

Phylogenetics and Phrogs: Methods, Morphology, and Conservation



Dr. Genevieve G. Mount
MVZ Seminar October 2022





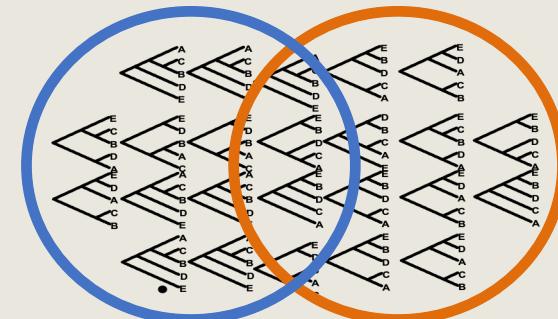
Phylogenetics and Phrogs



- Phylogenetics - Awesome and difficult
 - Leveraging variation
 - Featuring: Trees, Genes, and Morphology



- Phrogs (Frogs) - Just awesome
 - Evolution of trait loss and conservation genetics
 - **Not** featuring: vocal sacs



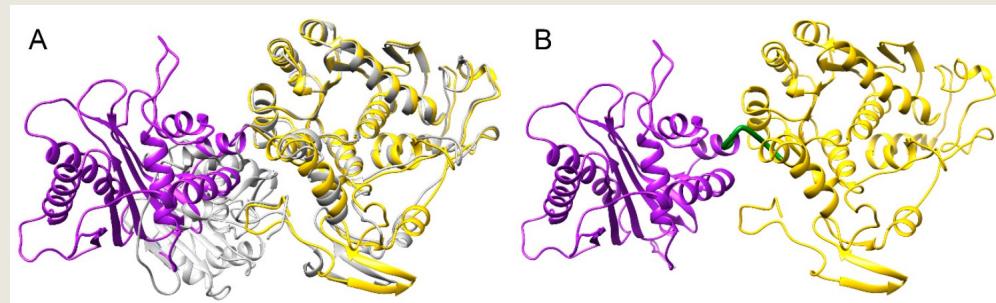
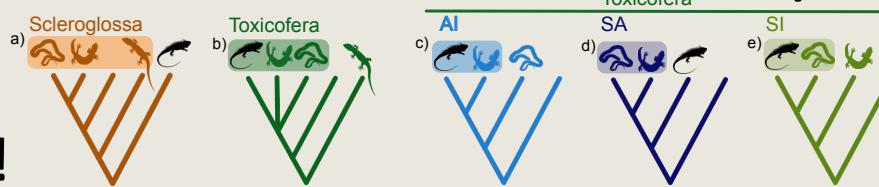


Phylogenetics – evolutionary biology and beyond

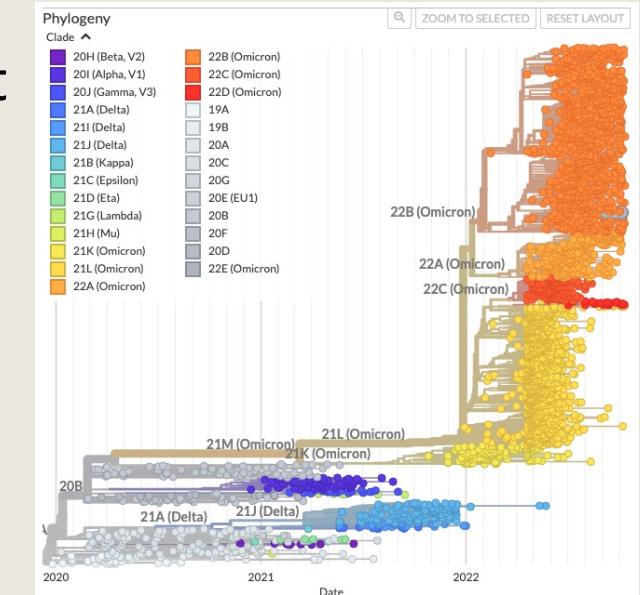
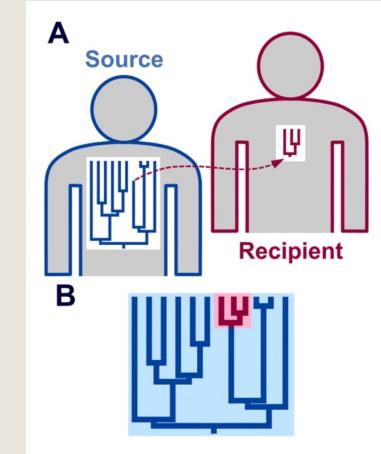
- Goal -> Reconstruction of evolutionary history

- Uses

- Evolution!
- Forensics – HIV transmission
- Virology/infectious diseases
- Drug development – identify molecular regions to target



Phylogenies
save lives



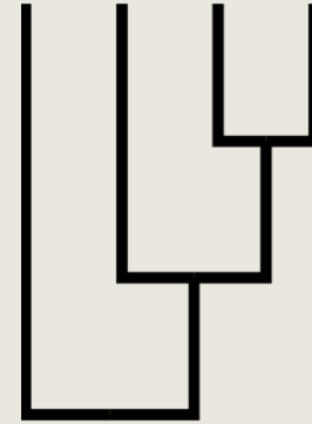
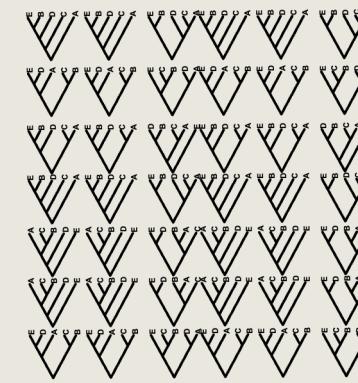


Goal: Reconstruct evolutionary history

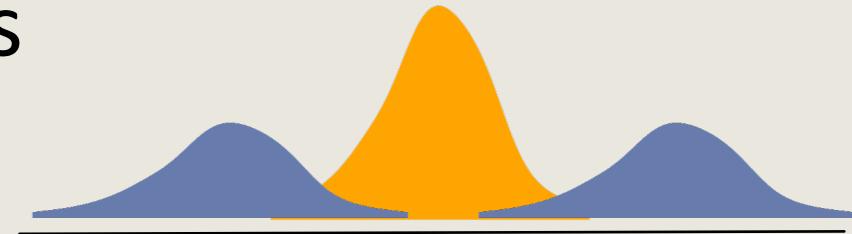


```
CCACACAAACA ACCAGGCCGGAAGCA AAAGCTAAA  
CCACACAAACA ACCAGGCCGGAAGCA AAAGCTAAA  
GATCCACACAAACA TACGAGCCGGAAACCA AAAGCTAAA  
TTCGCAACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
CCAATCA ACCACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
AGATCCACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
TTCGCAATCA GTCACACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
CCAATCA GTCACACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
GATCCACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
TTCGCAACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
CCAATCA GTCACACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA  
GAGCGTAA AGAGG TGGGGAAACCA AGCGAGCCGGAAACCA AAAGCTAAA  
CCACACAAACA AGCGAGCCGGAAACCA AAAGCTAAA
```

```
elif [ "$model" == "JC+I" ]  
cp JC+I.bayesblock $gene  
elif [ "$model" == "JC+G" ]  
then  
cp JC+G.bayesblock $gene  
else  
echo "The model ($model) you have chosen for $gene is not one of  
the standard models. You will need to create your own bayesblock file to add to  
done < $1  
for f in *.nex  
do  
base=`basename $f .nex`  
cd base  
count=1>bayesblock  
if [ $count -eq 1 ]  
then  
sed -i .tmp ">/data/$base/g" bayesblock  
rm bayesblock.tmp  
wdir=$(pwd)  
bb=$wdir/bayesblock  
echo $wdir"/$bb" > ./empDataList  
echo $base "is ready to be analyzed!"  
else  
echo "There is no bayesblock file in" $base  
fi  
done
```



- Single tree for downstream analyses
- Average across information in data
- Problem: Loss of information
- Is our tree missing relevant information?





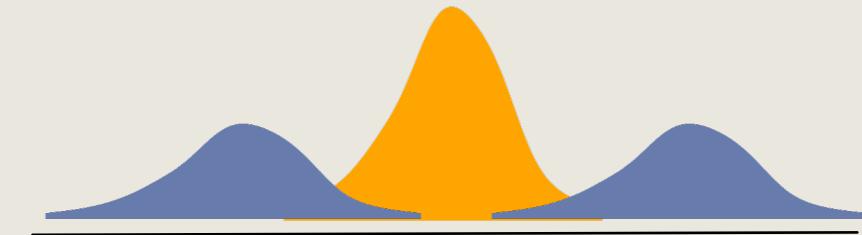
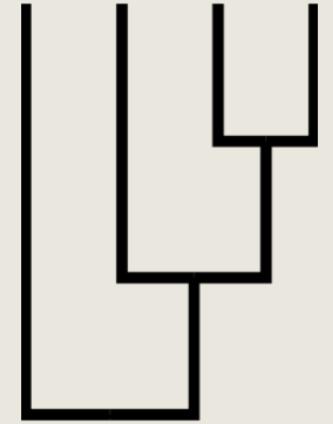
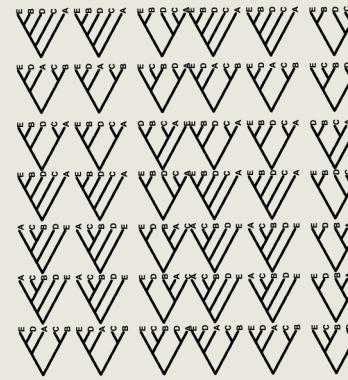
Is our single tree missing relevant information?



A stack of four DNA sequence alignments showing nucleotide variations across multiple genes. The sequences are color-coded by base (A, T, C, G).

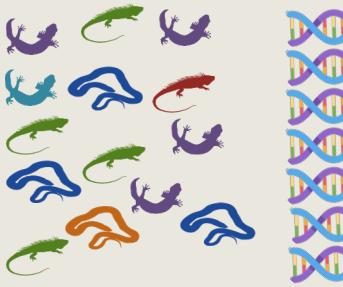
```
elif [ "$model" == "JC+I" ]
cp JC1.bayesblock $gene
elif [ "$model" == "JC+G" ]
then
cp JCIG.bayesblock $gene
else
echo "The model ($model) you have chosen for $gene is not one of
echo "You will need to create your own bayesblock file to add to
done < $1

for f in *.nex
do
base=${f%.nex}
cp $base
count=1>bayesblock | wc -l
if [ $count != 0 ]
then
rm $base.tmp >/dev/null
rm $base.bayesblock
rm $base.bayesblock.tmp
wc -w $base.bayesblock
else
$base>$base.bayesblock
echo $base" is ready to be analyzed"
else
echo "There is no bayesblock file in" $base
fi
done
```



- Big data = lots of variation
- Conflicting signal
 - Support for alternate topologies
- Source - biological & methodological
 - Convergent evolution, horizontal gene transfer, misidentified parologue, inappropriate models

So what if we miss information?

