



KU BIODIVERSITY INSTITUTE
The University of Kansas

Photos: Carl Hutter
Unless noted

FrogCap: a modular sequence capture probe set for phylogenomics and population genetics for Anurans, assessed across multiple phylogenetic scales


















Carl R. Hutter[1], Kerry A. Cobb[2], Dan Portik[3],
Scott Travers[1], Rafe M. Brown[1]

[1] University of Kansas [2] Auburn University [3] University of Arizona

carl.hutter@gmail.com www.github.com/chutter @CarlHutter

Reduced genomic representation

Method	Advantages	Disadvantages	Cost
RADSeq	High variability	Rapid loci drop-off	$
Ultra-conserved elements (UCE)	Designed for amniotes	Low variability, unknown function	$$
Exome capture	Medium variability	Requires taxon-specific design, genomic resources	$$$
Transcriptomes	Sequence many exons, estimate expression levels	Requires fresh tissue, expensive	$$$$

Major goals

- 200+ Myr radiation (Feng et al. 2017)
- Other frog probe sets
  - UCEs (Streicher et al. 2017)
  - Afrobatrachia (Portik et al. 2016)
  - Anchored (Hime et al., *in prep*)
- Backwards compatible
  - Sanger, ultra conserved elements, Feng et al. 2017, functional loci

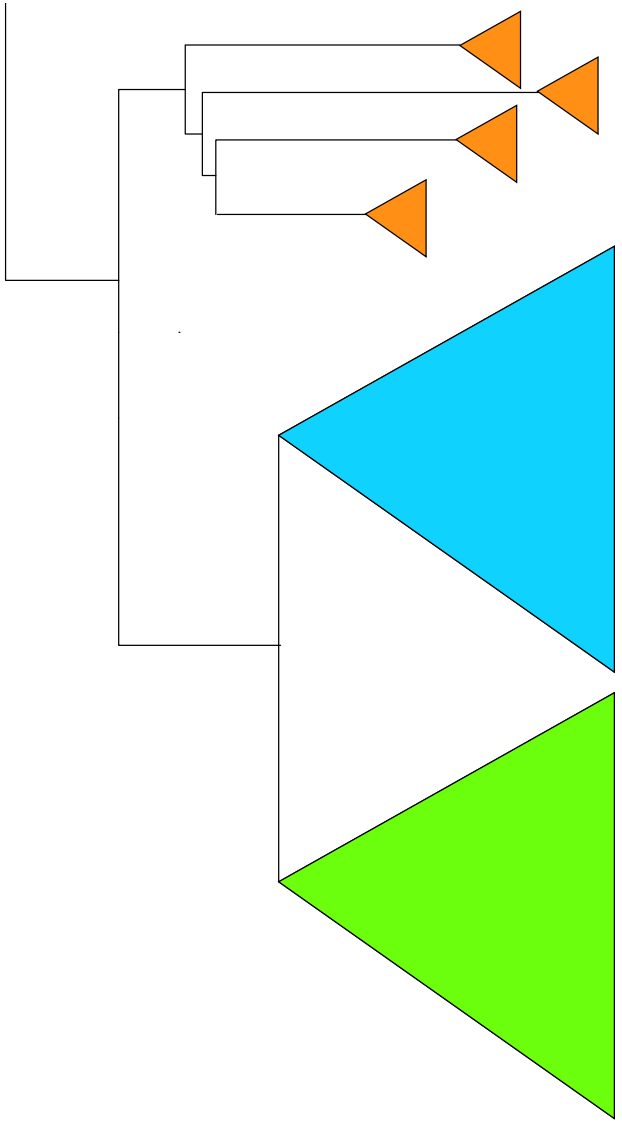
PNAS

Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary

Yan-Jie Feng[a], David C. Blackburn[b], Dan Liang[a], David M. Hillis[c], David B. Wake[d,1], David C. Cannatella[c,1], and Peng Zhang[a,1]

[a]State Key Laboratory of Biocontrol, College of Ecology and Evolution, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510006, China; [b]Department of Natural History, Florida Museum of Natural History, University of Florida, Gainesville, FL 32611; [c]Department of Integrative Biology and Biodiversity Collections, University of Texas, Austin, TX 78712; and [d]Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720

Anuran relationships



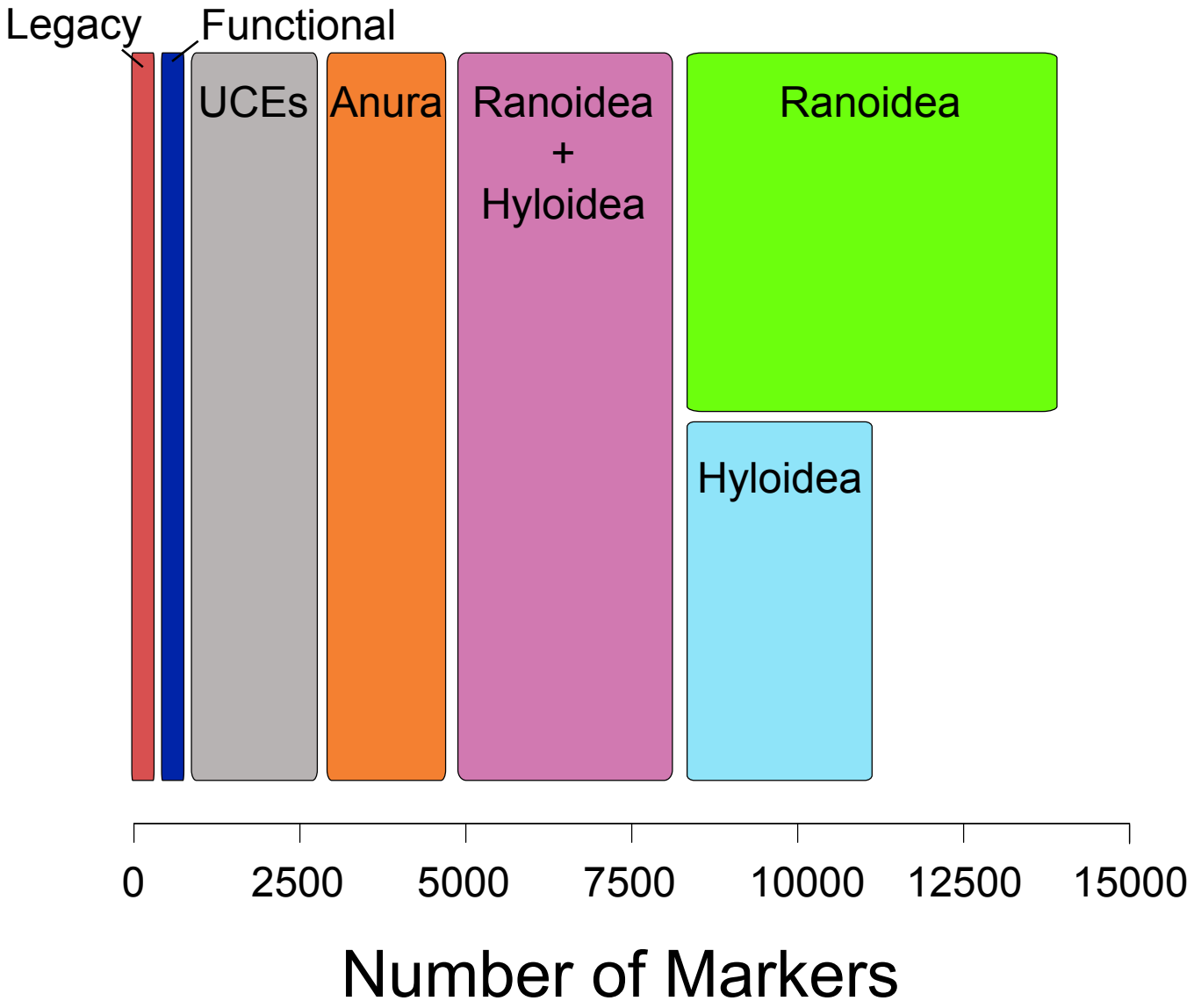
Archaeobatrachia (~400 species)
(e.g. *Xenopus, Bombina*)

