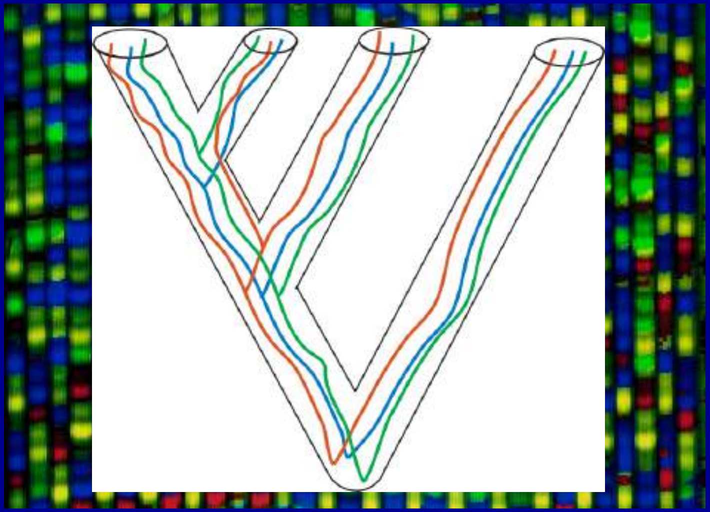
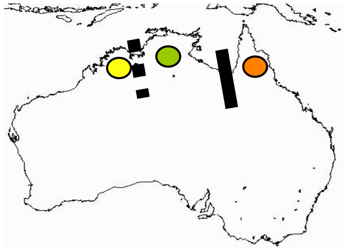
The continuum between phylogeography and phylogenetics

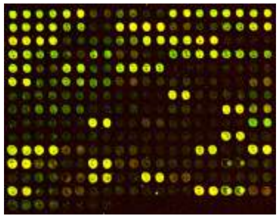




Scott V. Edwards
Department of Organismic and Evolutionary Biology
Harvard University
Cambridge, MA USA
https://edwards.oeb.harvard.edu/

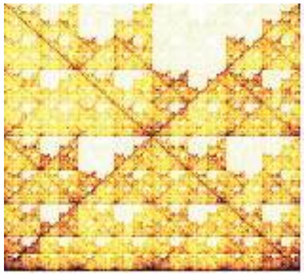
Edwards lab – birds, genomics and evolution



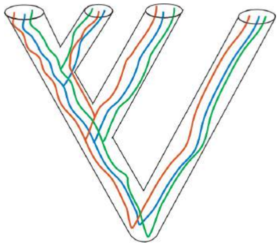
Phylogeography



Ecological Genomics



Comparative Genomics



Phylogenetic Methods

May 2025

I am a scientist, too!

Sunset Beach, OR

Plum Island, MA

76 days
3848 miles

@ScottVEdwards1

Adventure Cycling Association

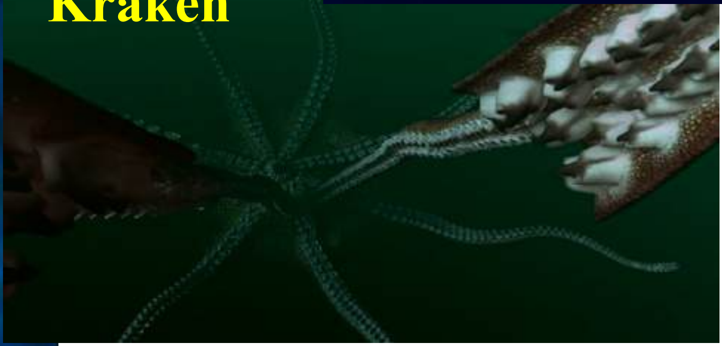
Beast Legends – Griffin episode

yap films off the leash tv

Griffin

Fijian shark god

Wild man



Kraken

Dragon

Terror

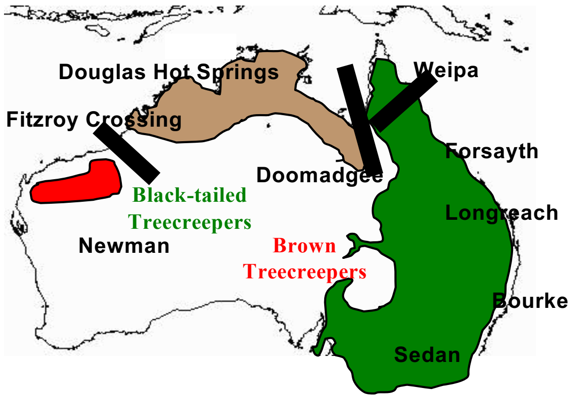
Graphics by Invisible Pictures, Inc

A 1980s conundrum in evolutionary biology

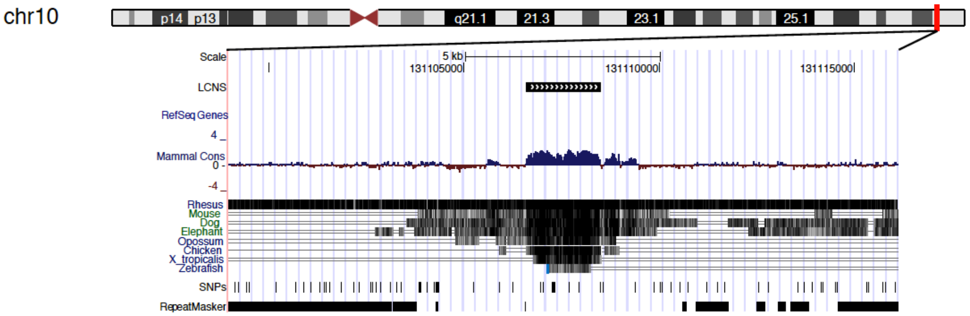
Oct. 2021,
Seattle

Talk overview

Part I: Reticulation and the emerging continuum between phylogeography and phylogenetics



Part II: PhyloG2P – Connecting genomes and phenotypic traits using phylogenies



Part III: Pangenomes: the future of evolutionary genomics

Phylogeographic case studies

from the wilds of Australia

Babblers

PhyloG2P

Paleognaths

Grassfinches

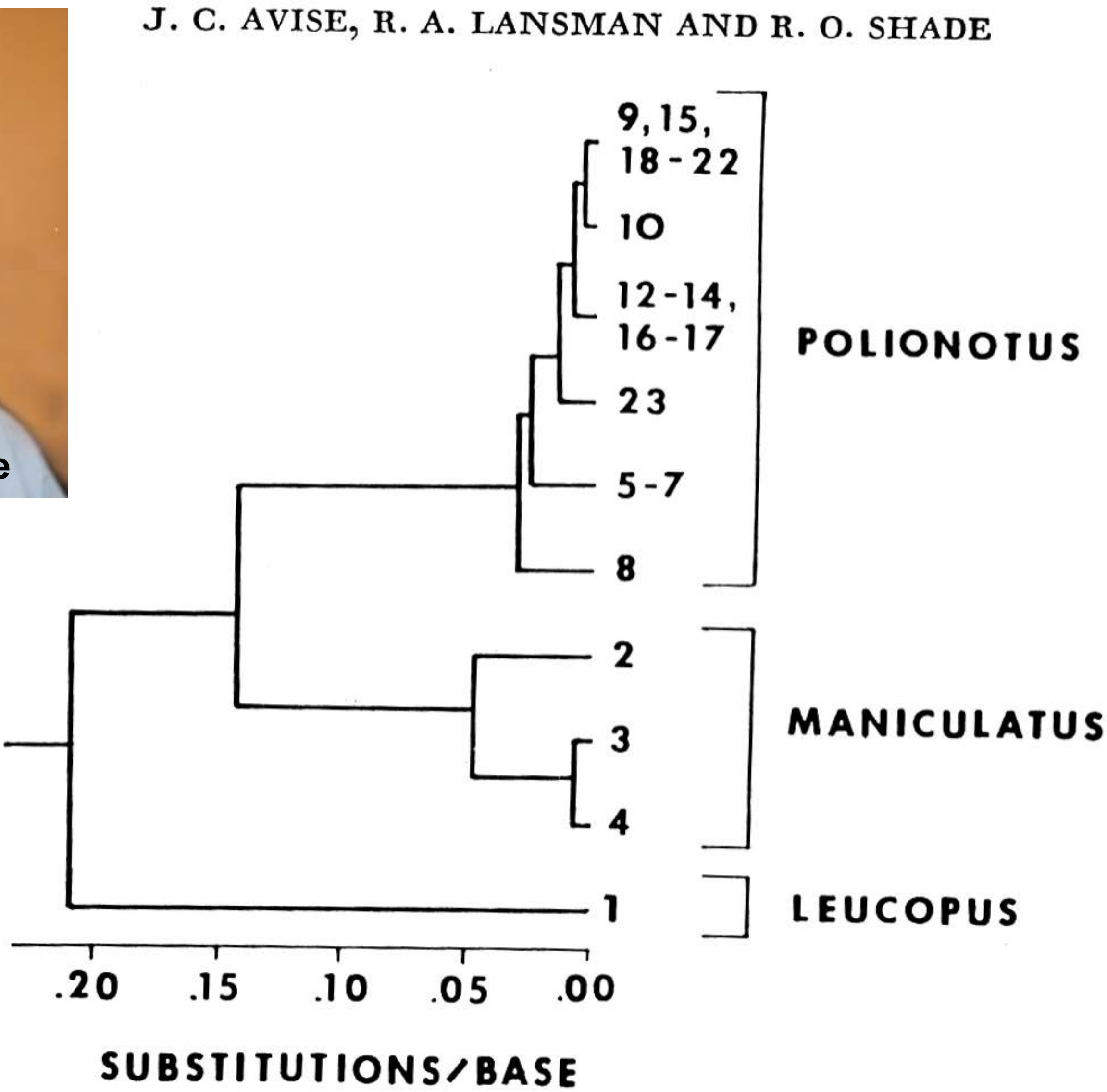
Honeyeaters

Treecreepers

Pangenomes

Scrub Jays

The first ‘gene tree’, 1979

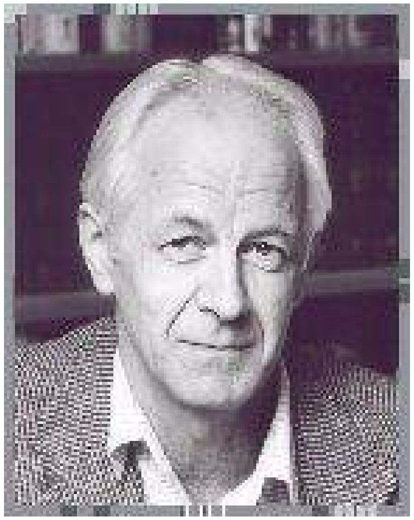


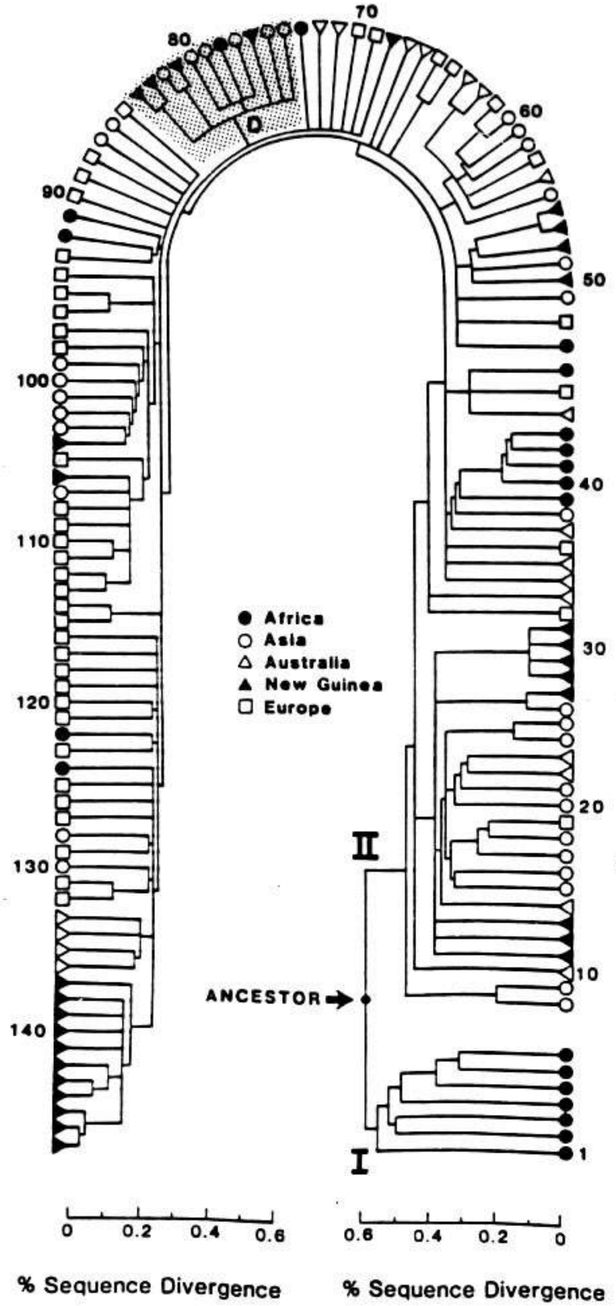
Reprinted from Nature, Vol. 325, No. 6099, pp. 31-36, 1 January 1987

Mitochondrial DNA and human evolution

Rebecca L. Cann*, Mark Stoneking & Allan C. Wilson

Department of Biochemistry, University of California, Berkeley, California 94720, USA

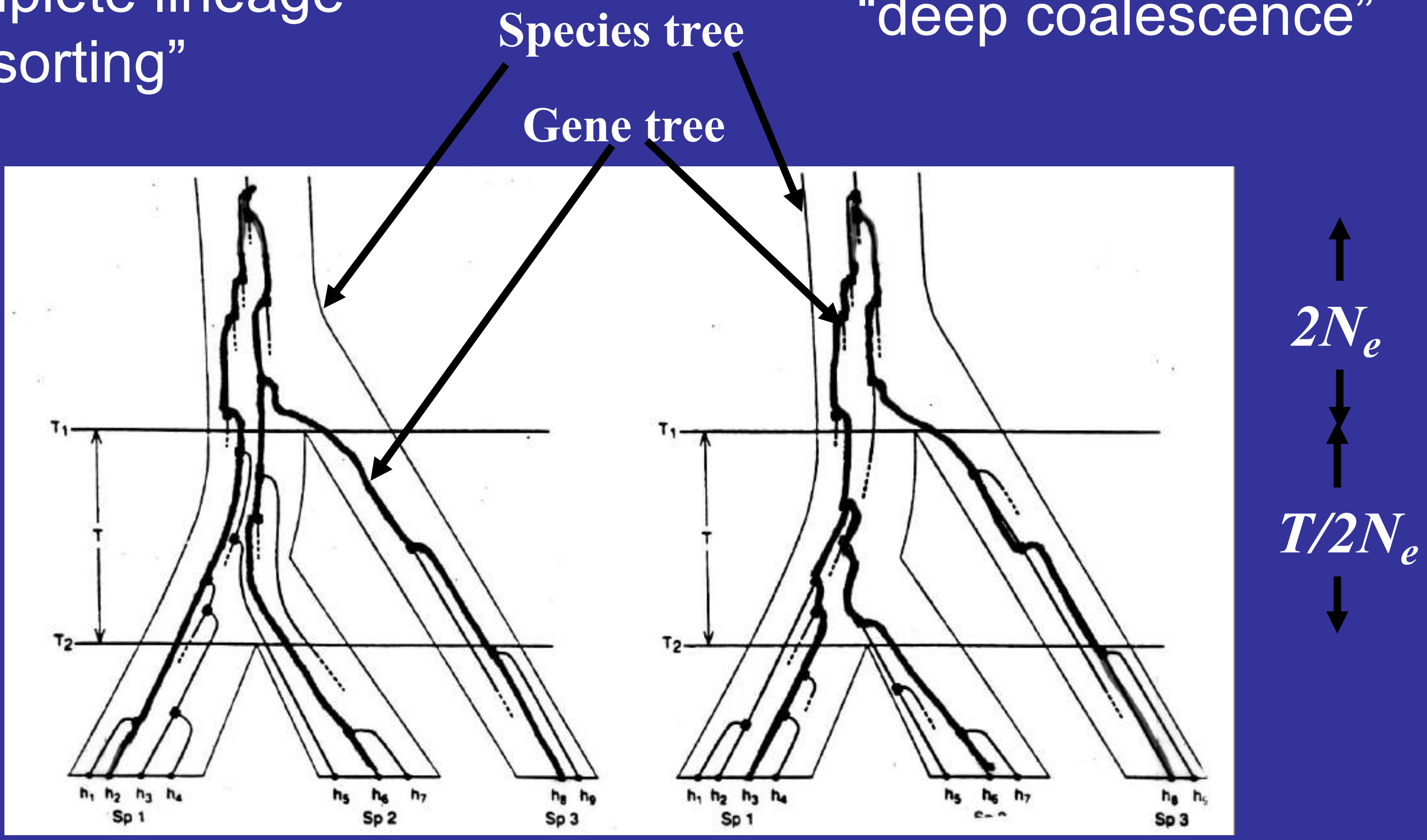




Stochastic gene tree variation during rapid radiations

“incomplete lineage sorting”

“deep coalescence”



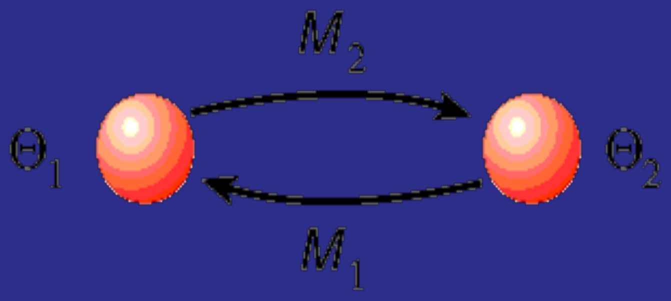
Species tree

Gene tree

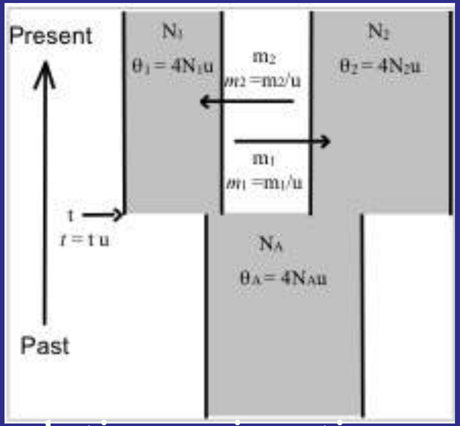
Hudson RR.1990. *Oxford Surv. Evol. Biol.* 7: 1-45