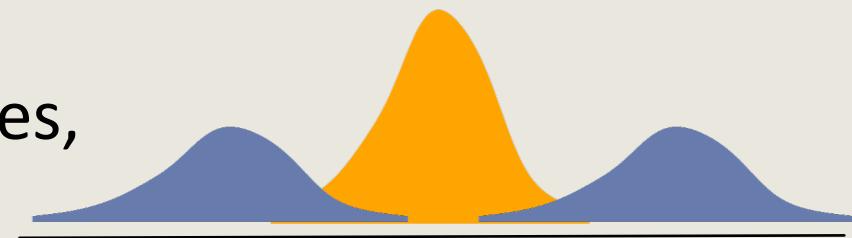
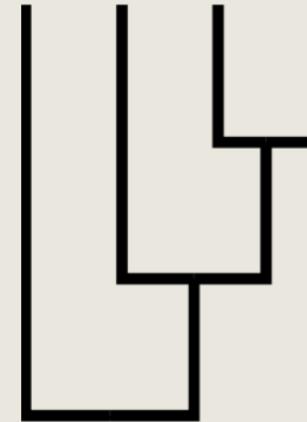
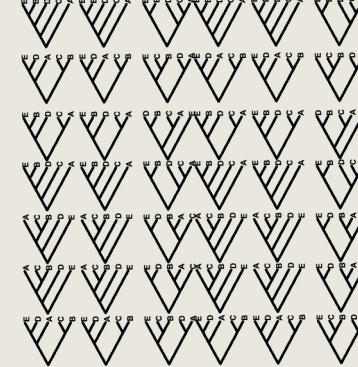


```

        elif [ "$model" == "JC+I" ]
        then
        cp JCI.bayesblock $gen
        elif [ "$model" == "JC+I+G" ]
        then
        cp JCIG.bayesblock $gen
        else
        echo "The model ( $model ) you have chosen for $gen is not one o
        echo "You will need to create your own bayesblock file to add to
        fi
done < $1

for f in *.nex
do
base=`basename $f .nex`
if [ $base = "base" ]
then
counts `ls -1 *bayesblock | wc -l`
if [ `$count` != 1 ]
then
rm -f tmp
rm -f bayesblock.tmp
rm -f bayesblock
wD=$base
base=$wD
base=$base"bayesblock"
echo $wD"/$base >> ../../empData.list
echo $base "is ready to be analyzed"
else
echo "There is no bayesblock file in" $base
fi
done
cd ..
done

```

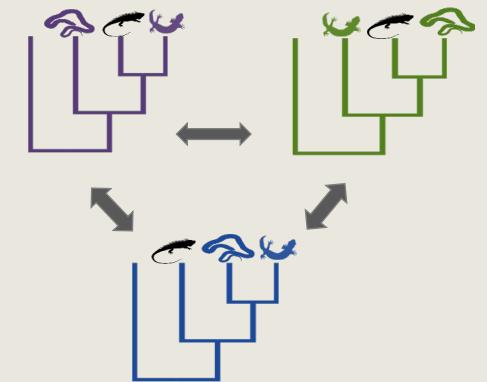


- Inaccurate inferences
 - Historical relationships, evolutionary processes, molecular evolution...
 - Extract larger patterns of variation – capture information



Identify and quantify variation

- Does our data support multiple topologies?
 - What cool information can we find?!?!
- Agnostic approaches
 - No *a priori* assumptions of processes that led to variation
 - Applicable to all systems

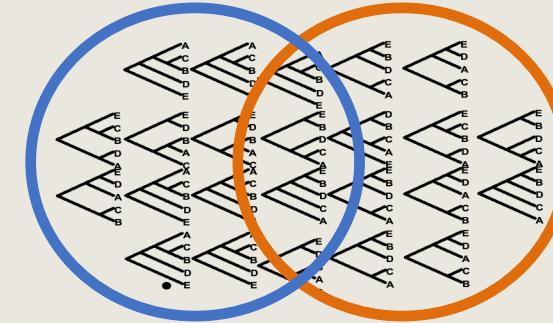




Identify and quantify variation: Is there support for multiple topologies?



- Trees
 - Identify broad patterns
- Genes
 - Quantify support for conflicting topologies & compare methods
- Morphology
 - Quantify support for conflicting topologies – in a different way

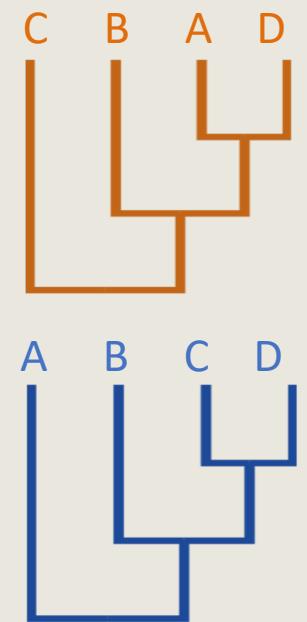
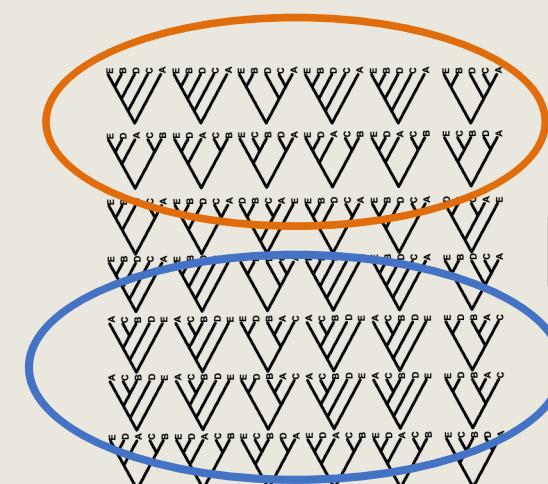
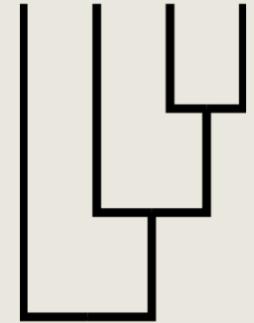
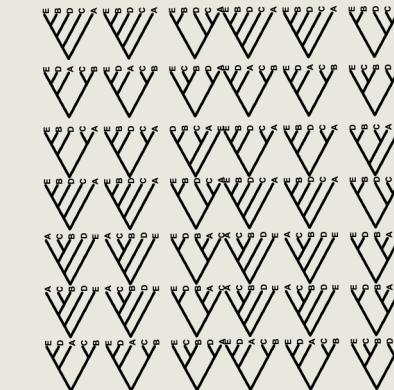




Trees

Can we identify broad variation in tree sets?

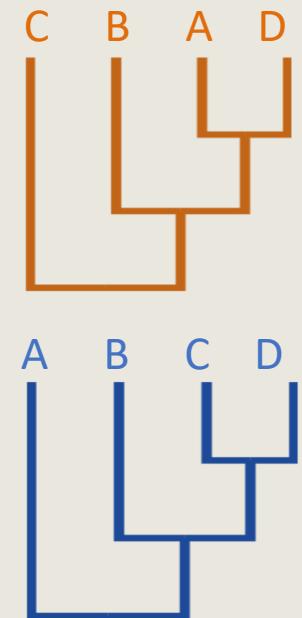
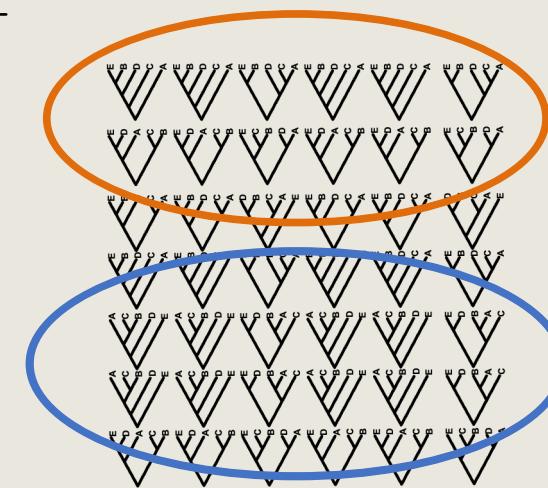
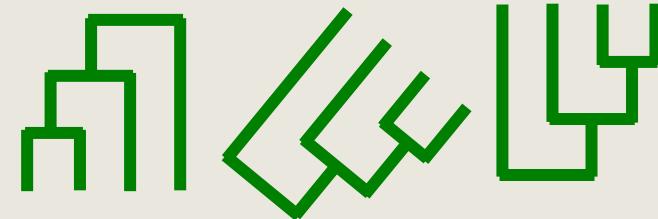
- Set of trees
 - Gene trees, multiple analyses, posterior distribution
- Group trees that contain similar information
 - Find conflicting signal, identify outlier loci, inspect posterior distribution, evaluate MCMC mixing





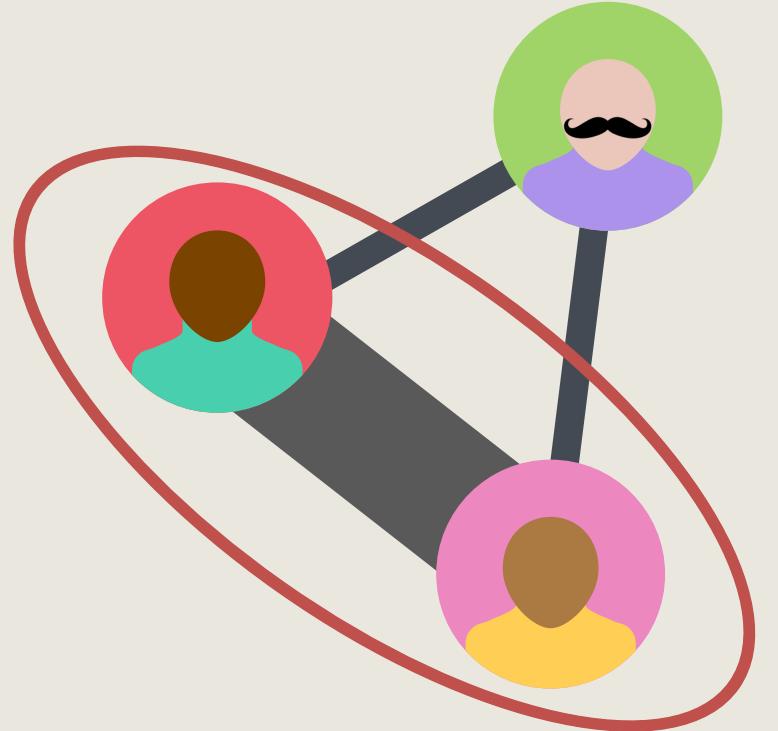
Trees are graphs

- Not a number
- Unknown expected distribution
- Multidimensional tree space
- Current methods
 - Multidimensional scaling, clustering
- How do we identify similar trees without losing information?



Network community detection – groups of similar trees

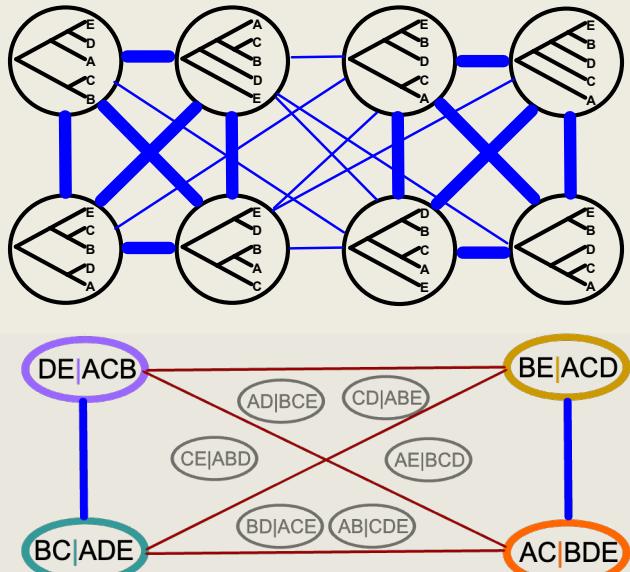
- CloudForest (TreeScaper)
- Networks
 - Node - information.
 - Edge - relationships between two nodes.
 - Preserve information
- Community detection
 - Extensive research outside of phylogenetics



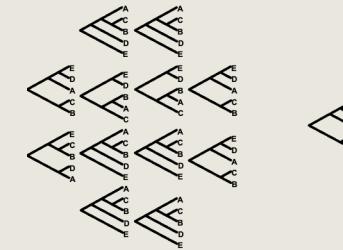


Can we use network community detection to identify variation in tree sets?