® KA Cobb

Hyloidea lineages (~3600 species)
(New World: toads, glassfrogs
Poison dart frogs, tree frogs)

Ranoidea lineages (~2800 species)
(Old World: ranids, microhylids,
flying frogs, mossy frogs)

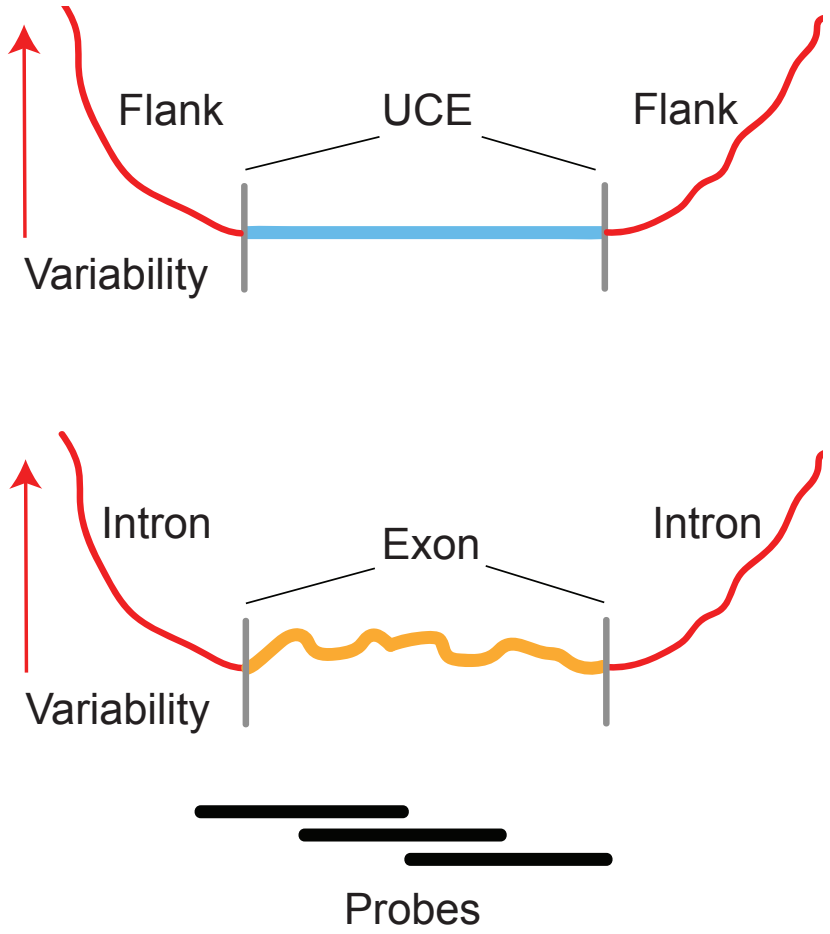
Modular probe selection



- Combine UCEs, exome capture, legacy
- Selectable modules
- Different probe sets for different questions
- Bait kit size (20k vs. 40k)
  - Depth of coverage vs. amount of data

Data types

- **Ultra-conserved elements**
  - Non-protein coding
  - Mostly unknown function
  - Conserved through tree of life
- **Exons**
  - Protein coding
  - Functional, under selection
  - Moderately conserved
- **Introns**
  - Non-protein coding
  - Flank the exons
  - High variability

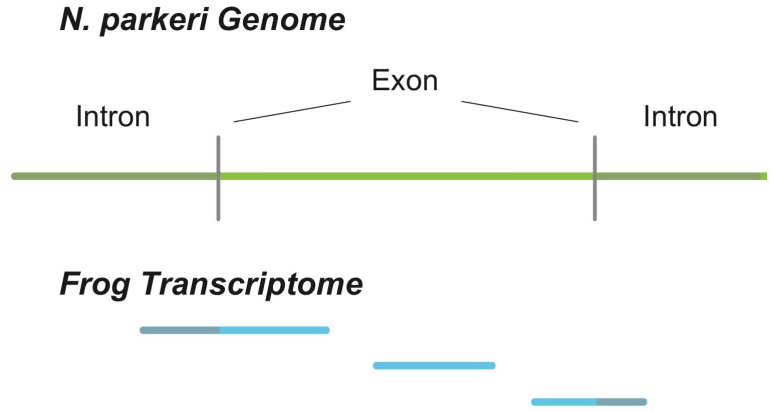


Questions

- **1) Do levels of gene discordance vary at different phylogenetic scales?**
  - Order, Superfamily, Family, Genus, Species
- **2) Do different data types affect phylogenetic relationships?**
  - Exon vs. intron vs. UCE

Probe design

- Based on *Nanorana parkeri* genome
- Reassembled 25 frog transcriptome datasets published on GenBank
- Pairwise matched frog transcriptomes to genome (w/ BLAT)
- Clustered proximate transcriptome matches together from the same exon