Multilocus models in phylogeography

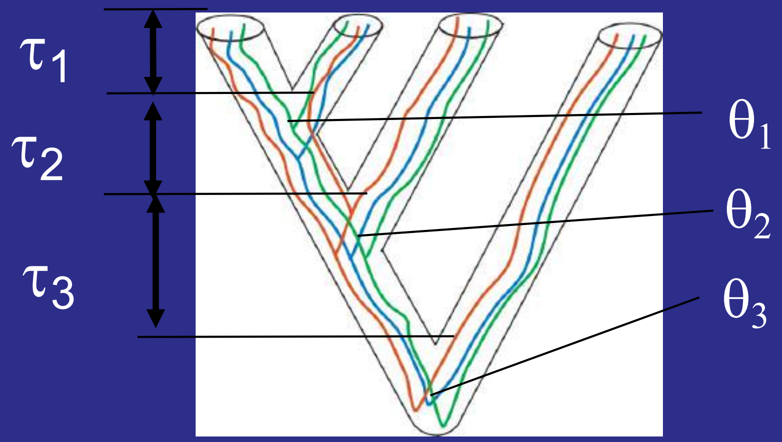
Population size: $\theta = 4N\mu$ Divergence time: $\tau = \mu t$ Gene flow: $M = m/\mu$



equilibrium migration model
MIGRATE: Beerli 2006 *Bioinformatics*



isolation-migration model
IM: Hey and Nielsen 2004 *Genetics*

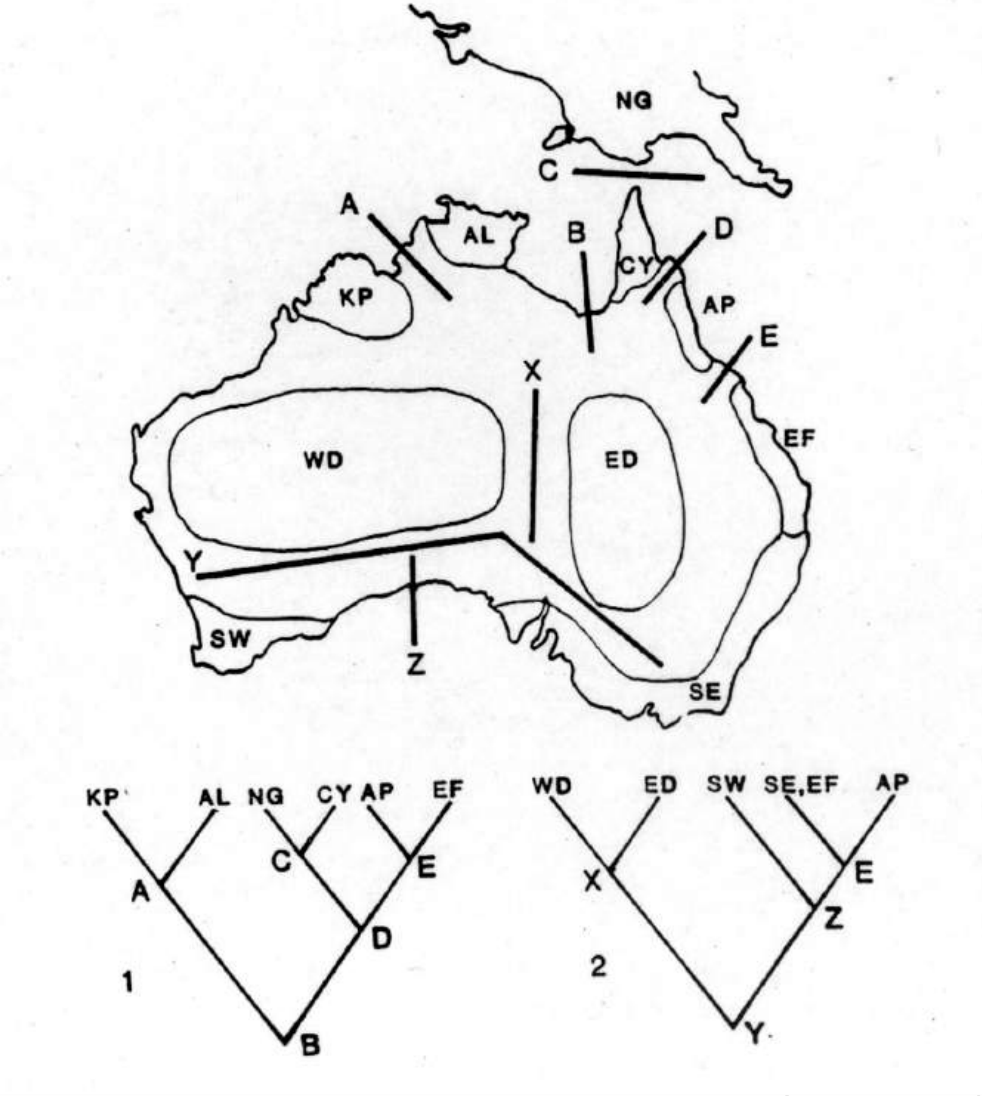


pure isolation
(phylogeny)
BEST: Liu and Pearl. 2007. *Syst. Biol*
BEAST*: Heled and Drummond. 2010. *Mol. Biol. Evol.*

~~Warm~~ welcome in the outback…

Stinking hot

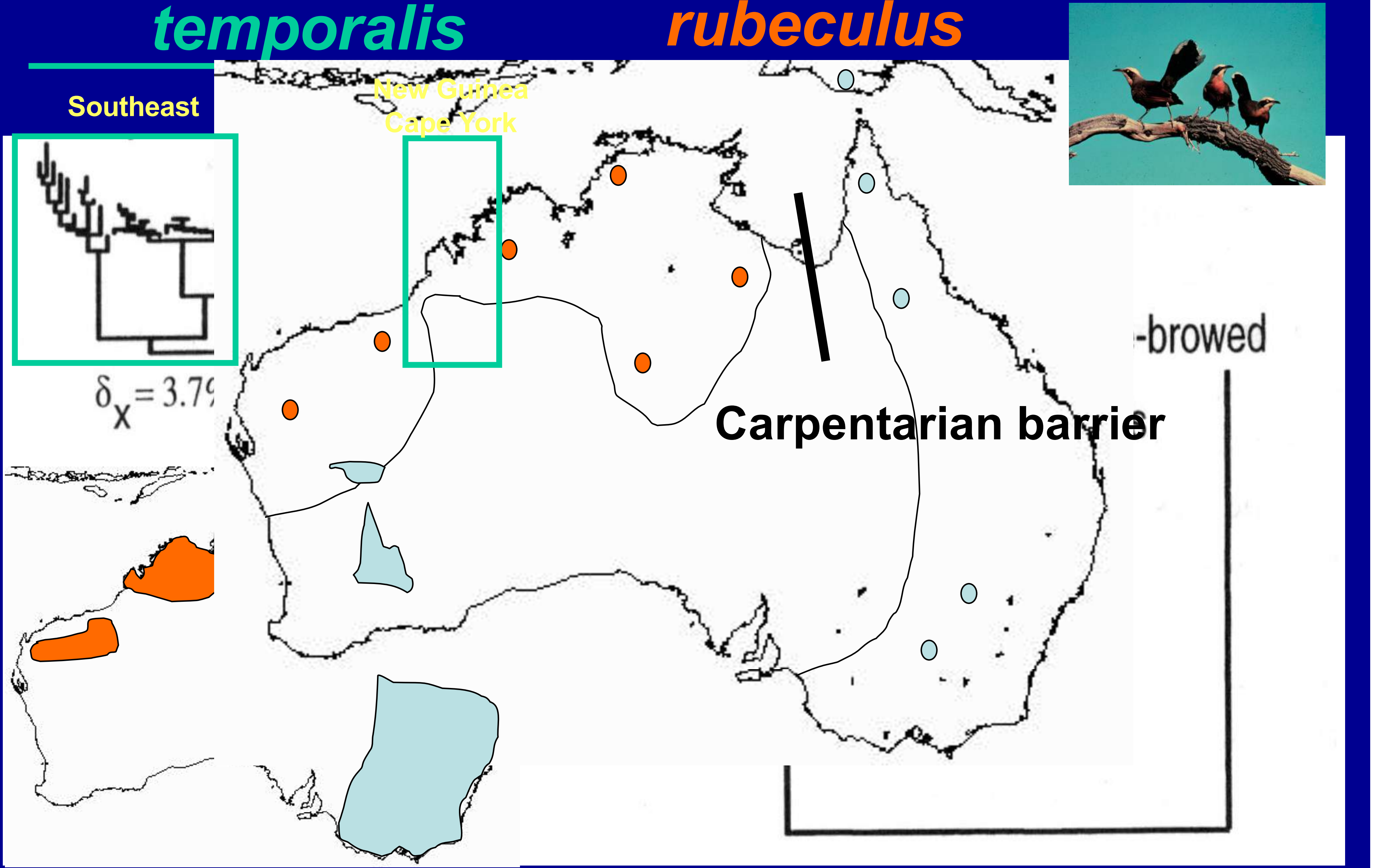
Carpentarian barrier (B) is deepest split in area cladograms of Australian biota



Cracraft 1986. *Evolution*

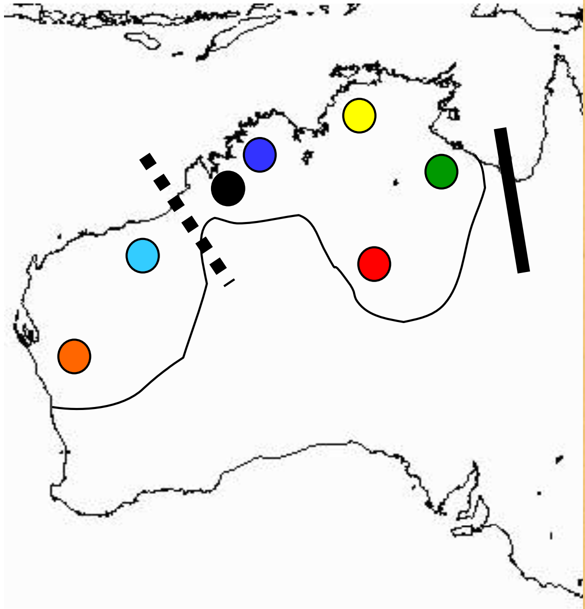
Mitochondrial gene tree for grey-crowned babblers

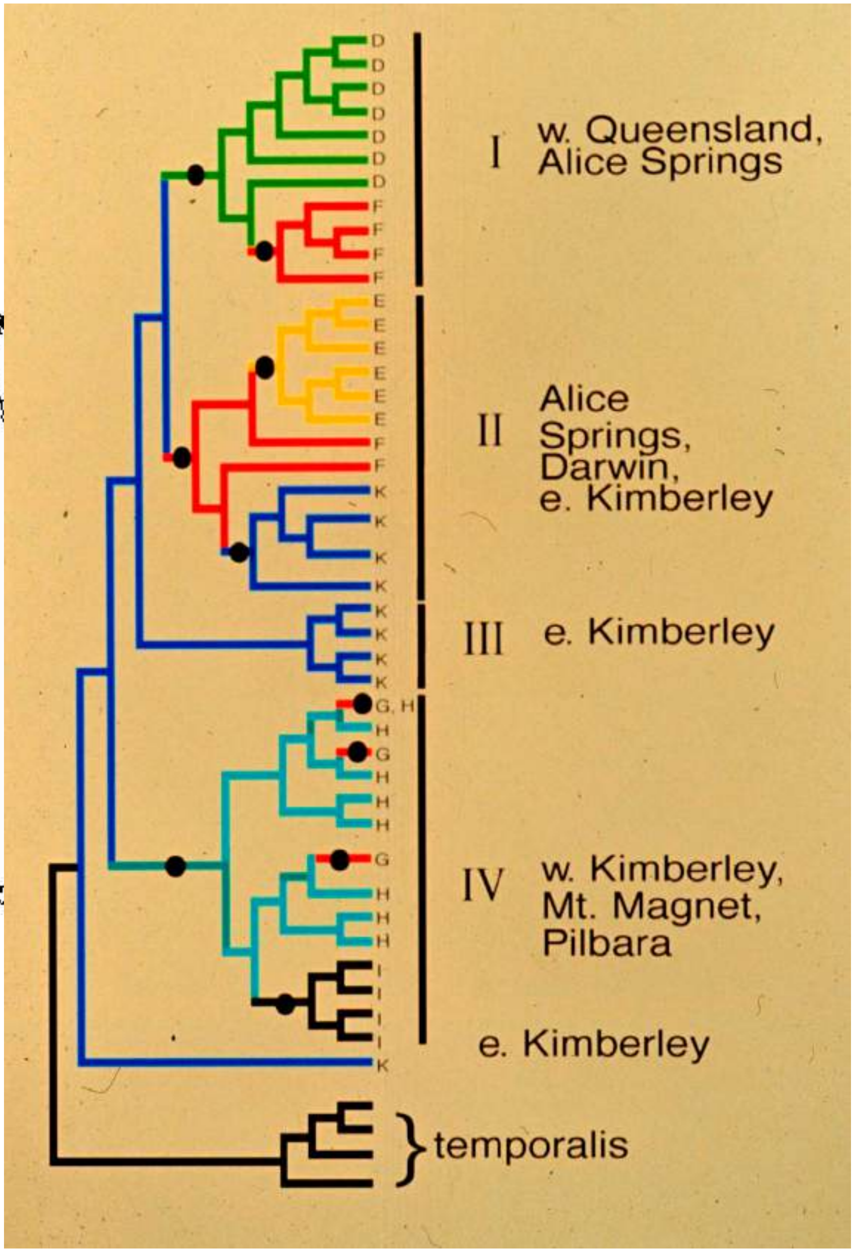
Sampling localities for grey-crowned babblers



Edwards, S. V. (1993) *Proc. R. Soc. Lond. B* 252, 177-185.

Gene flow erodes population monophyly





● Migration event reconstructed by parsimony

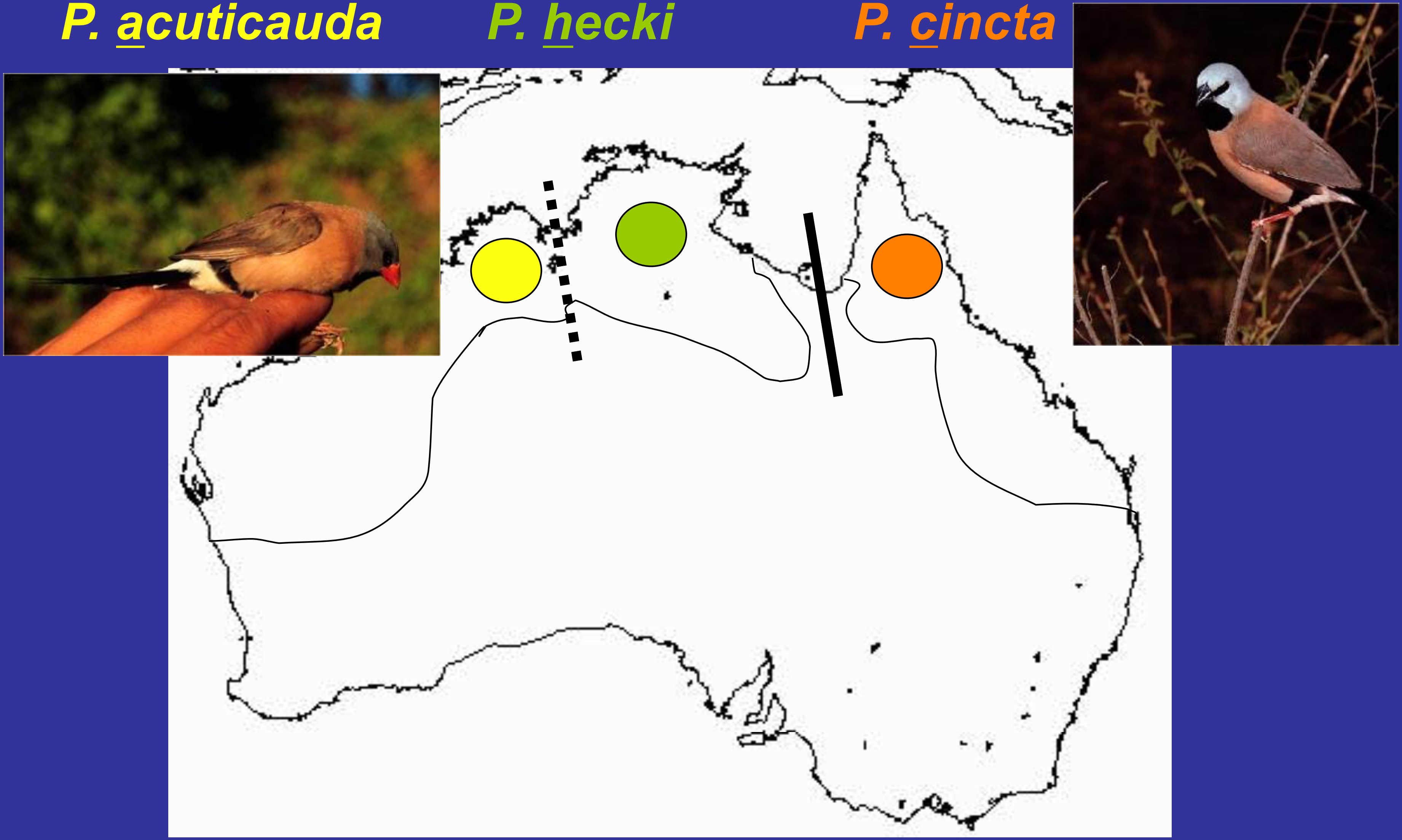
Bayesian approach: Advantages for estimating gene flow

- Parsimony approach
  - Assumes complete certainty of gene tree
  - Statistical testing of different hypotheses of gene flow is cumbersome
  - Confidence limits reflect only uncertainty in *s*

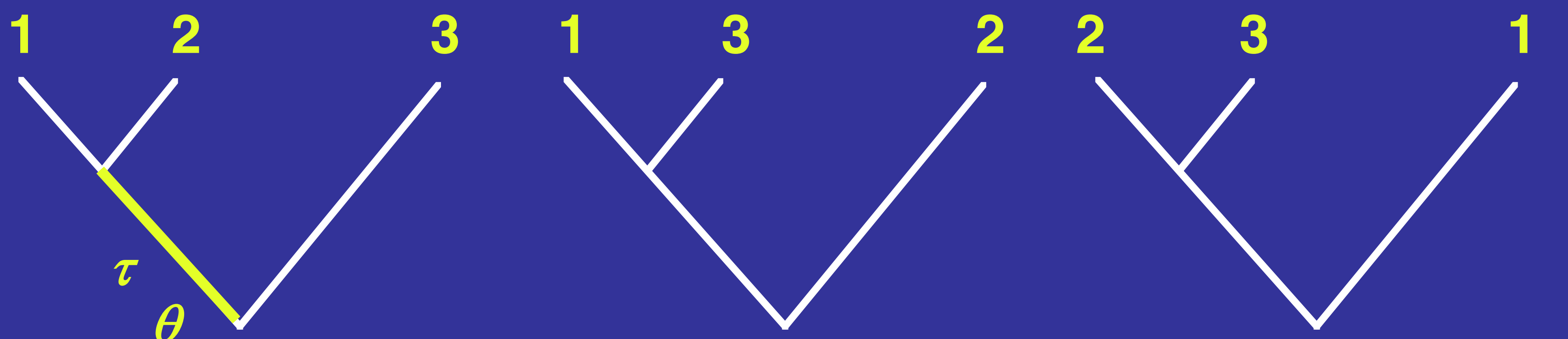
- Bayesian approach (migrate-n)
  - Integrates the two sources of uncertainty
    - Accommodates uncertainty in trees by integrating over all trees
    - Accommodates the range of *Nm* given these trees
  - Hypothesis testing easier

East-west pairs -- grassfinches (*Poephila*)

P. acuticauda ***P. hecki*** ***P. cincta***



Probability of gene trees in a rooted 3-tip species tree

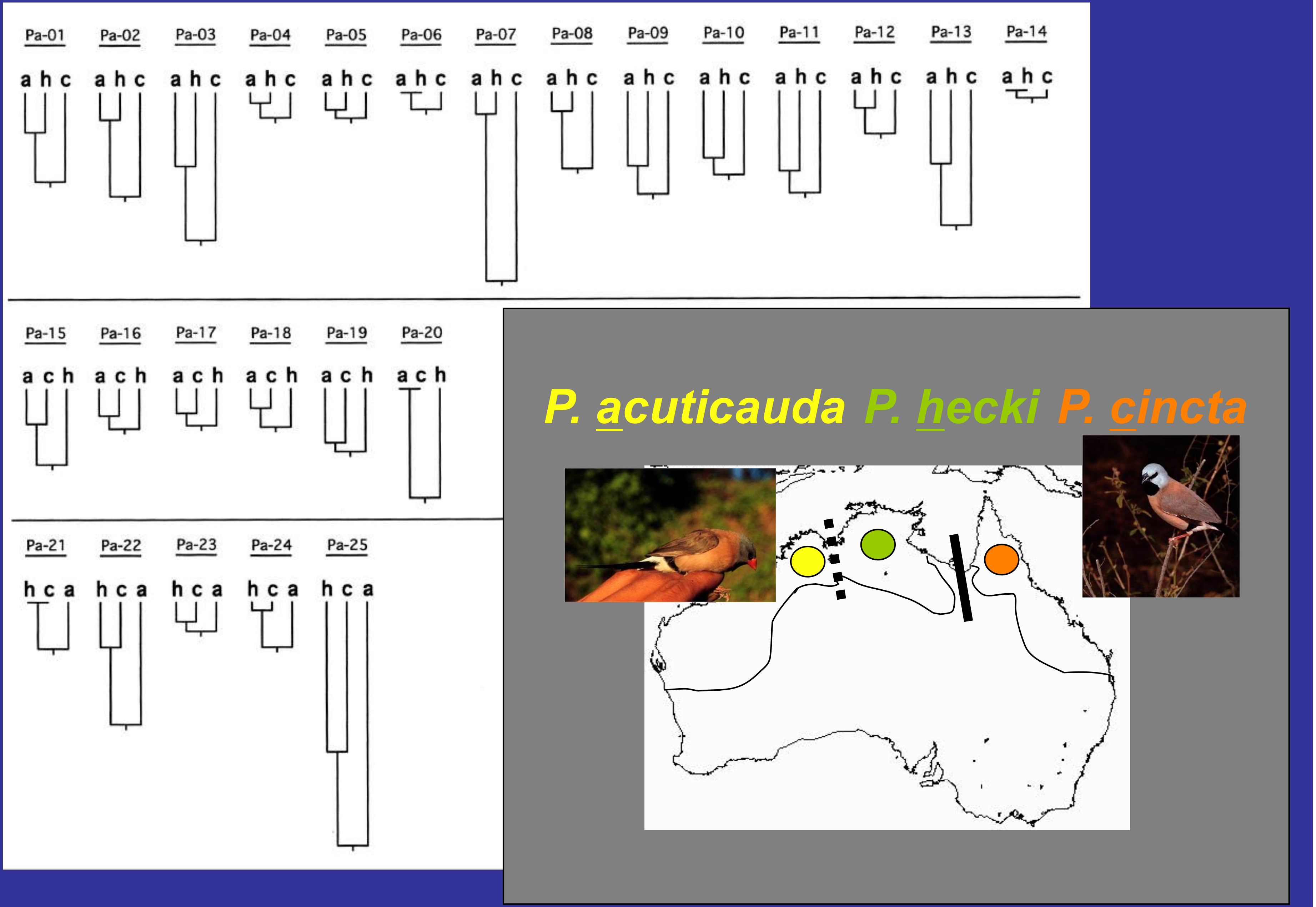


congruent gene tree **incongruent gene tree** **incongruent gene tree**

$$P_{incongruence} = (2/3)e^{-\tau/\theta}$$

Pamilo and Nei 1988. Mol. Biol. Evol. 5: 568-583
Liu, et al. 2010. BMC Evolutionary Biology 10:302

