



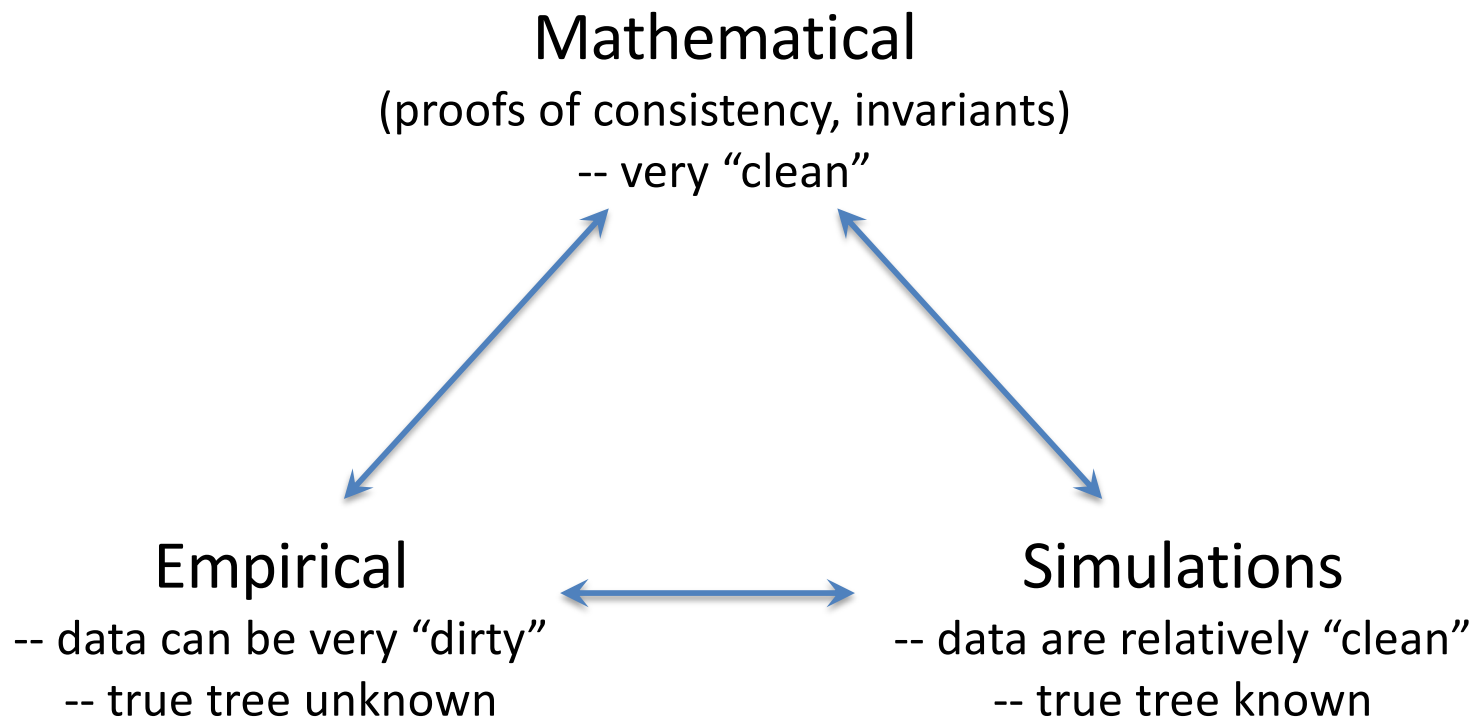
THE IMPLICATIONS OF EMPIRICAL SEQUENCE CAPTURE STUDIES IN BIRDS FOR LARGE-SCALE PHYLOGENETICS

Edward L Braun
Department of Biology
University of Florida

Seminar: ICERM, Brown University, 18 September 2024

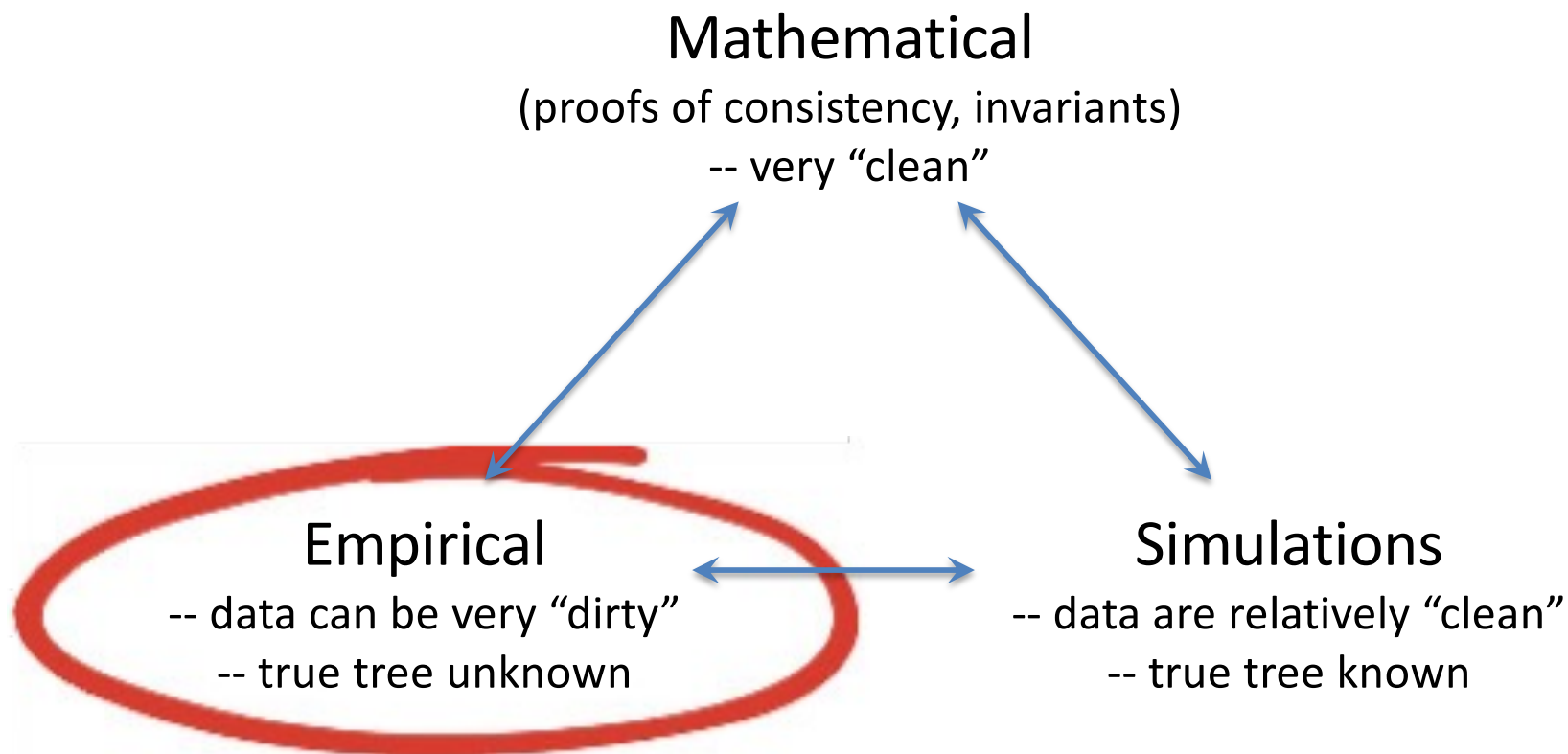
WAYS TO STUDY PHYLOGENETIC METHODS

- Different ways of studying methods of phylogenetic estimation should complement each other



WAYS TO STUDY PHYLOGENETIC METHODS

- Different ways of studying methods of phylogenetic estimation should complement each other



OUTLINE

- Why are birds interesting for phylogenetics?
- Biological and technical challenges
 - The “Usual Suspects”
 - Recombination landscape of genomes
 - Sequence assemblers
- Comparison of inference methods
 - Including distance methods (METAL)

WHY BIRDS? PRIOR INFORMATION!

○ Large number of phylogenomic studies

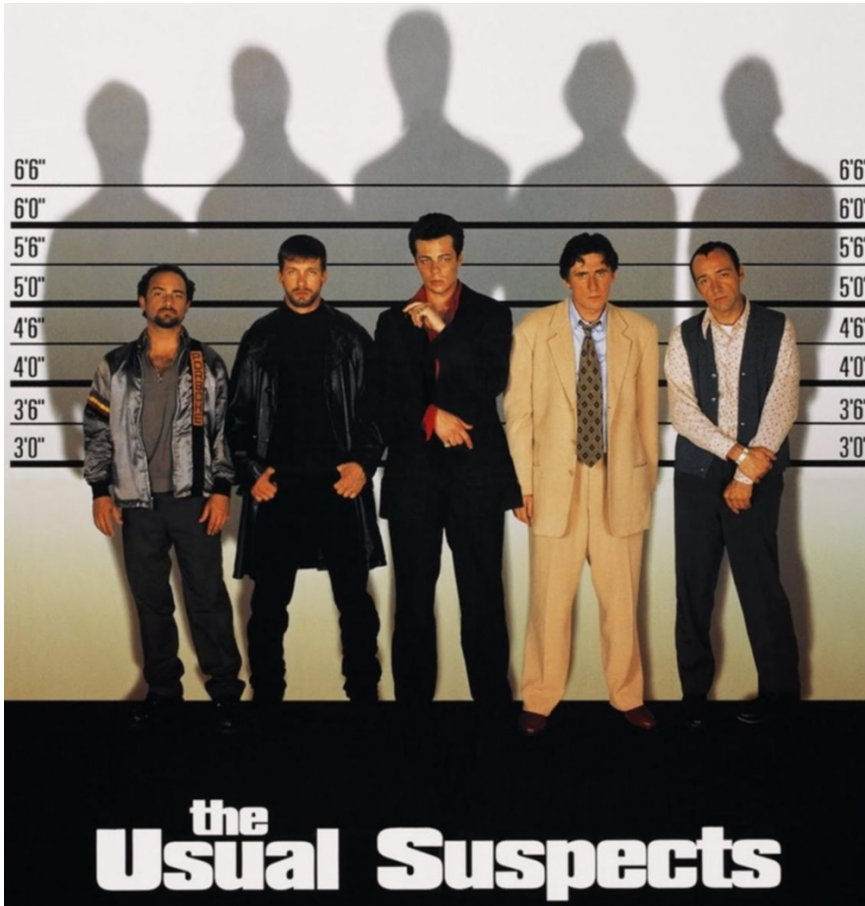
Citation	# of Loci	Data Type(s)	# of Taxa ²
Multigene¹			
Hackett et al. (2008)	19	Introns	169
Harshman et al. (2008)	20	Introns	14 (Palaeognathae)
Haddrath and Baker (2012)	27	Coding exons and introns	23 (Palaeognathae)
Wang et al. (2011)	30	Introns	28 (Telluraves)
Kimball et al. (2013)	31	Introns	77
Smith et al. (2013)	40	Introns	10 (Palaeognathae)
Reddy et al. (2017)	54	Introns	235
Liu et al. (2018)	63	Coding exons	48
Phylogenomic¹			
McCormack et al. (2013)	1,541	UCEs ³	32 (Neoaves)
Jarvis et al. (2014)	11,839	Whole genomes ⁴	48
Baker et al. (2014)	1,448	Coding exons and UCEs	198
Prum et al. (2015)	259	Coding exons	198
Suh et al. (2015)	2,118	TE insertions ⁵	43
Cloutier et al. (2019)	20,850	Whole genomes ⁶	15 (Palaeognathae)
White and Braun (2019)	4,243	UCEs	23 (Strisores)
Kuhl et al. (2021)	— ⁷	3' UTRs ⁸	429
Wang et al. (2022b)	— ⁷	Whole genomes	16 (Palaeognathae)
This study	4,307	UCEs	394
Recent phylogenomic¹			
Wu et al. (2024)	25,460	Whole genomes ⁹	124
Stiller et al. (2024)	63,430	Whole genomes ¹⁰	363

WHY BIRDS? PRIOR INFORMATION!

○ Large number of phylogenomic studies

Citation	# of Loci	Data Type(s)	# of Taxa ²
Multigene¹			
Hackett et al. (2008)	19	Introns	169
Harshman et al. (2008)	20	Introns	14 (Palaeognathae)
Haddrath and Baker (2012)	27	Coding exons and introns	23 (Palaeognathae)
Wang et al. (2011)	30	Introns	28 (Telluraves)
Kimball et al. (2013)	31	Introns	77
Smith et al. (2013)	40	Introns	10 (Palaeognathae)
→ Reddy et al. (2017)	54	Introns	235
Liu et al. (2018)	63	Coding exons	48
Phylogenomic¹			
McCormack et al. (2013)	1,541	UCEs ³	32 (Neoaves)
Jarvis et al. (2014)	11,839	Whole genomes ⁴	48
Baker et al. (2014)	1,448	Coding exons and UCEs	198
Prum et al. (2015)	259	Coding exons	198
Suh et al. (2015)	2,118	TE insertions ⁵	43
Cloutier et al. (2019)	20,850	Whole genomes ⁶	15 (Palaeognathae)
White and Braun (2019)	4,243	UCEs	23 (Strisores)
Kuhl et al. (2021)	— ⁷	3' UTRs ⁸	429
Wang et al. (2022b)	— ⁷	Whole genomes	16 (Palaeognathae)
→ This study	4,307	UCEs	394
Recent phylogenomic¹			
→ Wu et al. (2024)	25,460	Whole genomes ⁹	124
→ Stiller et al. (2024)	63,430	Whole genomes ¹⁰	363

WHY BIRDS?



○ Avian phylogeny exhibits

❖ Rapid radiation

- Many short branches
- Discordance among gene trees (incomplete lineage sorting)

❖ Rate differences among taxa

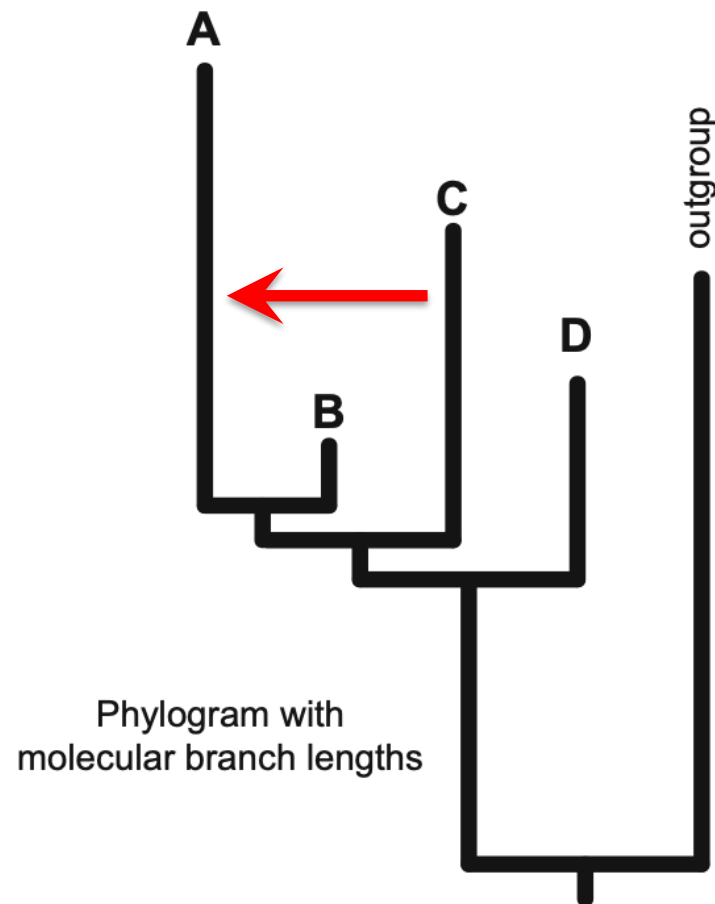
- Potential for long branch attraction

❖ Shifts in evolutionary models

- Convergence in GC content

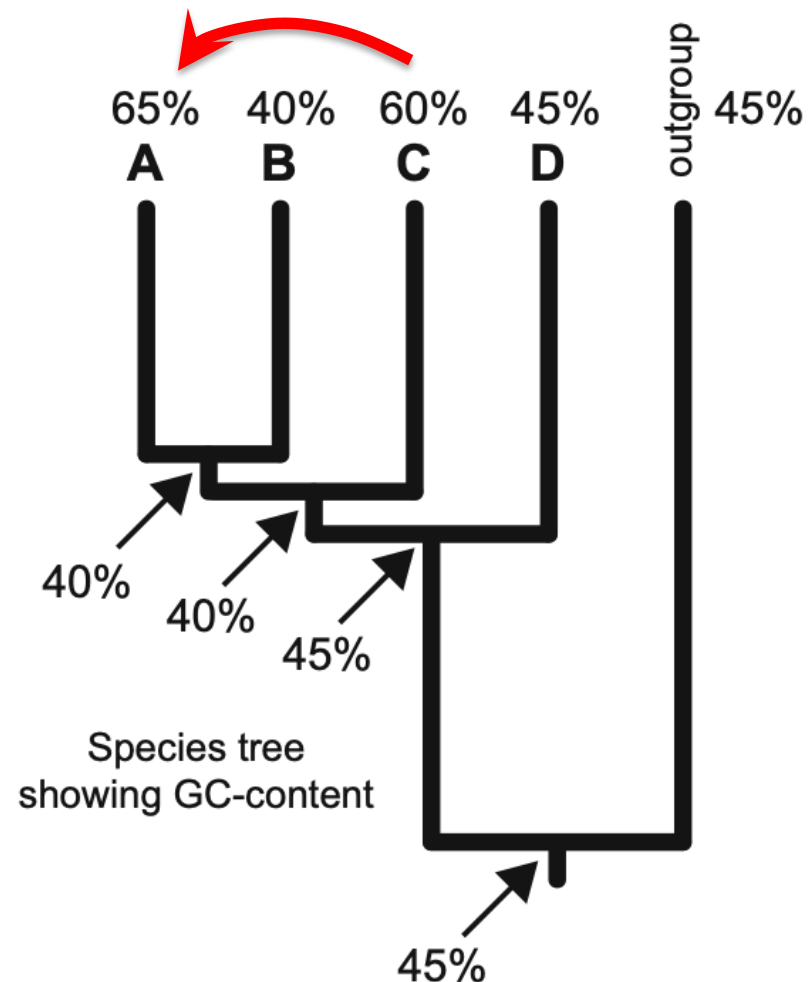
WHAT ARE THE "USUAL SUSPECTS"?

- Long-branch attraction (an oldie but a baddie...)