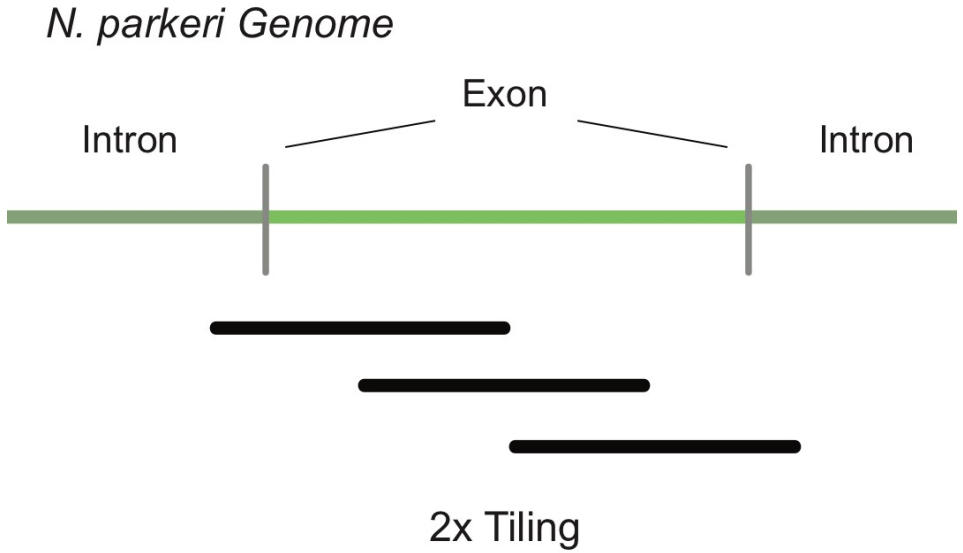
PNAS

Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes

Yan-Bo Sun[a,1], Zi-Jun Xiong[b,c,d,1], Xue-Yan Xiang[b,c,d,e,1], Shi-Ping Liu[b,c,d,f], Wei-Wei Zhou[a], Xiao-Long Tu[a,g], Li Zhong[h], Lu Wang[h], Dong-Dong Wu[a], Bao-Lin Zhang[a,h], Chun-Ling Zhu[a], Min-Min Yang[a], Hong-Man Chen[a], Fang Li[b,d], Long Zhou[b,d], Shao-Hong Feng[b,d], Chao Huang[b,d,f], Guo-Jie Zhang[b,d,i], David Irwin[a,j,k], David M. Hillis[l,2], Robert W. Murphy[a,m], Huan-Ming Yang[d,n,o], Jing Che[a,2], Jun Wang[d,n,p,q,r,2], and Ya-Ping Zhang[a,h,2]

Probe design

- Gathered statistics to filter:
  - Optimal GC content ~40-50%
  - Parsimony informative sites > 15%
  - Maximize phylogenetic representation
  - No duplicate matches
  - 120bp length baits, 20bp intron buffer



Final probe sets

- 40k bait Ranoidea set
  - *N. parkeri* genome
  - 700 UCEs (Alexander et al. 2016)
  - = ~13,000 markers
- 40k bait Hyloidea + Archeobatrachia set
  - Redesigned 2300 UCEs (Streicher et al. 2017)
  - New transcriptome consensus sequences
  - = ~9,500 markers

Libraries and sequencing

- 101 samples
  - 77: Ranoidea (40,000 baits)
  - 24: Hyloidea (40,000 baits)
    - 1 Salamandra
    - 2 Archaeobatrachia
- MYcroarray (Arbor Biosciences) MyBaits, library preparation
- Illumina TruSeq libraries
- Hybridization-based enrichment
- Multiplex 8 samples / pool
- Illumina Hi-Seq 3000 150bp PE

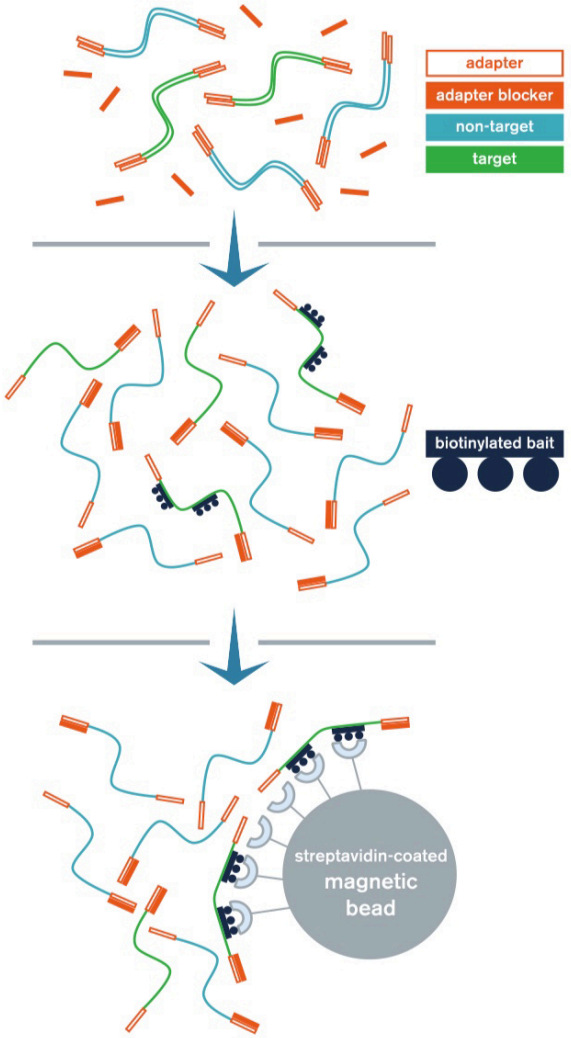
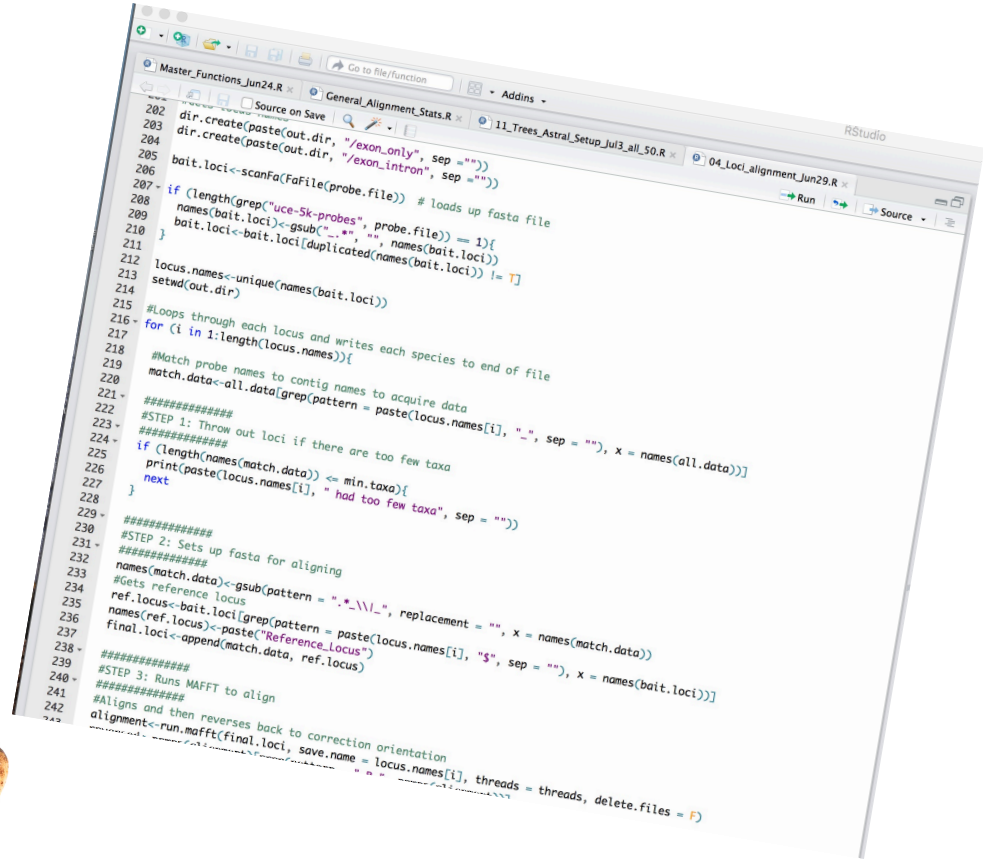