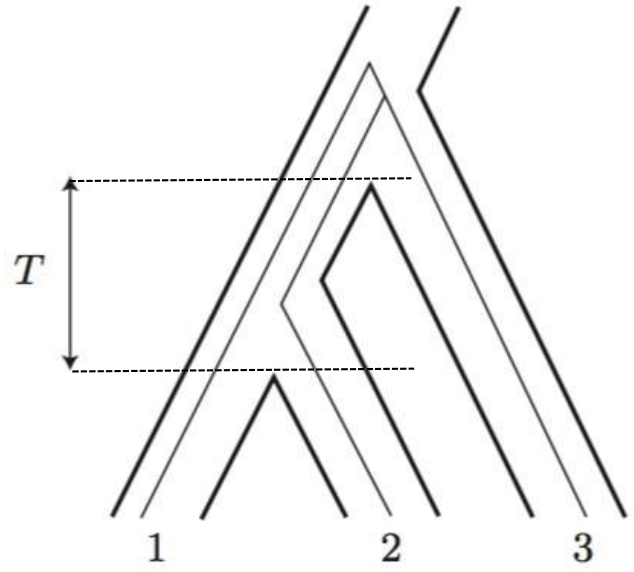
30 gene trees from Australian finches

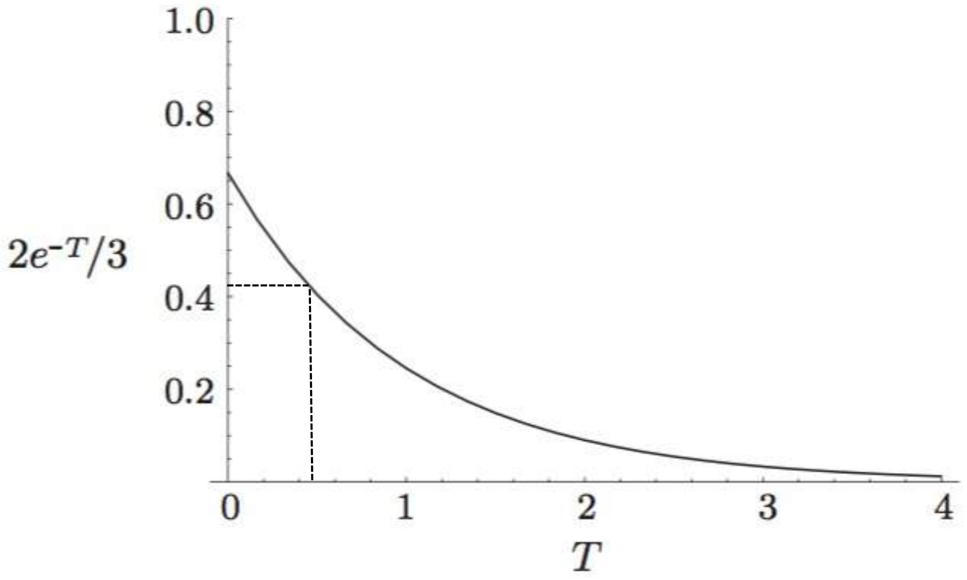


P. acuticauda P. hecki P. cincta

Jennings & Edwards (2005) *Evolution* 59, 2033-2047

Probability of discordance between gene tree and species tree





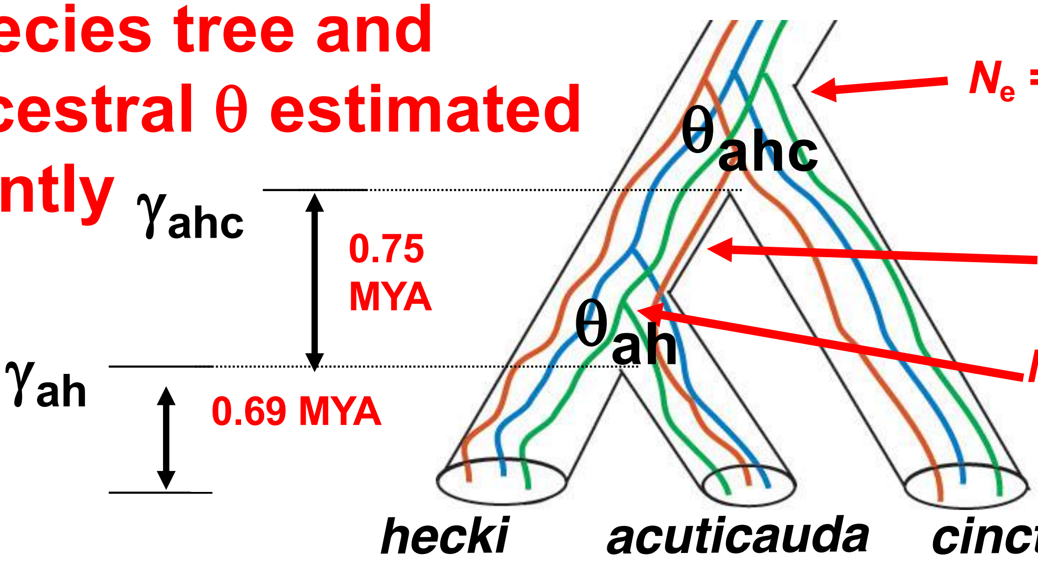
***T* = internode length = 2μt/4Nμ = t/2N generations long**

$$P\{\text{discordant}\} \quad = \quad \frac{2}{3}e^{-T}$$

= 12/28

Species tree of *Poephila* grassfinches

Species tree and ancestral θ estimated jointly



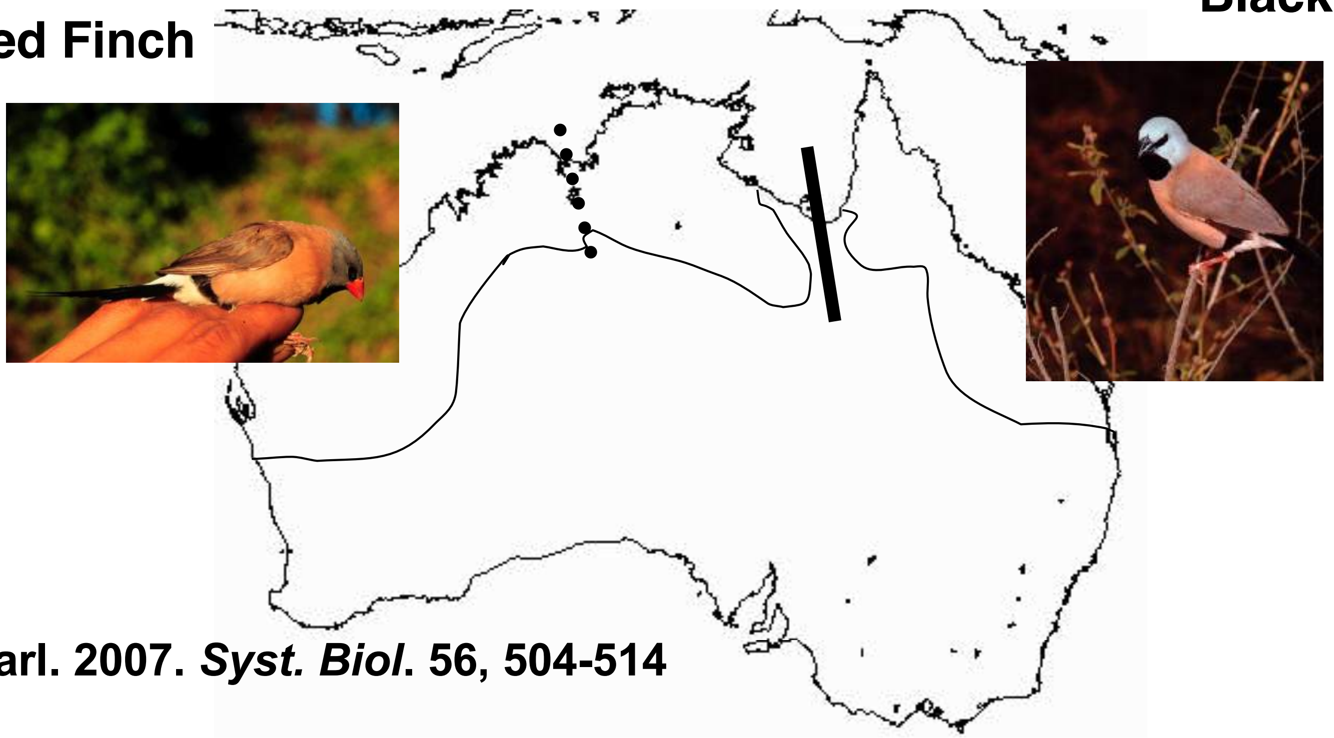
N_e = ~350,000

Posterior probability of branch in species tree= 0.88

N_e = ~200,000

Long-tailed Finch

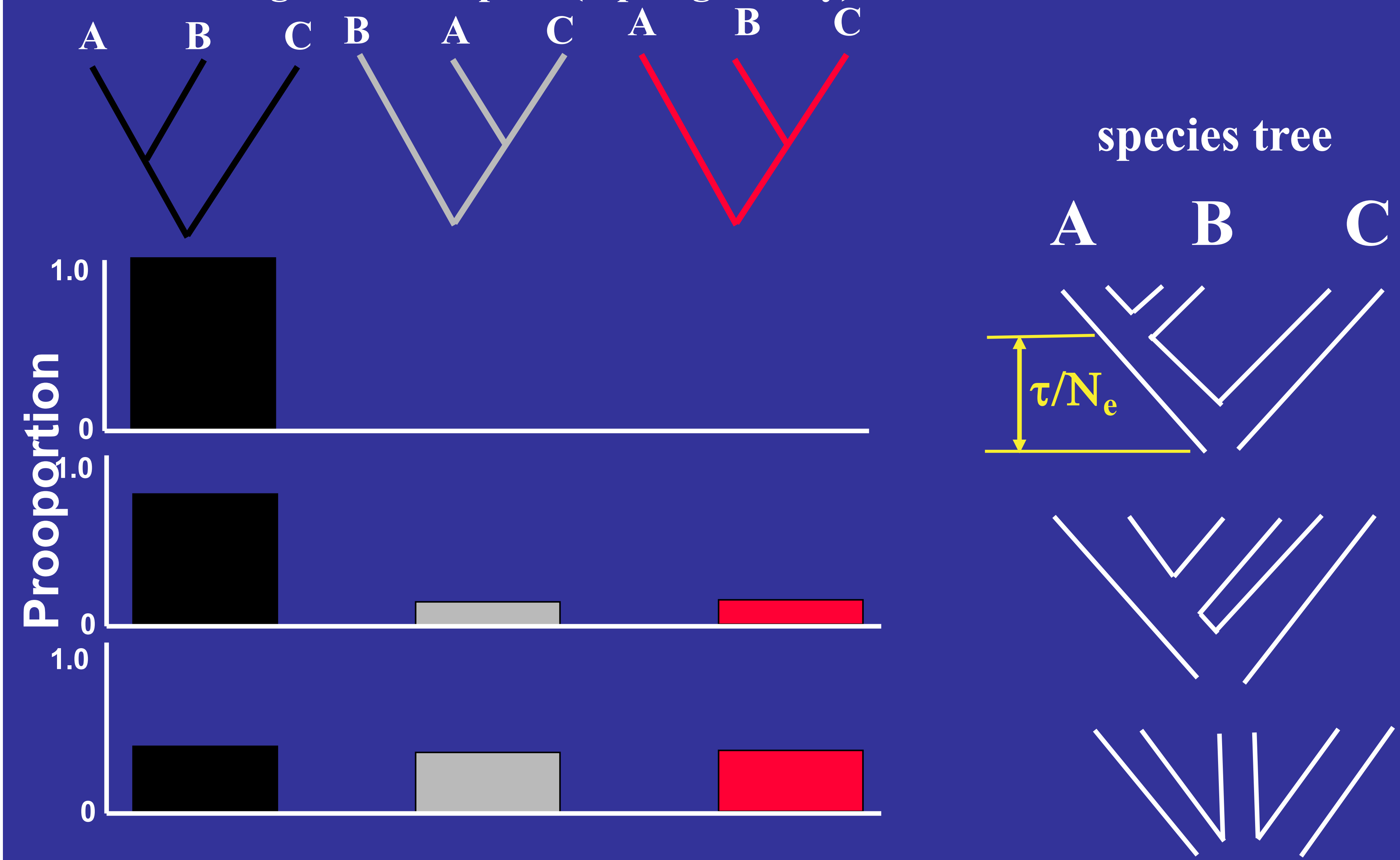
Black-throated Finch



Liu and Pearl. 2007. *Syst. Biol.* 56, 504-514

Maximum (pseudo) likelihood method for species trees

Rooted gene tree triplets (topologies only)

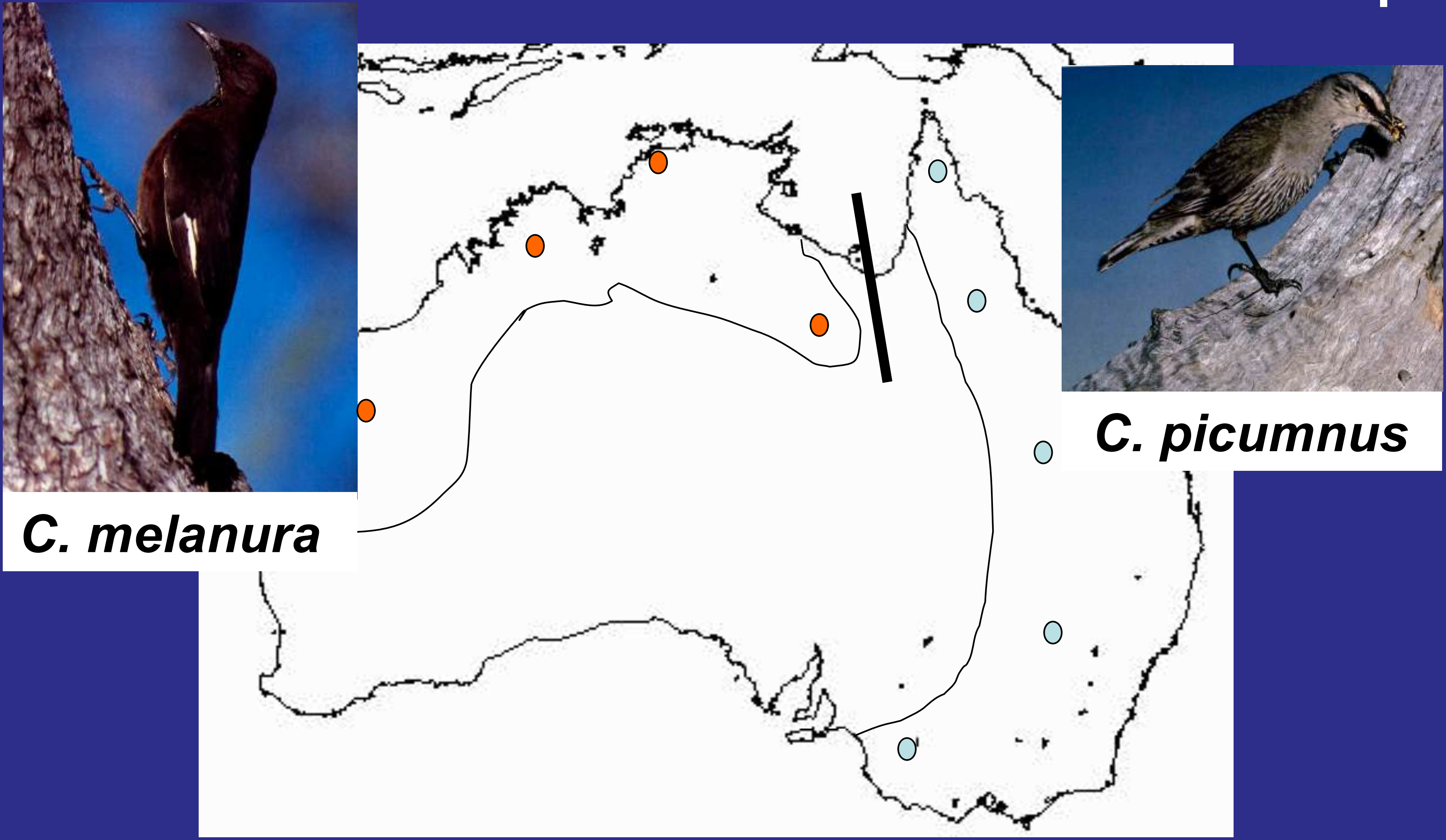


Liu, Li & Edwards. 2010. *BMC Evol. Biol.*

Treecreepers (*Climacteris*)