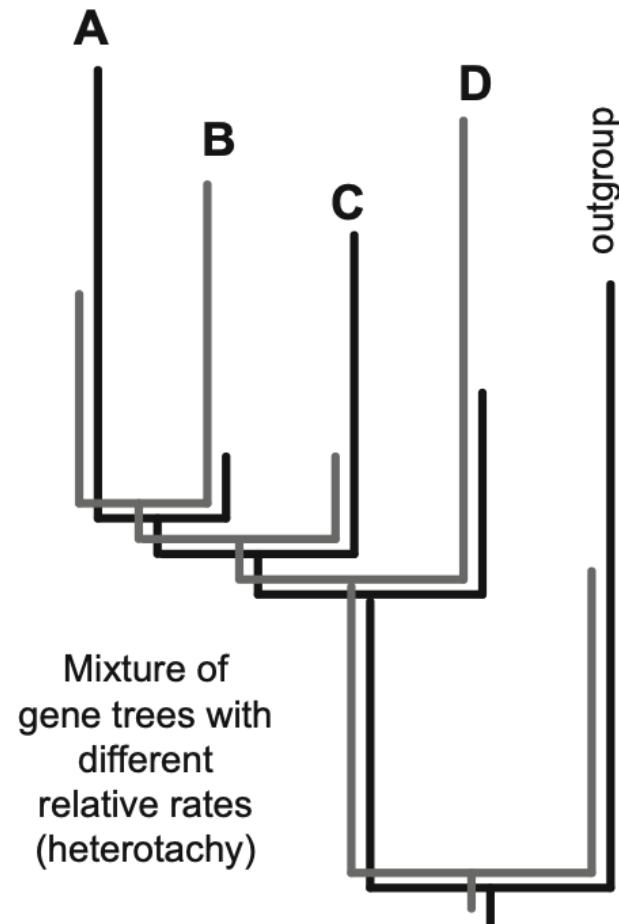
WHAT ARE THE "USUAL SUSPECTS"?

- Convergence in base composition



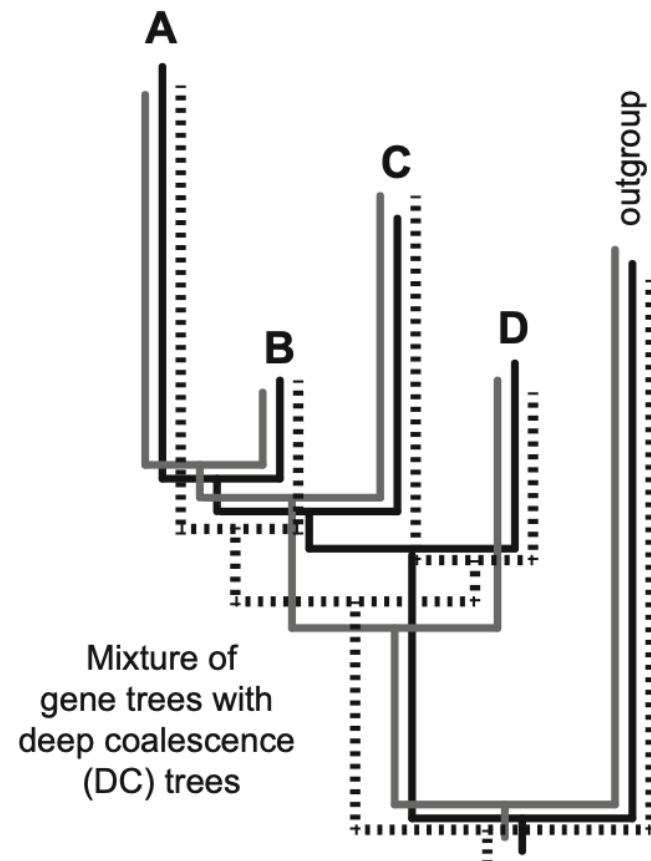
WHAT ARE THE "USUAL SUSPECTS"?

- Heterotachy – mixtures of branch lengths

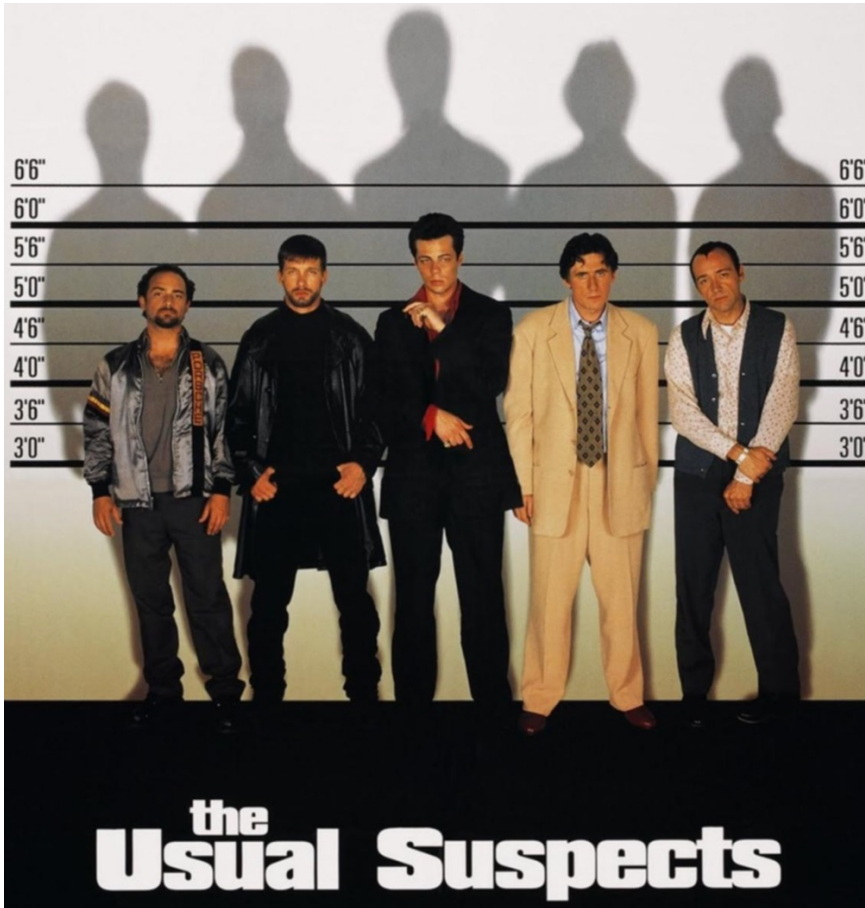


WHAT ARE THE "USUAL SUSPECTS"?

- Incomplete lineage sorting (ILS)
 - ❖ Deviation from strict multispecies coalescent is possible



WHY BIRDS?

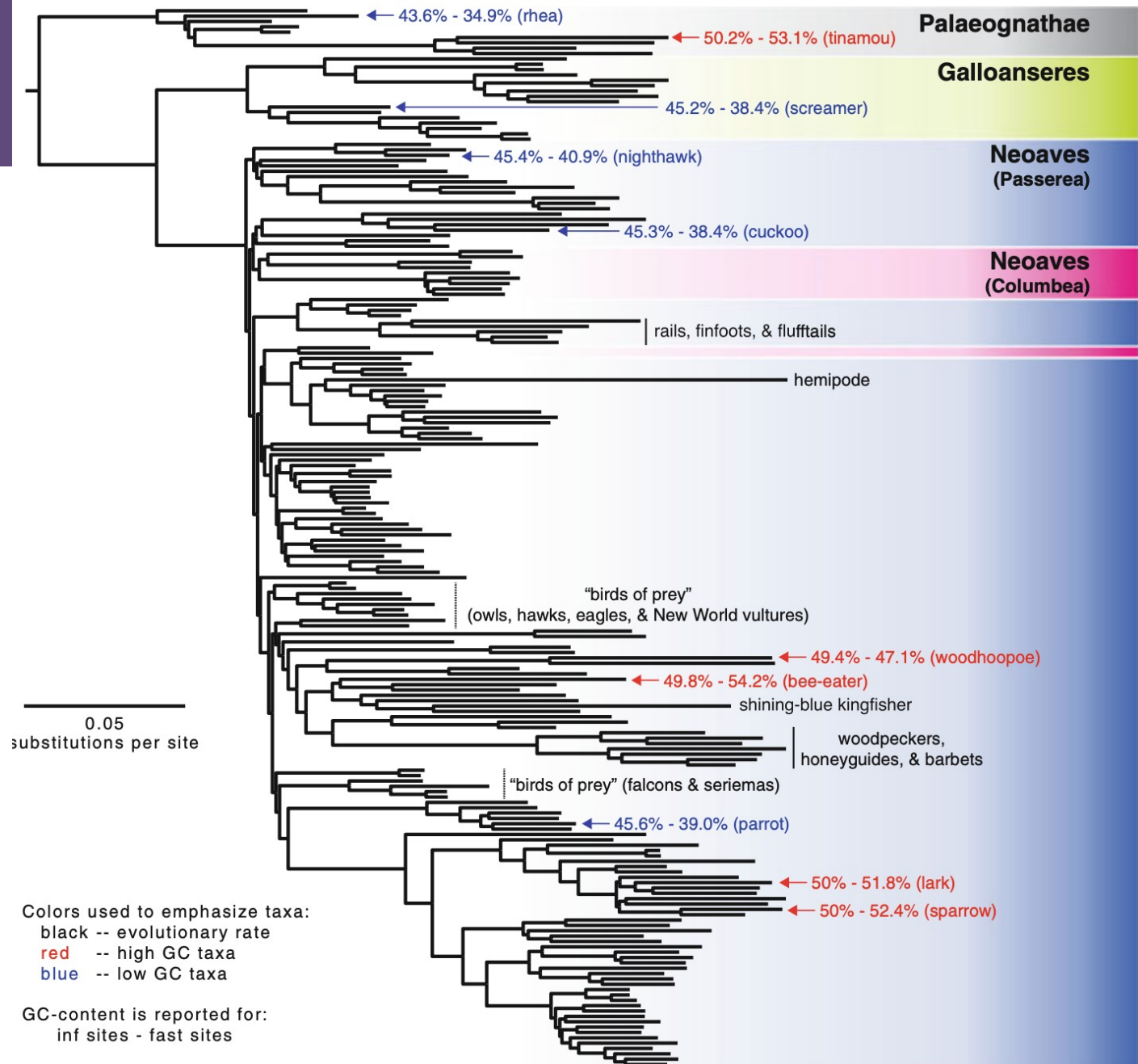


- But the problems don't include all the usual suspects...
 - ❖ Limited gene duplication and loss
 - ❖ Conservation of synteny
- ✓ Orthology is (relatively) easy to establish

RATES (& GC CONTENT)

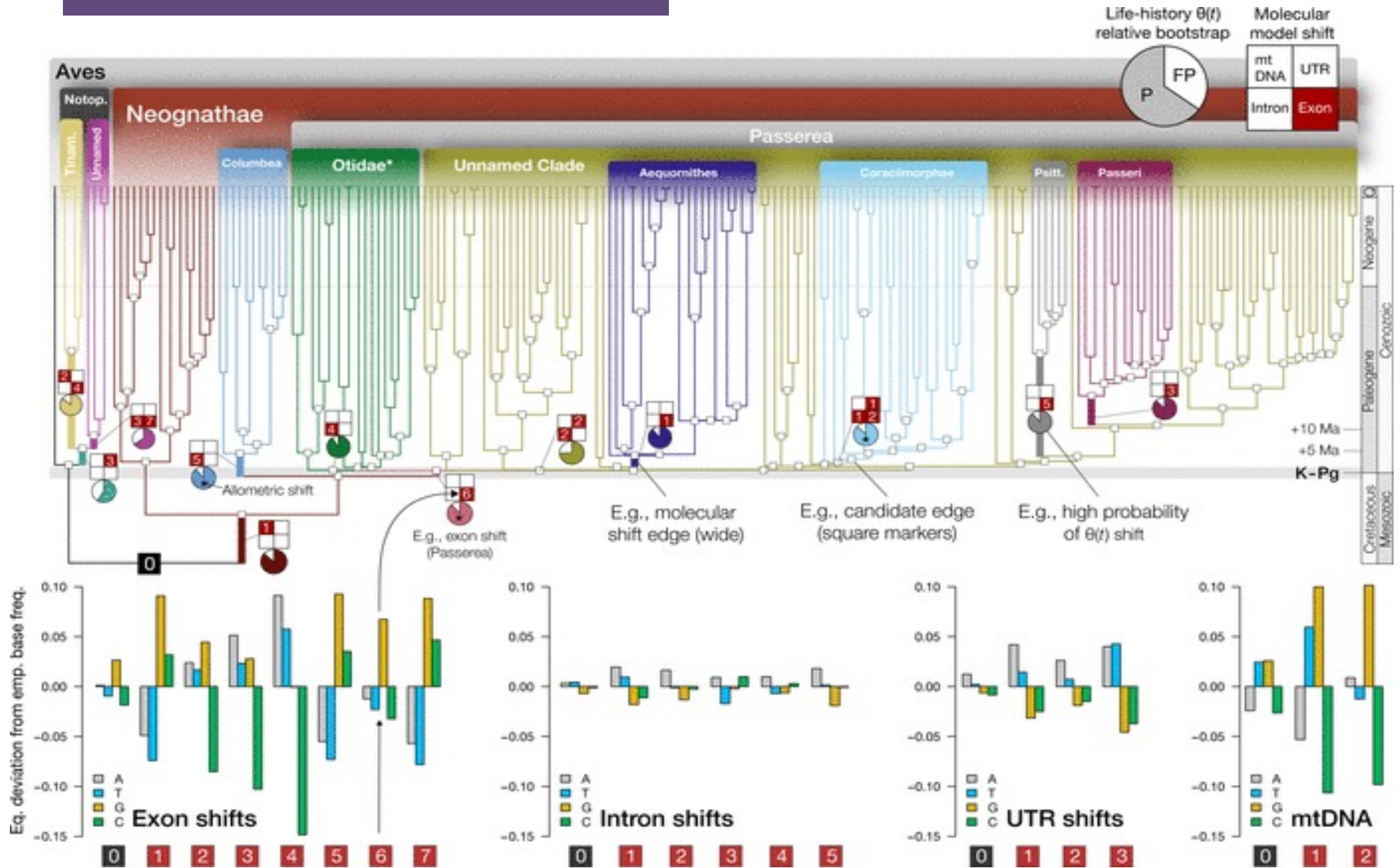
- Substantial branch length variation
- Rapid radiation (short branches at base of blue & pink clade)
- GC-content variation

From Braun et al. (2019)
chapter in *Avian Genomics*
- data are from Prum et al.
(2015) *Nature*



Book chapter: https://doi.org/10.1007/978-3-030-16477-5_6

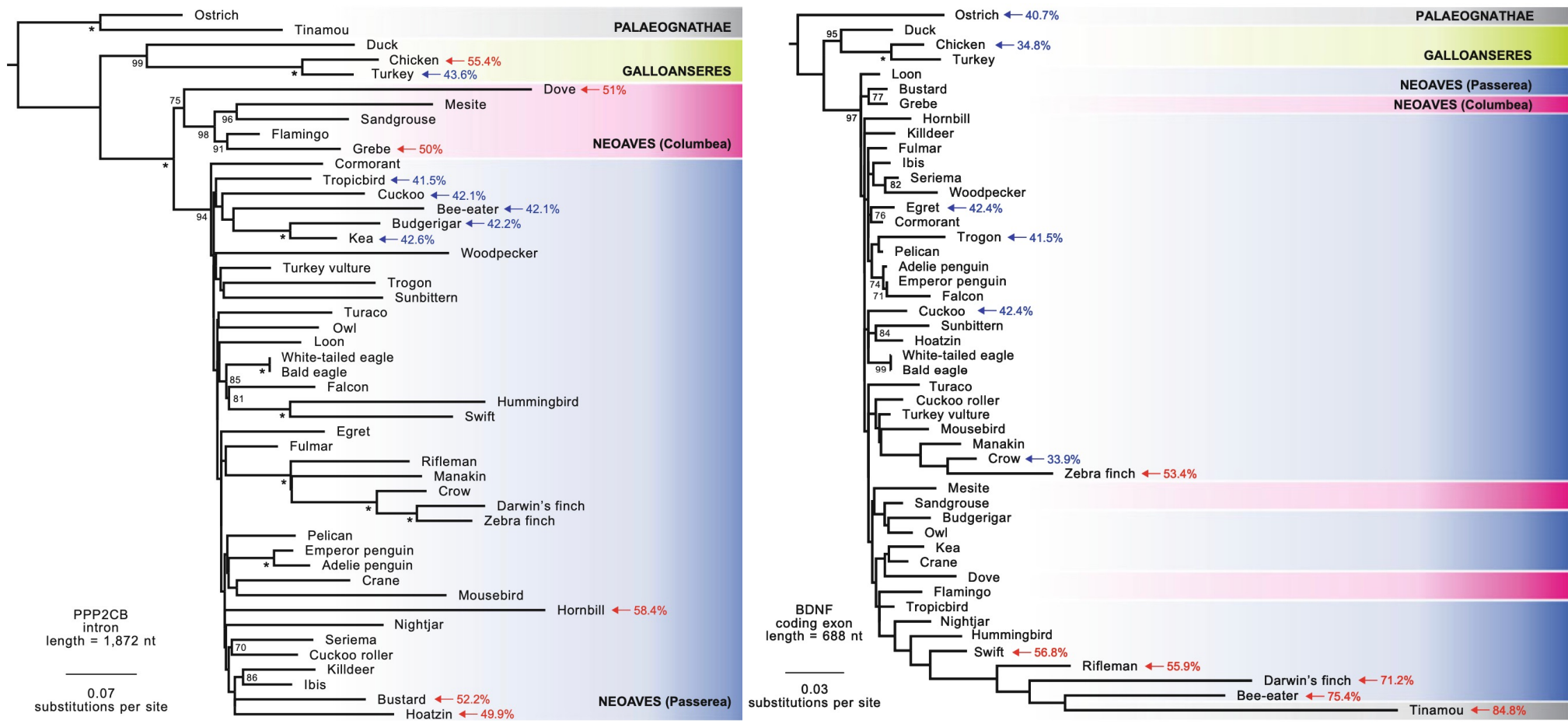
MODEL SHIFTS



Berv et al. 2024. Genome and life-history evolution link bird diversification to the end-Cretaceous mass extinction, Volume: 10, Issue: 31, DOI: (10.1126/sciadv.adp0114)