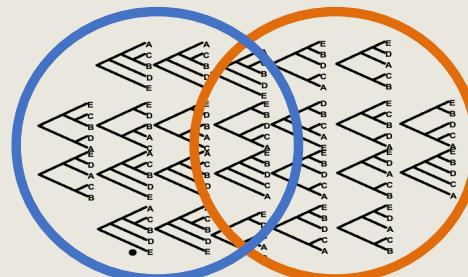
- Two types of information
 - Trees
 - Bipartitions



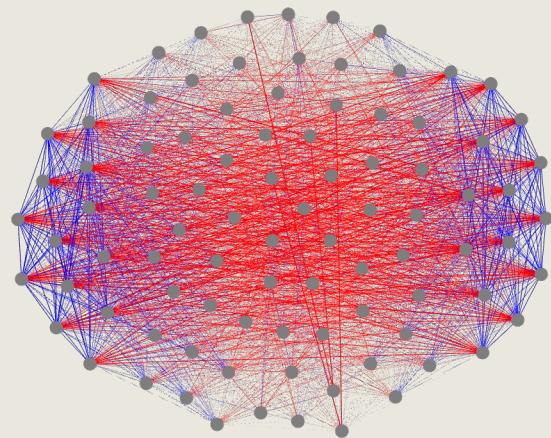
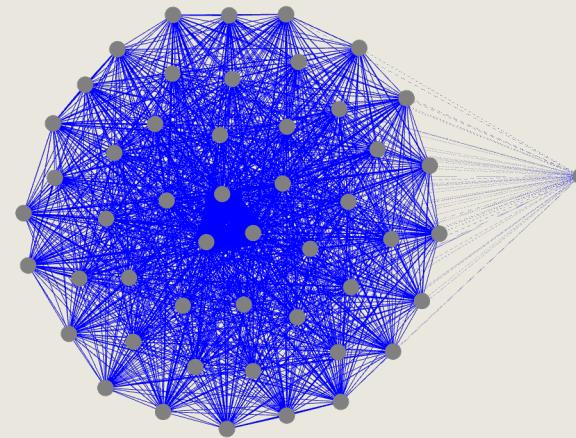
- Simulate difficult scenarios
 - Outliers



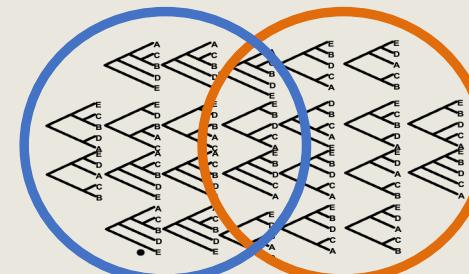
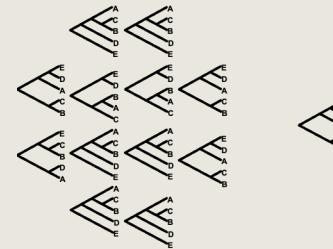
- Subtle differences



Accurate identification of simulated groups

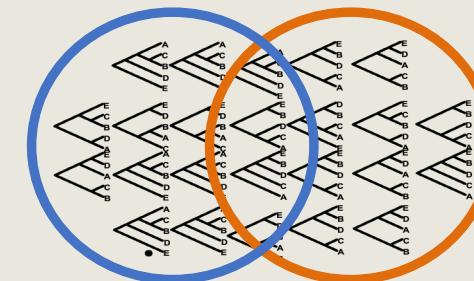
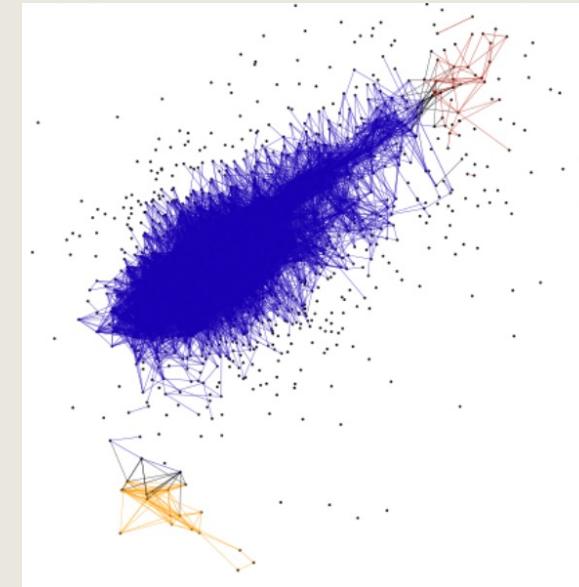


- Difficult scenarios
 - Outliers – tree networks
- Subtle differences – bipartition networks



Network community detection

- Big cat phylogeny (Li et al 2019)
 - Identified low recombination region on X chromosome
- Interesting and important biological patterns
 - No previous knowledge of system necessary

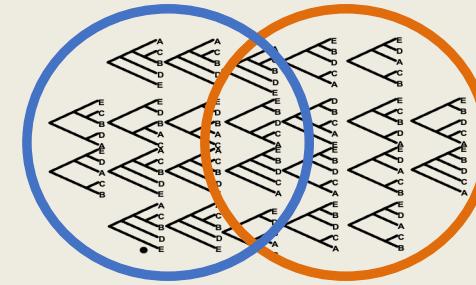




Identify and quantify variation: Is there support for multiple topologies?



- Trees
 - Identify broad patterns
 - Network community detection
 - Lose less information
 - Tackle difficult problems

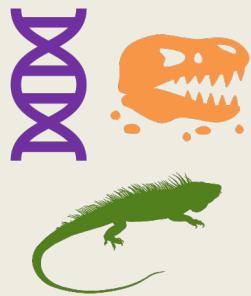
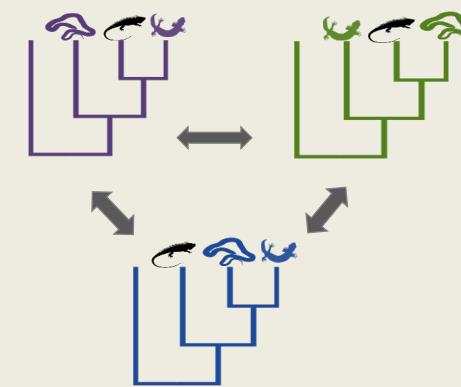




Identify and quantify variation: Is there support for multiple topologies?



- Trees
 - Identify broad patterns



- Genes & Morphology
 - Known: Conflicting topologies
 - Unknown: Support for each topology



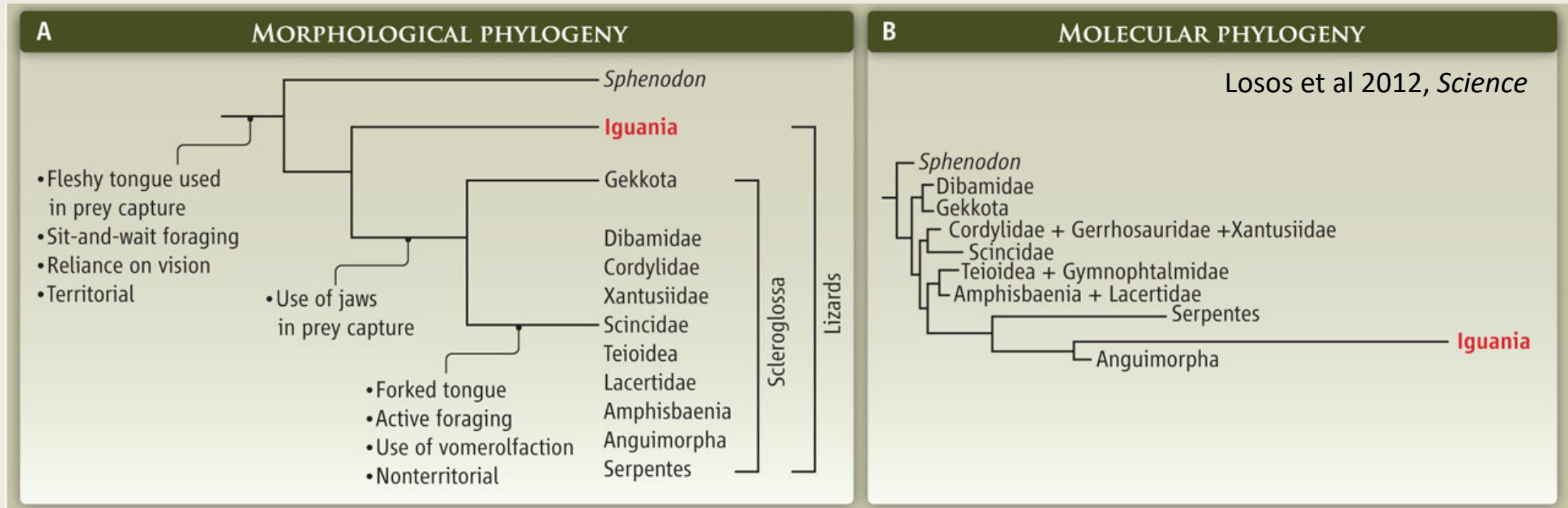


Squamates!





Conflict within and across data types





Toxicofera relationships

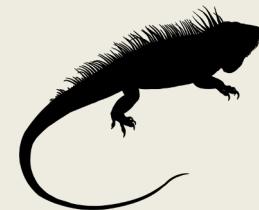
Inconsistent support across genetic studies



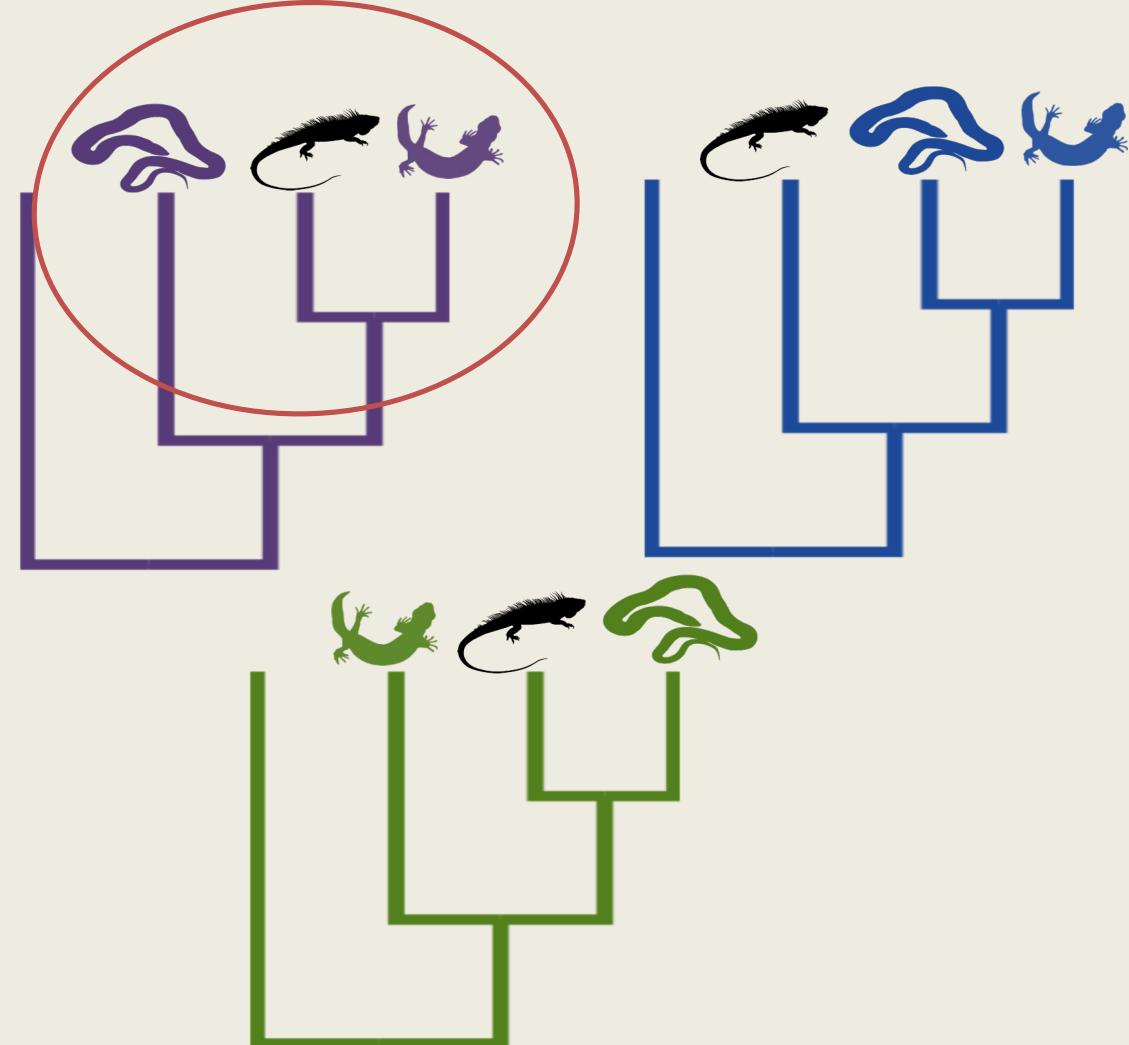
Snakes



Iguania



Anguimorphs





Identify and quantify variation: Is there support for multiple topologies?