Black-tailed treecreeper

Brown treecreeper



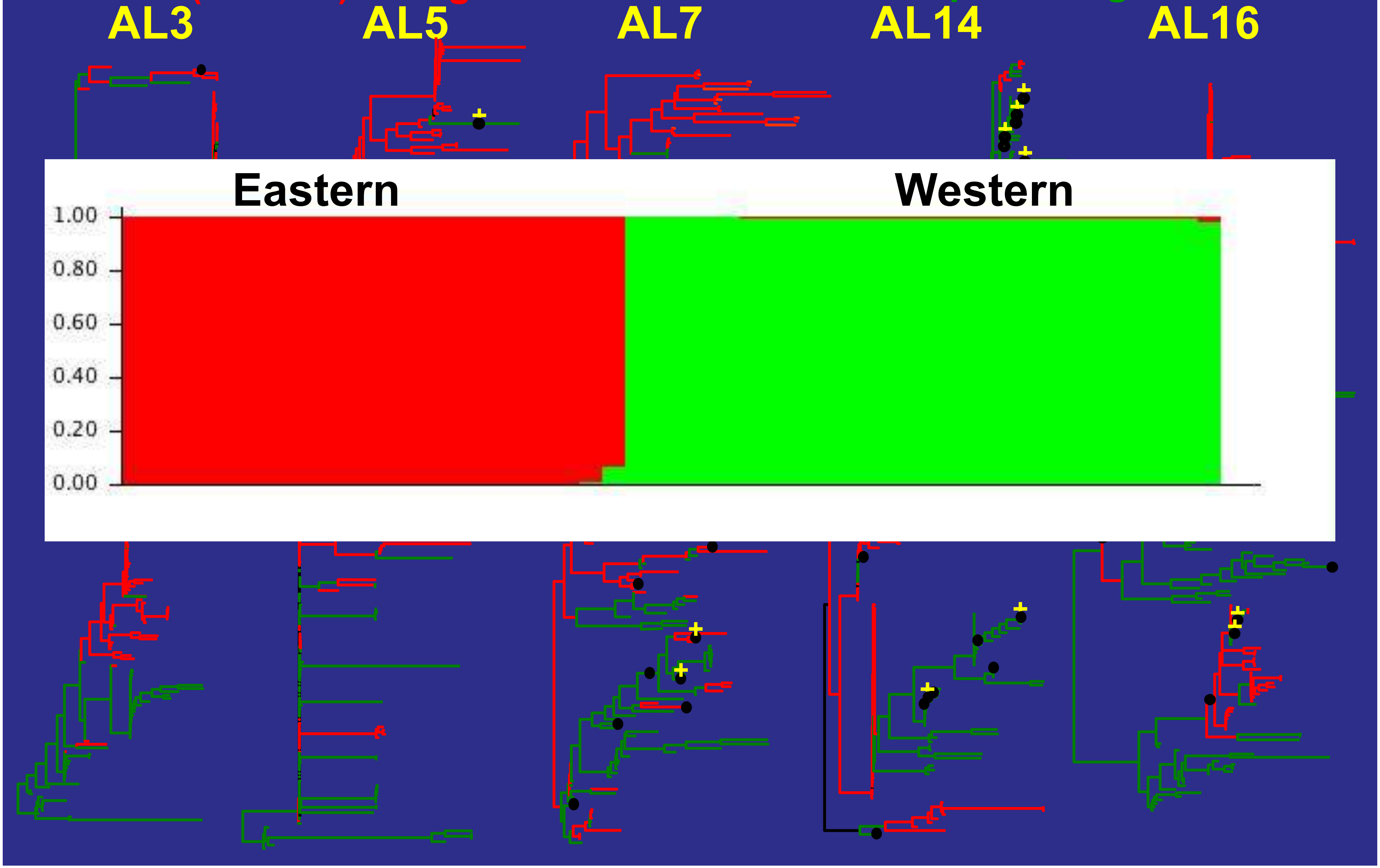
C. melanura

C. picumnus

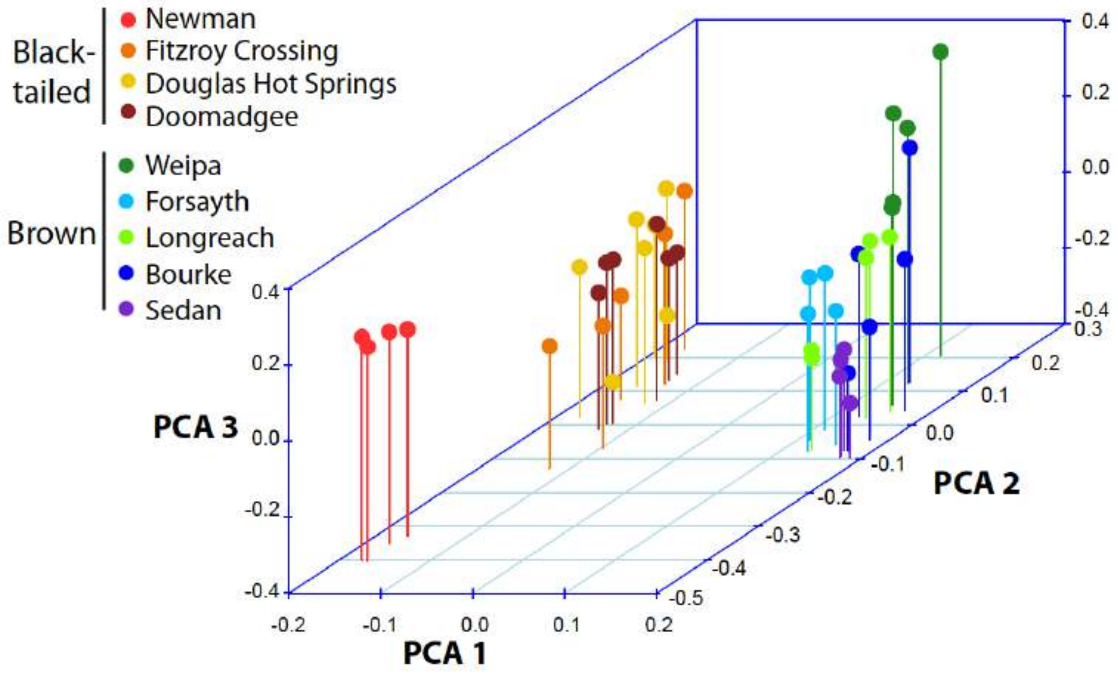
12 gene trees in Australian treecreepers

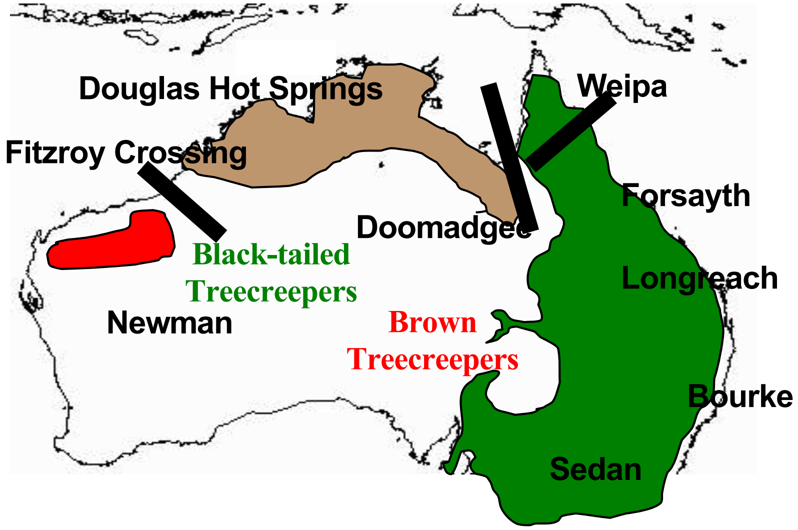
Brown Treecreeper (eastern) lineages

Black-tailed (western) Treecreeper lineages



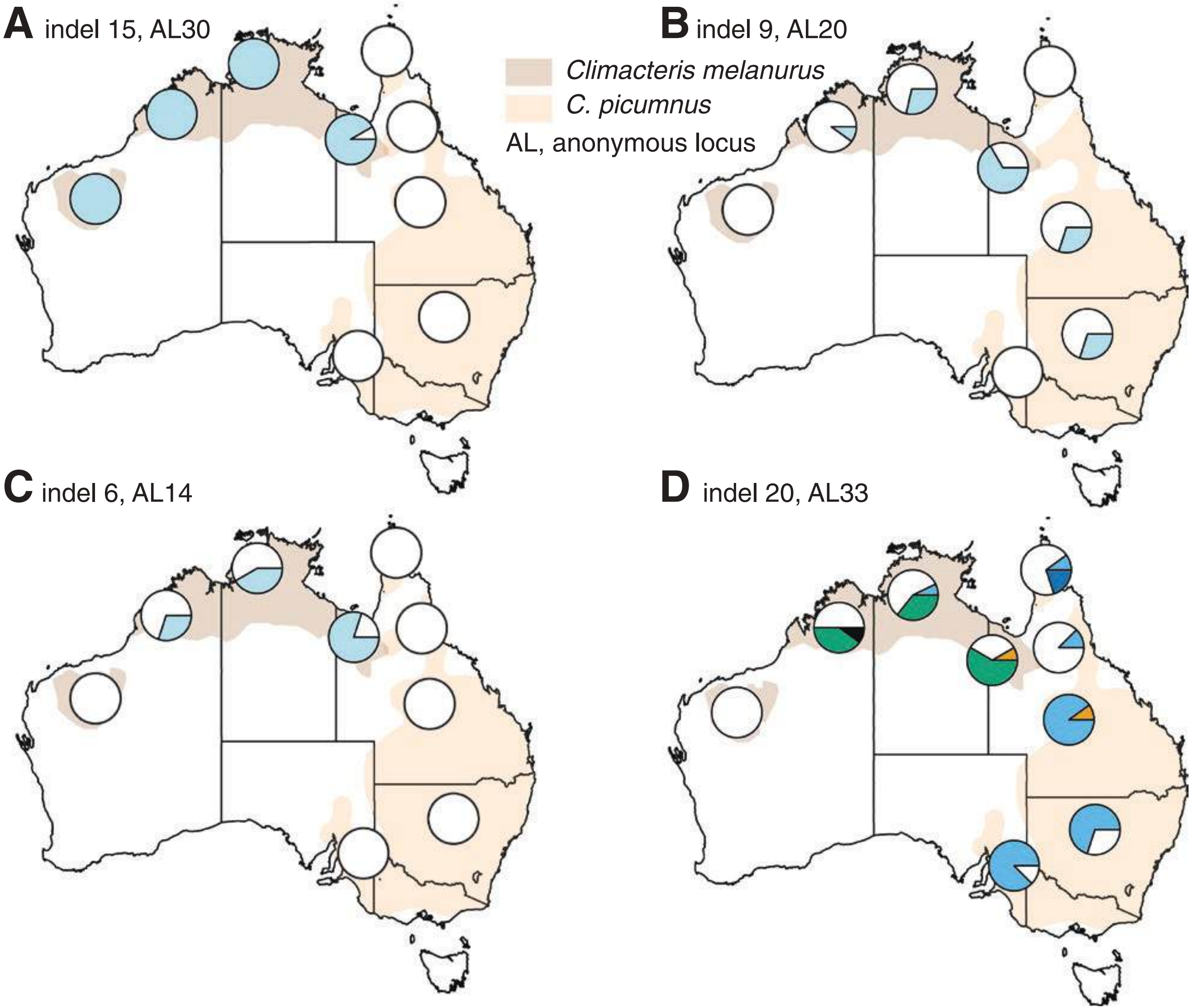
Correspondence between PCA and geography





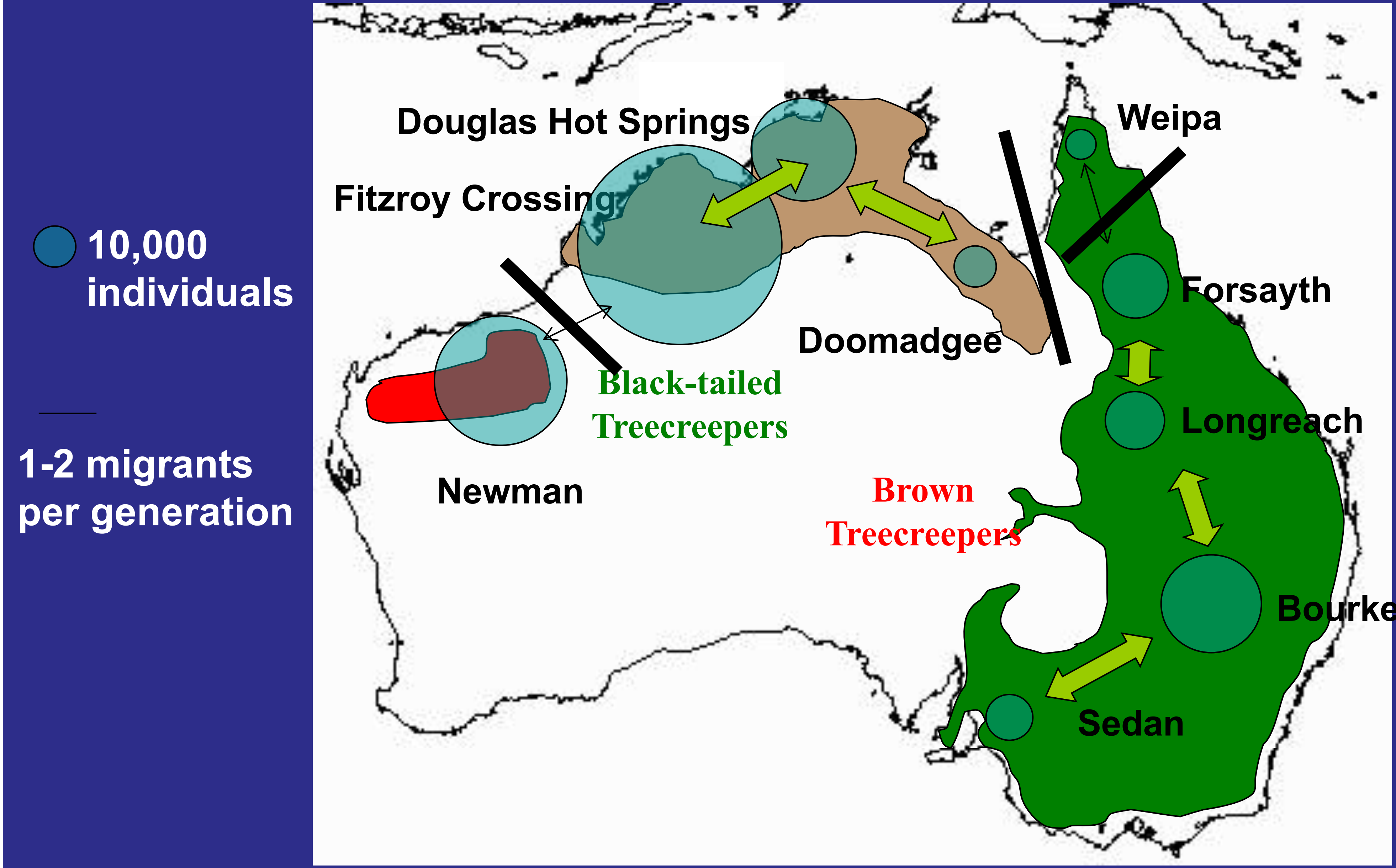
Edwards, Tonini, McInerney, Welch & Beerli. 2022*. Biol. J. Linn. Soc.*

Insertions-deletions provide significant phylogeographic signal



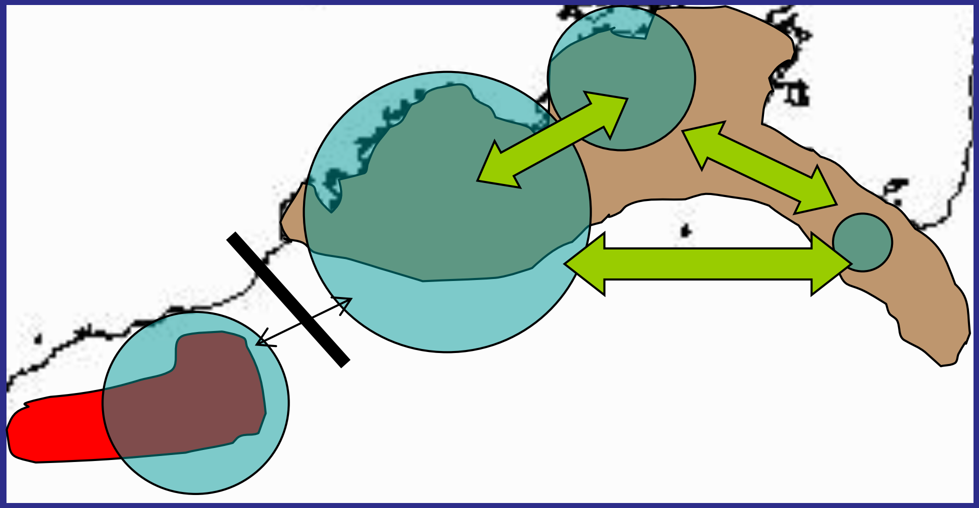
Edwards, Tonini, McInerney, Welch & Beerli. 2022. *Biol. J. Linn. Soc.*

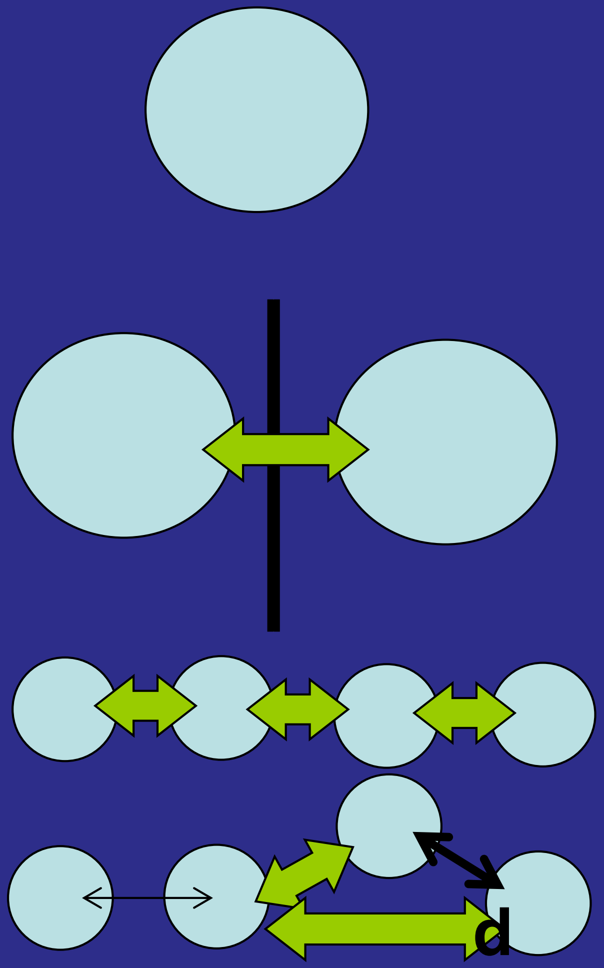
Treecreeper populations are connected but variable in size (MIGRATE)



Rotzel, Edwards and Beerli, unpubl. data

Using Bayes Factors to evaluate phylogeographic models





$2(mL_1 – mL_2)$

-98.60

-83.14

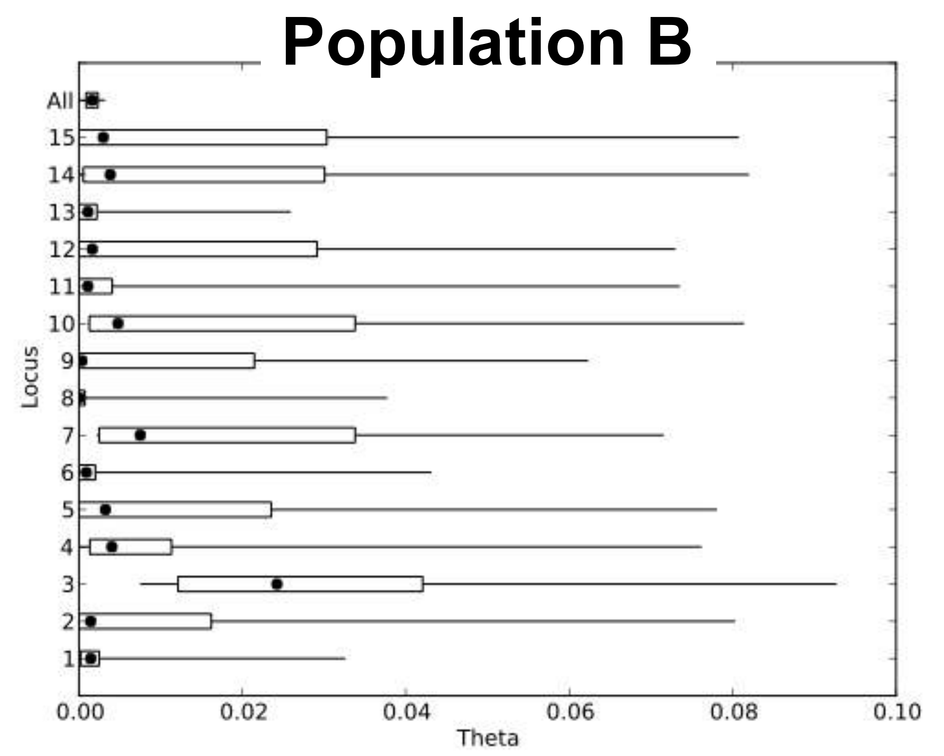
-54.98

0.00

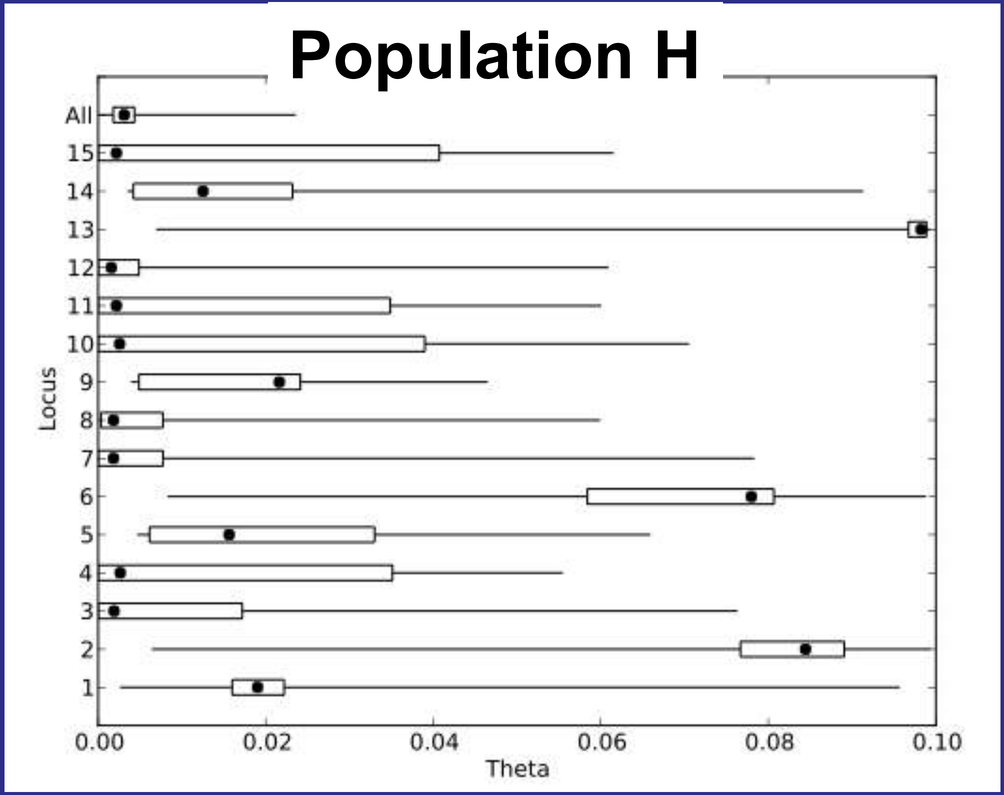
mL = marginal likelihood

Edwards, Tonini, McInerney, Welch & Beerli. 2022. *Biol. J. Linn. Soc.*

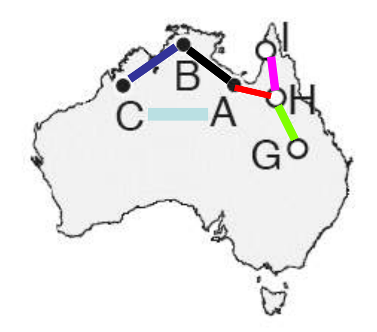
Higher precision estimates of demographic parameters (θ) with more loci



