAGACCTGGGGGAATCATGGCGAACG
TCCACACAACAACTACGGAGCCGGAAACCAAAACCGCAAA

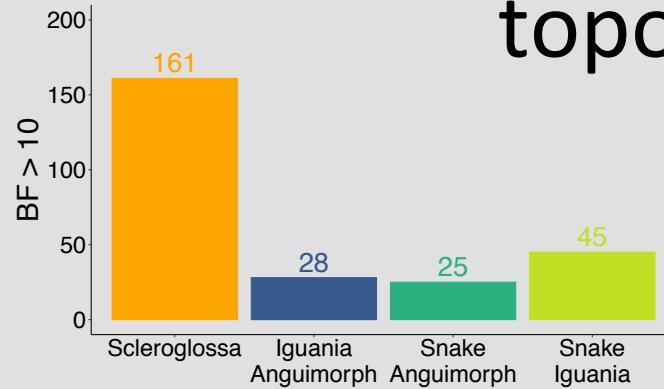
-25

Distribution of relative support values

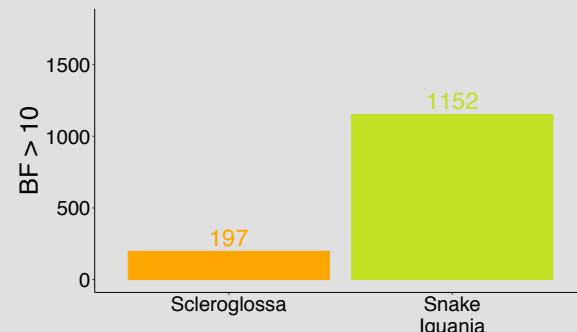
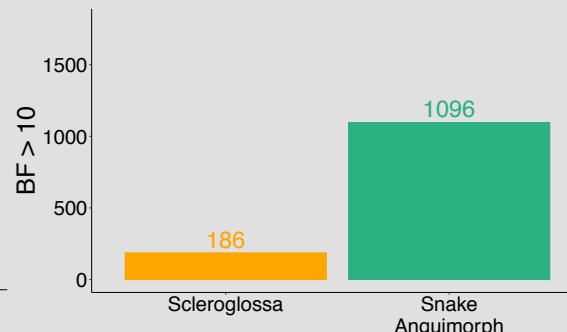
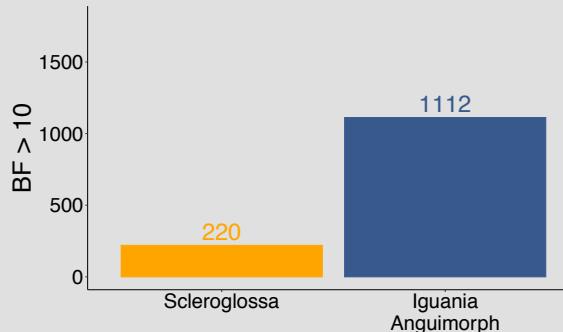


Sorted support values for all loci

Pairwise comparisons give different than all topology comparison



- All Toxicofera hypotheses preferred for pairwise GGI & BF prefer a Toxicofera



Streicher