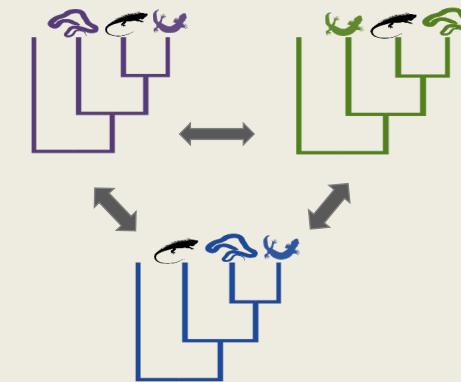


- Trees
 - Identify broad patterns



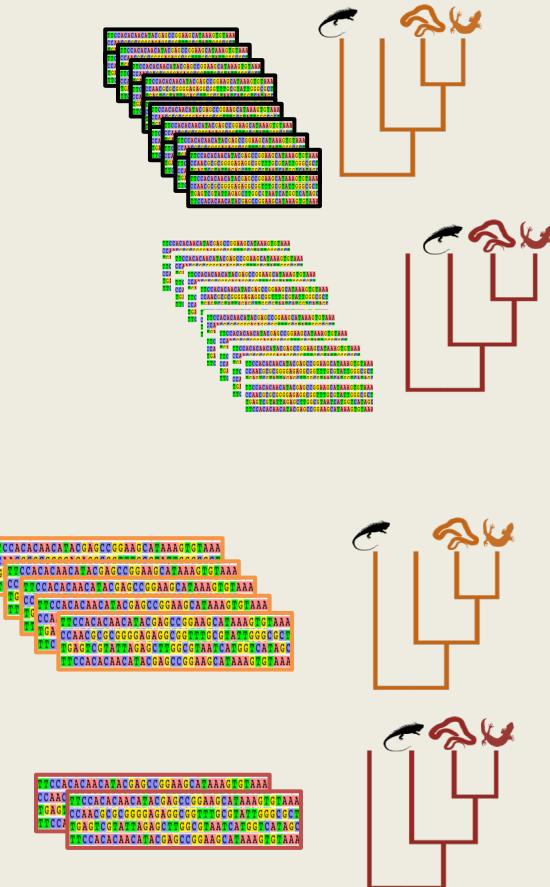
- Genes
 - Support for each topology?
 - Consistency across methods?

- Quantify support
- Does the amount of support vary across methods?



Phylogenetic support in big data

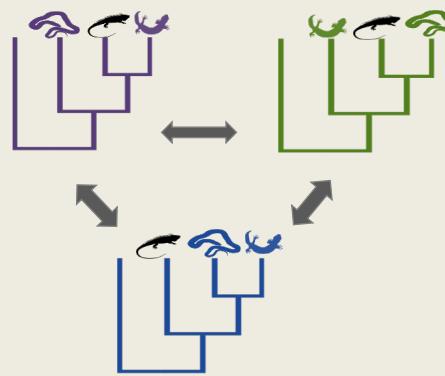
- Genome scale data maximizes traditional support metrics
- Locus by locus approach
 - Compare relative support for each topology
- Multiple types of analyses
 - Bayes factors (BF), difference in gene log likelihood (dGLS)
 - Thresholds for “strong” support
 - Count loci for each topology
- Compare support across methods
 - Are values correlated?



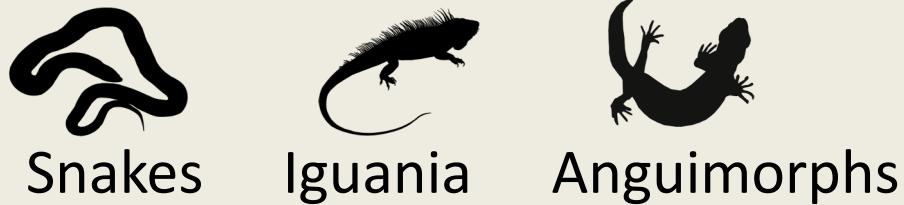
Is there conflicting support?

- Data

- Pairwise comparisons
- Number of loci
- Magnitude of support values

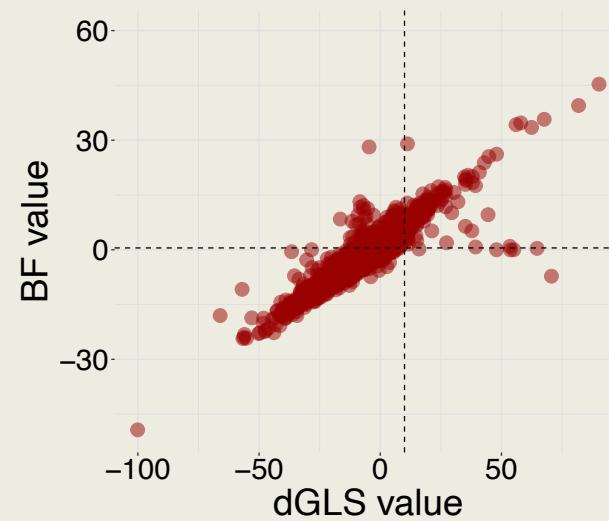


Dataset	Taxa	Loci	Locus Type
Burbrink	289	227	Anchored Hybrid Elements (AHE)
Reeder	161	46	Legacy Sanger Loci
Singhal	69	4926	AHE, UCE, Sanger
Streicher	32	3961	Ultraconserved Elements (UCE)



- Methods

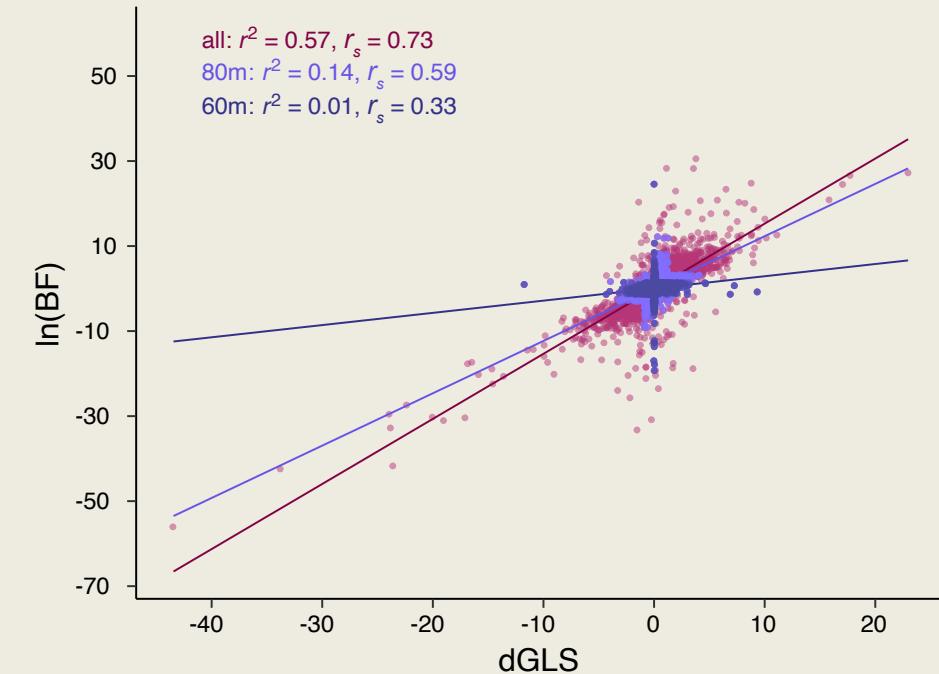
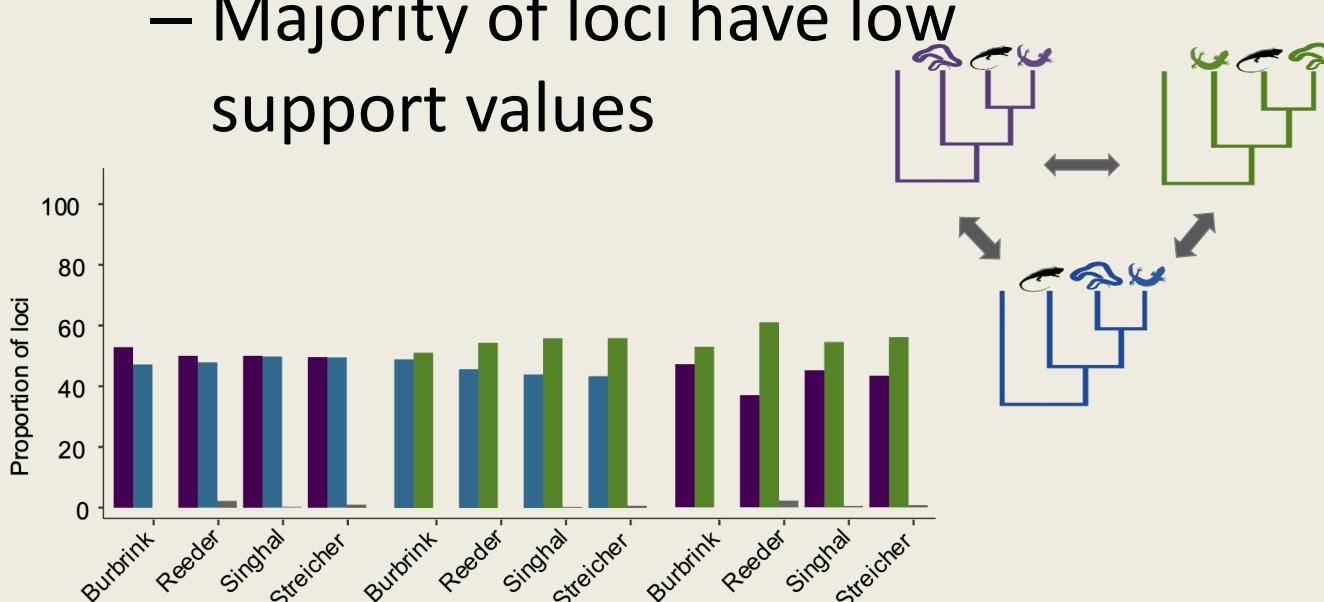
- Bayes factors(BF) vs difference in gene log likelihoods(dGLS)



Is there conflicting support?



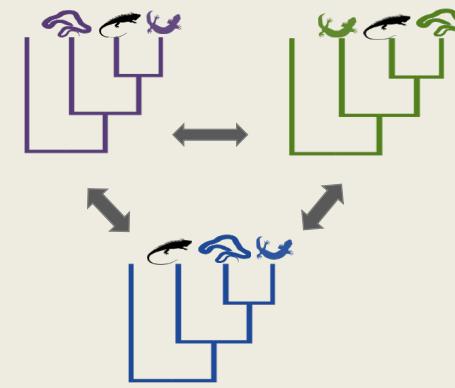
- No strong support for any one topology
 - Equal number of loci
 - Majority of loci have low support values
- Low support values not correlated across methods
 - High values correlated



Identify and quantify variation: Is there support for multiple topologies?



- Trees
 - Identify broad patterns
- Genes - Conflicting topologies
 - Questions:
 - Support for each topology
 - Consistency across methods
 - Answers:
 - Low, similar support for all possible toxicofera topologies
 - Methods consistent for high support values, not low values





Identify and quantify variation: Is there support for multiple topologies?