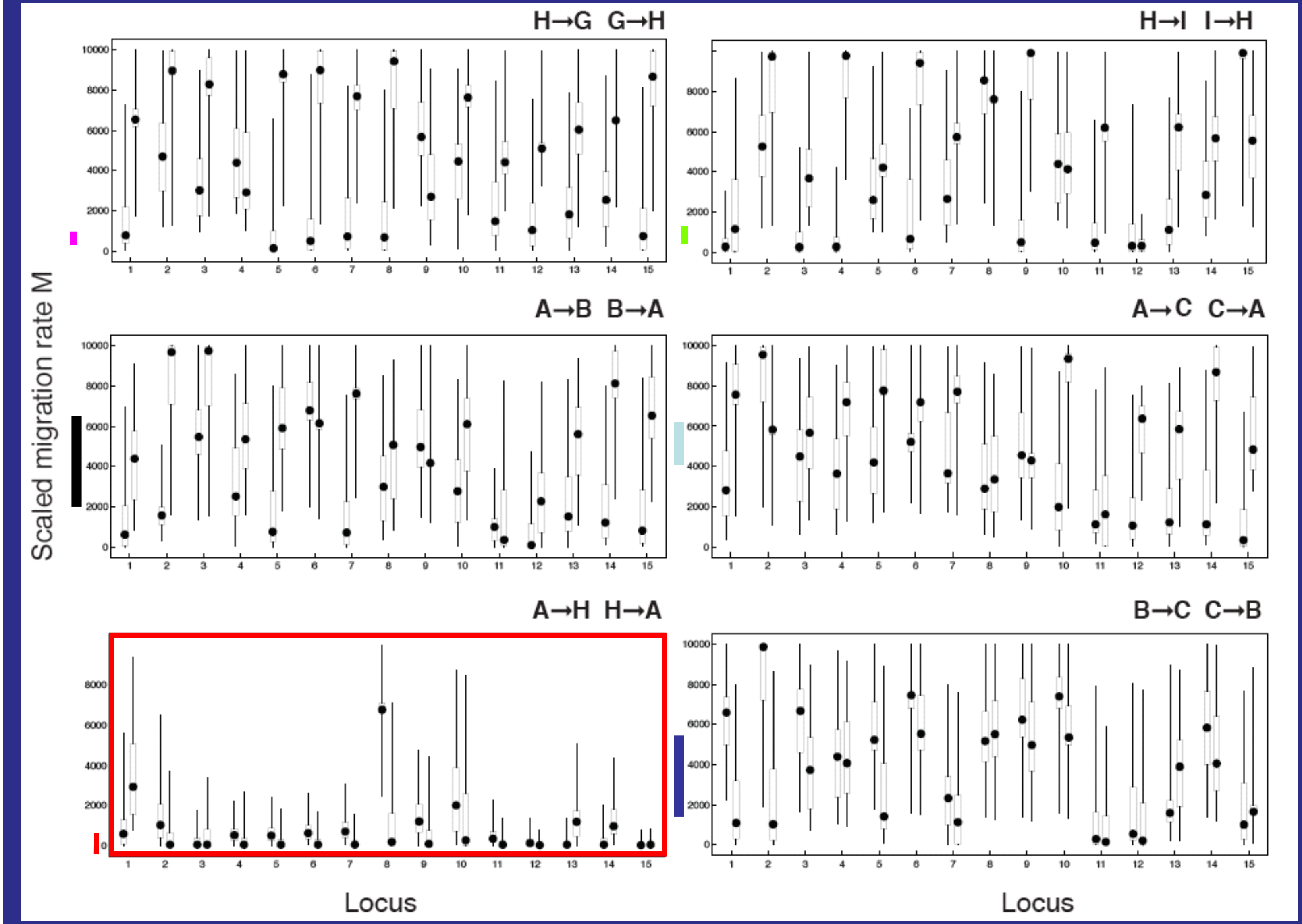
Population B

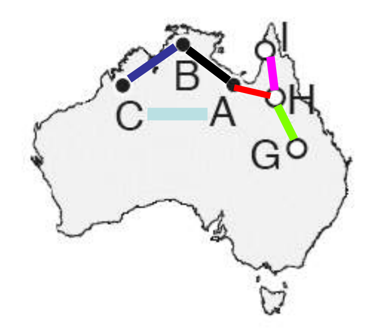


Population H

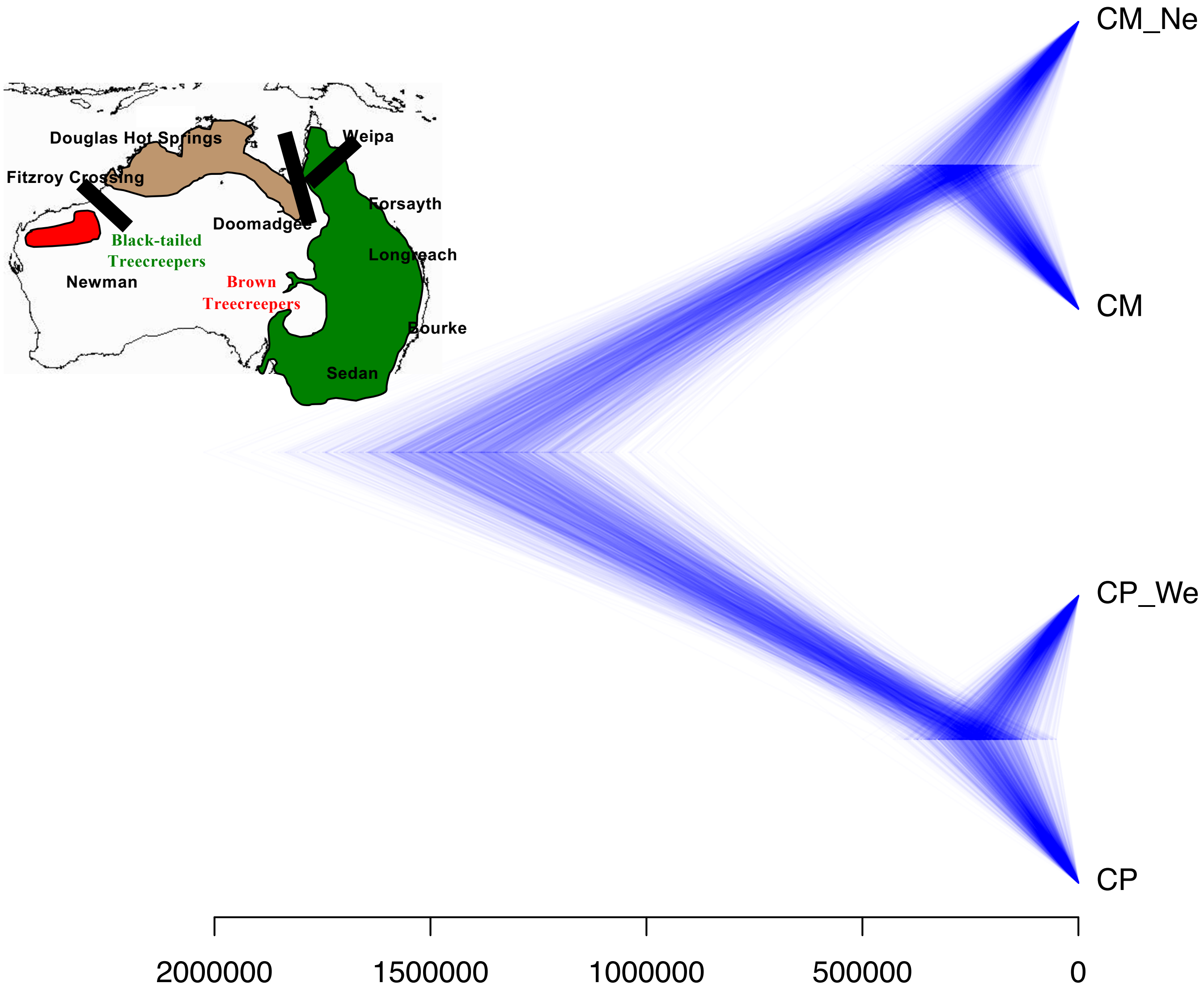


Multilocus estimates of migration rate





Phylogeny and divergence times



Divergence time (years), gen = 1 yr., μ = 2.2e-09

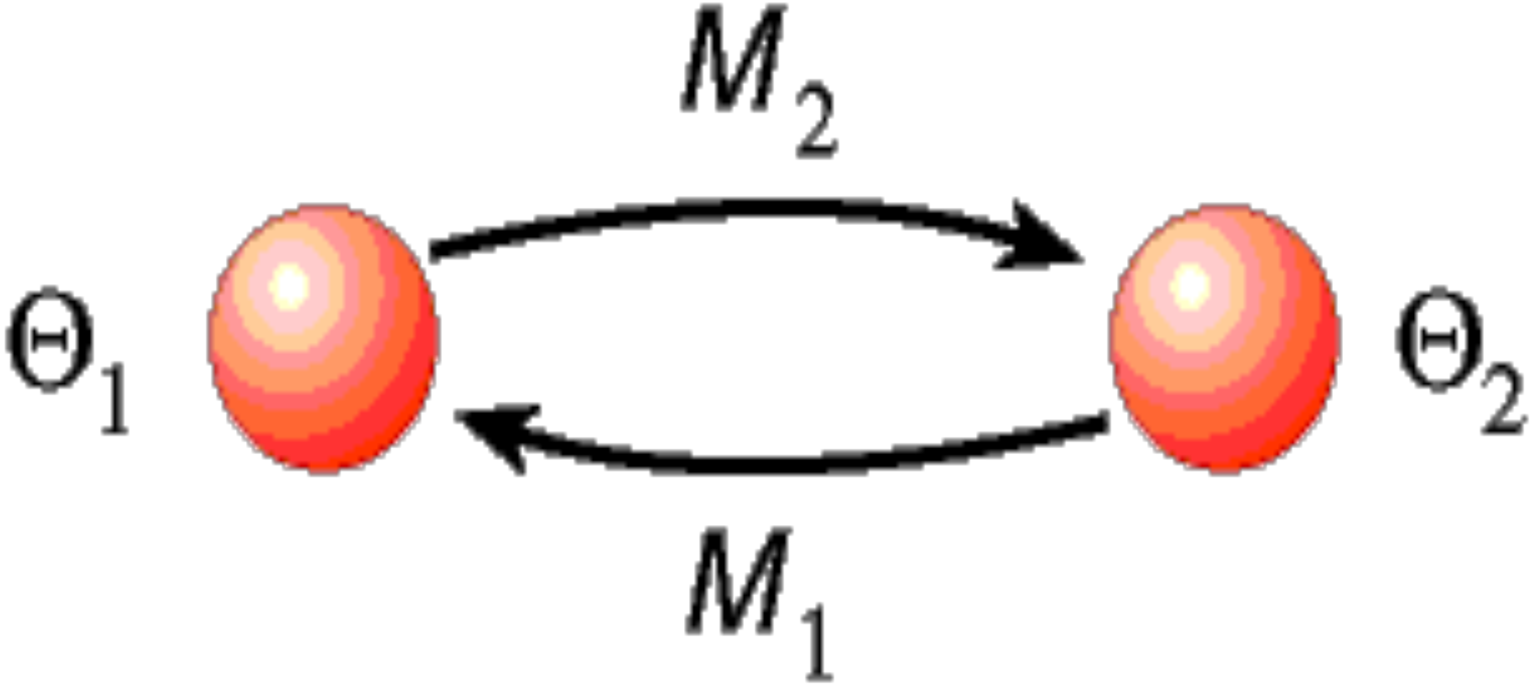
Model selection using Migrate

1

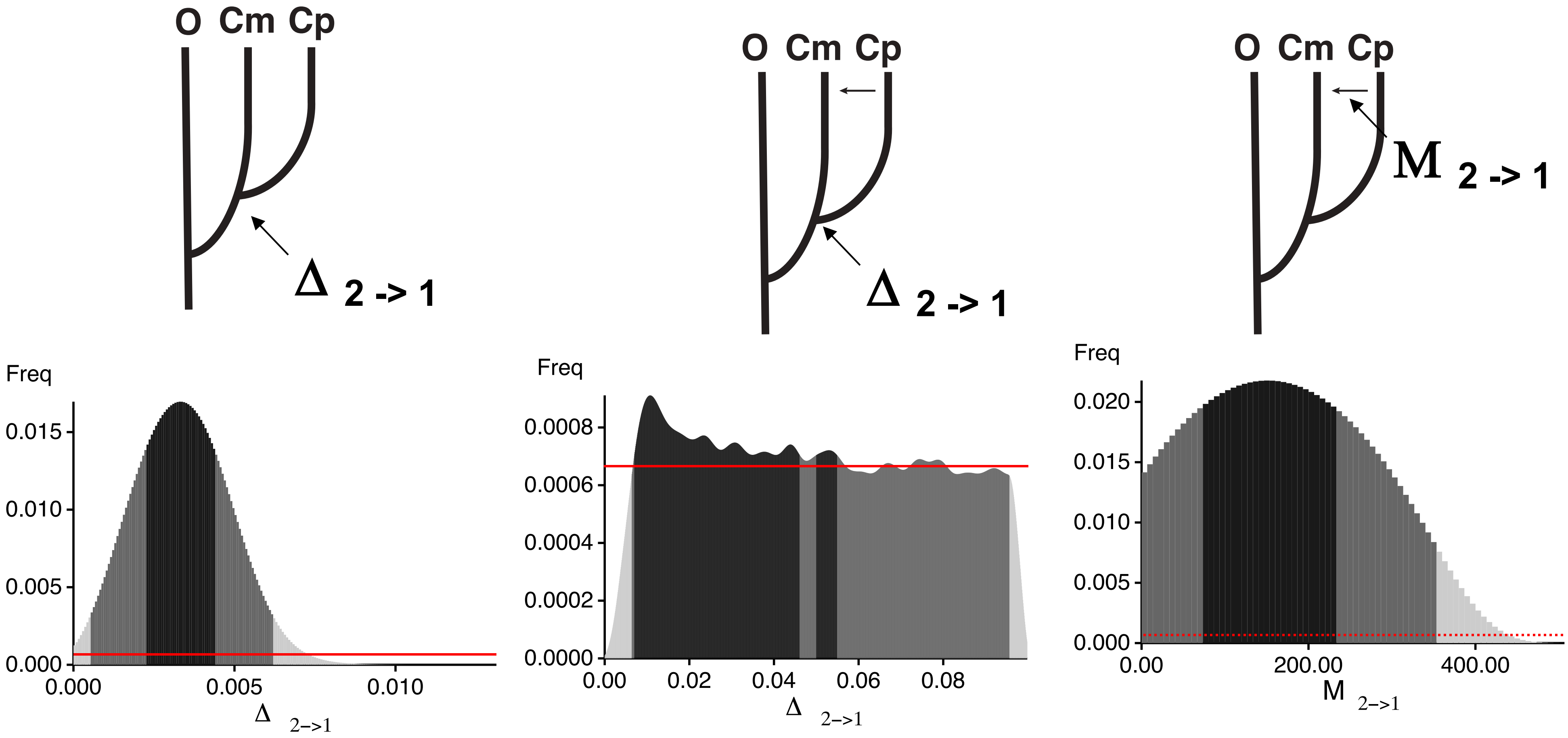
- 0 0 0
0 - 0 d
0 0 - d
d 0 0 -

2 Best model!

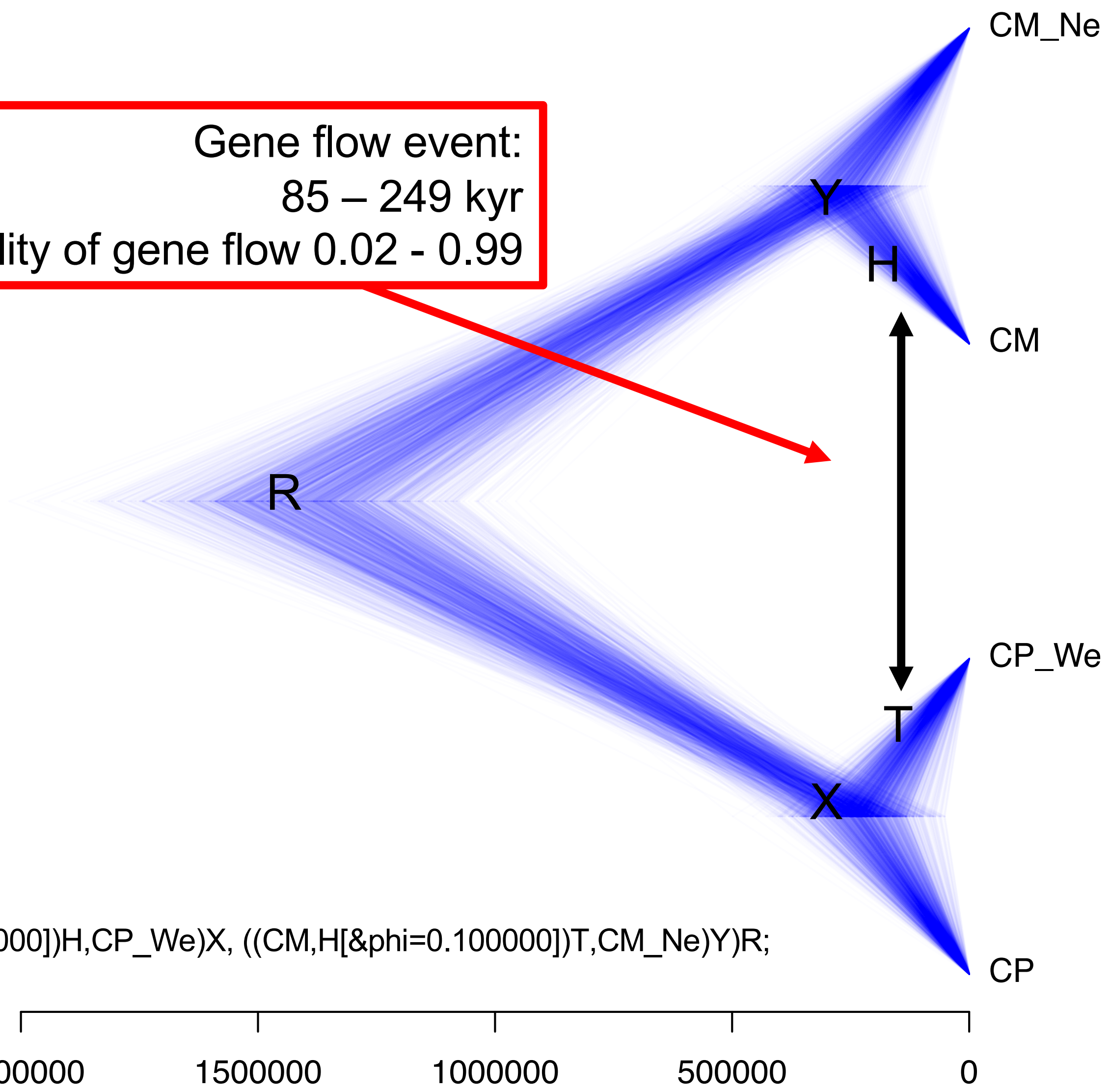
- 0 0
0 - d
d 0 -



Challenges estimating isolation-migration parameters



Phylogeny and gene flow with bpp



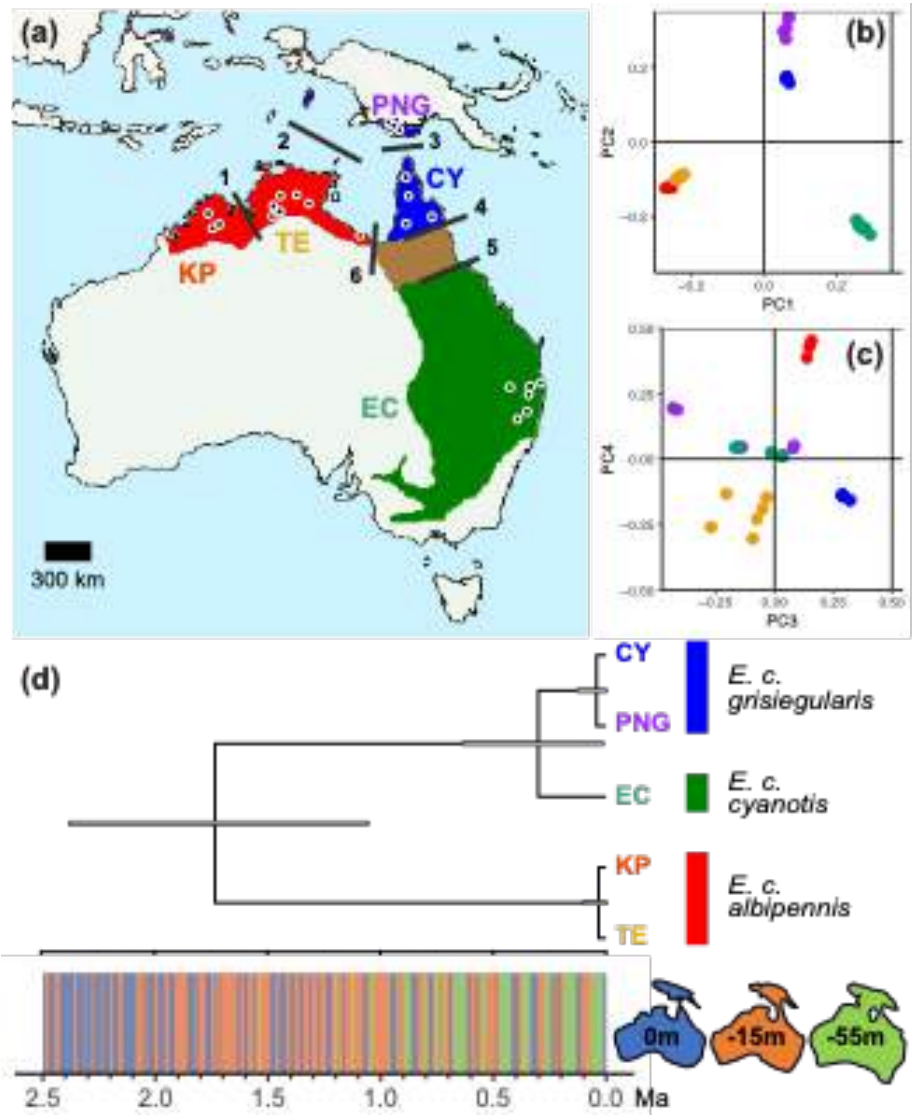
Gene flow event:
85 – 249 kyr
Probability of gene flow 0.02 - 0.99

(((CP,T[&phi=0.200000])H,CP_We)X, ((CM,H[&phi=0.100000])T,CM_Ne)Y)R;

Divergence time (years), gen = 1 yr., μ = 2.2e-09

Whole-genome phylogeography of a widespread Australian honeyeater

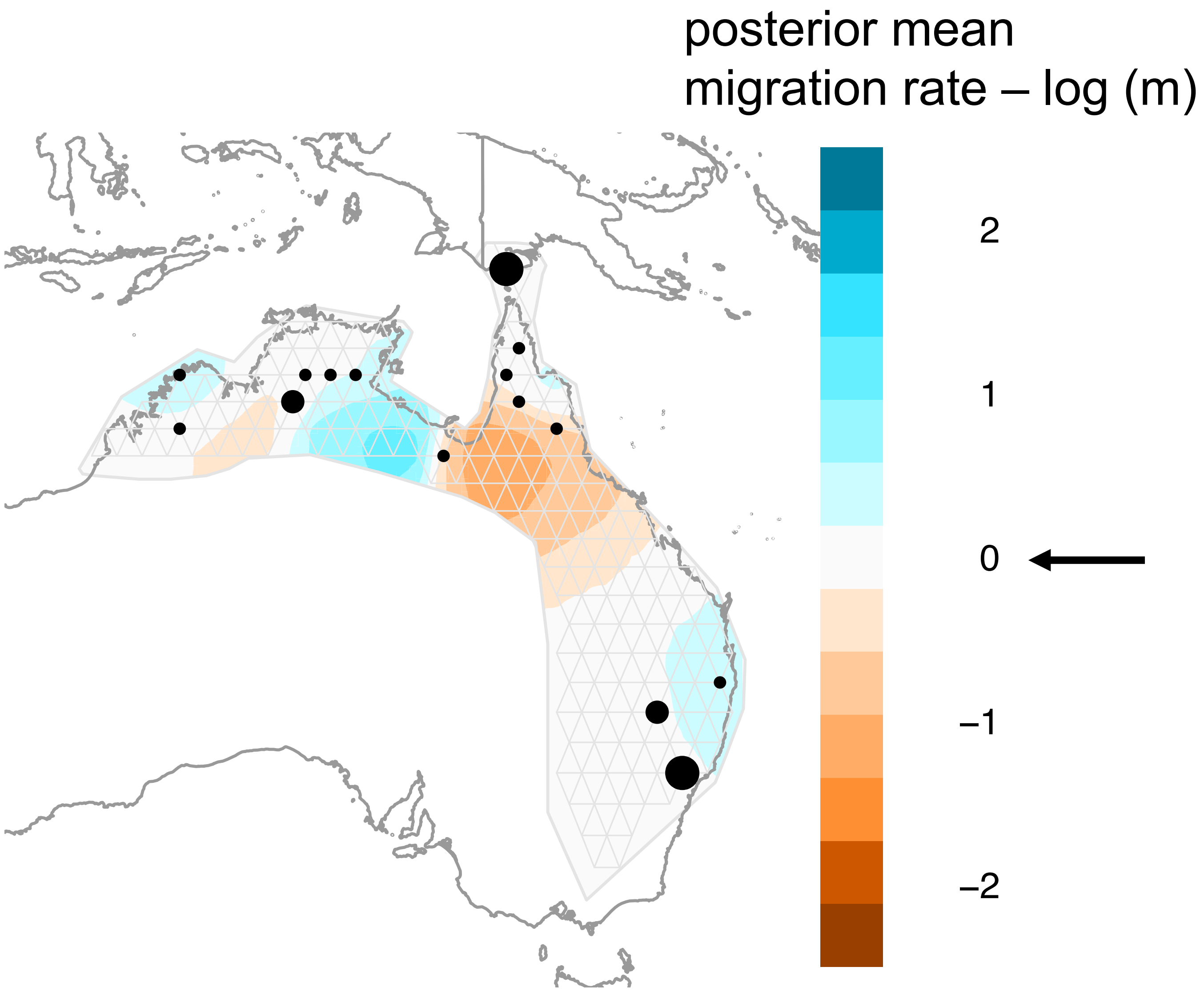
Blue-faced honeyeater
Entomyzon cyanotis



- **24 samples**
- **2 outgroups**
- **64X reference genome**
- **7-12X population resequencing**
- **GATK variant calling**