Figure from MyBaits manual

Pre-processing

- Custom pipeline w/ R
- Trimming barcodes, adapter contamination (AfterQC)
- Remove contamination
  - Match reads to genomes (BBMap)
  - (e.g. bacteria, *Drosophilia*, *C. elegans*)
- Paired-end read merging (BBTools)
- Assembled with SPADES
  - 10 k-mer values



Alignment and cleaning

- Matched probes to assembled contigs (BLAST)
- Aligned with MAFFT 7.3
- Separated loci into three categories:
  - 1) Exons
  - 2) Introns
  - 3) UCEs
- Trimming and cleanup
  - 1) Maintain open reading frames
  - 2) TRIMAL (automatic1)
  - 3) End trimming

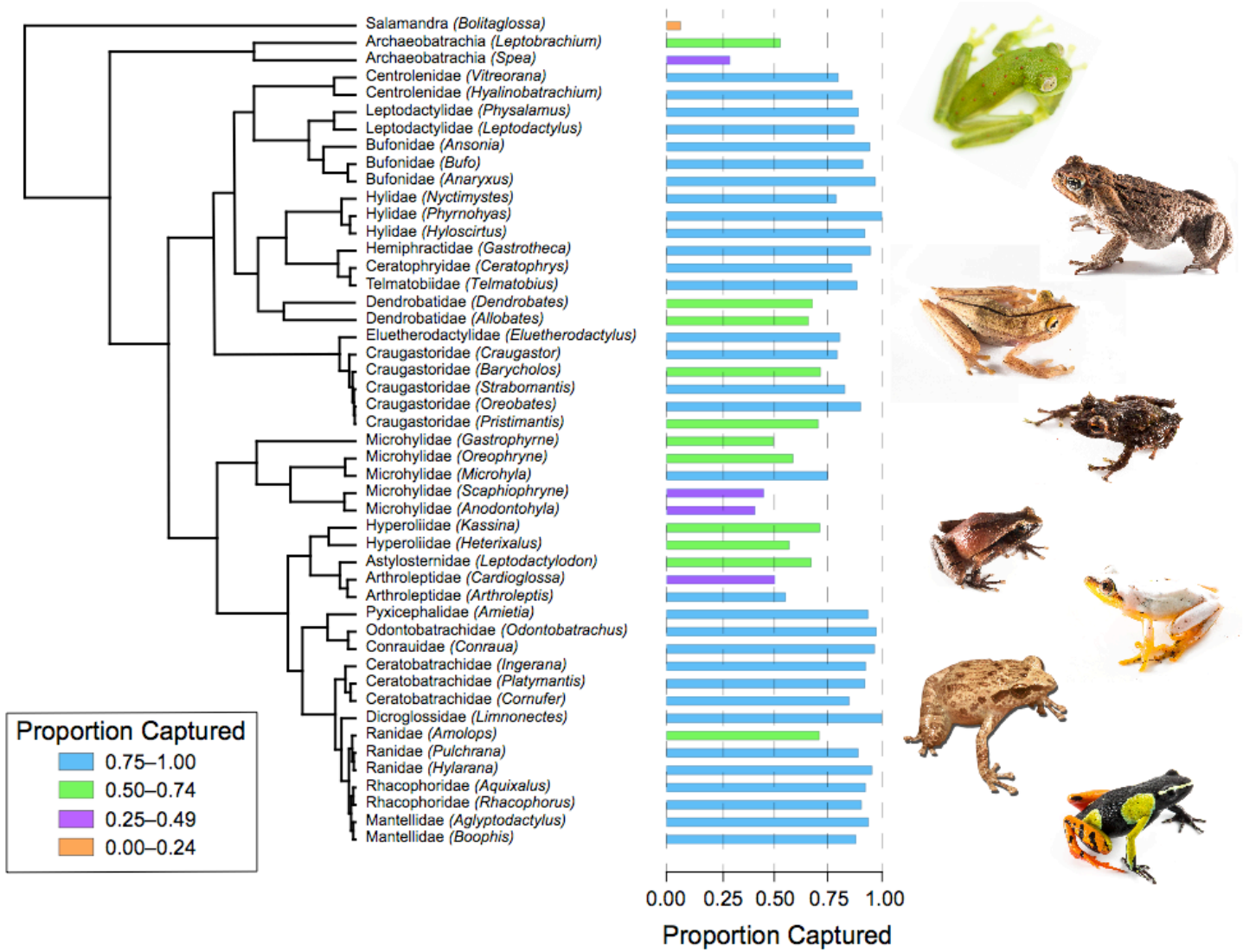
Phylogeny estimation

- Gene trees
  - IQTree w/ ModelFinder
  - 1000 ultrafast bootstrap replicates
- Astral-III summary species trees
  - Used polytomies for poor support (<50 bootstrap)
  - Exon, Intron, UCE separately
- Robinson-Fould's Distance
  - Normalized
  - Compared data type within groups