- Trees
 - Identify broad patterns

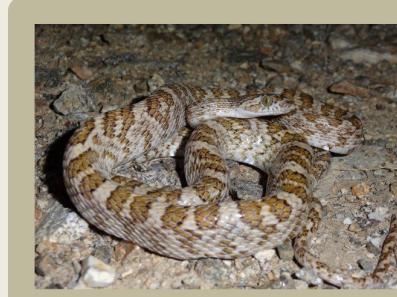
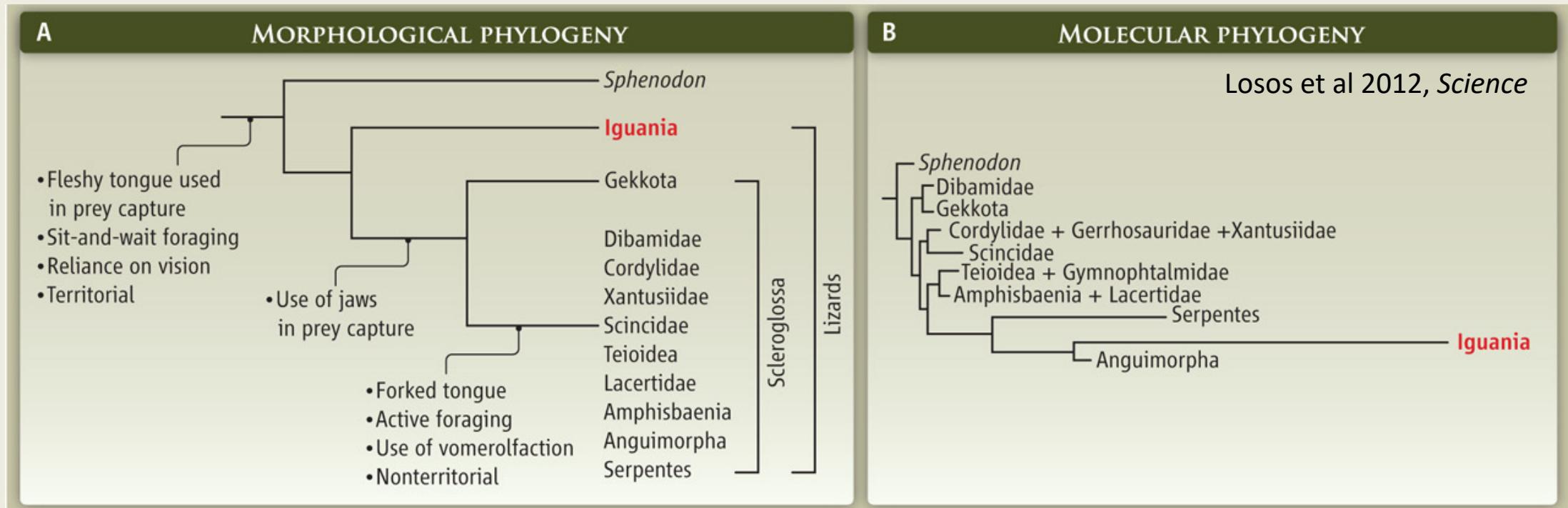


- Genes
 - Quantify support & compare methods
- Morphology – conflicting topologies



- Quantify support
- Learn more about morphological evolution

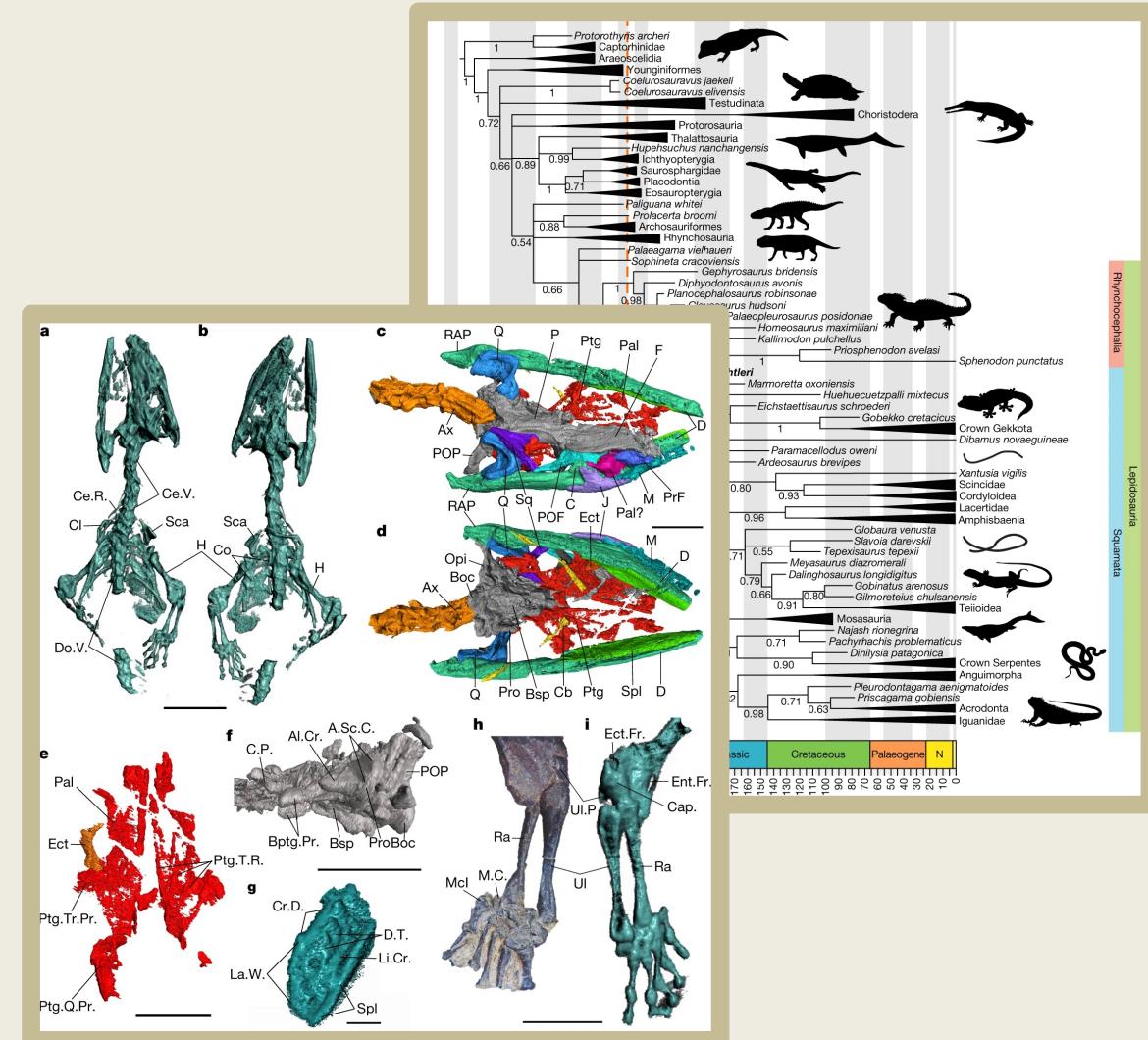
Conflict within and across data types





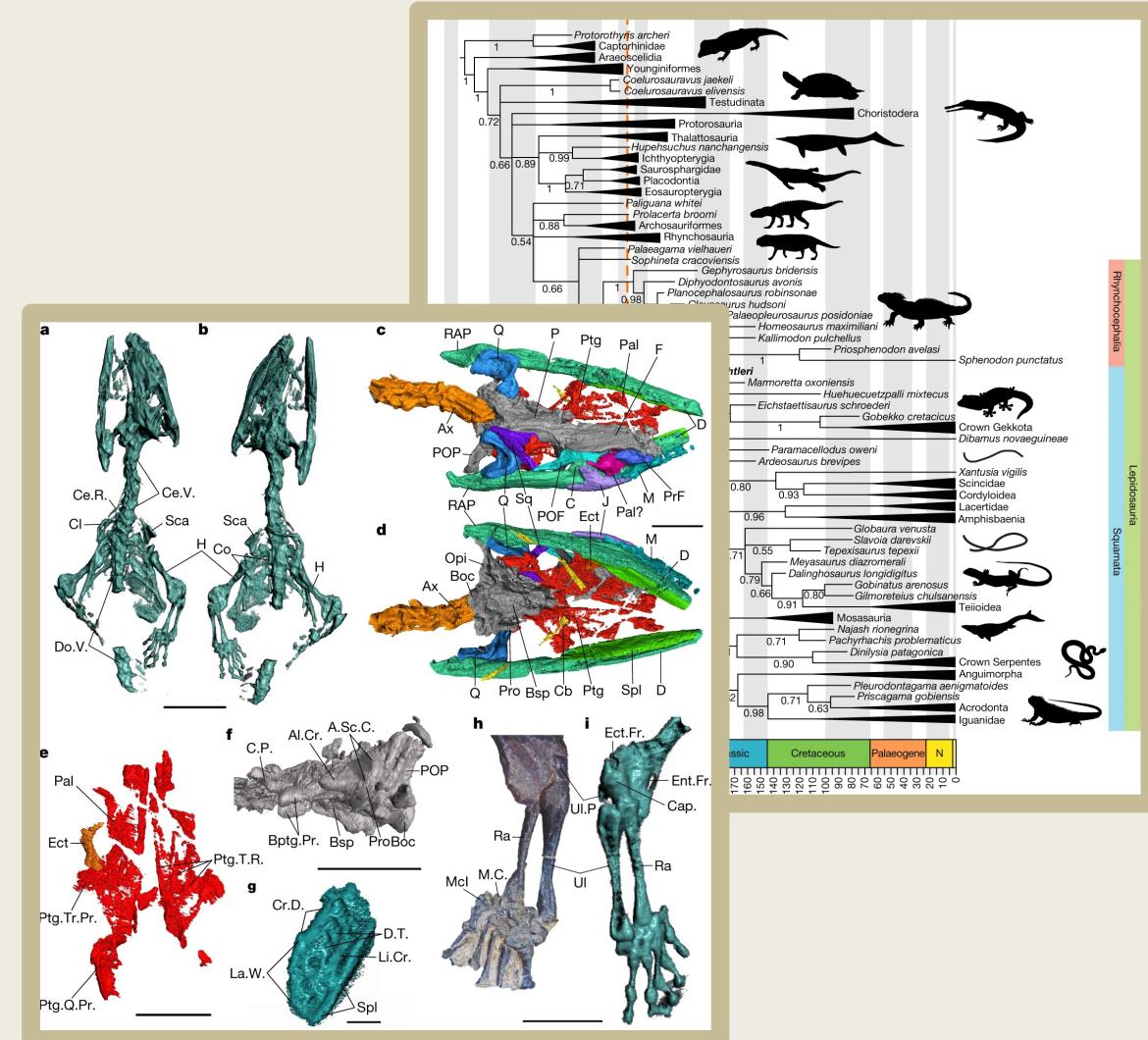
Morphological phylogenetics is important

- Placement/dating of fossils
- Morphological evolution



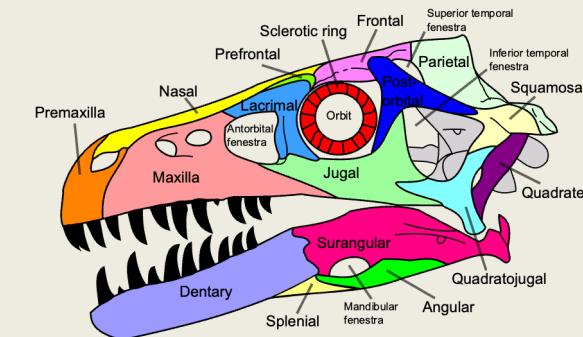
Quantify support for conflicting topologies

- Group traits with similar evolutionary histories
 - Complex models
 - Morphological equivalent of genes?
 - Modules of evolution -unknown
- Can we quantify support AND identify modules of evolution?