- **ANGSD (Korneliussen et al. 2014. *BMC Genomics*)**

Burley et al. 2022. *Molecular Ecology* 32, 1248-1270

Effective migration surface



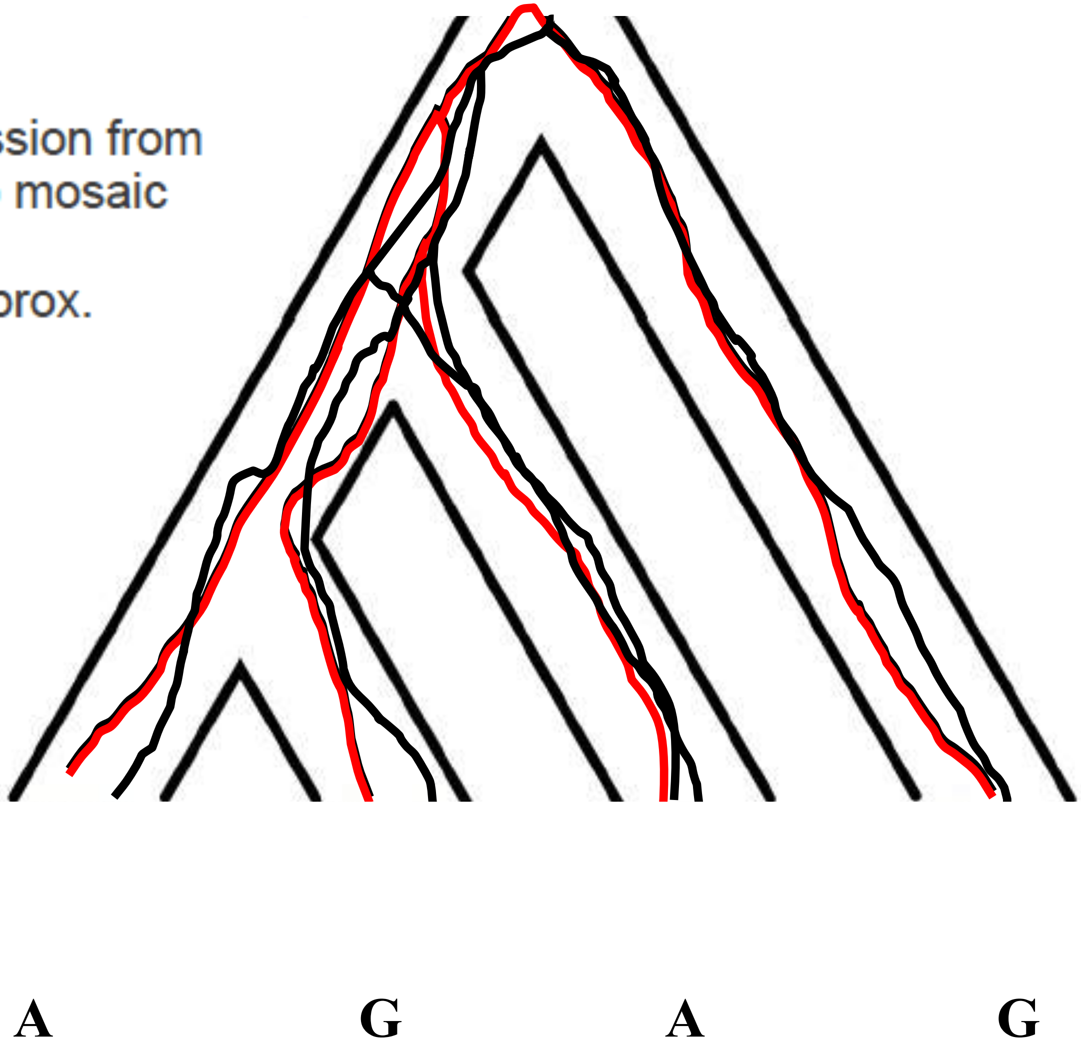
Average level of gene flow

Burley et al. 2022. *Molecular Ecology* 32, 1248-1270

ABBA-BABA: Genome-wide test of introgression

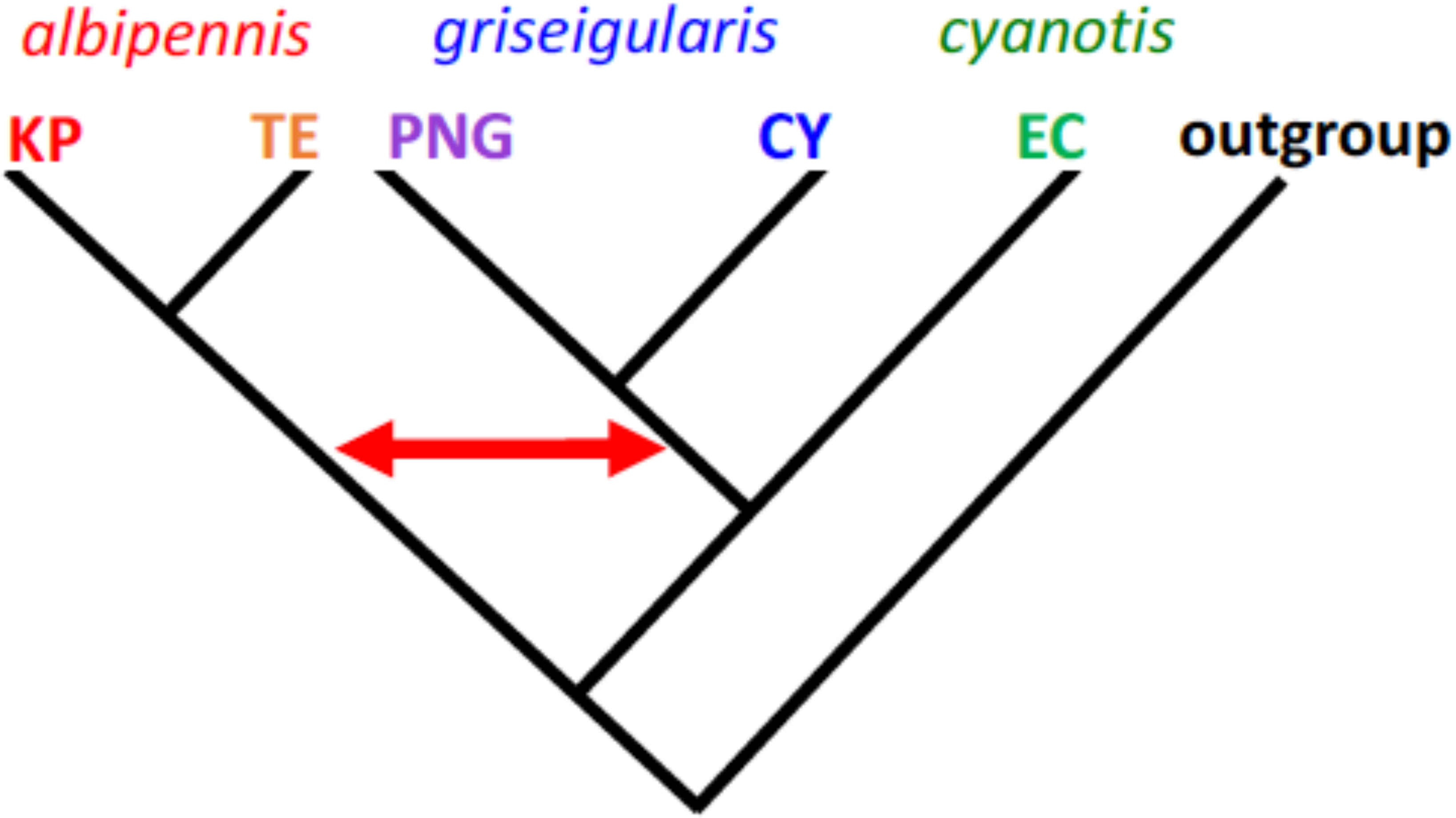
$$D = \frac{Num.\ ABBAs - Num.\ BABAs}{Num\ SNPs}$$

- *D* > 0: introgression from *Z. chrysops* into mosaic population
- (*D* assumed approx. Normal)



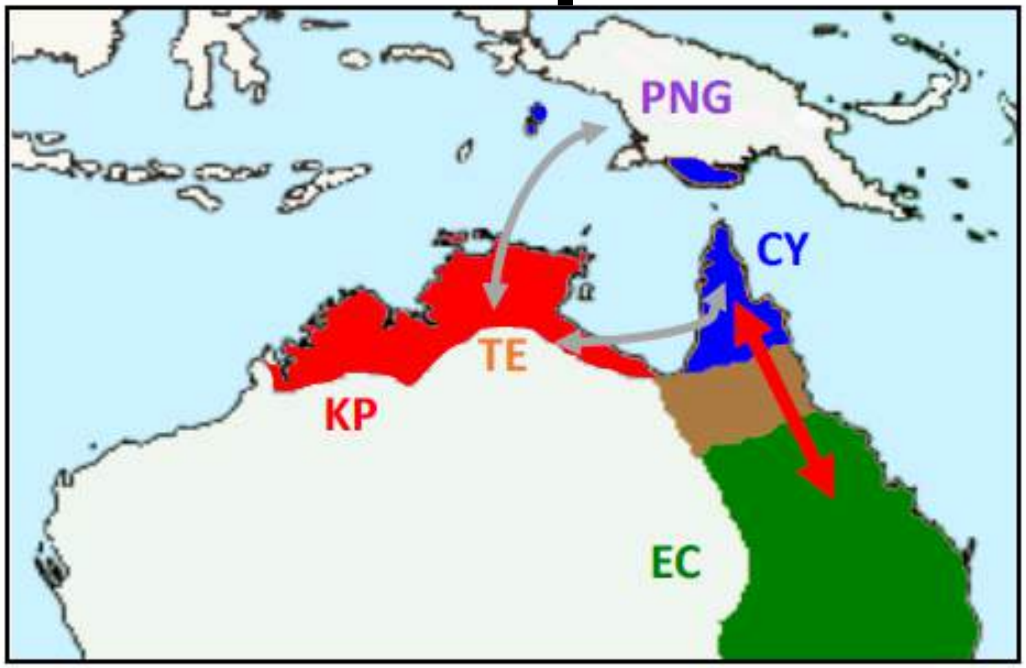
ABBA-BABA tests reveal signals of past introgression

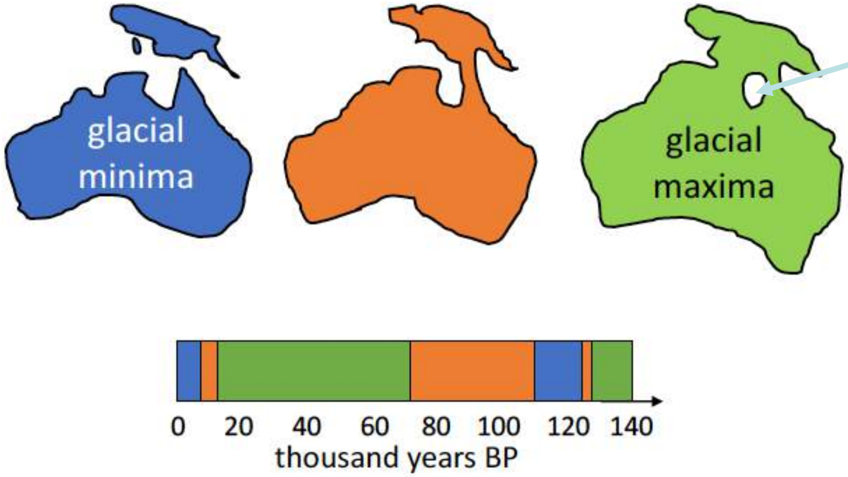
(Dsuite, Malinsky et al. 2021. *Mol. Ecol. Res.*)



Burley et al. 2022. *Molecular Ecology* 32, 1248-1270

Phylogeographic shadows of Lake Carpentaria





Pleistocene Lake Carpenta

Trends in Ecology & Evolution

CellPress
REVIEWS

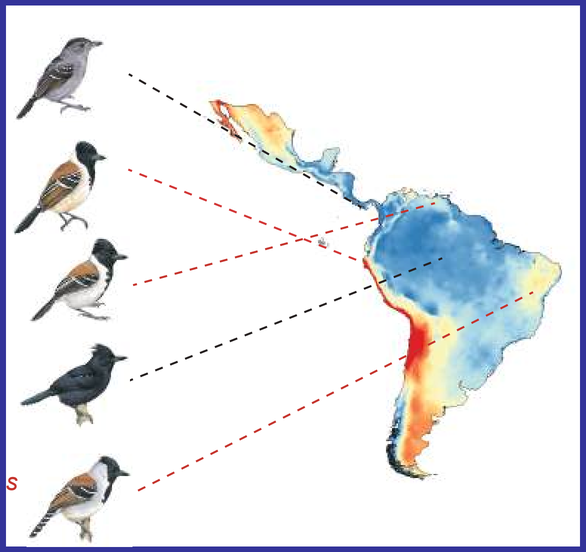
Opinion

Phylogenetics is the New Genetics (for Most of Biodiversity)

Stacey D. Smith,[1,6,*,@] Matthew W. Pennell,[2] Casey W. Dunn,[3] and Scott V. Edwards[4,5]