Scales

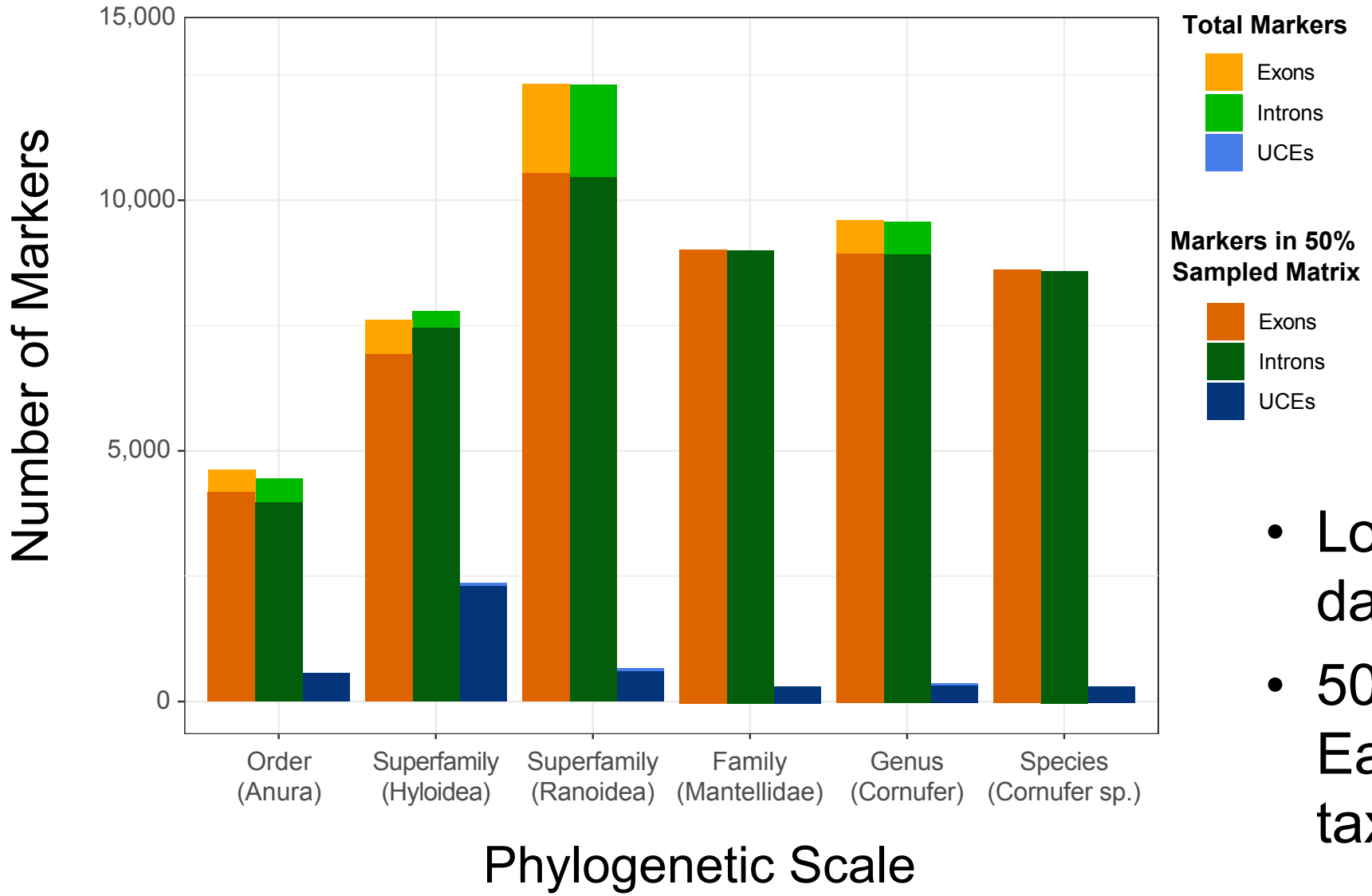
- 1) Order: Anura, 48 samples
- 2) Superfamily: Hyloidea, 24 samples
- 3) Superfamily: Ranoidea, 24 samples
- 4) Family: Mantellidae, 12 samples
- 5) Genus: *Cornufer,* 24 samples
- 6) Species: *Cornufer* sp., 16 samples

Markers by sample



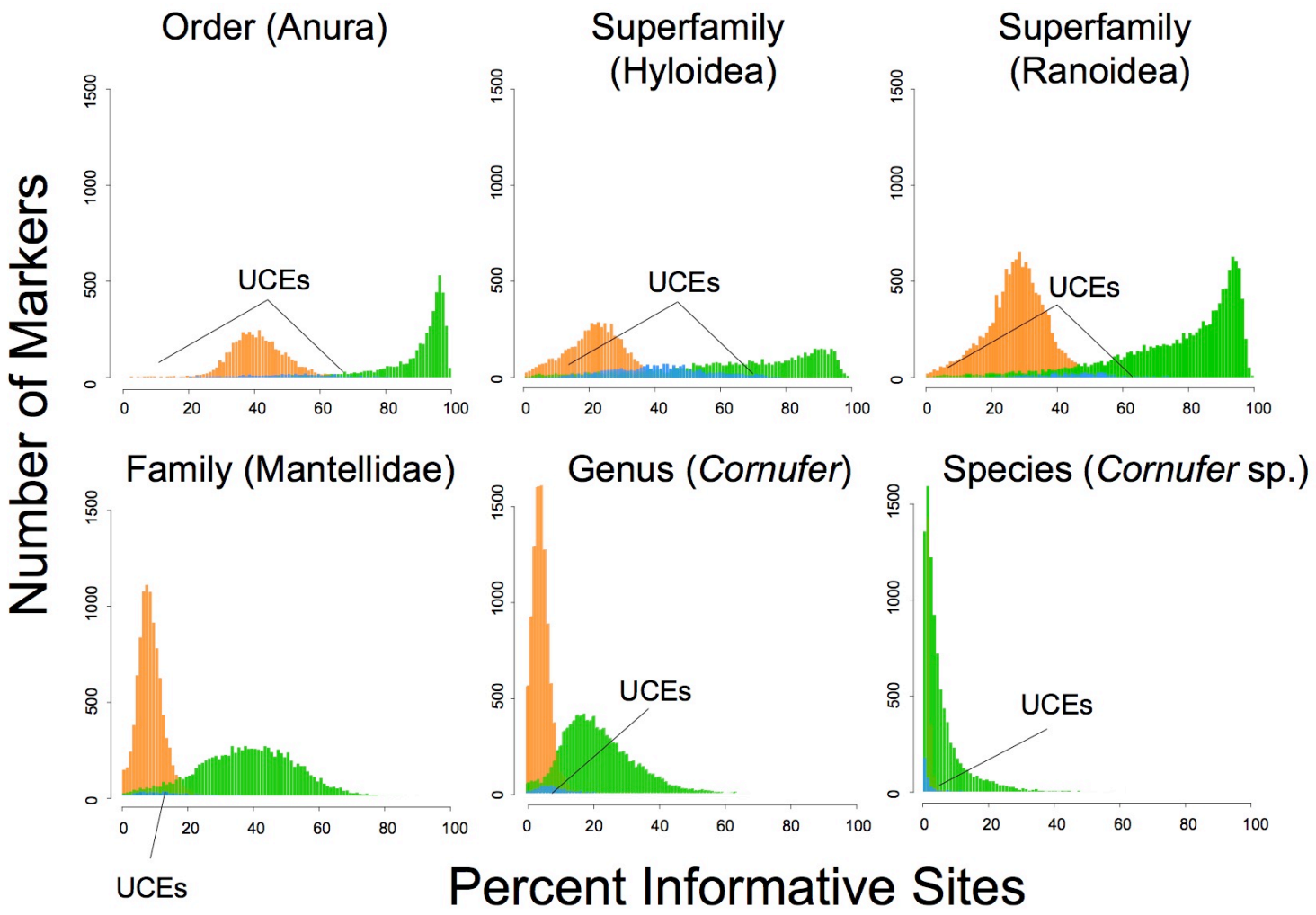
- Loci captured per sample
- Blue most successful
- Salamander did poorly