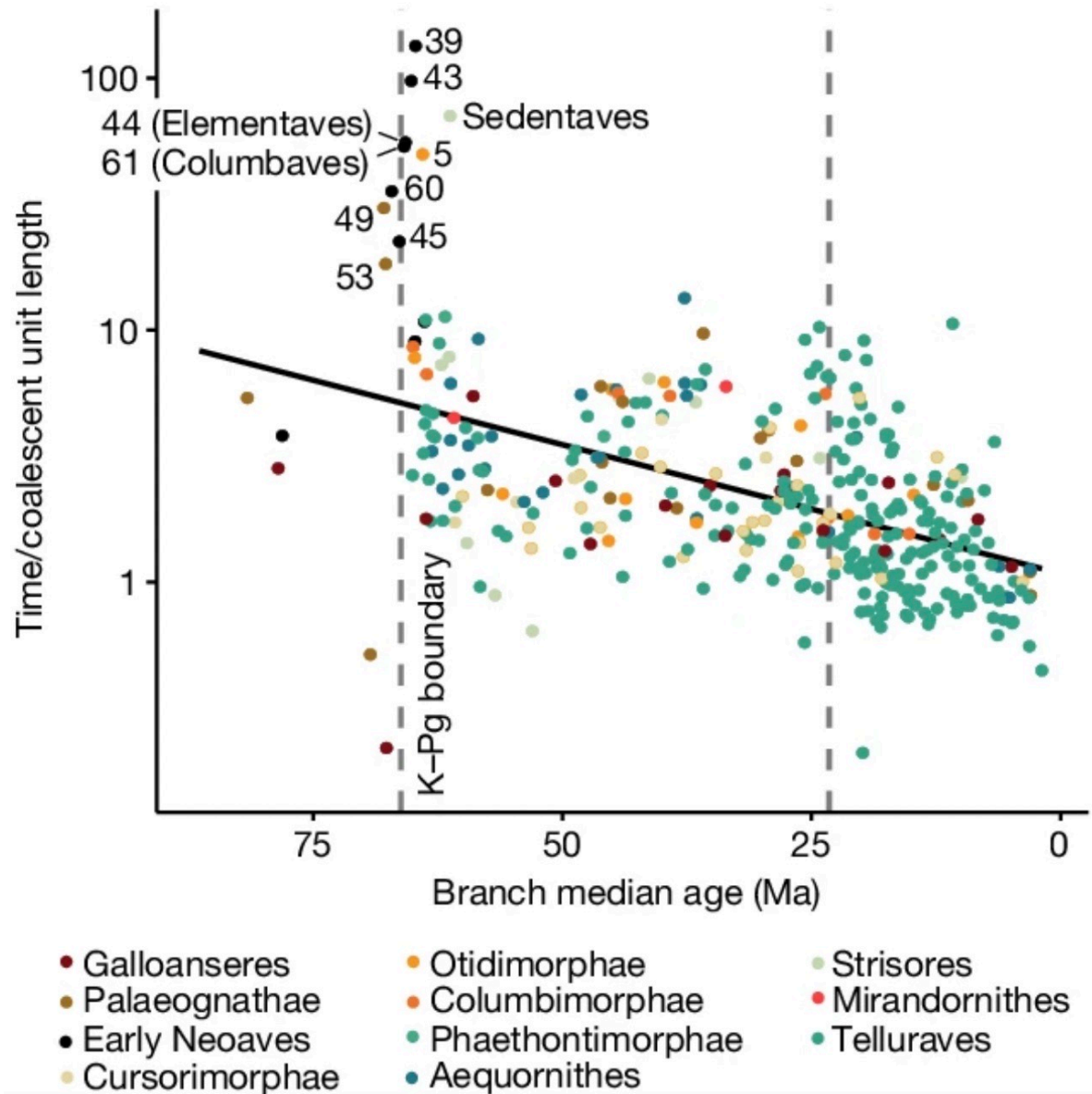
HETEROTACHY+ -- PPP2CB vs BDNF

- Comparison of two gene trees – PPP2CB (left) vs BDNF (right)
 - Heterotachy + base composition variation (+ BDNF long-branch attraction)



CHANGES IN EFFECTIVE POPULATION SIZE

- Note outliers in graph of timetree branch lengths divided by coalescent branch lengths (CBLs)
- Simplest explanation for elevated time/CBL values is **increased N_e**
 - Generation time shifts or errors in gene trees could explain as well
- Houde et al. (2019) *Diversity* came to similar conclusions regarding N_e
 - Sparse taxon sample, but accounted for generation time and used long indels to establish bipartitions in gene trees



CHALLENGES FOR AVIAN PHYLOGENOMICS

- Very short branches in some parts of the tree
- Shifts in the relative rate of evolution
 - ✓ Potential for long-branch attraction
- Shifts in the model of sequence evolution
 - ✓ Base composition convergence (especially GC-content)
 - ✓ Model shifts are most extreme in coding exons
- Gene tree discordance due to incomplete lineage sorting
 - ✓ Potentially exacerbated by periods of increased N_e

BIRD PHYLOGENY DOES NOT PRESENT A UNIQUE CHALLENGE

- Challenging relationships probably exist in all clades



New Results

The danger zone: the joint trap of incomplete lineage sorting and long-branch attraction in placing Rafflesiaceae

 Liming Cai,  Liang Liu, Charles C. Davis

doi: <https://doi.org/10.1101/2024.08.07.606681>

This article is a preprint and has not been certified by peer review [what does this mean?].



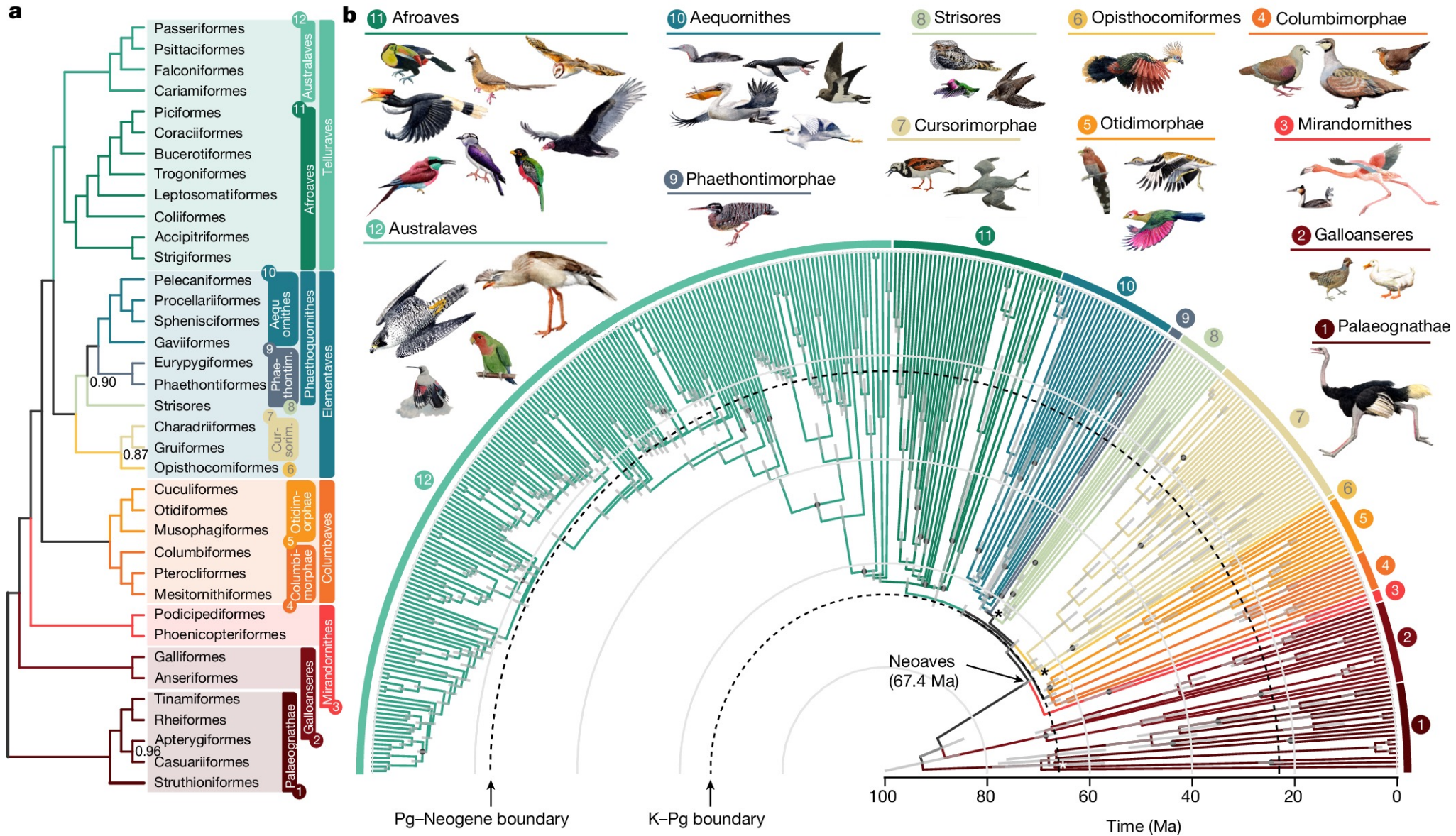
- But we have a lot of information about bird phylogeny

SOME “UNUSUAL SUSPECTS” ...

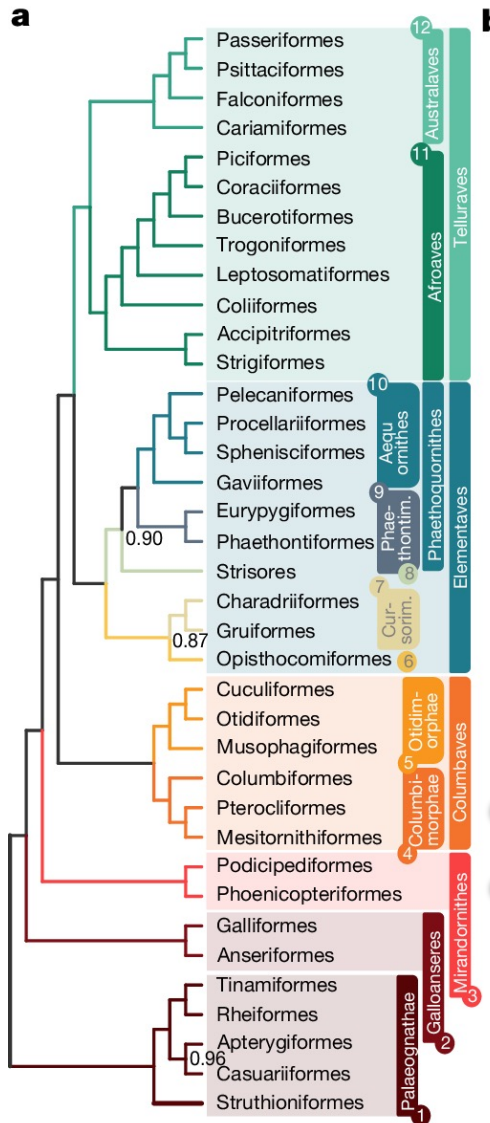
- Recombination landscape
- Sequence assembly and data quality



CURRENT BEST ESTIMATE OF THE BIRD TREE



CURRENT BEST ESTIMATE OF THE BIRD TREE



Columbimorphae (doves, mesites, and sandgrouse)

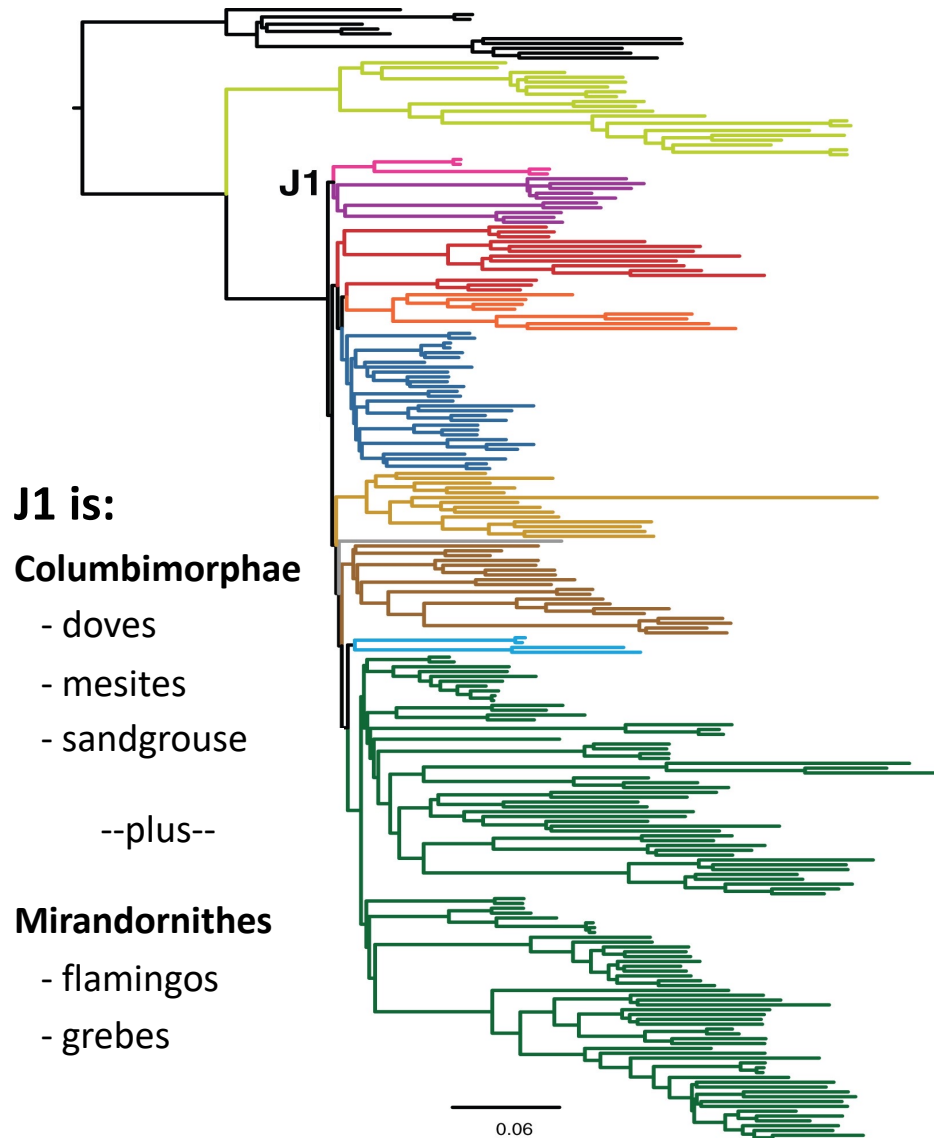
- **United in some earlier trees (J1)**

Mirandornithes (flamingos and grebes)



EARLIER (REDDY ET AL. 2017) PHYLOGENY

- 54-locus tree
 - Strong support for clade J1
- Recent phylogenies reject clade J1



Analysis and data matrix

Clade J1

	EB2	KIM	JAR
--	-----	-----	-----

RAxML, fast bootstrap, GTR+ Γ model

Early Bird II data, fully partitioned
 Early Bird II data, optimal rcluster (PF)
 Early Bird II data, unpartitioned

99	99	99
97	97	100
98	97	98

Non-coding data, fully partitioned

95	95	99
----	----	----

RAxML standard bootstrap, GTR+ Γ model

Early Bird II data, optimal rcluster (PF)

96	97	98
----	----	----

IQ-TREE ultrafast bootstrap, I+ Γ rates

Early Bird II data, fully partitioned
 Early Bird II data, optimal rcluster (PF)
 Early Bird II data, unpartitioned