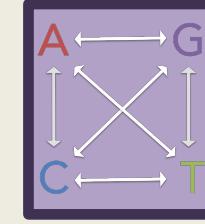
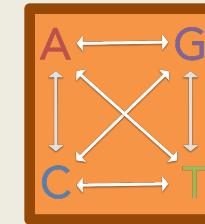


Quantify support and modules of evolution

- Group traits with similar evolutionary histories
- Bayesian framework
- Mixture model
 - Class – sub-sets of model
 - Estimate parameter values
 - Preference of character to class
 - CAT model – class - different equilibrium frequencies



```
ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA  
CCAACGGCGGGGGAGAGGGGGTTTGC GTAA TGGGGCTT CCAACGGCGGGGGAGAGGGGGTTTGC GTAA TGGGGCTT  
TGAGCTCGTAATAGAGCTTGGCGTAAACAGGCA TAGC TGAGCTCGTAATAGAGCTTGGCGTAAACAGGCA TAGC  
ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA
```

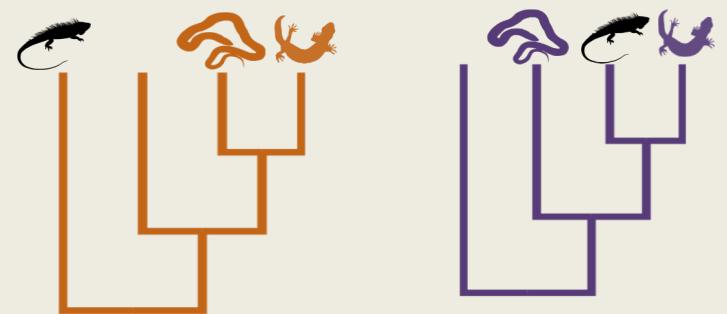


<pre>ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA CCAACGGCGGGGGAGAGGGGGTTTGC GTAA TGGGGCTT TGAGCTCGTAATAGAGCTTGGCGTAAACAGGCA TAGC ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA</pre>	<pre>ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA CCAACGGCGGGGGAGAGGGGGTTTGC GTAA TGGGGCTT TGAGCTCGTAATAGAGCTTGGCGTAAACAGGCA TAGC ATCCACACAACA TACGAGCGGGAAAGCA AAAGTG AAA</pre>
--	--

Quantify support and modules of evolution

- Group traits with similar evolutionary histories
- Mixture model
 - Class – sub-sets of model
- Topological mixture models
 - Class – topology or topological constraint
 - Identify recombination breakpoints
 - BUCKy – clusters genes by topology

TTCCACACAAACA|ACCGAGCCGGAAAGCA|AAACTG|AAA
CCAACCGCGGGGAGAGGGGG|TGGC|AT|GGGGCGC|
TGAGCT|AT|AGAGC|TGGC|AA|CAGGC|CATAGC|
TTCCACACAAACA|ACCGAGCCGGAAAGCA|AAACTG|AAA
CCAACCGCGGGGAGAGGGGG|TGGC|AT|GGGGCGC|
TGAGCT|AT|AGAGC|TGGC|AA|CAGGC|CATAGC|



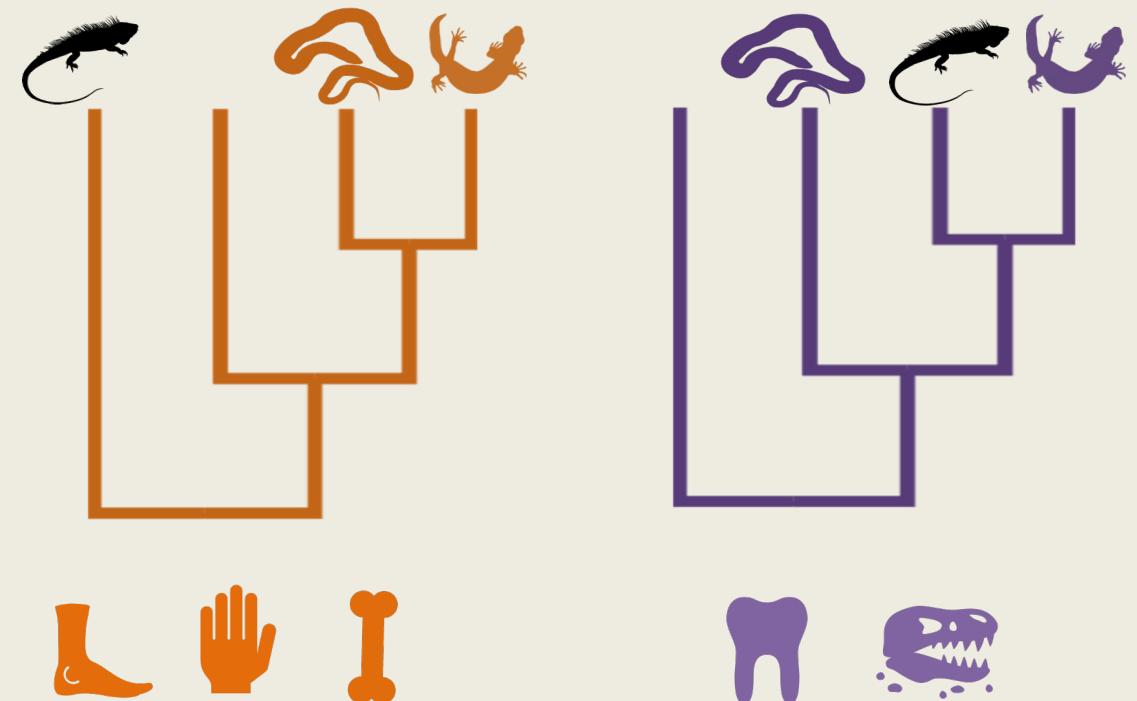
TTCCACACAAACA|ACCGAGCCGGAAAGCA|AAACTG|AAA
CCAACCGCGGGGAGAGGGGG|TGGC|AT|GGGGCGC|
TGAGCT|AT|AGAGC|TGGC|AA|CAGGC|CATAGC|
TTCCACACAAACA|ACCGAGCCGGAAAGCA|AAACTG|AAA

Topological mixture model - estimate tree and character assignment to class

- Class = topology
- Estimate all parameters for each class separately
- Character preference – time spent in class during analysis
- Bayesian framework
- Data – Simões et al. 2018

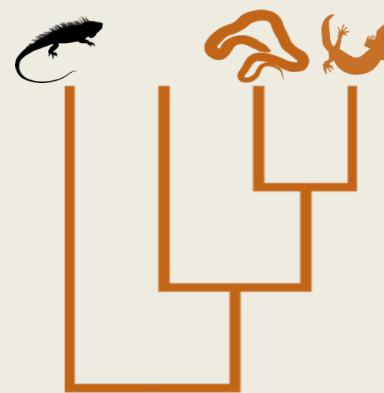


RevBayes

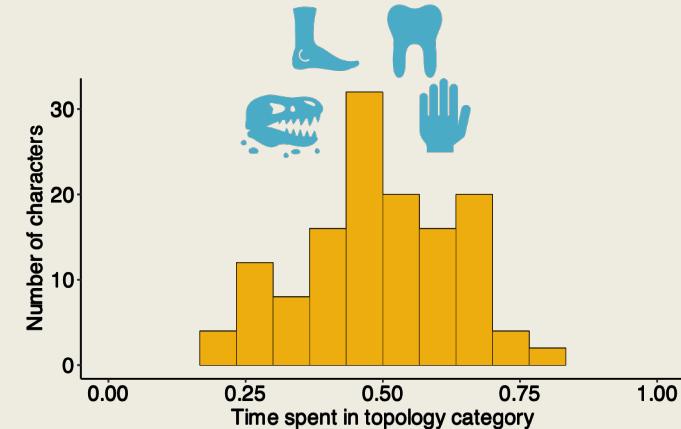


Interpreting character assignment to class

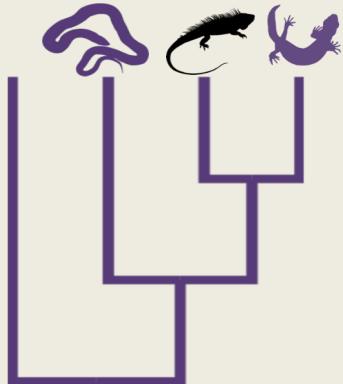
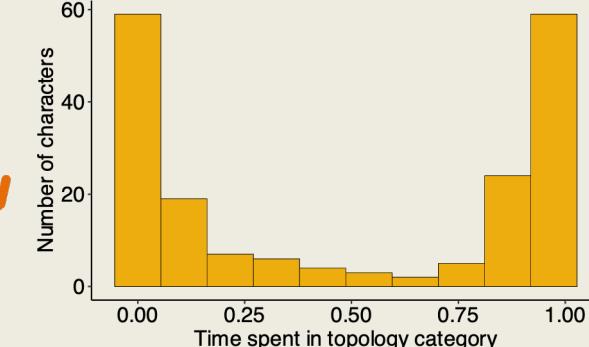
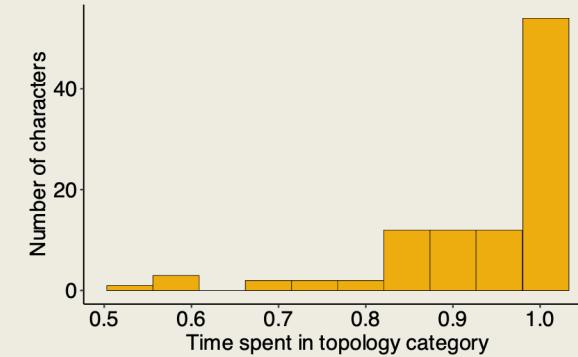
- No preference
 - Equal time in each class



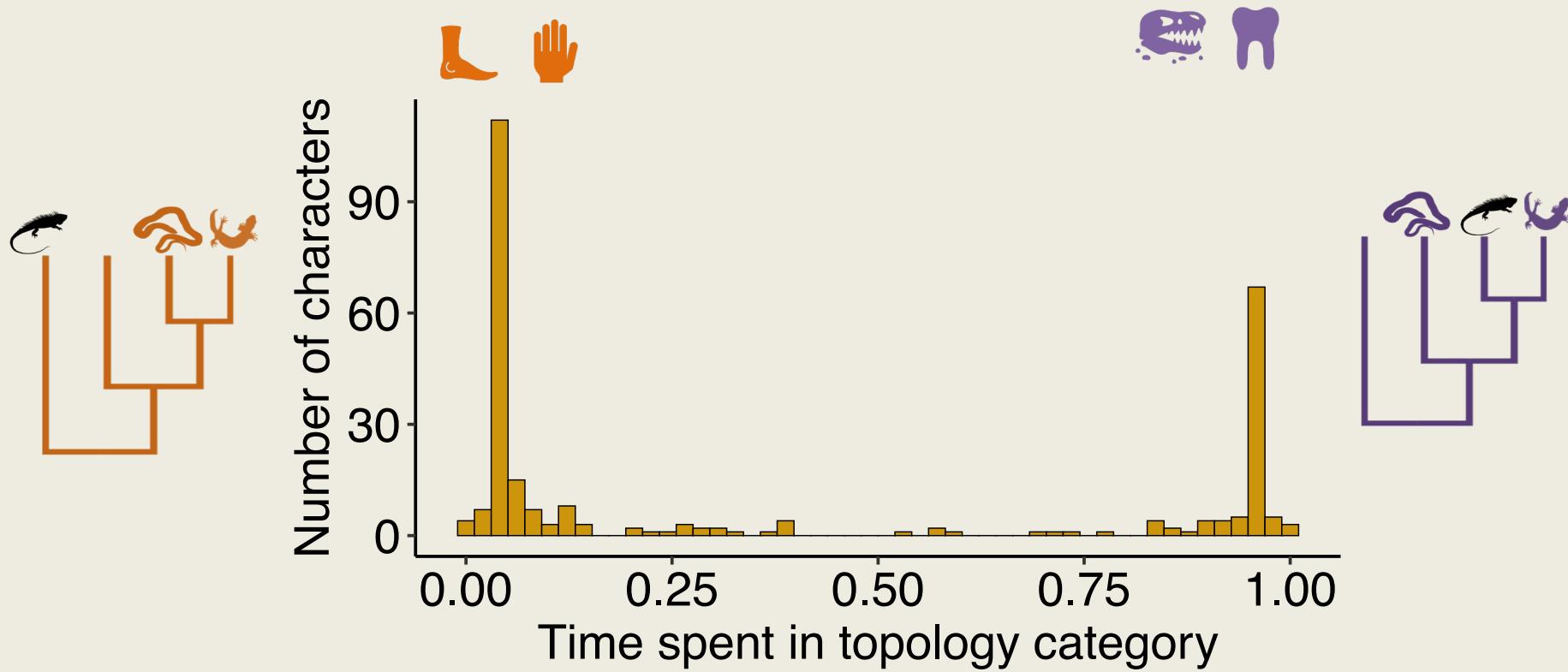
- Strong preference - agree
 - Majority of characters in one class