Alignment summary



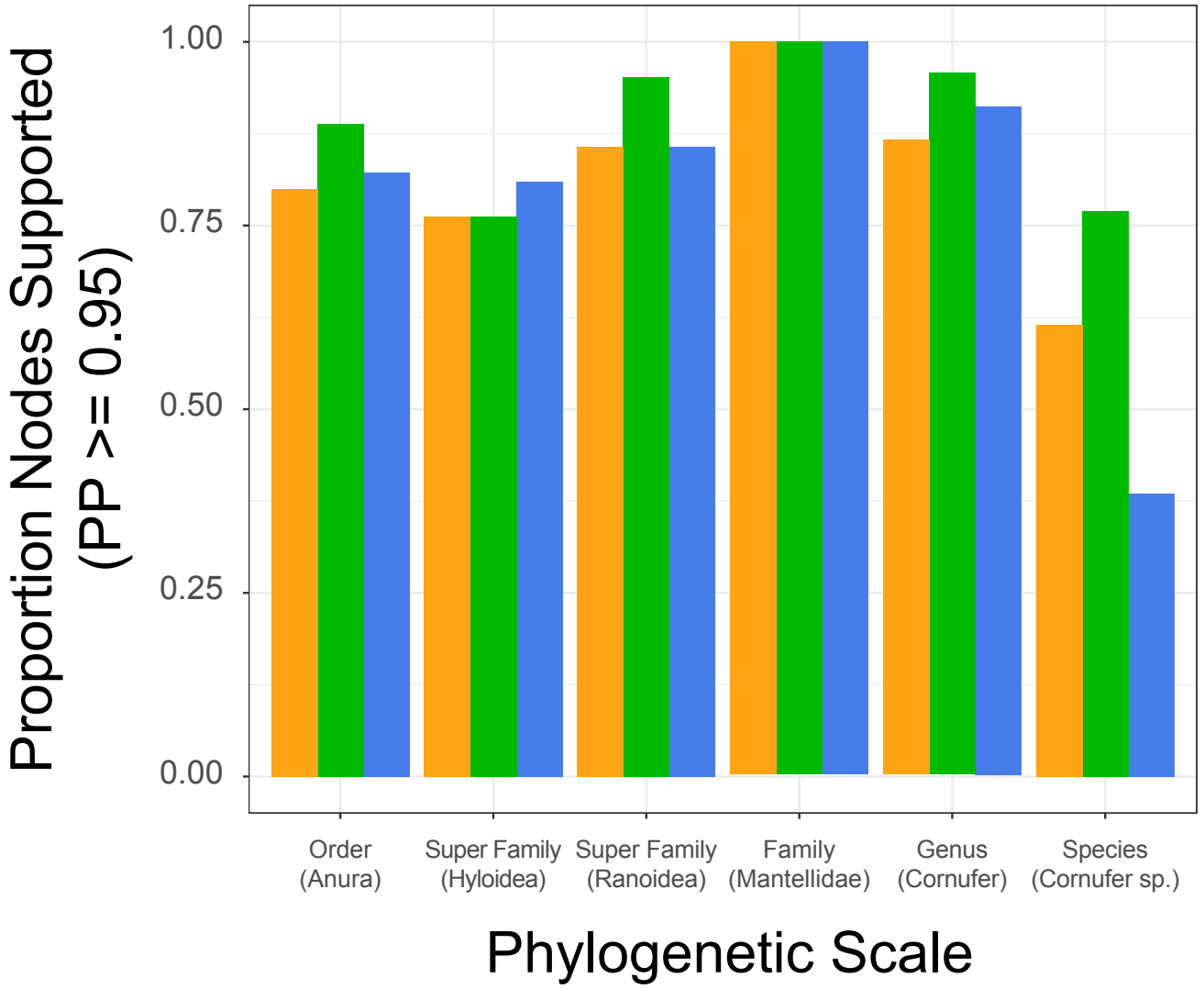
- Loci captured for each data type within groups
- 50% Sampled Matrix = Each alignment > 50% taxon sampling