100	99	99
100	100	100
100	100	100

Non-coding data, fully partitioned

100	100	100
-----	-----	-----

IQ-TREE ultrafast bootstrap, free rates

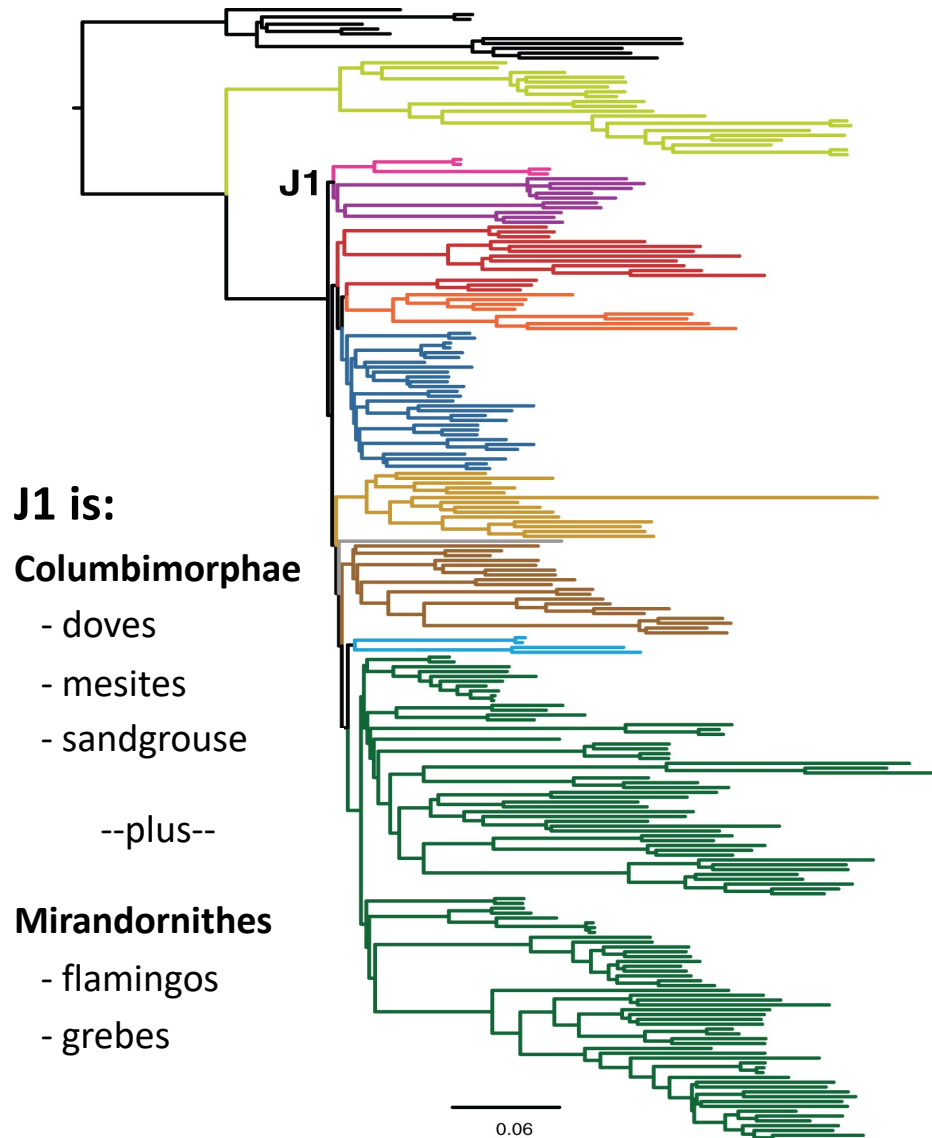
Early Bird II data, fully partitioned

100	100	100
-----	-----	-----

Non-coding data, fully partitioned

100	100	100
-----	-----	-----

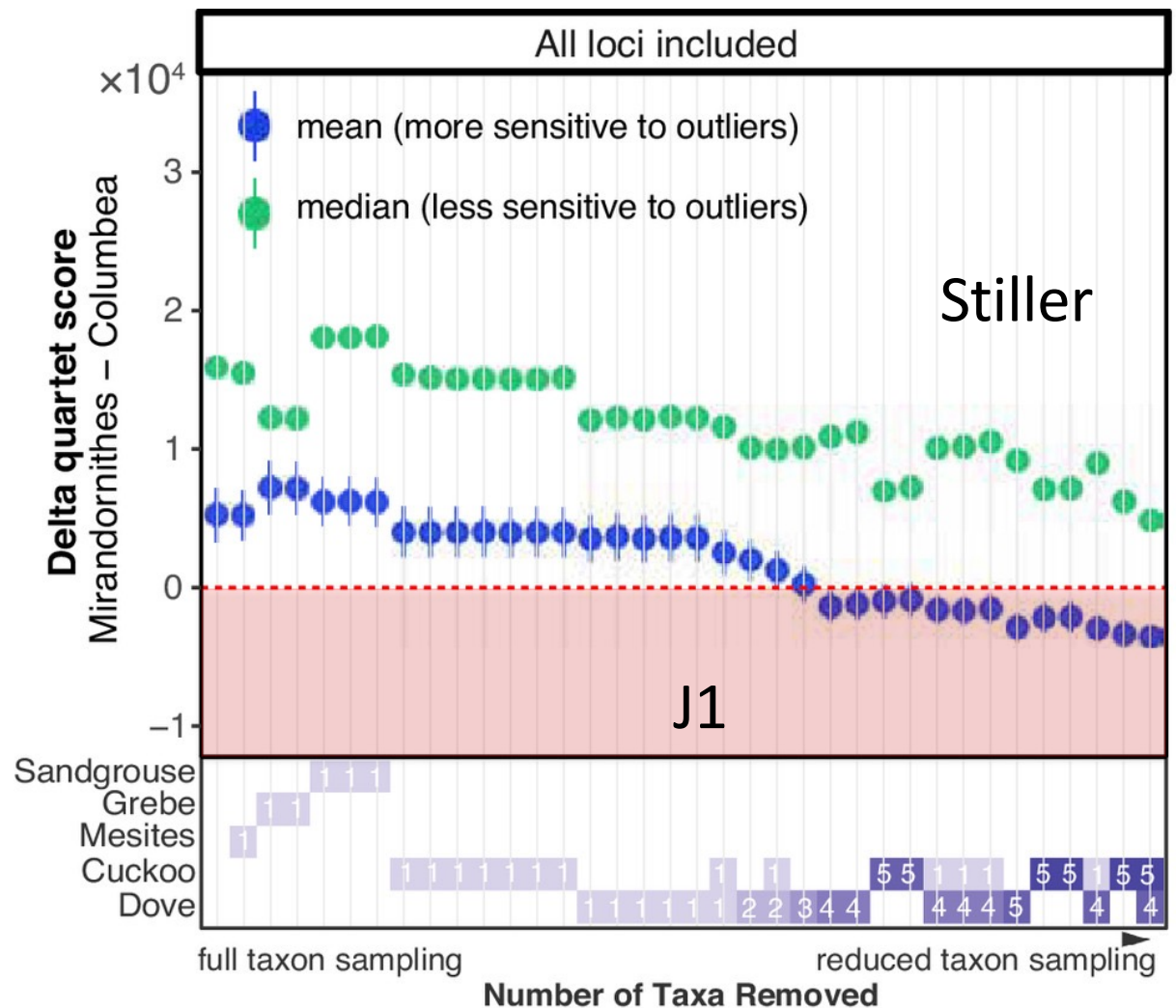
WHY DOES REDDY SUPPORT J1?



- Few large-scale trees support J1
 - Most phylogenomic trees (e.g., Prum et al. 2015; Wu et al. 2024) have low support at the base
 - Stiller et al. 2024 has high support, but it contradicts J1
- **Exception** - Jarvis et al. (2014) has 100% bootstrap support
 - Jarvis has a limited taxon sample
 - But Reddy has many taxa
- Why is support so strong in Reddy and Jarvis?

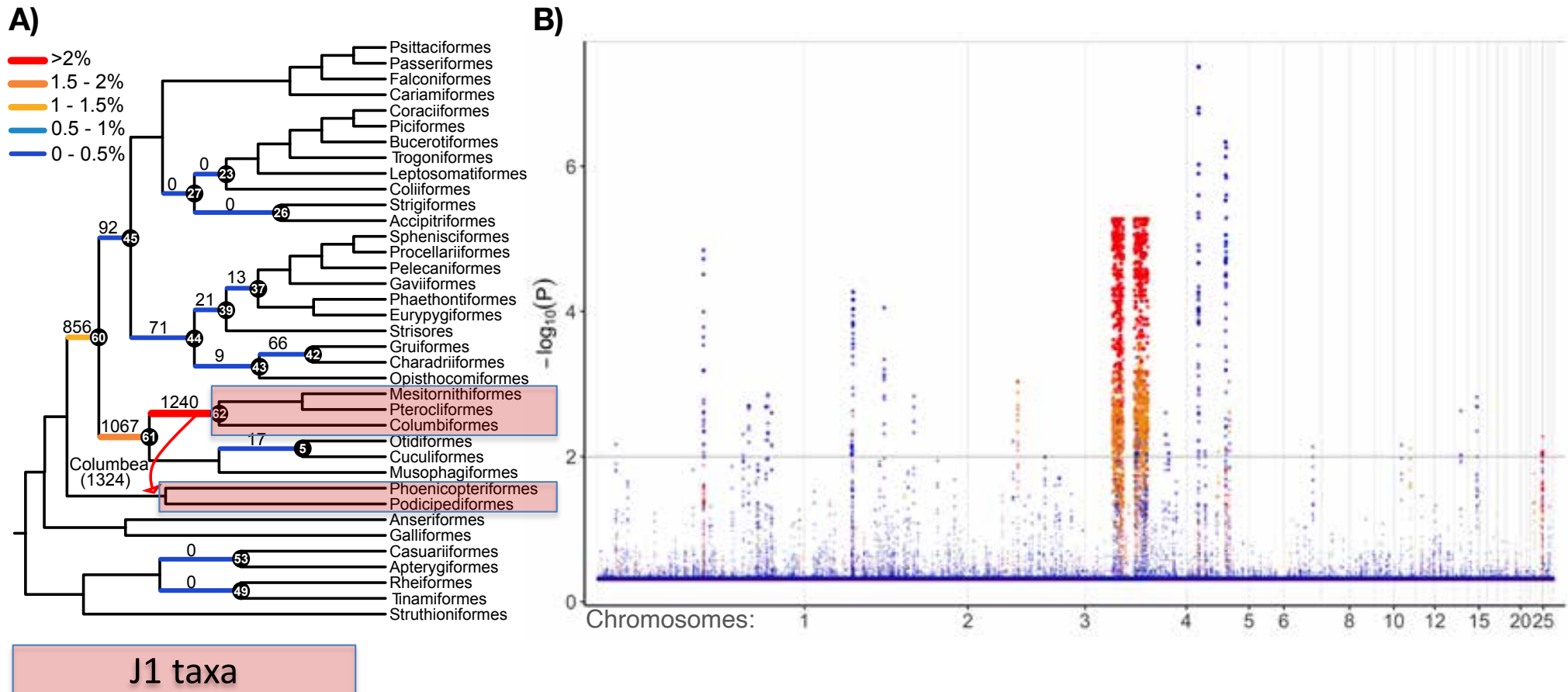
TAXON SAMPLING MATTERS

- Smaller taxon samples tend to support J1
 - Explains Jarvis result
 - But not Reddy!
- **Median quartet score supports Stiller** (not J1)
- Produced smaller taxon samples by pruning trees
 - Gene trees estimated using >300 taxa
- Mean vs median quartet scores differ



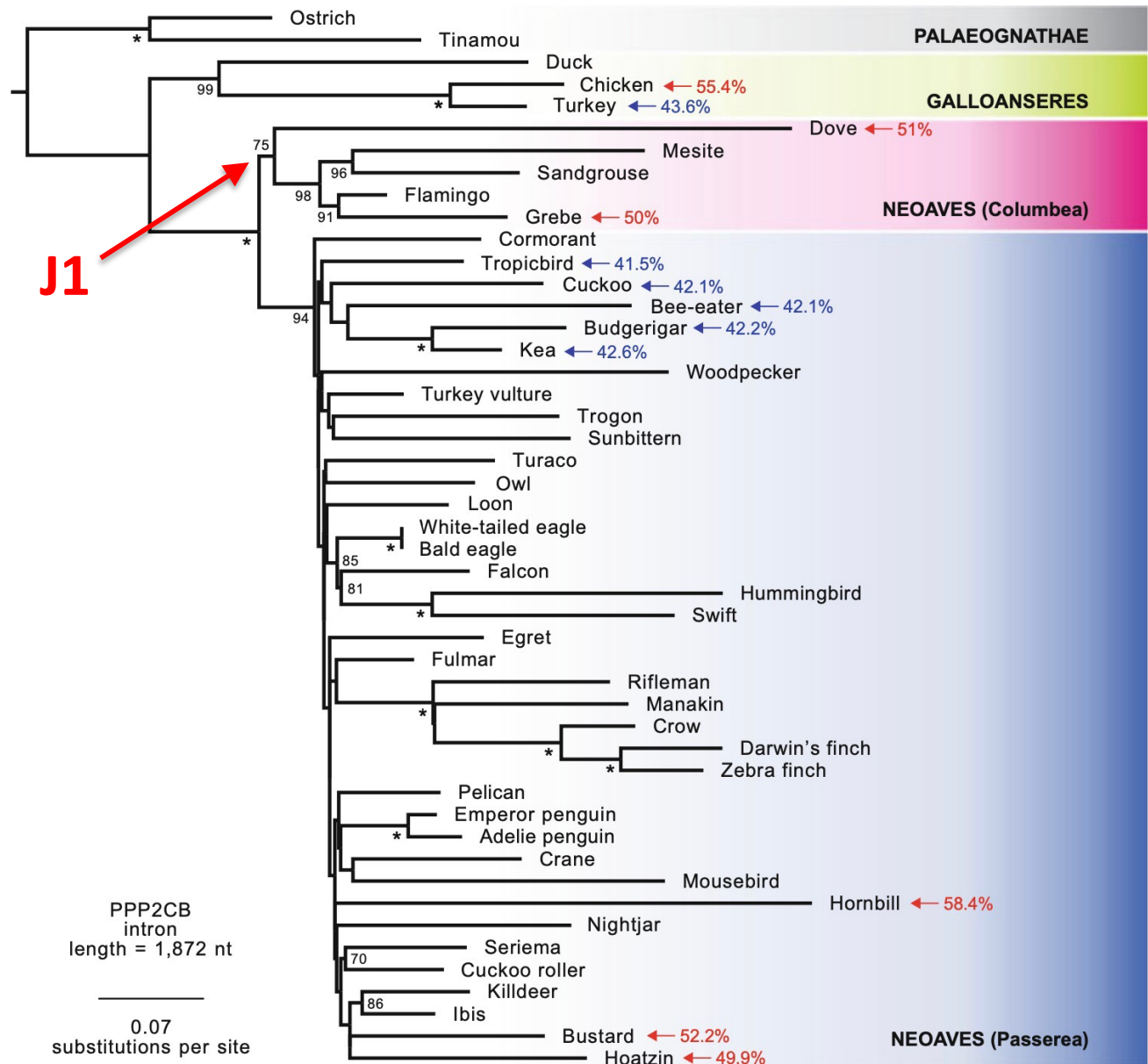
GENE TREES SUPPORTING J1 ARE CLUSTERED

- Approximately 2% of the genome strongly supports clade J1
- Manhattan plot shows $-\log(P)$ that 20-locus (~200 kb) windows have unusual quadripartition quartet support



PPP2CB –REDDY ET AL. OUTLIER LOCUS

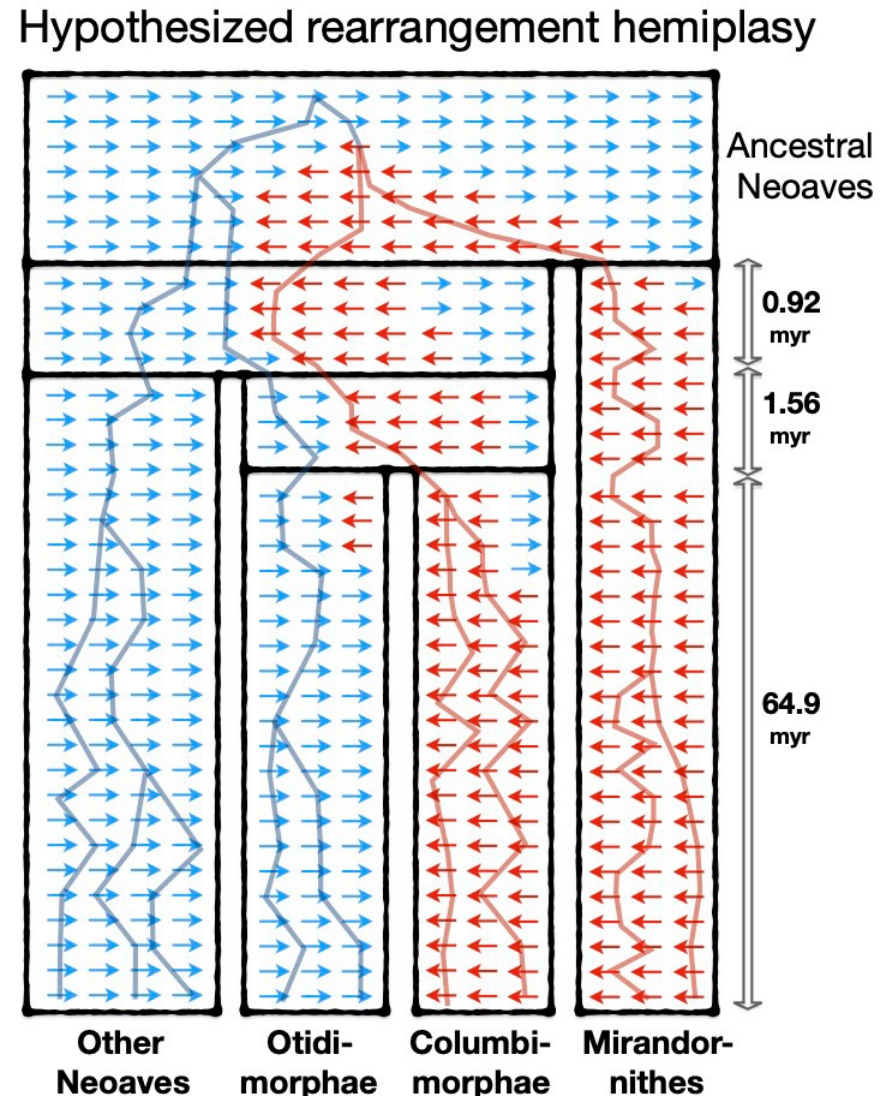
- Short region but 75% support for clade J1
 - J1 = “Columbea”
- Result attributed to limited variation in base composition
 - PPP2CB has a “better” tree than BDNF
- BUT.. PPP2CB lies in the outlier region
 - Supports J1



Braun et al. (2019) chapter in “Avian Genomics” book

MODEL FOR CHROMOSOME 4 EVOLUTION

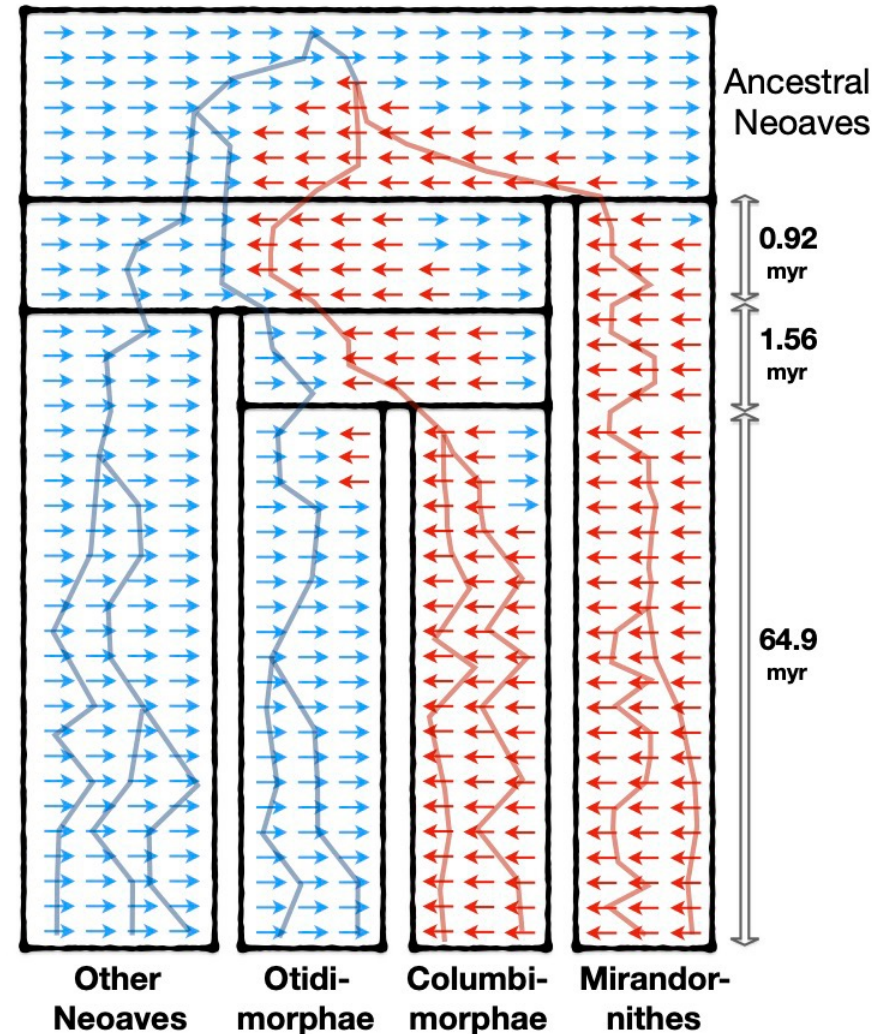
- ~ 2% of the bird genome have a specific gene tree
 - Recombination suppressed in the “outlier region”
 - Shifts estimates of the avian species tree to include clade J1
 - (for limited taxon samples)
 - ASTRAL and concatenation
- Reddy et al. (2017) **recovered J1** because **one of 54 loci** is in the outlier region