- Strong preference – conflict
 - Different characters in different classes

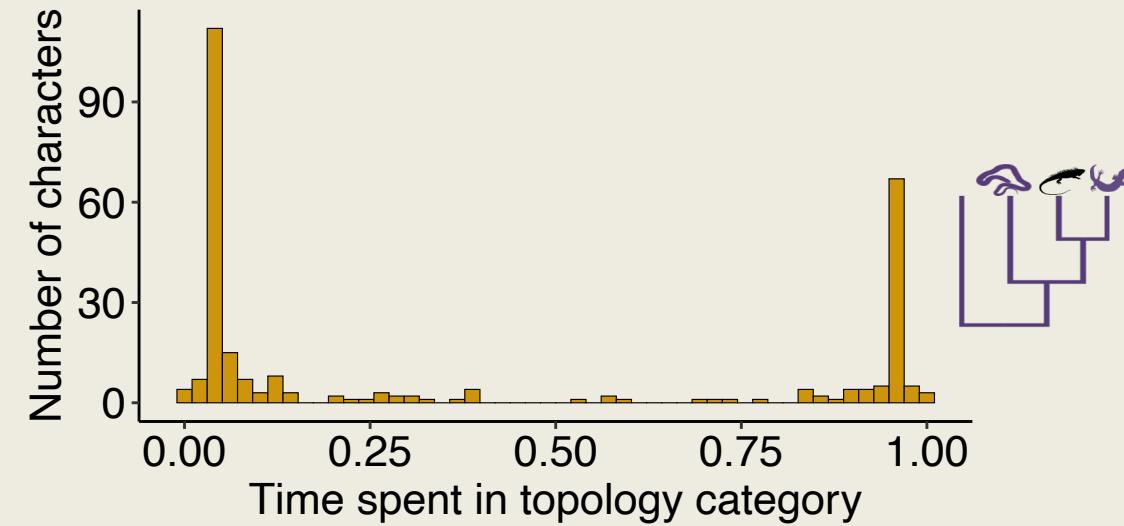
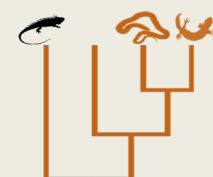
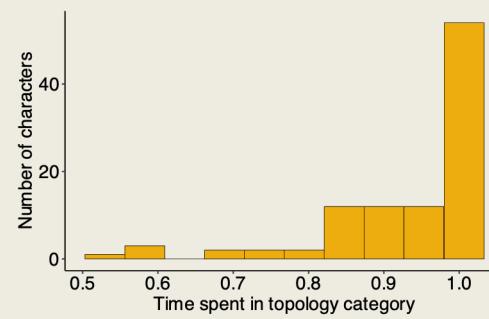
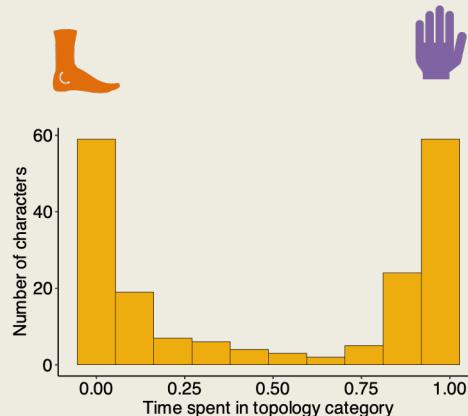


Squamate morphology – strong preference: conflict



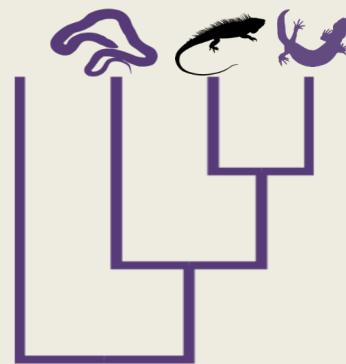
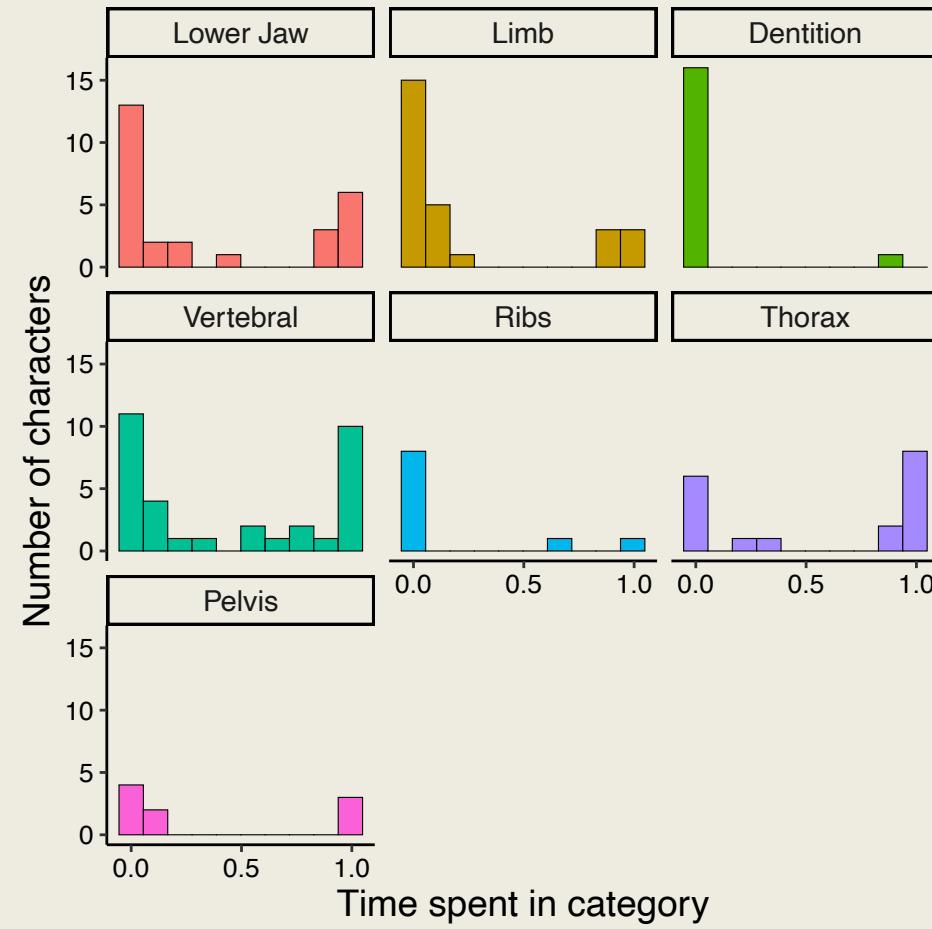
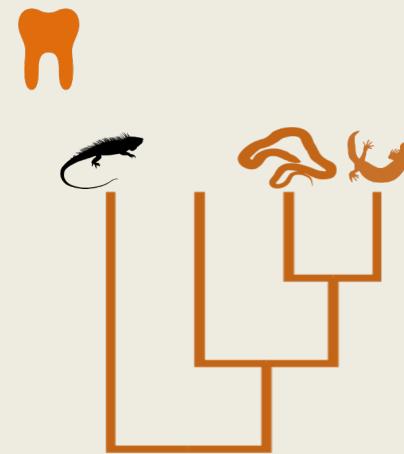
Anatomical modules in squamates?

- Do subsets of characters all prefer the same topology?
- Anatomical subsets
 - Lower jaw, limbs, dentition etc.



Anatomical modules in squamates?

- Not really
 - Except dentition
- Proof of concept
 - Predicted modules not realistic





Identify and quantify variation: Is there support for multiple topologies?



- Trees
 - Identify broad patterns



- Genes
 - Quantify support & compare methods
- Morphology



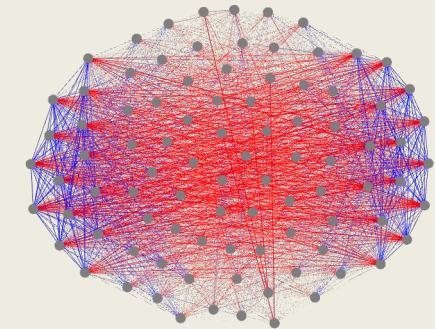
- Quantify support, test modularity, AND estimate morphological topology
- Applicable to poorly understood systems!



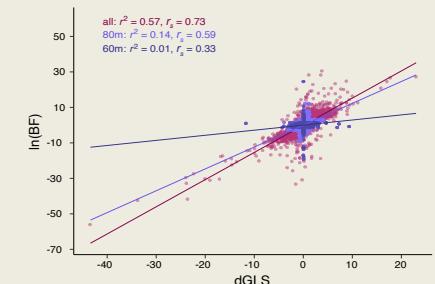
Identify and quantify variation: So much information in our data!



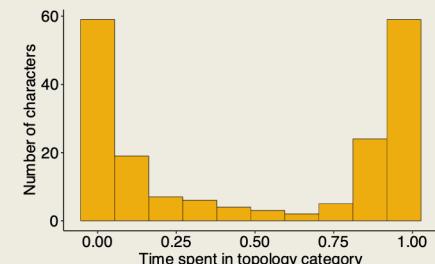
- Trees
 - Identify broad patterns – Community detection - networks



- Genes
 - Quantify support & compare methods - Some support values correlated



- Morphology
 - Quantify support & test modularity - Mixture models





Phylogenetics and Phrogs



- Phylogenetics - Awesome and difficult



- Leveraging variation in your data
 - Featuring: Trees, Genes, and Morphology



- Phrogs (Frogs) - Just awesome

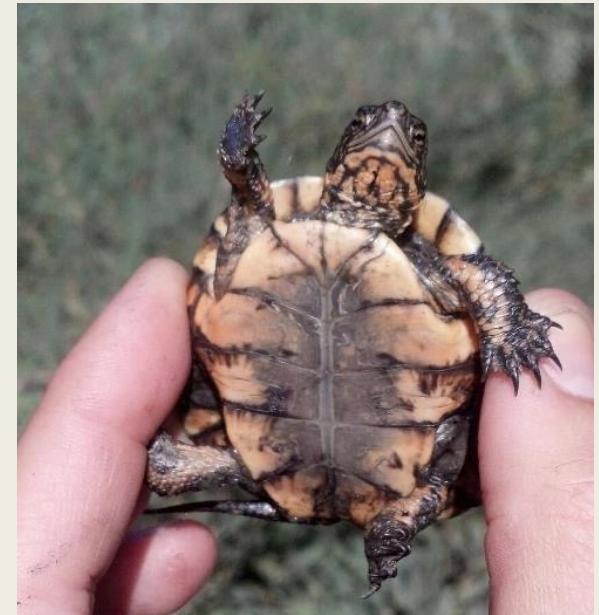


- Evolution of trait loss and conservation genetics
 - **Not** featuring: vocal sacs



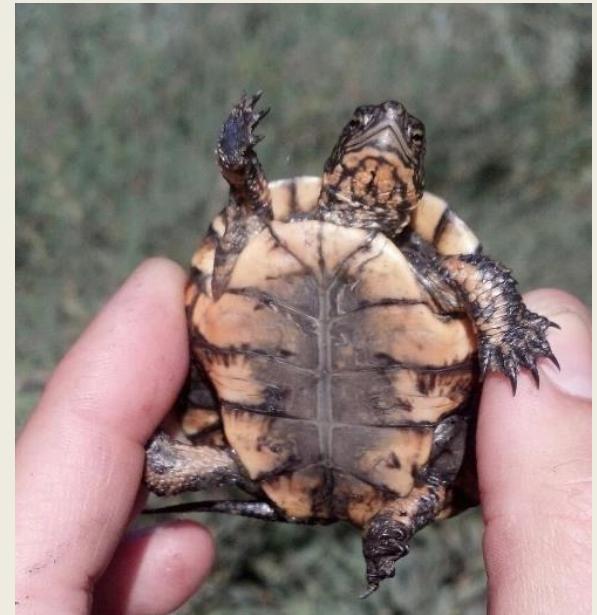
Computational biologist origin story

- Phylogenetics, conservation genetics, amphibians
- Agnostic approach to methods



Return to amphibians

- Phylogenetics from the other side – character trait evolution
- Conservation genetics





Phrogs (Frogs)



- Is the trill gone?
 - Repeated loss of sexually selected trait (vocal sacs)



- Conservation genetics
 - Declining populations in Utah



Boreal toad / Western toad / California toad

Vocal sacs enable loud calls

- Secondary sexual characteristic
 - Only males
- Attract females
- Defend territory



Sexually selected traits are lost? How, where, when, why?