Variability



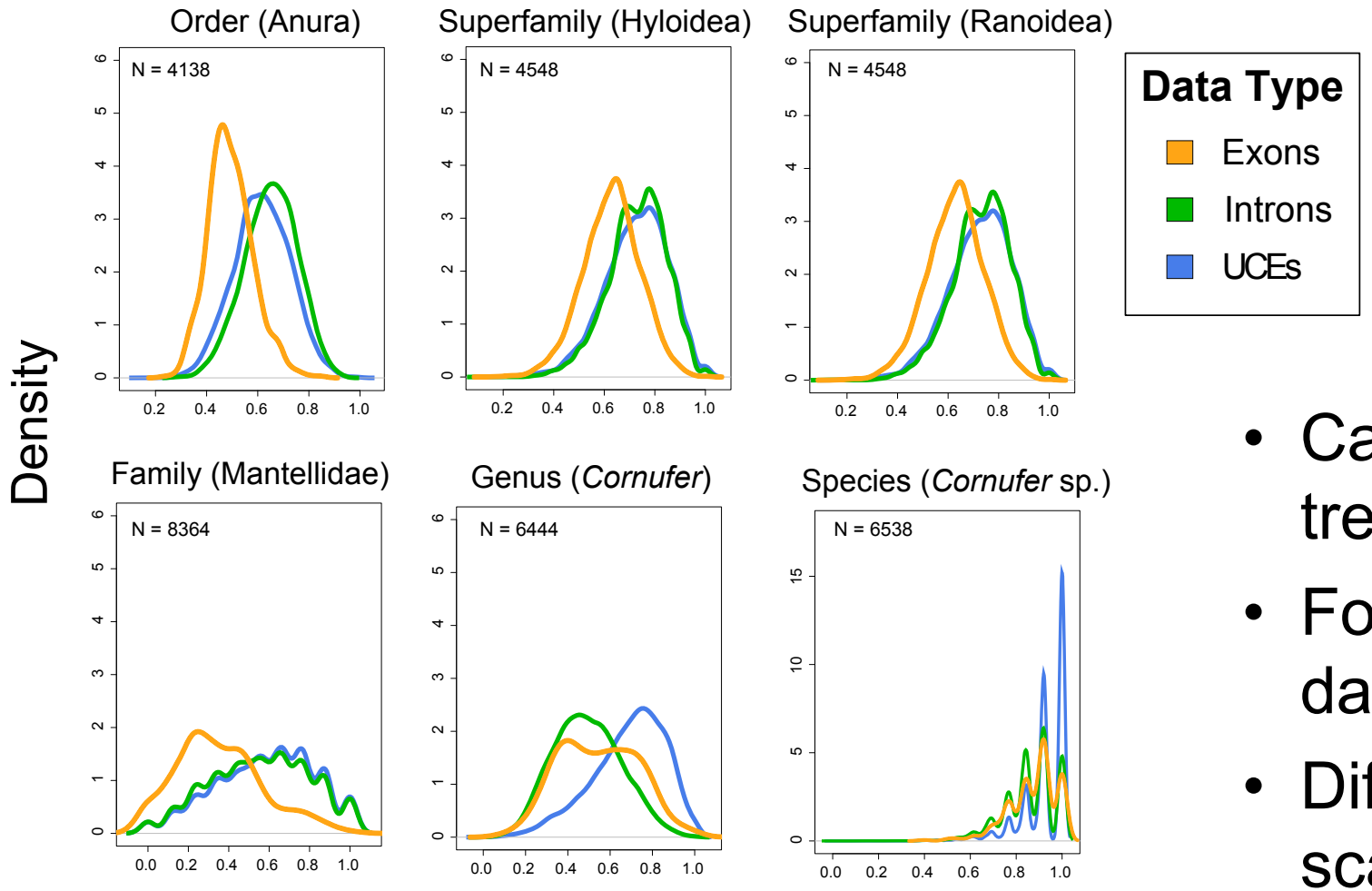
- Introns most variable
- Exons moderate
- UCEs moderate to low

Support value comparison



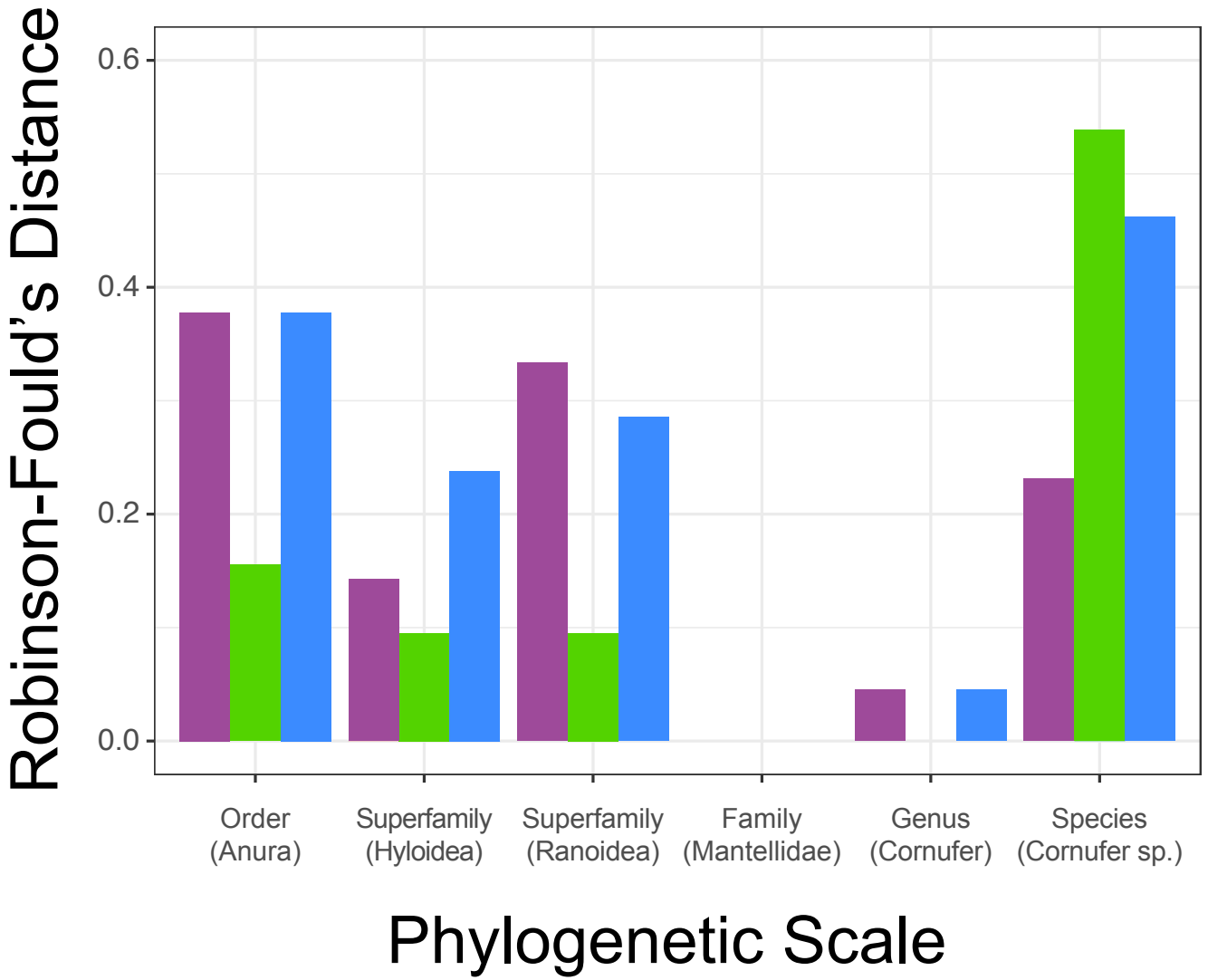
- From summary Astral trees
- Most groups greater than 75% of nodes strongly supported
- Species level poor support