IMPACT OF CHROMOSOME 4 EVOLUTION

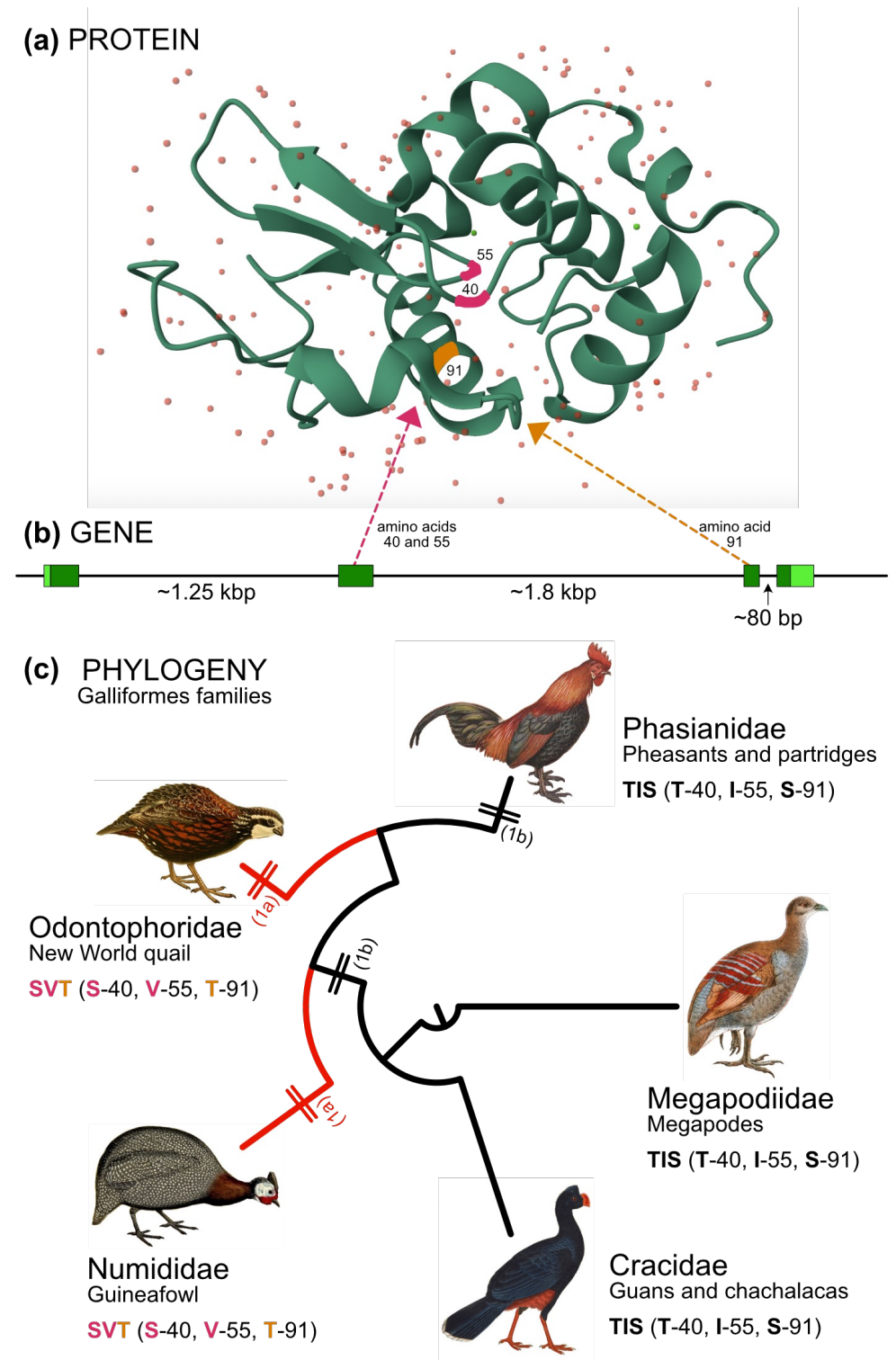
- Ancient (>60 mya) chromosomal rearrangement was polymorphic for a few million years
 - Detected by **absence of ILS** in the region
 - For birds, recombination will limit recombination free segments to a mean length of 5(!) to 5000 bp
- Persistence in population could reflect neutral evolution
 - Alternatively, could be balancing selection (weak evidence)

Hypothesized rearrangement hemiplasy



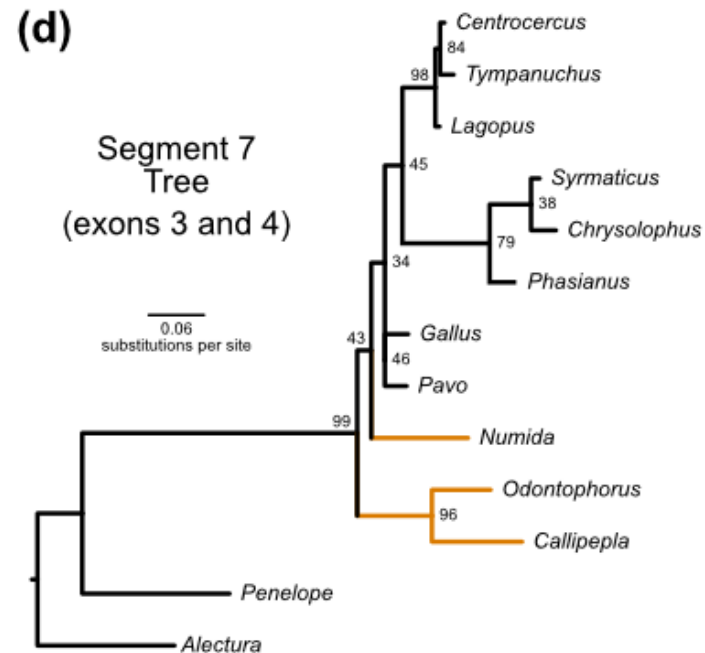
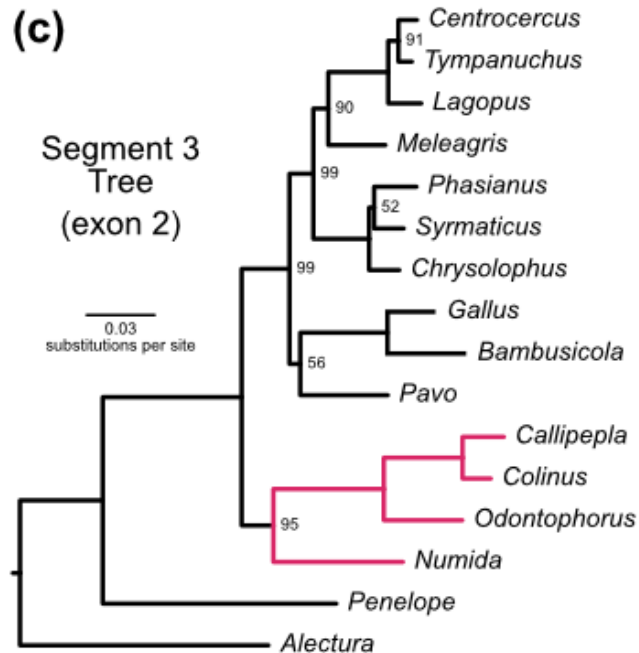
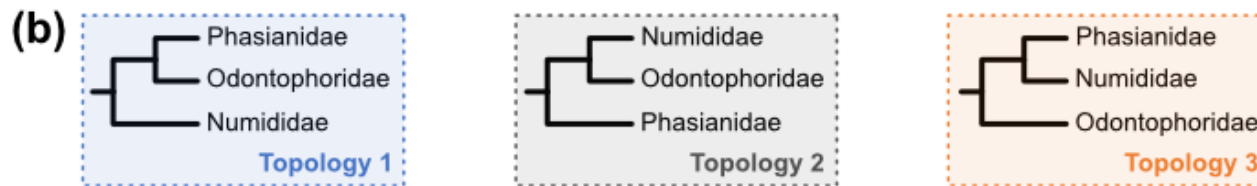
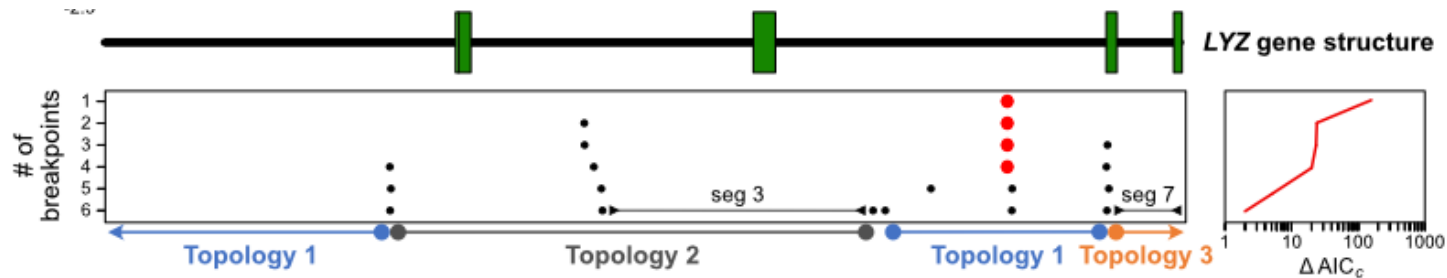
RECOMBINATION

- Estimates of recombination-free region lengths are ~5 kb for large chromosomes
 - Are the estimates realistic?
- Protein with changes at three functionally important amino acids
 - Encoded by gene on chr 1 (largest chromosome)
- Changes would have to occur twice given the species tree



RECOMBINATION

- Recombination-free segments are ~1 kb on chr 1

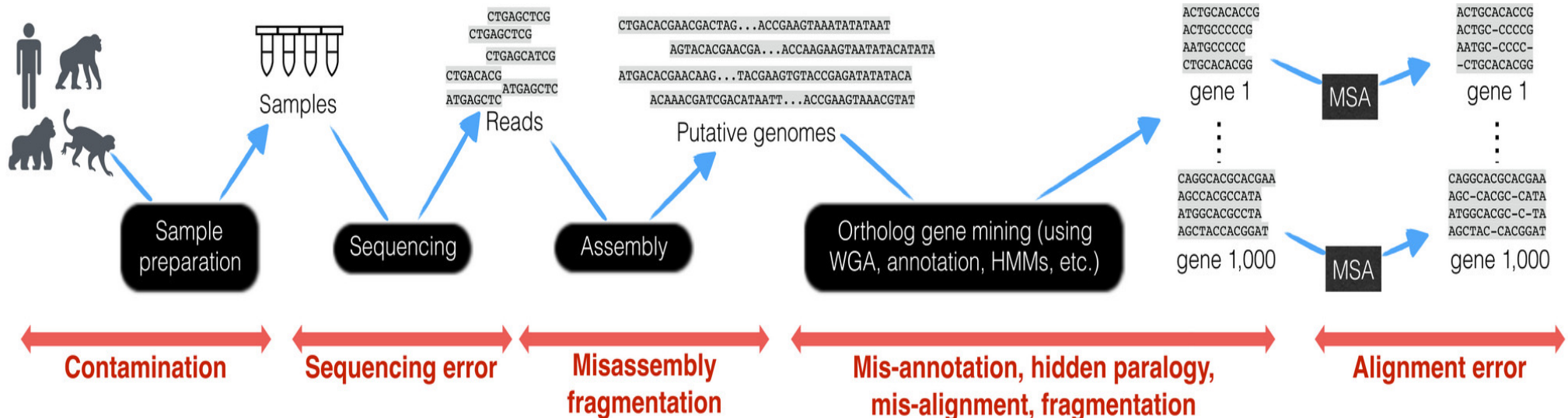


TECHNICAL ISSUES

- Sequence alignment
 - Including the extraction of orthologous data
- Sequence assembly

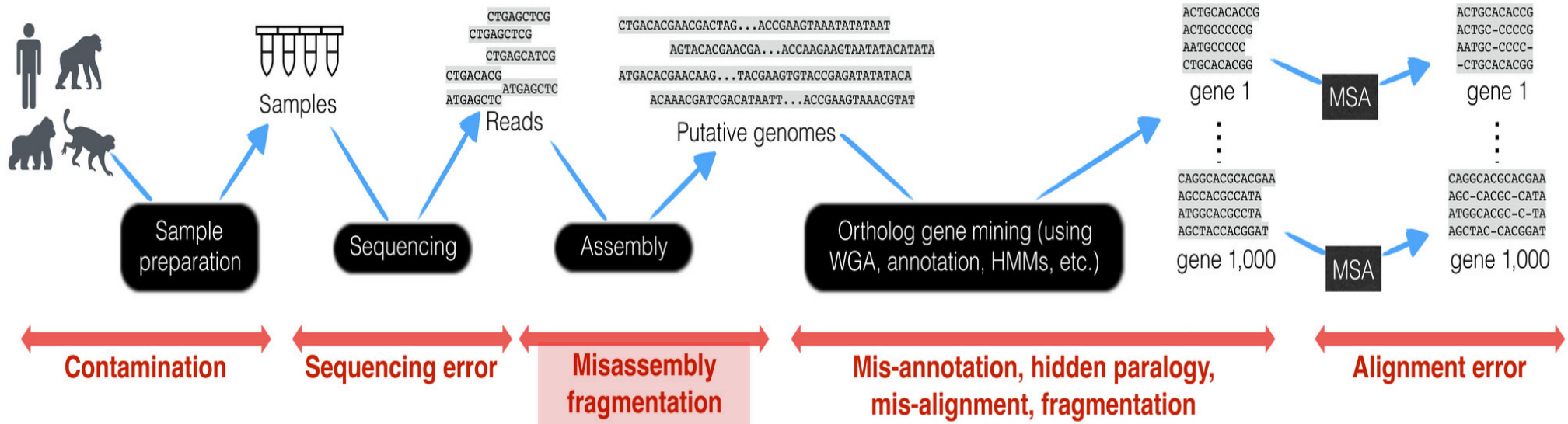
TECHNICAL ISSUES

- There are many (potentially) problematic steps in phylogenomic pipelines



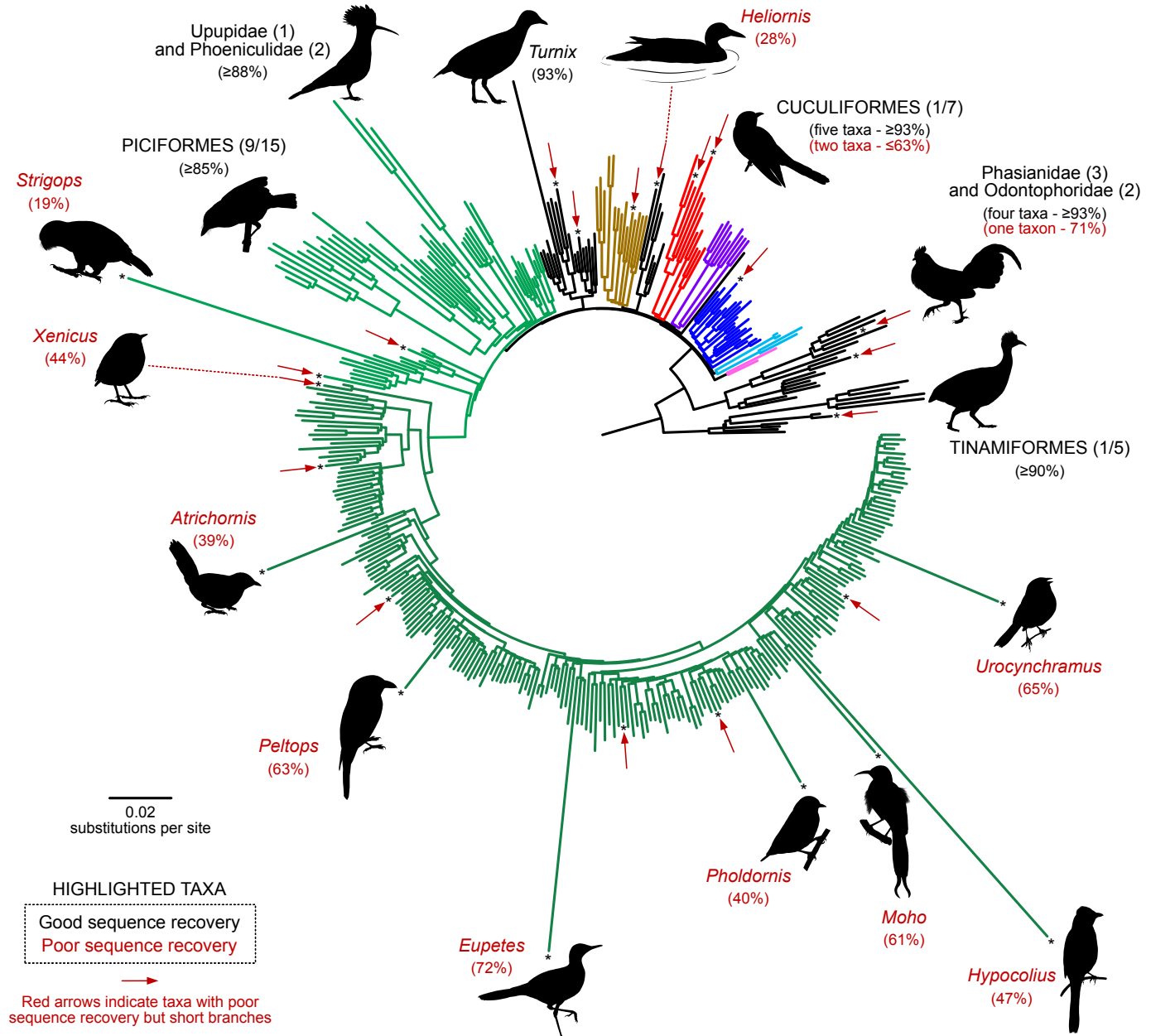
SEQUENCE ASSEMBLY

- What is the potential for misassembly?