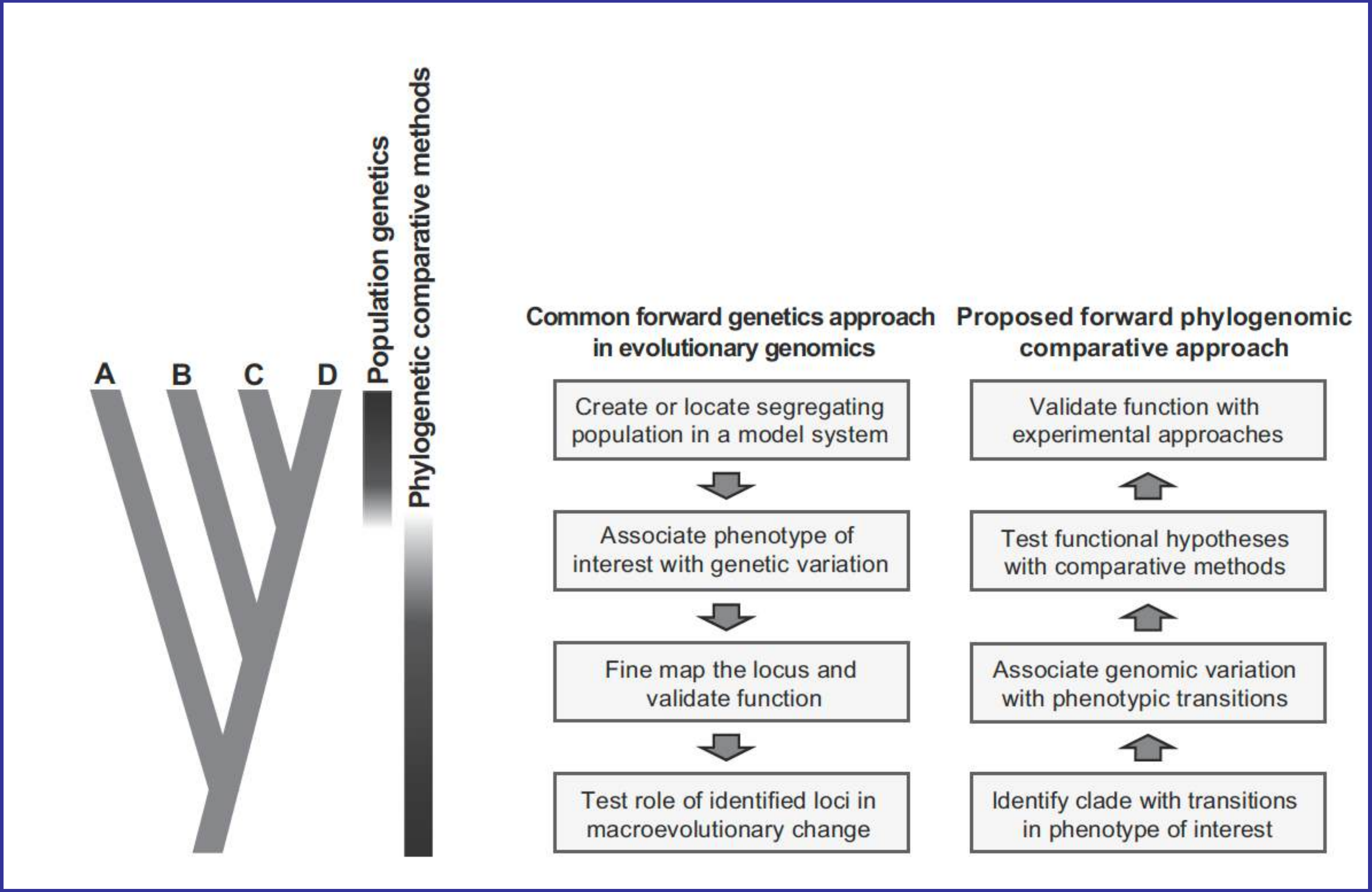
Smith, et al. 2020. *TREE* 35: P415-525

- Recent ornithological applications of the PhyloG2P approach:
  - Beak size in birds
    - Yusuf et al. 2020. *Genome Res. 2020. 30: 553-565*
  - Loss of flight in birds
  - Sackton et al. 2019. *Science* 364: 74-78.
  - Egg architecture and genomic adaptations to dry habitats
    - Gustavo Bravo, Harvard, in prep.



Gustavo Bravo, in prep.
Antbirds, Thamnophilidae

Using phylogenies to connect genotype to phenotype



Common forward genetics approach in evolutionary genomics

- Create or locate segregating population in a model system
- Associate phenotype of interest with genetic variation
- Fine map the locus and validate function
- Test role of identified loci in macroevolutionary change

Proposed forward phylogenomic comparative approach

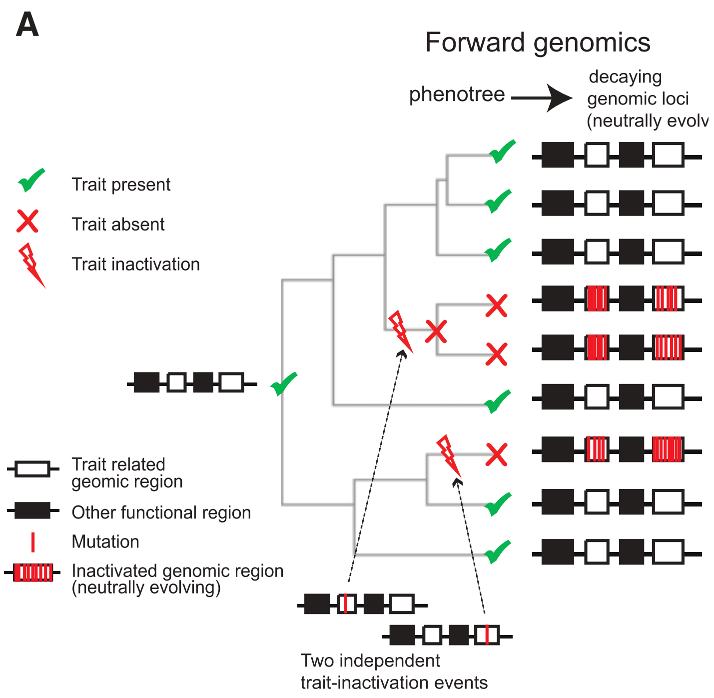
- Validate function with experimental approaches
- Test functional hypotheses with comparative methods
- Associate genomic variation with phenotypic transitions
- Identify clade with transitions in phenotype of interest

Smith, et al. 2020. *TREE* 35:

Genomic signatures of trait associations

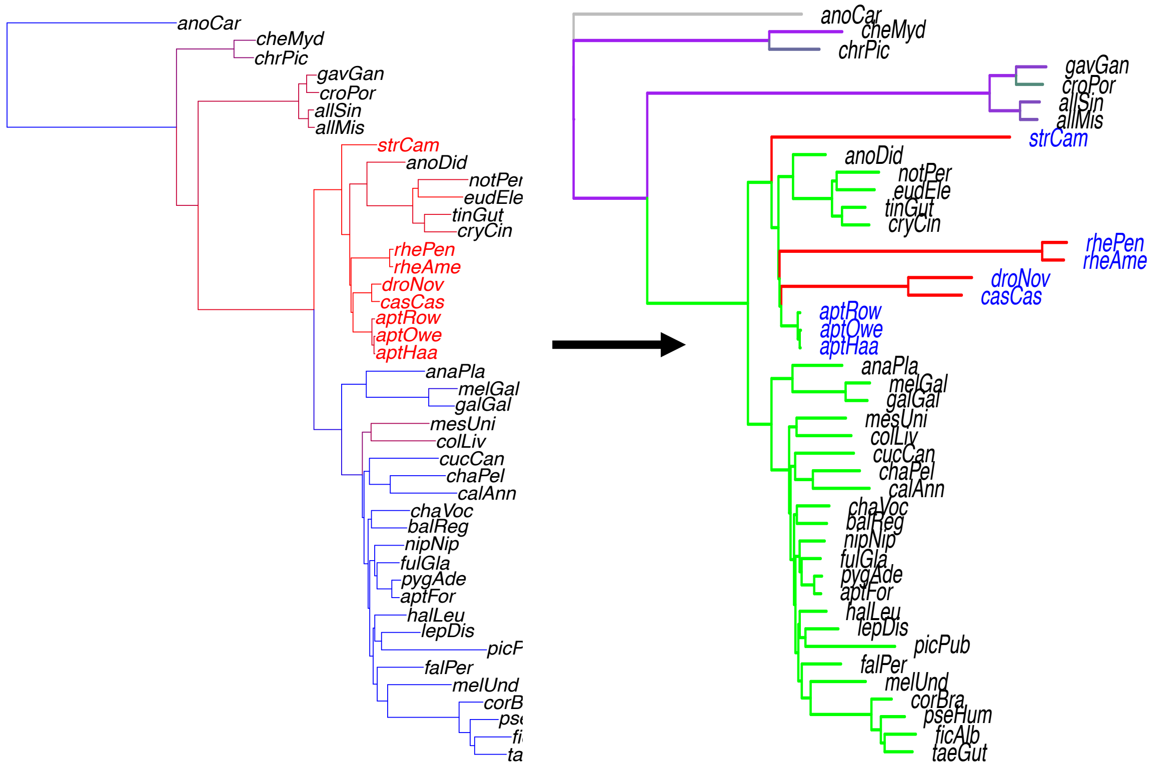
Deletion or inactivation of conserved noncoding elements

Acceleration of clade-wide conserved noncoding elements



Genome-wide neutral tree

Single locus accelerated tree



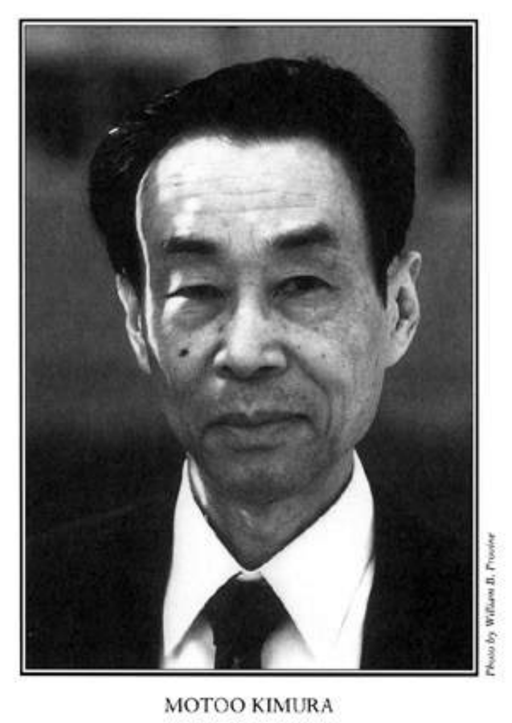
Marcovitz et al. 2016

Mol Biol Evol, 33: 1358–1369

Target lineages in red

Accelerated lineages in red

Acceleration suggests change, loss or relaxation of function in the neutral theory of molecular evolution



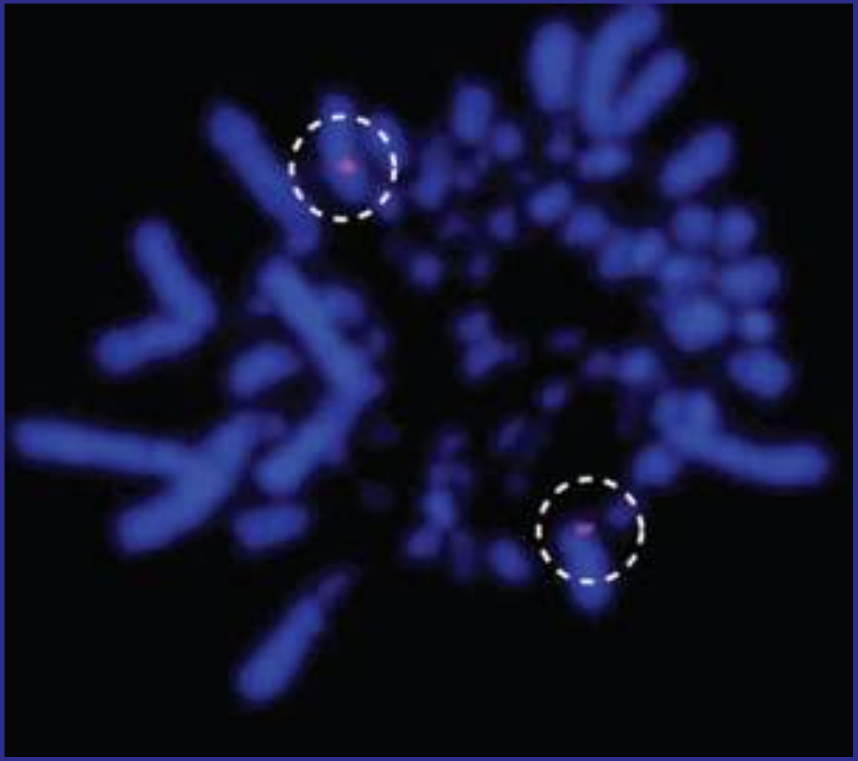
MOTOO KIMURA

Motoo Kimura (1924-1994)

Functionally less important parts of genomes will have a higher substitution rate than functionally important ones

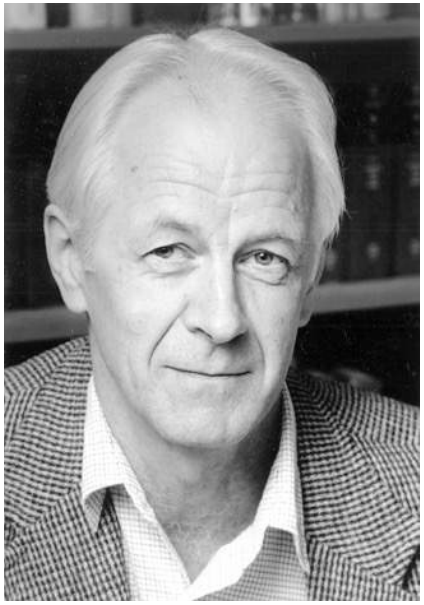
Functionally less important molecules or parts thereof will have a higher rate of substitution than functionally important ones

Non-coding 'Dark matter' of the genome: a regulatory network?



Karyotype of an Emu

Evolutionary change: genes or gene regulation?



Evolution at Two Levels in Humans and Chimpanzees

Their macromolecules are so alike that regulatory mutations may account for their biological differences.

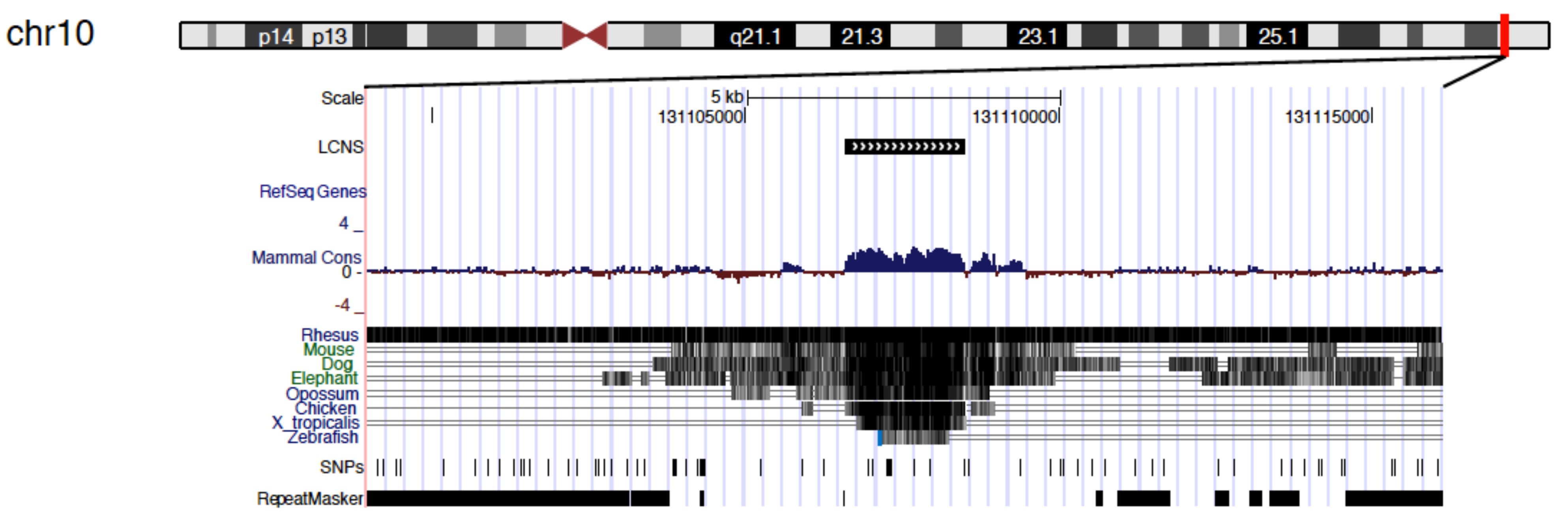
Mary-Claire King and A. C. Wilson

SCIENCE

11 April 1975, Volume 188, Number 4184

CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements

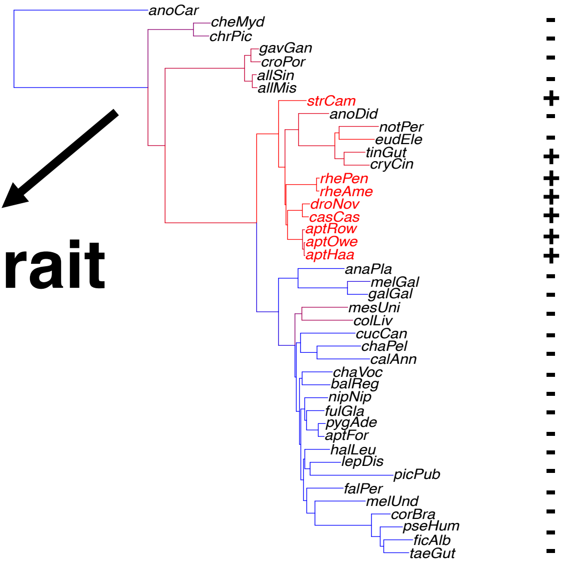


View of a segment of human chromosome 10 using UCSC Genome Browser

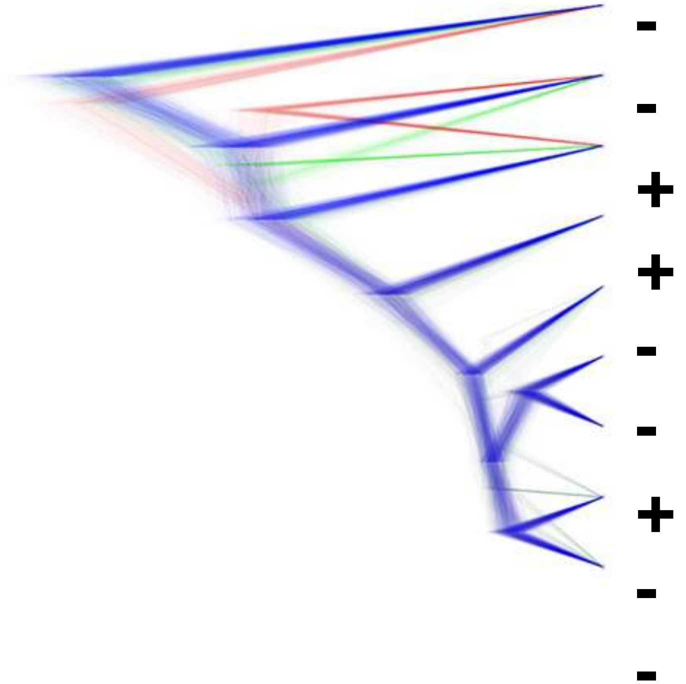
Janes et al. (2011) *Genome Biol. Evol.* 3:102–113

The growing PhyloAcc software family connects genomic and trait variation via phylogenies