Robinson-Fould's density plots



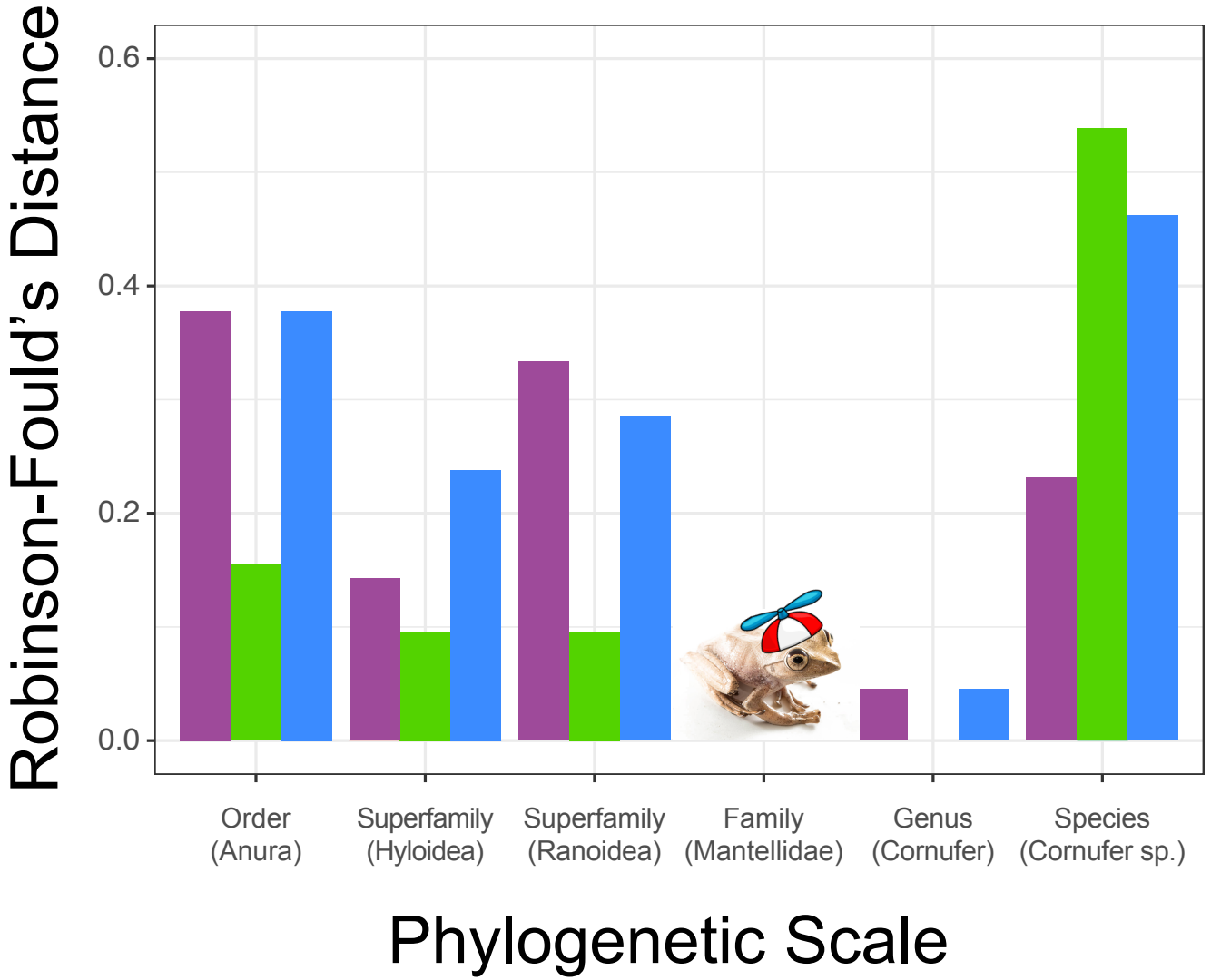
- Calculated RF from species tree to every gene tree
- For species tree from each data type (exon, intron, UCE)
- Differs across phylogenetic scales and data type

Data type tree comparison



- From summary Astral trees
- Higher discord at larger scales
- Introns differed strongly from UCE and exons
- Genus and Family very similar

Data type tree comparison



- From summary Astral trees
- Higher discord at larger scales
- Introns differed strongly from UCE and exons
- Genus and Family very similar

Conclusions

- Results summary
  - Informativeness: Intron > Exon > UCE
  - Data type disagreement at larger scales
  - Family and genus level concordant
  - Species level = gene flow
- Future directions
  - Assess in species delimitation context
  - (e.g. Structure, SVDQuartets)
  - Additional sampling for Anuran phylogeny

Probe set and data availability

- The FrogCap Sequence Capture probe set and processing pipeline will soon be available on GitHub
- Contact me if you want to use it now
- **github.com/chutter/FrogCap-Sequence-Capture**
- All data (Raw reads, contigs, alignments) will be made available upon manuscript acceptance
  - NCBI Sequence Read Archive
  - Open Science Framework
    - https://osf.io/gvbr5/

CENTER FOR OPEN SCIENCE

carl.hutter@gmail.com www.github.com/chutter @CarlHutter

Questions?

Acknowledgements

Advisors: Rafe Brown and Rich Glor

Tissues / Samples: Chan Kin Onn, Aaron Bauer, Dave Blackburn, Bill Duellman, Juan Guayasamin, Miguel Vences, Frank Glaw, Rich Glor

Funding sources: University of Kansas, NSF GRFP, SSE Rosemary Grant Award

Graduate Research Fellowship Program

SOCIETY *for the* STUDY *of* EVOLUTION

carl.hutter@gmail.com

www.github.com/chutter

@CarlHutter