A UCE TREE (“ALLFAM”)

- Typical patterns of branch length heterogeneity are evident...
- ...but there are also additional long branches evident
 - Taxa with **red names** had poor sequence recovery
 - The **red arrows** also indicate poor sequence recovery

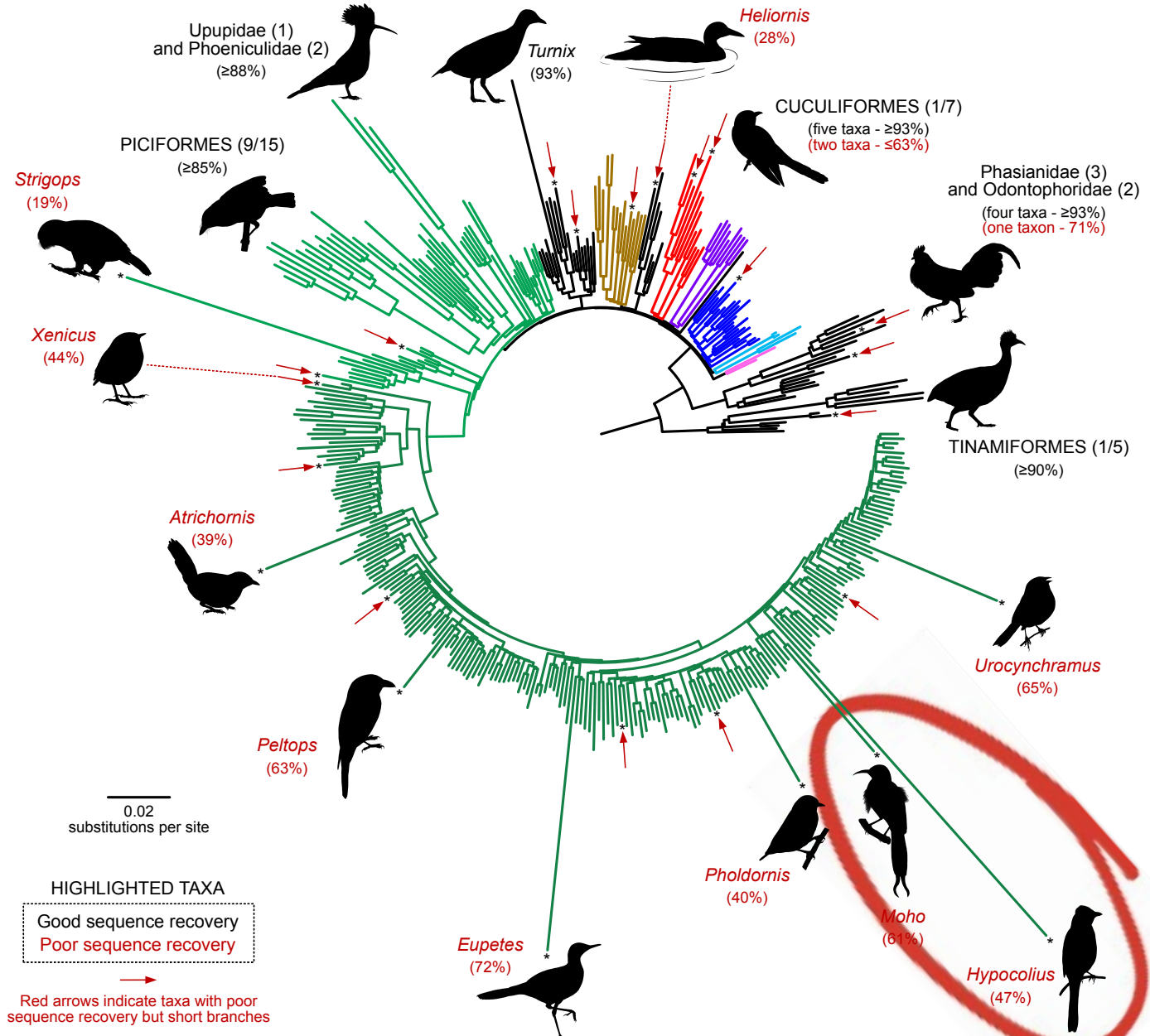


A UCE TREE (“ALLFAM”)

- Both circled taxa were from historical samples

- *Moho nobilis** (Hawai’i ‘ō’ō)
- *Hypocolius ampelinus* (Grey hypocolius)
- Sister group:
 - *Hylocitrea bonensis* (yellow-flanked whistler)
- Short branch

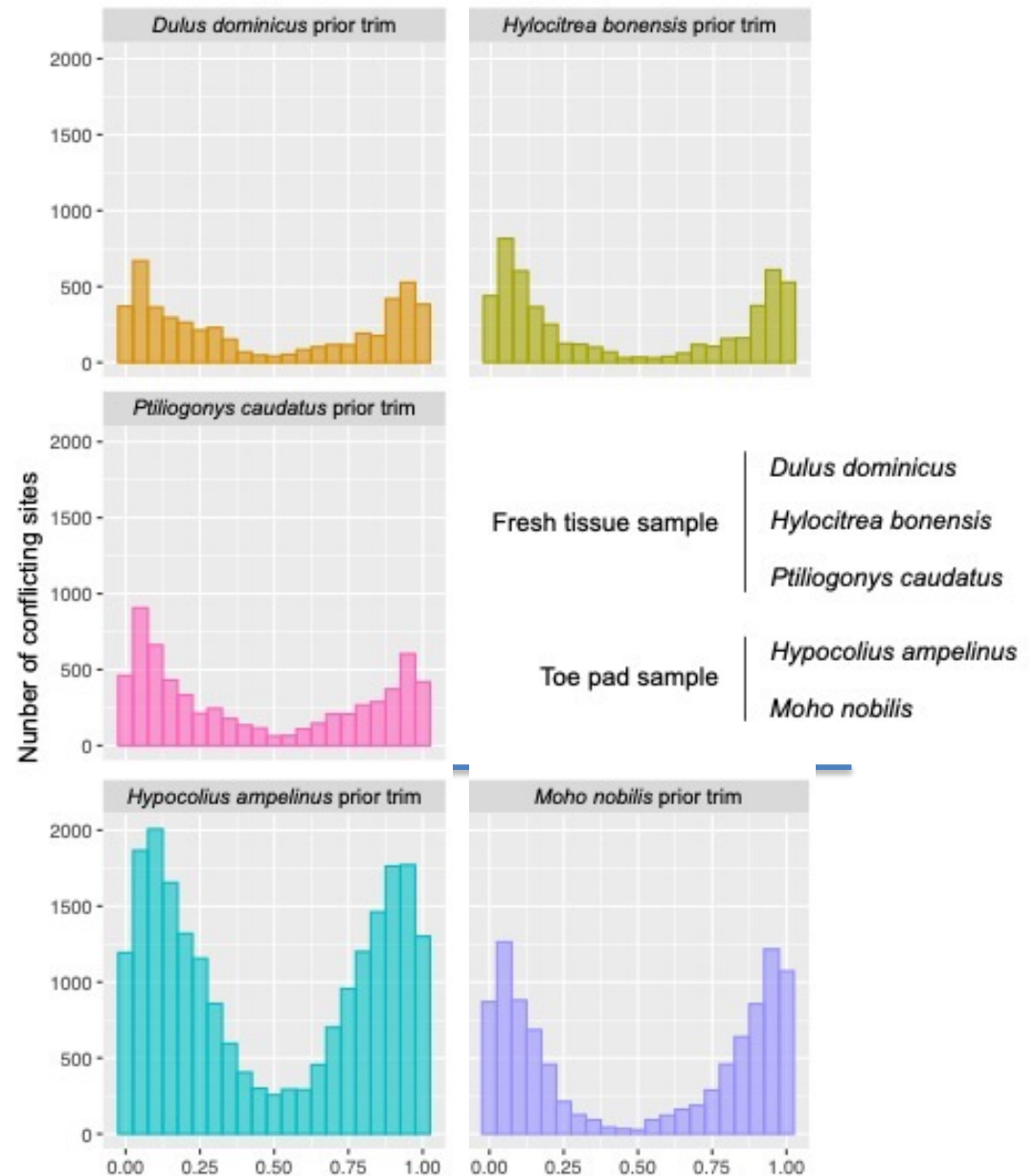
*extinct



CONFLICTS BETWEEN ASSEMBLERS

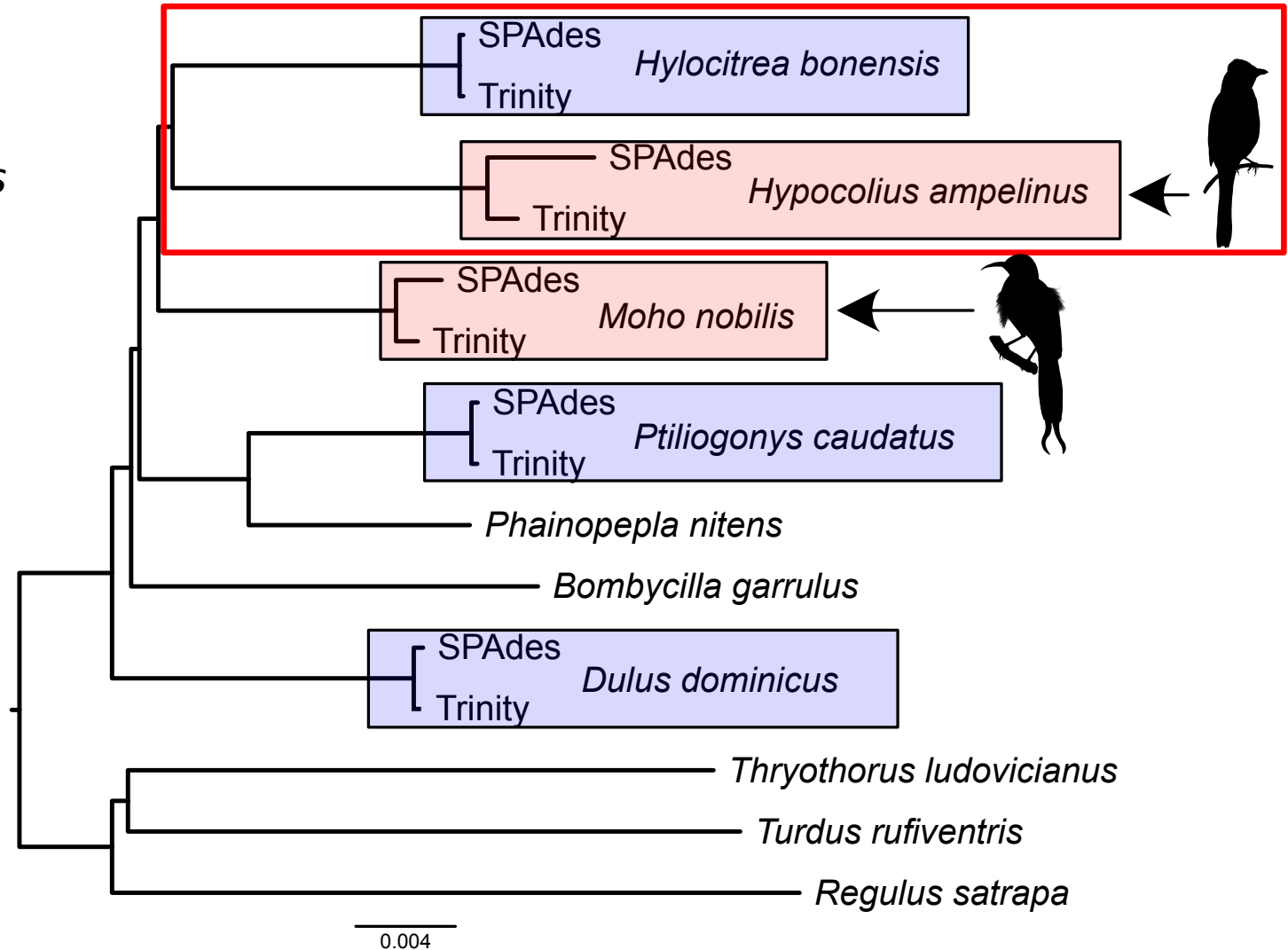
- Re-assemble data using two different programs
 - SPAdes and Trinity
 - Original assemblies were SPAdes (different settings)
- Two historical
 - *Moho nobilis* (extinct)
 - *Hypocolius ampelinus*
 - Form a clade in “allfam”
- Three recent
 - *Dulus dominicus*
 - *Hylocitrea bonensis**
 - *Ptilogonys caudatus*

* Sister to “historical taxa clade”

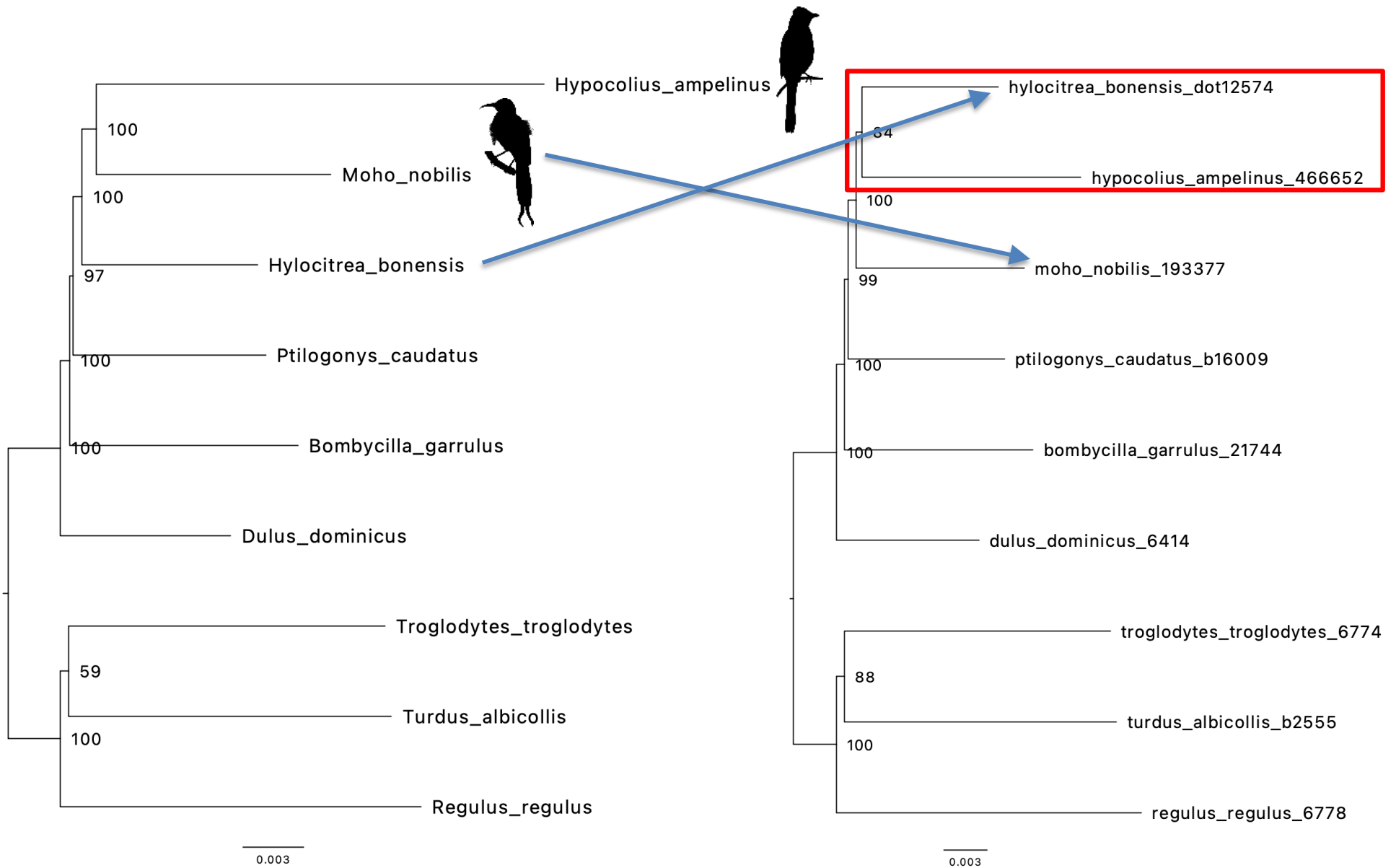


LONG BRANCHES REFLECT ASSEMBLY

- Tree with different assemblies treated as taxa
- *Moho* and *Hypocolius* have shorter root-to-tip branches
 - Comparable to *Hylocitrea*
- Different topology (note red box)
- **Terminal branches longer for historical samples!**



ALIGNMENTS FROM BRAUN ET AL. 2024 (LEFT) VS REALIGNED DATA USING THE SAME ASSEMBLIES (RIGHT)



ASSEMBLY AND ALIGNMENT

- So...what are the relationships for *Hypocolius*, *Moho*, and *Hylocitrea*?
 - *Hypocolius* and *Hylocitrea* form a clade (with *Moho* sister) with all assemblies (and all taxon samples) where *Hypocolius* has a short branch
 - *Hypocolius* + *Hylocitrea* found if conflicts between assemblers recoded as N's
- Long branches are artifactual
- Interaction between assembly quality and alignment

PERFORMANCE OF VARIOUS METHODS

- Avian phylogeny has received extensive study
- It is possible to establish a set of “reliable clades” likely to be in the true bird species tree
- Can be divided into subsets based on their “easy”
 - Range from “easy” (typically supported by individual genes) to “hard” (only recovered in analyses of phylogenomic data)

	Easy	Medium	Hard	TOTAL
Families ¹	93	—	—	93
Superfamilies ²	3 / 1	14	4	21 / 1
Orders ^{1,2}	36 / 1	3	—	39 / 1
Superorders ²	7 / 5	12 / 11	9 / 8	28 / 24
TOTAL ²	139 / 7	29 / 11	13 / 8	181 / 26

Reliable clades that can be scored for “allfam” (left) and Jarvis (right) taxon samples

PERFORMANCE OF VARIOUS METHODS

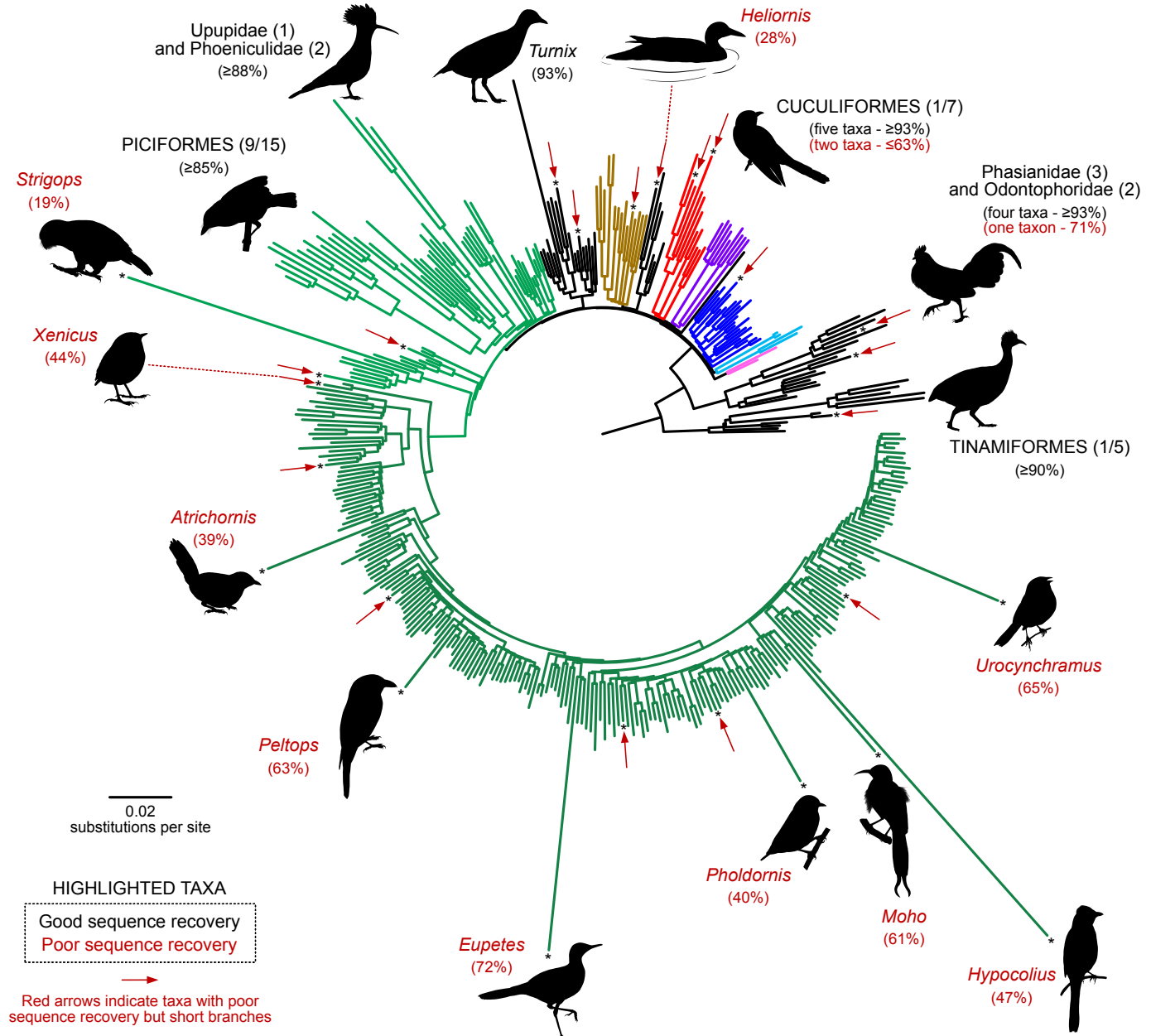
- The number of reliable clades recovered can provide a way to evaluate methods
 - Imperfect because we don't have knowledge of the true tree
 - Less controlled than simulations...
 - ...but still a useful method