- Often big and showy
- Research focuses on gain rather than loss
- Few studies across large clades
- Vocal sacs
 - Variable across anurans
 - Who/where/when?



Photo credit – Matt Brady

Is the trill gone?

- Who?
 - Presence/absence across species
- Where?
 - Ancestral state reconstruction
- Why?
 - Identify correlated characters



Why is the trill gone?

- Breakdown in signal transmission
 - Loud habitat – torrential
 - Loss of ears
- Change in female preference
 - Loss of multiple secondary sex characteristics – nuptial pads





Data collection

- ~~Examine Bufonidae museum specimens~~
- Literature search across all anurans
- Ten undergraduate researchers, three campuses
- Total recognized anurans
 - ~7000 species
 - 456 genera
 - 54 families
- ~1500 species with vocal sac data
 - 17% species lack vocal sacs (268/1546)
 - 16% genera variable (35/225)
 - 35% families variable (17/48)





Phrogs (Frogs)

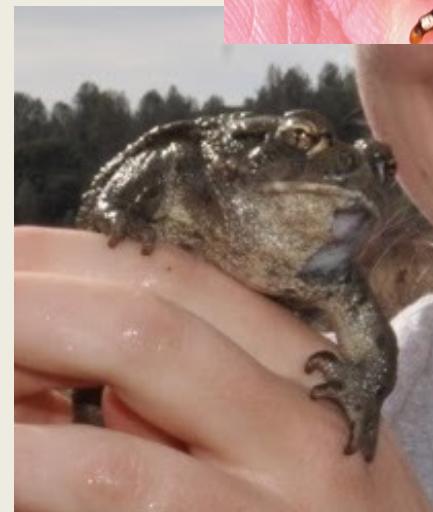


- Is the trill gone?

- Repeated loss of sexually selected trait (vocal sacs)
- Data collection
- Next – put data on trees



- Conservation genetics



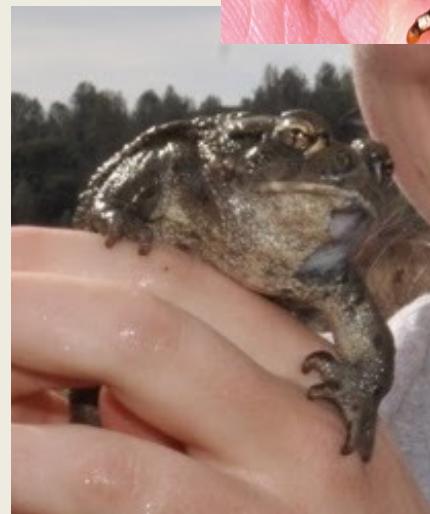
Boreal toad / Western toad
Anaxyrus boreas



Phrogs (Frogs)

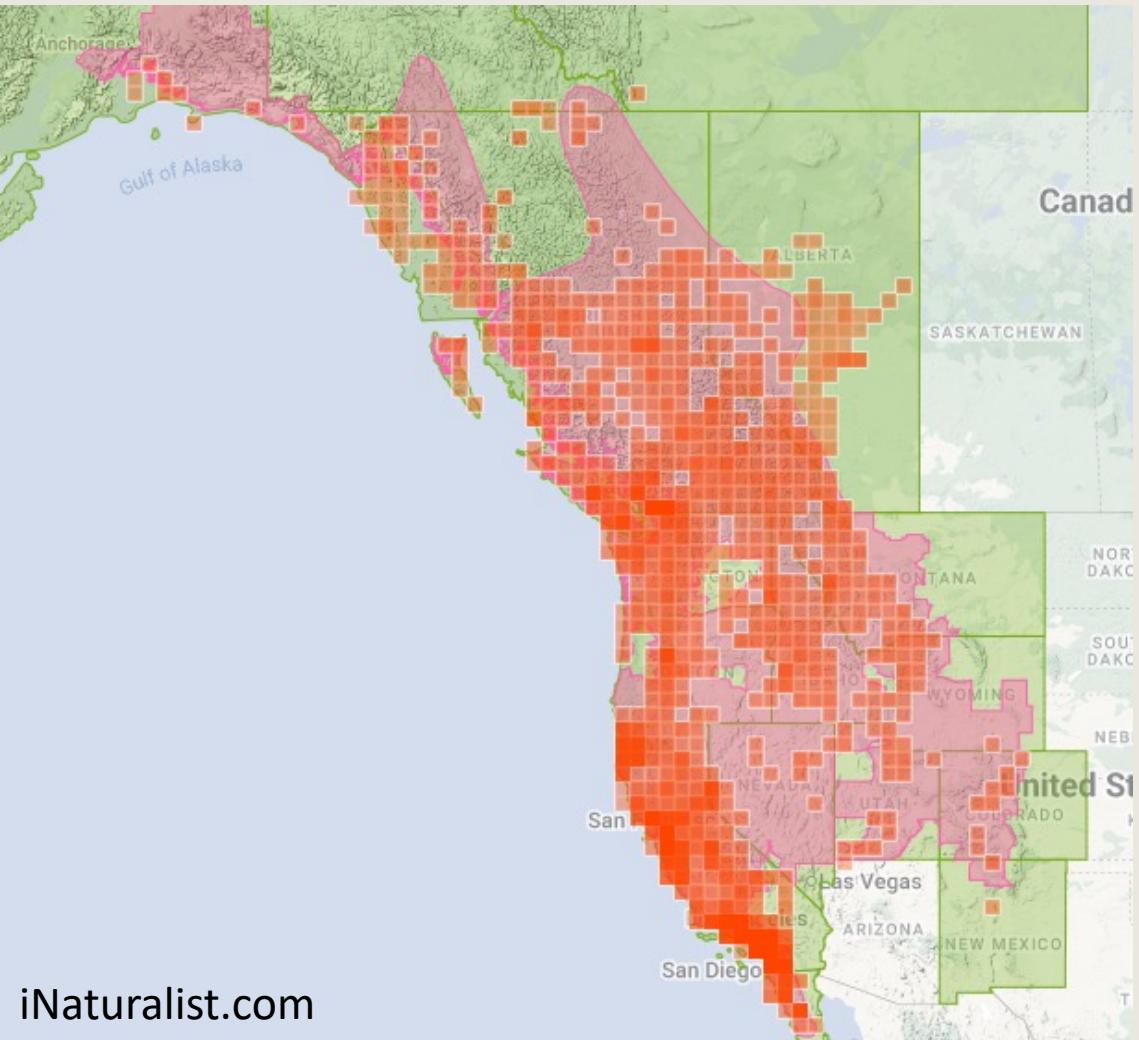


- Is the trill gone?
 - Repeated loss of sexually selected trait (vocal sacs)
- Conservation genetics
 - Boreal toad (*Anaxyrus boreas*)
 - Isolated populations in Utah
 - Captive breeding



Boreal toad / Western toad
Anaxyrus boreas

Boreal toads – understudied species complex



- Isolated populations
- Recently described species
- Majority of studies mtDNA
- Declining populations in Utah
- Endangered Species
Mediation Fund
 - Utah Division of Wildlife
Resources (UDWR)

Genetics of Utah toads

- Evidence for genetic structure within Utah
- Which populations are most closely related?
- Data
 - Collecting with UDWR
 - First MVZ specimens with genetics from Utah
 - Swab captive toads at Utah Hogel Zoo
- Sequencing – FrogCap ~80 samples





Phrogs (Frogs)



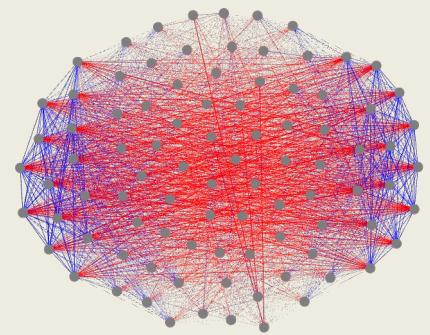
- Is the trill gone?
 - Repeated loss of sexually selected trait (vocal sacs)
- Conservation genetics
 - Genetics currently sequencing
 - Inform captive breeding



Boreal toad / Western toad
Anaxyrus boreas



Phylogenetics and Phrogs

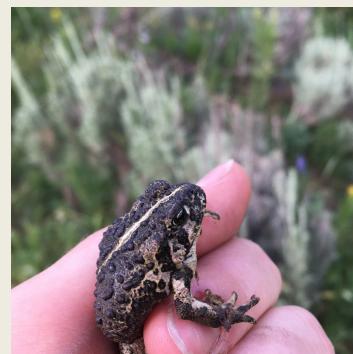
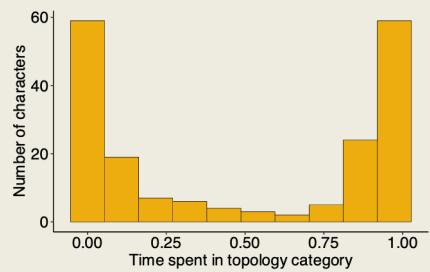
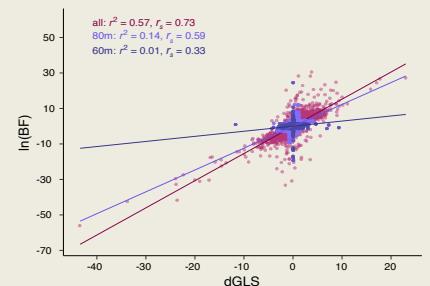


• Phylogenetics

- Networks identify broad patterns in tree space
- Quantify support - genetics & morphology
- Genetics – compare methods
- Morphology – test modularity

• Frogs

- Loss of sexually selected trait – vocal sacs
- Conservation genetics in Utah



Acknowledgements

Phylogenies

Jeremy Brown, Chris Austin

LSU Museum Community

Zach Rodriguez, Subir Shakya

Frogs

Molly Womack, Rebecca Tarvin & labs

UDWR - Drew Dittmer, Paul Thomson

Utah Hogel Zoo - Kayleigh Mullen



LSU

Center for Computation
& Technology



ENDANGERED SPECIES
MITIGATION FUND



HPC | PCP | HIGH
PERFORMANCE
COMPUTING

Questions?

Phylogenies

Jeremy Brown, Chris Austin

LSU Museum Community

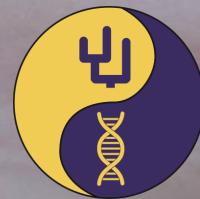
Zach Rodriguez, Subir Shakya

Frogs

Molly Womack, Rebecca Tarvin & labs

UDWR - Drew Dittmer, Paul Thomson

Utah Hogel Zoo - Kayleigh Mullen



LSU | Center for Computation
& Technology



ENDANGERED SPECIES
MITIGATION FUND

HPC | PCP | HIGH
PERFORMANCE
COMPUTING