	Easy	Medium	Hard	TOTAL
Families ¹	93	—	—	93
Superfamilies ²	3 / 1	14	4	21 / 1
Orders ^{1,2}	36 / 1	3	—	39 / 1
Superorders ²	7 / 5	12 / 11	9 / 8	28 / 24
TOTAL ²	139 / 7	29 / 11	13 / 8	181 / 26

Reliable clades that can be scored for “allfam” (left) and Jarvis (right) taxon samples

THE ALLFAM TREE

- Phylogram based on ML analysis of concatenated data
- 4307 UCE loci
- 2,042,823 aligned sites
- 852,768 informative sites
- Initial “species tree” analyses yielded trees with many fewer reliable clades than concatenation



THE ALLFAM TREE

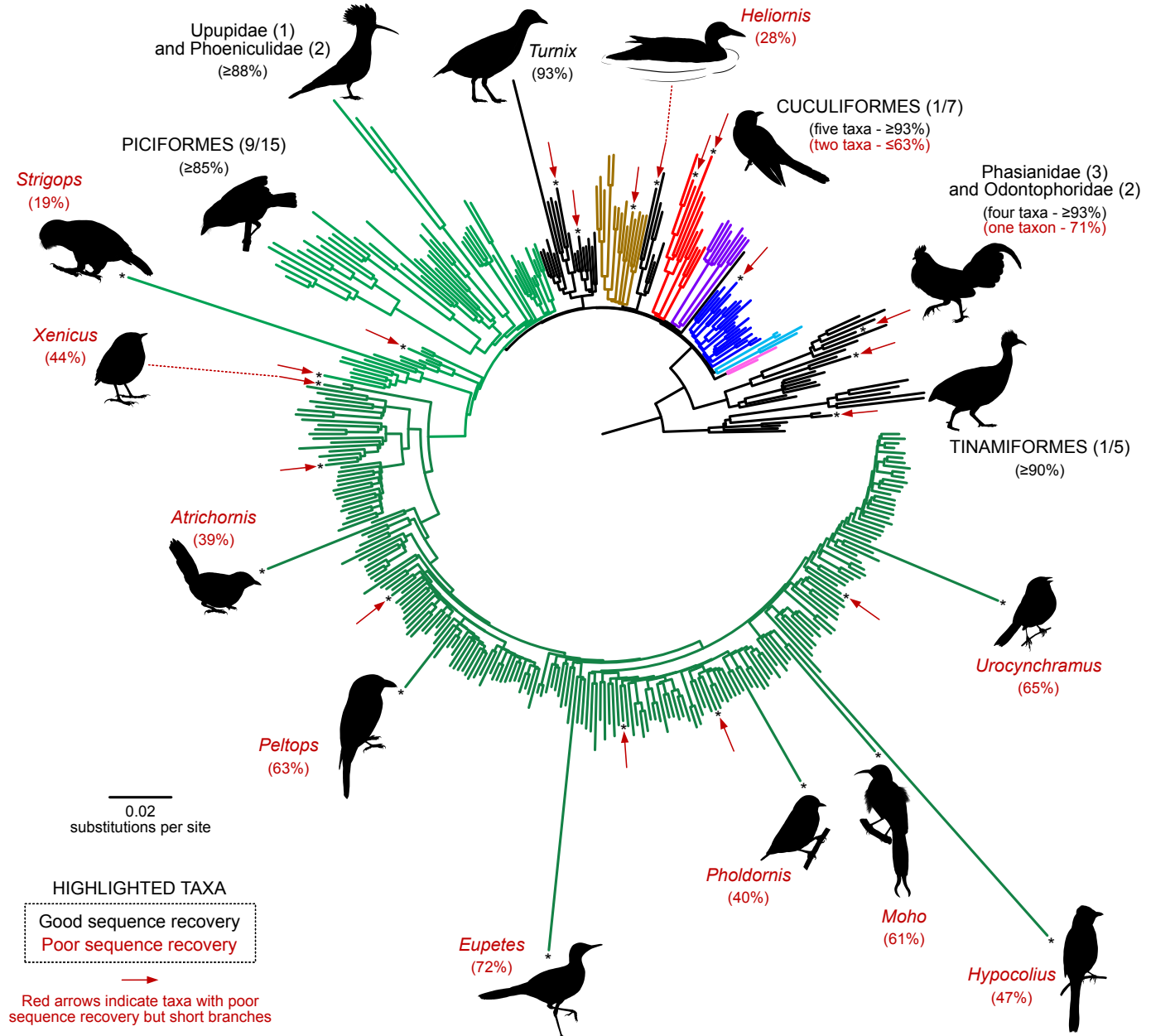
○ Decided to try a broad suite of species tree analyses

- ASTRAL
- ASTRID
- SVD quartets
- METAL (distances)

-- Dasarathy et al. 2015
IEEE/ACM TCBB 12: 422-32

-- Allman et al. 2019 *SIAM J Appl Algebra Geometry*, 3: 107-27.

Braun et al. (2024) bioRxiv
<https://doi.org/10.1101/2024.02.28.582627>



CLADE RECOVERY

-- Broken in subsets by "ease" of recovery

- Easy at top
- Uncertain at the bottom
- Bold clade have TE support

-- Only method with consistently good performance was ML concatenation

-- Potentially unfair due to missing data

(a)

#	Clade	ML concat (ExaML)	ML concat (IQ-TREE)	weighted ASTRAL	weighted ASTRID	METAL (NJ)						
						SVD quartets	p-distances	ML distances	Trace distances	logdet i.r.	logdet-inv	logdet inf sites
1	PALAEOGNATHAE											
2	Notopalaeognathae											
4	NEOGNATHAE											
5	Galloanserae											
6	Neoaves											
7	Mirandornithes (VII)											
16	Daedalornithes											
33	Passeriformes											
34	Eupasserres											
		0	0	0	0	1	2	2	2	2	2	0
3	Novaeratitae											
12	Strisores (V)											
19	Aequornithes (II)											
20	Feraequornithes											
21	Procellariimorphae											
24	Telluraves (I)											
26	Cavitaves											
27	Eucavitaves											
28	Picocoraciae											
30	Australaves											
31	Eufalconimorphae											
32	Psittacopasserae											
		1	1	1	0	5	8	9	9	8	8	5
8	Columbimorphae (VI)											
9	Pteroclimesites											
11	Musophagotides											
17	Phaethoquornithes											
18	Phaethontimorphae (III)											
22	Pelecanimorphae											
23	Pelecanes											
25	Coraciimorphae											
29	Picodynastornithes											
		1	1	1	1	5	7	7	7	7	7	5
10	Otidimorphae (IV)											
13	Vanescaves											
14	Sedentaves											
15	Letornithes											
-	Afroaves											
-	Accipitriformes+Strigiformes											
-	Coraciimorphae+Australaves											

(b)

Tree	Non-monophyletic	
	Orders	Families
ML concatenation		
IQ-TREE	0	0
ExaML	0	0
MSC methods		
weighted ASTRAL	1	6
weighted ASTRID	0	4
SVD quartets	3	1
METAL (NJ)		
p-distances	4	4
ML distances	5	4
Trace distances	5	4
logdet i.r.	4	4
logdet-inv	4	4
logdet inf sites	0	3

(c)

Clade	ML concat (IQ-TREE)	ML concat (ExaML)	weighted ASTRAL	weighted ASTRID	METAL (NJ)						
					SVD quartets	p-distances	ML distances	Trace distances	logdet i.r.	logdet-inv	logdet inf sites
Easy (+++)											
Tyranni (suboscines)											
Passeri (oscines)											
	0	0	0	0	0	0	0	0	0	0	0
Medium (++)											
Eurylaimides											
Tyrannides											
Furnariida											
Tyrannida											
Malaconotoidea											
Corvoidea											
Sylviida											
Sylvioidea											
Muscicapida											
Bombycilloidea											
Muscipapoidea											
Certhioidea											
Passerida											
Emberizoidea											
	0	0	4	3	2	5	5	5	5	5	5
Hard (+)											
Corvides											
Orioloidea											
Passerides											
Aegithaloidea											
	0	0	3	1	0	3	3	3	3	3	3
Locustelloidea											

CLADE RECOVERY

-- Eliminating taxa with large amounts of missing data did improve all methods

-- Is the issue simply the amount of data

-- NJ analyses had worst performance

(a)

	#	Clade	METAL (NJ)																			
			ML concat (IQ-TREE)	weighted ASTRAL	weighted ASTRID	SVD quartets (sub)	SVD quartets (ex)	p-distances	ML distances	Trace distances	logdet i.r.	logdet-inv	logdet inf sites									
Easy (+++)	1	PALAEOGNATHAE																				
	2	Notopalaeognathae																				
	4	NEOGNATHAE																				
	5	Galloanserae																				
	6	Neoaves																				
	7	Mirandornithes (VII)																				
	16	Daedalornithes																				
33	Passeriformes																					
34	Eupasserres																					
			0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Medium (++)	3	Novaeratitae																				
	12	Strisores (V)																				
	19	Aequornithes (II)																				
	20	Feraequornithes																				
	21	Procellariimorphae																				
	24	Telluraves (I)																				
	26	Cavitaves																				
	27	Eucavitaves																				
	28	Picocoraciae																				
	30	Australaves																				
	31	Eufalconimorphae																				
	32	Psittacopasserae																				
			1	1	0	5	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Hard (+)	8	Columbimorphae (VI)																				
	9	Pteroclimesites																				
	11	Musophagotides																				
	17	Phaethoquornithes																				
	18	Phaethontimorphae (III)																				
	22	Pelecanimorphae																				
	23	Pelecanes																				
	25	Coraciimorphae																				
29	Picodynastornithes																					
			1	4	2	5	4	5	4	3	5	4	3									
Uncertain (-)	10	Otidimorphae (IV)																				
	13	Vanescaves																				
	14	Sedentaves																				
	15	Letornithes																				
	-	Afroaves																				
	-	Accipitriformes+Strigiformes																				
	-	Coraciimorphae+Australaves																				

(b)

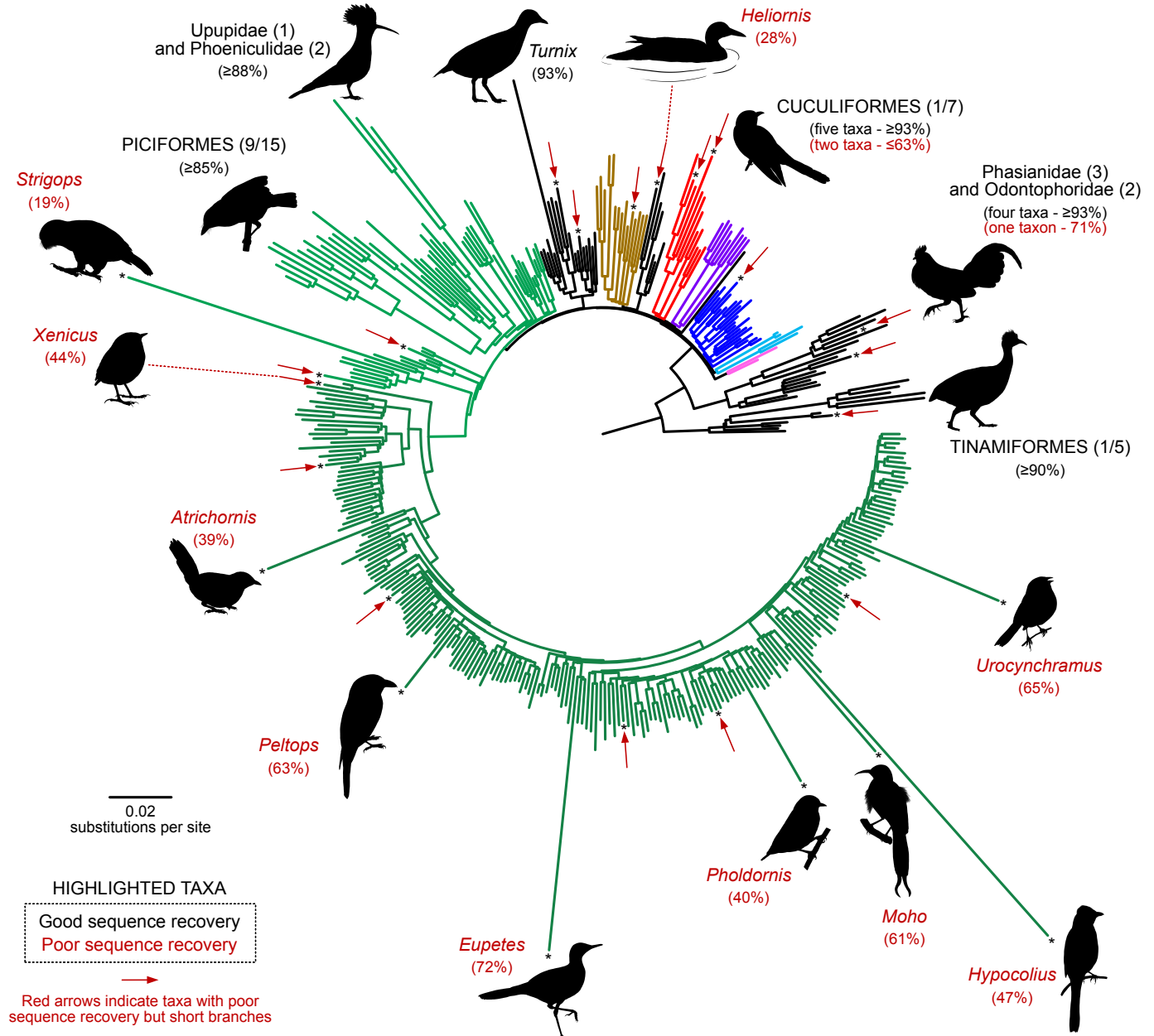
Tree	Non-monophyletic	
	Orders	Families
ML concatenation		
IQ-TREE	0	0
MSC methods		
weighted ASTRAL	0	1
weighted ASTRID	0	0
SVD quartets (sub)	3	0
SVD quartets (ex)	3	0
METAL (NJ)		
p-distances	0	0
ML distances	0	0
Trace distances	0	0
logdet i.r.	0	0
logdet-inv	0	0
logdet inf sites	0	0

(c)

	Clade	METAL (NJ)																				
		ML concat (IQ-TREE)	ML concat (ExAML)	weighted ASTRAL	weighted ASTRID	SVD quartets	p-distances	ML distances	Trace distances	logdet i.r.	logdet-inv	logdet inf sites										
Easy (+++)	Tyranni (suboscines)																					
	Passeri (oscines)																					
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Medium (++)	Eurylaimides																					
	Tyrannides																					
	Furnariida																					
	Tyrannida																					
	Malaconotoidea																					
	Corvoidea																					
	Sylviida																					
	Sylvioidea																					
	Muscicapida																					
	Bombycilloidea																					
	Muscicapoidae																					
Certhioidea																						
Passerida																						
Emberizoidea																						
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hard (+)	Corvides																					
	Orioloidea																					
	Passerides																					
	Aegithaloidea																					
			0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Locustelloidea																						

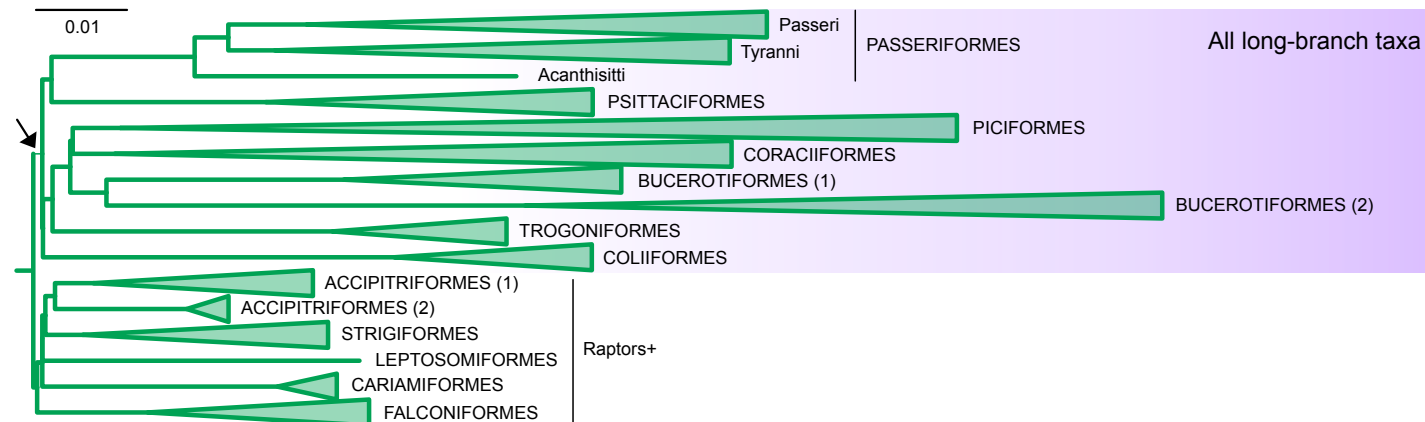
LONG BRANCH ATTRACTION IN "LANDBIRDS"? (THE GREEN CLADE)

- Note the branch length heterogeneity -- even if we ignore the artifactual long branches

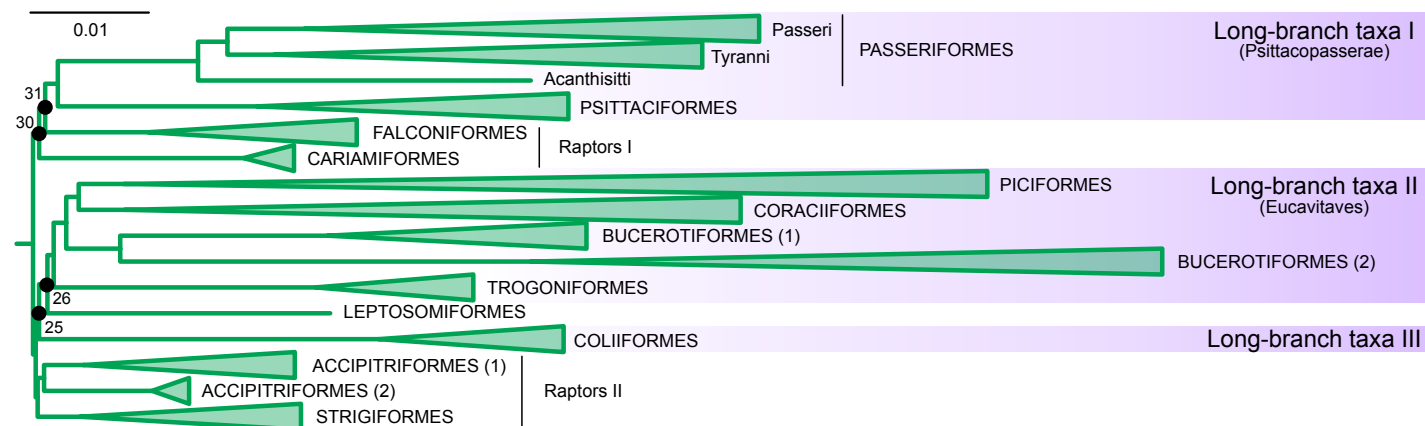


IS METAL PRONE TO LONG BRANCH ATTRACTION?

(a) NJ of logdet inf sites distances



(b) ML concatenation



- Long branch taxa clustered in NJ analysis of “landbirds” (Telluraves)

LARGER DATASET

-- UCE data from Jarvis et al. (2014)

■ Fewer loci (3158) but longer flanks

■ 9.2 million sites

■ 4.5x the number of informative sites as allfam

-- Pink clades are non-monophyletic if long-branch taxa in landbirds are clustered

	#	Name	Jarvis et al. (2014)					METAL (NJ)					
			ML concat (IQ-TREE)	weighted ASTRAL	weighted ASTRID	SVD quartets	p-distances	ML distances	Trace distances	logdet i.r.	logdet-inv	logdet inf sites	
Easy (+++)	1	PALAEOGNATHAE											
	4	NEOGNATHAE											
	5	Galloanserae											
	6	Neoaves											
	7	Mirandornithes (VII)											
	33	Passeriformes											
	34	Eupasserres											
		# non-monophyletic in allfam	0	0	0	0	0	0	0	0	0	0	
			-	0	0	0	1	0	0	0	0	0	
Medium (++)	12	Strisores (V)											
	19	Aequornithes (II)											
	20	Feraequornithes											
	21	Procellariimorphae											
	24	Telluraves (I)											
	26	Cavitaves											
	27	Eucavitaves											
	28	Picocoraciae											
	30	Australaves											
	31	Eufalconimorphae											
32	Psittacopasserae												
		# non-monophyletic in allfam	0	0	0	0	4	6	2	2	4	4	1
			-	0	1	0	4	3	3	3	3	3	3
Hard (+)	8	Columbimorphae (VI)											
	9	Pteroclimesites											
	11	Musophagotides											
	17	Phaethoquornithes											
	18	Phaethontimorphae (III)											
	23	Pelecanes											
	25	Coraciimorphae											
29	Picodynastornithes												
		# non-monophyletic in allfam	2	1	0	0	5	5	2	2	3	3	1
			-	1	4	2	4	5	4	3	5	4	3
Uncertain (-)	10	Otidimorphae (IV)											
	-	Afroaves											
	-	Accipitriformes+Strigiformes											
	-	Coraciimorphae+Australaves											

LARGER DATASET

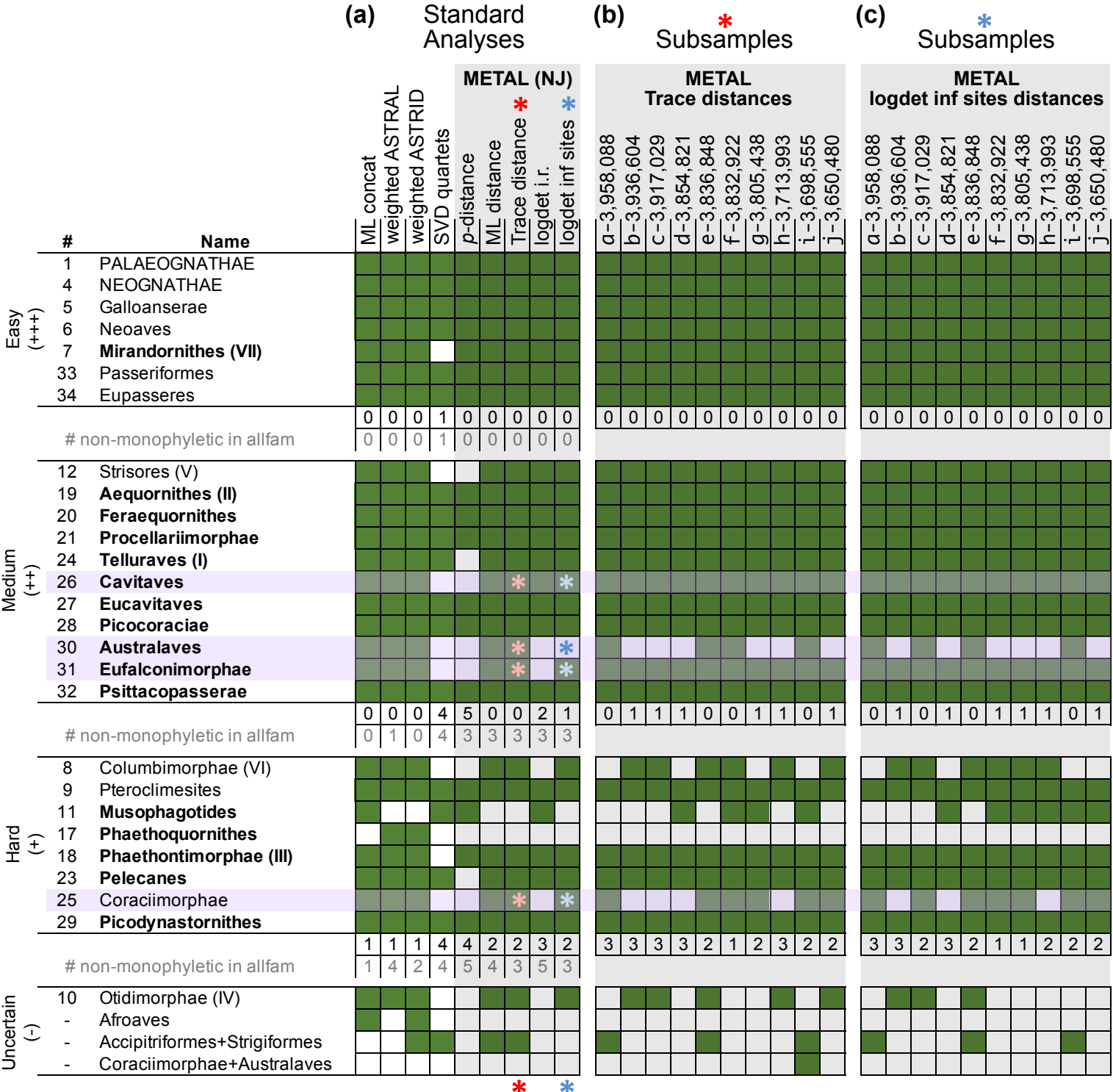
-- Jarvis et al. (2014) intronic data
 -- 3x more parsimony informative sites than Jarvis UCEs

	#	Name	METAL (NJ)																																																																																																																																																																																																																																																																																																																																			
			ML concat	weighted ASTRAL	weighted ASTRID	SVD quartets	p-distance	ML distance	Trace distance *	logdet i.r.	logdet inf sites *																																																																																																																																																																																																																																																																																																																											
Easy (+++)	1	PALAEOGNATHAE																																																																																																																																																																																																																																																																																																																																				
	4	NEOGNATHAE																																																																																																																																																																																																																																																																																																																																				
	5	Galloanserae																																																																																																																																																																																																																																																																																																																																				
	6	Neoaves																																																																																																																																																																																																																																																																																																																																				
	7	Mirandornithes (VII)				■																																																																																																																																																																																																																																																																																																																																
	33	Passeriformes																																																																																																																																																																																																																																																																																																																																				
34	Eupasserres										# non-monophyletic in allfam			0	0	0	1	0	0	0	0	0	Medium (++)	12	Strisores (V)				■	■					19	Aequornithes (II)										20	Feraequornithes										21	Procellariimorphae										24	Telluraves (I)					■					26	Cavitaves							*	*		27	Eucavitaves										28	Picocoraciae										30	Australaves							*	*		31	Eufalconimorphae							*	*		32	Psittacopasserae										# non-monophyletic in allfam			0	0	0	4	5	0	0	2	1	# non-monophyletic in allfam			0	1	0	4	3	3	3	3	3	Hard (+)	8	Columbimorphae (VI)				■	■					9	Pteroclimesites										11	Musophagotides			■							17	Phaethoquornithes										18	Phaethontimorphae (III)										23	Pelecanes										25	Coraciimorphae							*	*		29	Picodynastornithes										# non-monophyletic in allfam			1	1	1	4	4	2	2	3	2	# non-monophyletic in allfam			1	4	2	4	5	4	3	5	3	Uncertain (-)	10	Otidimorphae (IV)				■	■					-	Afroaves										-	Accipitriformes+Strigiformes										-	Coraciimorphae+Australaves									
# non-monophyletic in allfam			0	0	0	1	0	0	0	0	0																																																																																																																																																																																																																																																																																																																											
Medium (++)	12	Strisores (V)				■	■																																																																																																																																																																																																																																																																																																																															
	19	Aequornithes (II)																																																																																																																																																																																																																																																																																																																																				
	20	Feraequornithes																																																																																																																																																																																																																																																																																																																																				
	21	Procellariimorphae																																																																																																																																																																																																																																																																																																																																				
	24	Telluraves (I)					■																																																																																																																																																																																																																																																																																																																															
	26	Cavitaves							*	*																																																																																																																																																																																																																																																																																																																												
	27	Eucavitaves																																																																																																																																																																																																																																																																																																																																				
	28	Picocoraciae																																																																																																																																																																																																																																																																																																																																				
	30	Australaves							*	*																																																																																																																																																																																																																																																																																																																												
	31	Eufalconimorphae							*	*																																																																																																																																																																																																																																																																																																																												
32	Psittacopasserae										# non-monophyletic in allfam			0	0	0	4	5	0	0	2	1	# non-monophyletic in allfam			0	1	0	4	3	3	3	3	3	Hard (+)	8	Columbimorphae (VI)				■	■					9	Pteroclimesites										11	Musophagotides			■							17	Phaethoquornithes										18	Phaethontimorphae (III)										23	Pelecanes										25	Coraciimorphae							*	*		29	Picodynastornithes										# non-monophyletic in allfam			1	1	1	4	4	2	2	3	2	# non-monophyletic in allfam			1	4	2	4	5	4	3	5	3	Uncertain (-)	10	Otidimorphae (IV)				■	■					-	Afroaves										-	Accipitriformes+Strigiformes										-	Coraciimorphae+Australaves																																																																																																																																															
# non-monophyletic in allfam			0	0	0	4	5	0	0	2	1																																																																																																																																																																																																																																																																																																																											
# non-monophyletic in allfam			0	1	0	4	3	3	3	3	3																																																																																																																																																																																																																																																																																																																											
Hard (+)	8	Columbimorphae (VI)				■	■																																																																																																																																																																																																																																																																																																																															
	9	Pteroclimesites																																																																																																																																																																																																																																																																																																																																				
	11	Musophagotides			■																																																																																																																																																																																																																																																																																																																																	
	17	Phaethoquornithes																																																																																																																																																																																																																																																																																																																																				
	18	Phaethontimorphae (III)																																																																																																																																																																																																																																																																																																																																				
	23	Pelecanes																																																																																																																																																																																																																																																																																																																																				
25	Coraciimorphae							*	*																																																																																																																																																																																																																																																																																																																													
29	Picodynastornithes										# non-monophyletic in allfam			1	1	1	4	4	2	2	3	2	# non-monophyletic in allfam			1	4	2	4	5	4	3	5	3	Uncertain (-)	10	Otidimorphae (IV)				■	■					-	Afroaves										-	Accipitriformes+Strigiformes										-	Coraciimorphae+Australaves																																																																																																																																																																																																																																																																
# non-monophyletic in allfam			1	1	1	4	4	2	2	3	2																																																																																																																																																																																																																																																																																																																											
# non-monophyletic in allfam			1	4	2	4	5	4	3	5	3																																																																																																																																																																																																																																																																																																																											
Uncertain (-)	10	Otidimorphae (IV)				■	■																																																																																																																																																																																																																																																																																																																															
	-	Afroaves																																																																																																																																																																																																																																																																																																																																				
	-	Accipitriformes+Strigiformes																																																																																																																																																																																																																																																																																																																																				
	-	Coraciimorphae+Australaves																																																																																																																																																																																																																																																																																																																																				

-- METAL has relatively good performance
 -- Best distances are:
 ■ GTR+G
 ■ logdet of inf sites
 ■ Indicated with asterisks

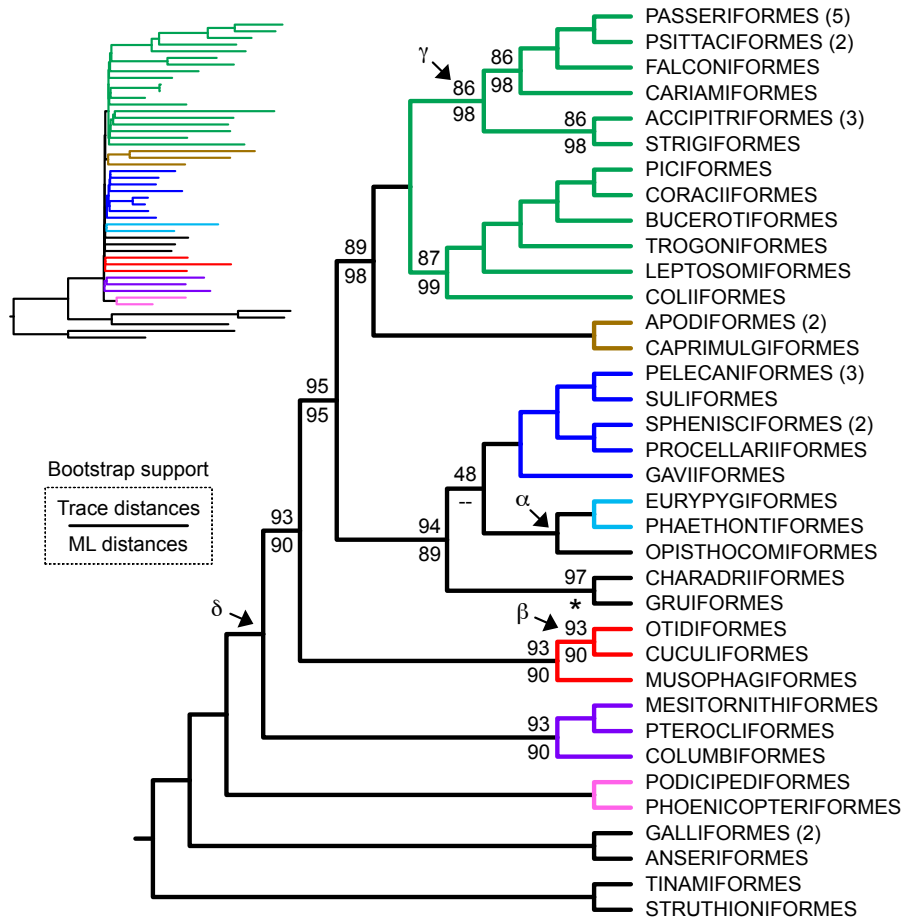
LARGER DATASET

-- Subsampling to yield datasets comparable to the Jarvis UCE dataset yield results that resemble the Jarvis UCE results

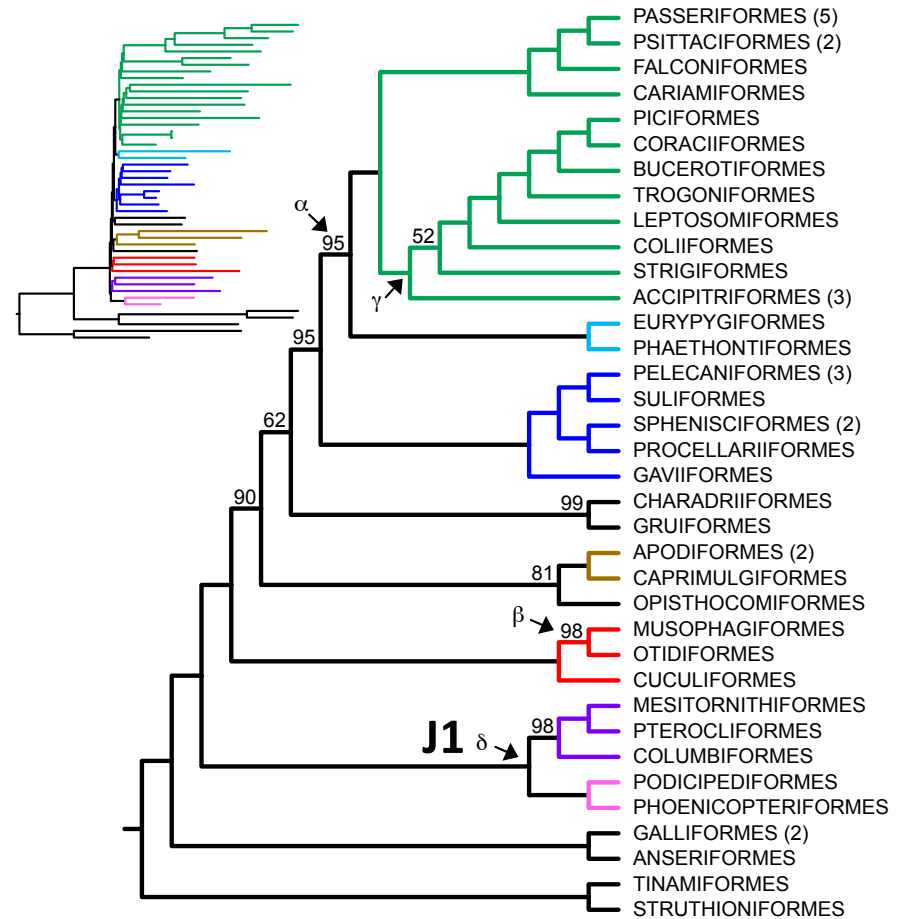


METAL ANALYSIS OF INTRON DATASET

(a) NJ of GTR+ Γ distances



(b) ML concatenation (IQ-TREE)



- METAL does not recover “J1” (Neoaves root matches Stiller et al. 2024)
- ML concatenation does recover J1

THANKS!

- Rebecca Kimball and Kimball-Braun lab
 - Min Zhao, Dawson Houghtaling
- “OpenWings” (UCE data collection)
 - Brian Smith, Brant Faircloth, Robb Brumfield, Carl Oliveros
- Many other collaborators
 - Especially the B10K consortium
 - Siavash Mirarab, Chao Zhang, Josefin Stiller, Mike Braun, Jake Berv
- NSF grant DEB-1655683



OPEN QUESTIONS

- What is the best way to deal with variation over time in the recombination landscape?
- Are there principled ways to address problems caused by misassembly?
 - Including the interaction of incorrect assembly and alignment
- Can the performance of METAL be improved?
 - Reduce data requirement (generalized least squares?)
 - Improve the selection of distance metrics
 - Deal with among-sites rate heterogeneity