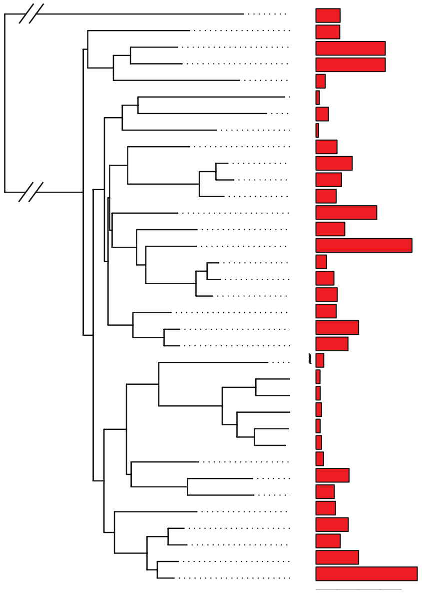
PhyloAcc **binary trait**



PhyloAcc-GT
binary trait



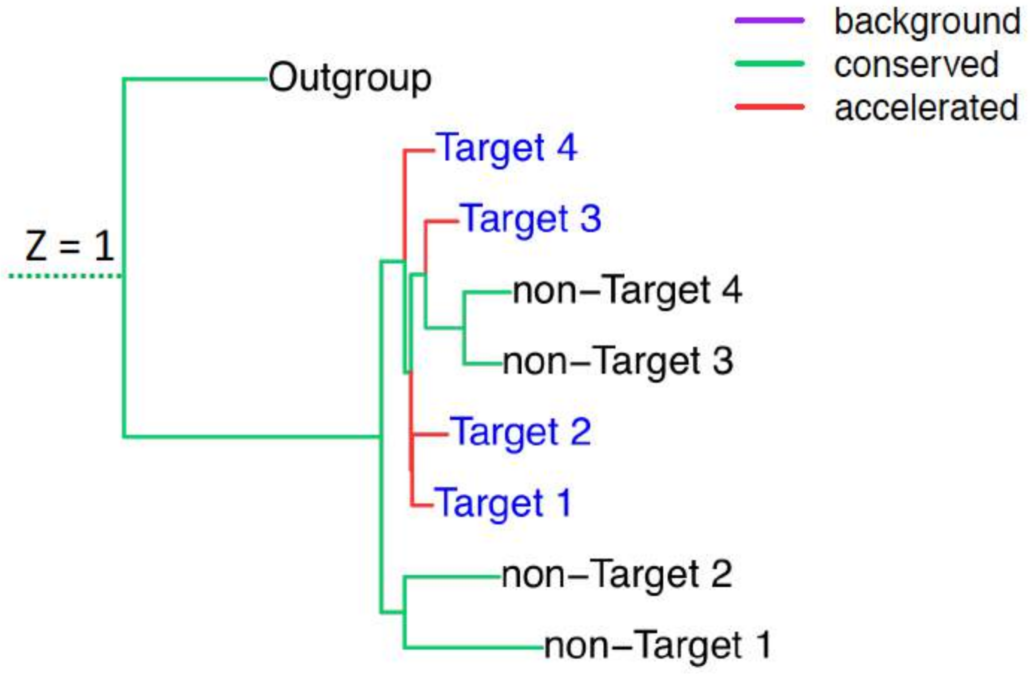
PhyloAcc-C
continuous trait



Hu et al. 2019. *Mol. Biol. Evol.* 36: 1086

Yan et al. 2023. bioRxiv. https://doi.org/10.1101/2022.12.23.521765

PhyloG2P: connecting genomes to phenotypes using PhyloAcc



for noncoding element *i*

$$Z = \begin{bmatrix} 1-\alpha_i & \alpha_i & 0 \\ 0 & 1-\beta_i & \beta_i \\ 0 & 0 & 1 \end{bmatrix}$$

α = probability of gain of conserved state

β = probability of loss of conserved state

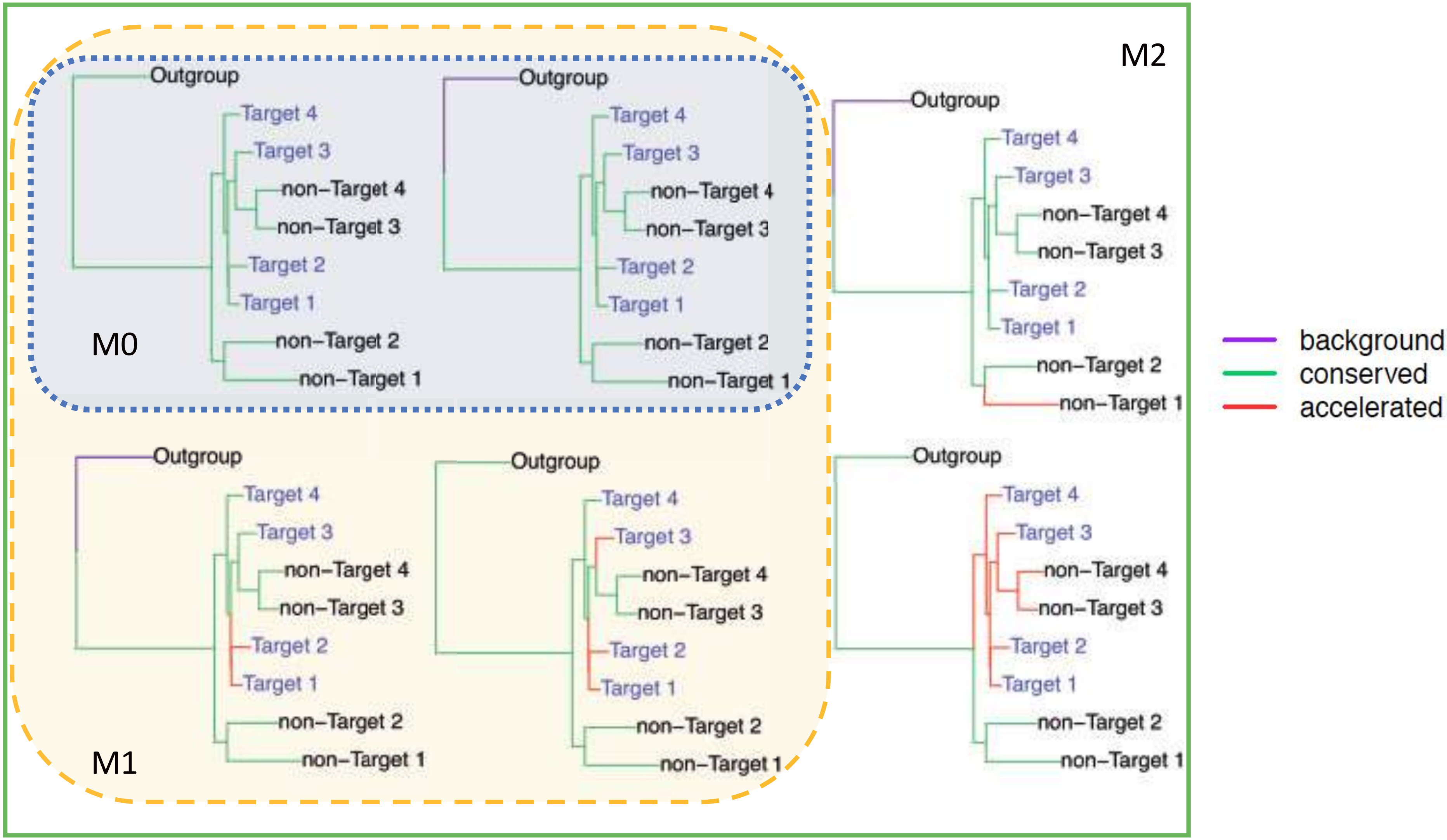
$$\text{For branch s,} \begin{cases} r_s = r_1, & \text{if } Z_s = 1, \text{ conserved} \\ r_s = r_0 = 1, & \text{if } Z_s = 0, \text{ background} \\ r_s = r_2, & \text{if } Z_s = 2, \text{ accelerated} \end{cases}$$

https://phyloacc.github.io/

Hu, Z., et al. 2019. *Mol. Biol. Evol*. 36: 1086

Two Bayes Factor tests discriminate competing hypotheses

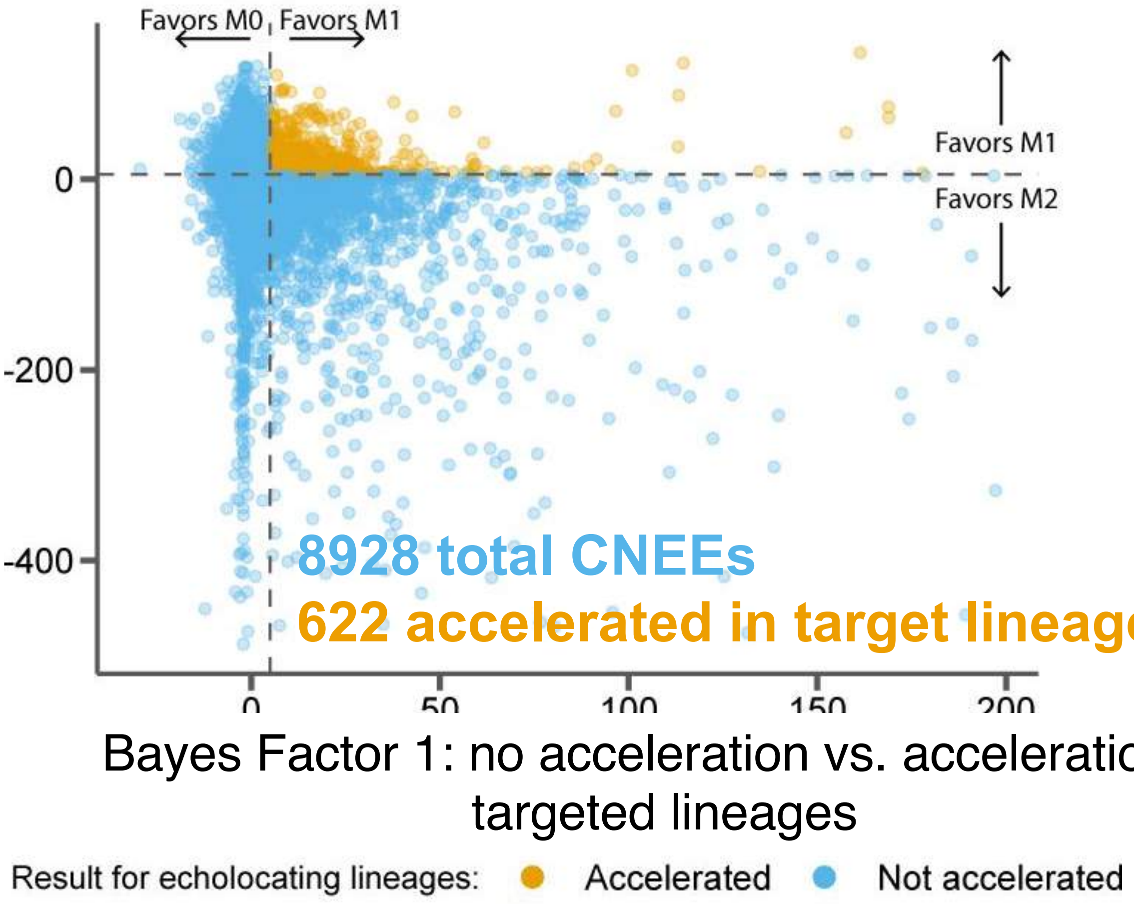
$$BF1 = \frac{P(Y|M_1)}{P(Y|M_0)} \text{ and } BF2 = \frac{P(Y|M_1)}{P(Y|M_2)}.$$



Hu, Z., et al. 2019. *Mol. Biol. Evol*. 36: 1086

Identifying CNEEs evolving on echolocating mammal lineages

https://github.com/phyloacc/phyloacc-workflows

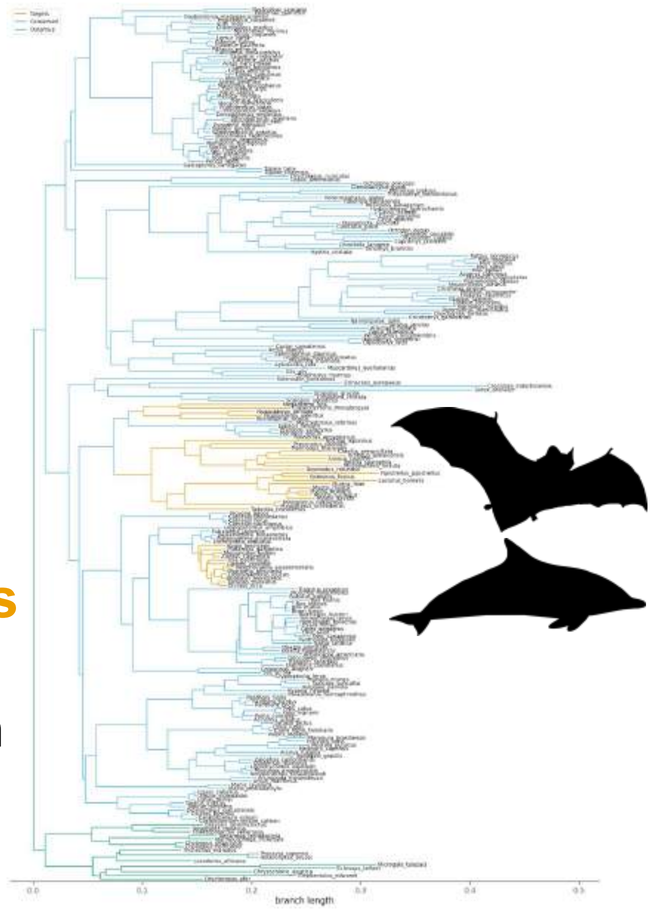


Bayes Factor 2: target lineage model vs. unrestricted model

8928 total CNEEs

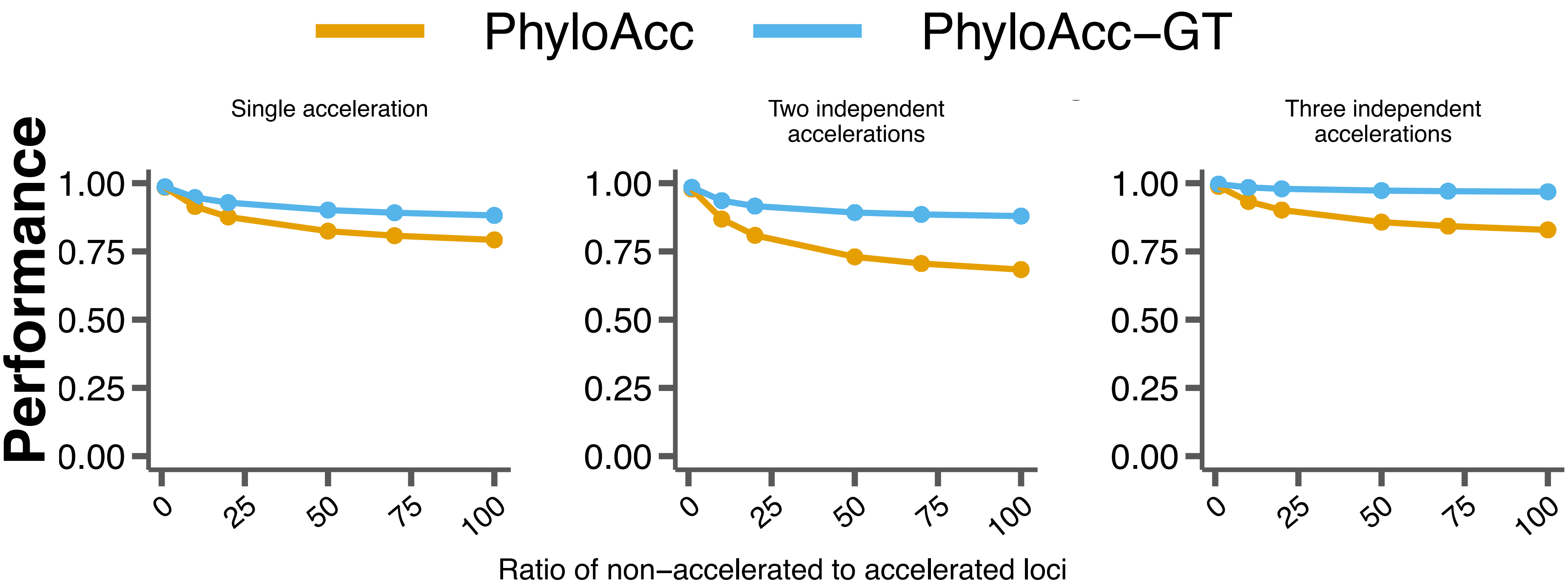
622 accelerated in target lineages

Bayes Factor 1: no acceleration vs. acceleration targeted lineages



Thomas et al. 2024. *Integrative and Comparative Biology*

PhyloAcc-GT: Detecting accelerations with gene tree variation

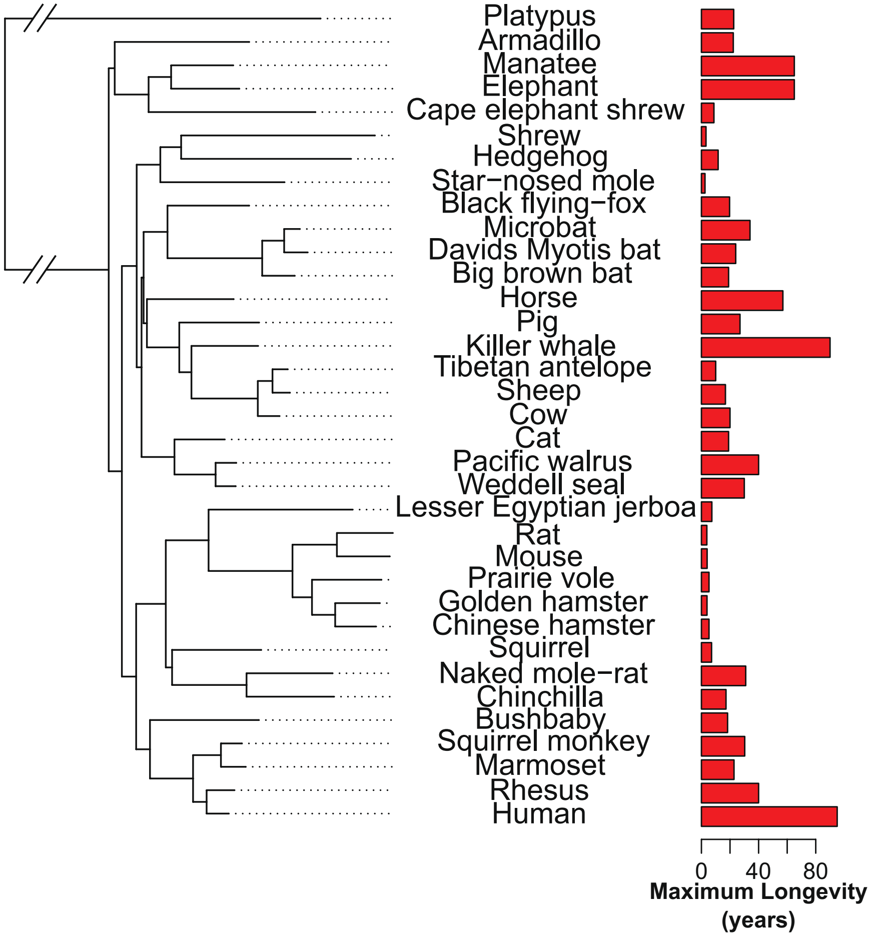


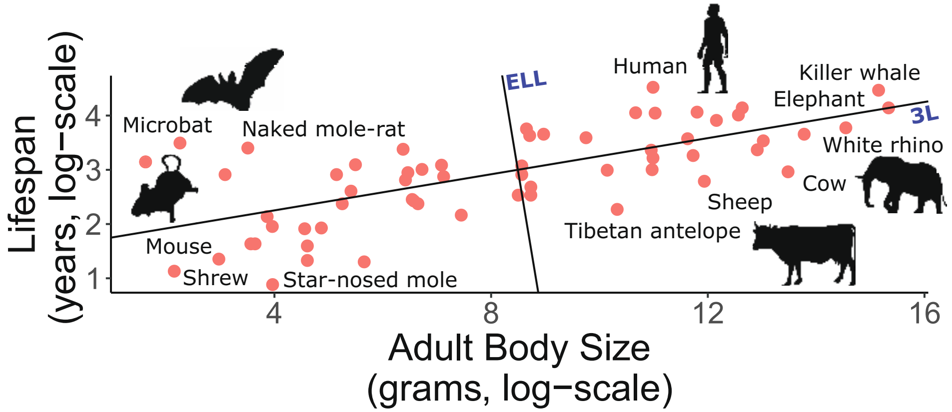
Yan et al. 2023. *Mol Biol Evol*. 40(9):msad195.

PhyloAcc-C: Models connecting molecular rates and continuous traits

Patrick Gemmell

longevity in mammals





https://github.com/phyloacc/PhyloAcc-C

Kowalczyk et al. 2020. *eLife*.

PhyloAcc-C: Rate multipliers for CNEEs and traits

Molecular rates **Phenotypic rates**

$$r_0 = 1 \quad if\ Z_s = 0\ \text{background}$$

$$r_1 < 1 \quad if\ Z_s = 1\ \text{conserved}_v = \begin{cases} \boldsymbol{\sigma}^2 if\ Z_s = 0\ \textbf{background} \\ \boxed{\beta_2}\ \boldsymbol{\sigma}^2 \quad if\ Z_s = 1\ \textbf{conserve} \\ \boxed{\beta_3}\ \boldsymbol{\sigma}^2 \quad if\ Z_s = 2\ \textbf{accelerat} \end{cases}$$

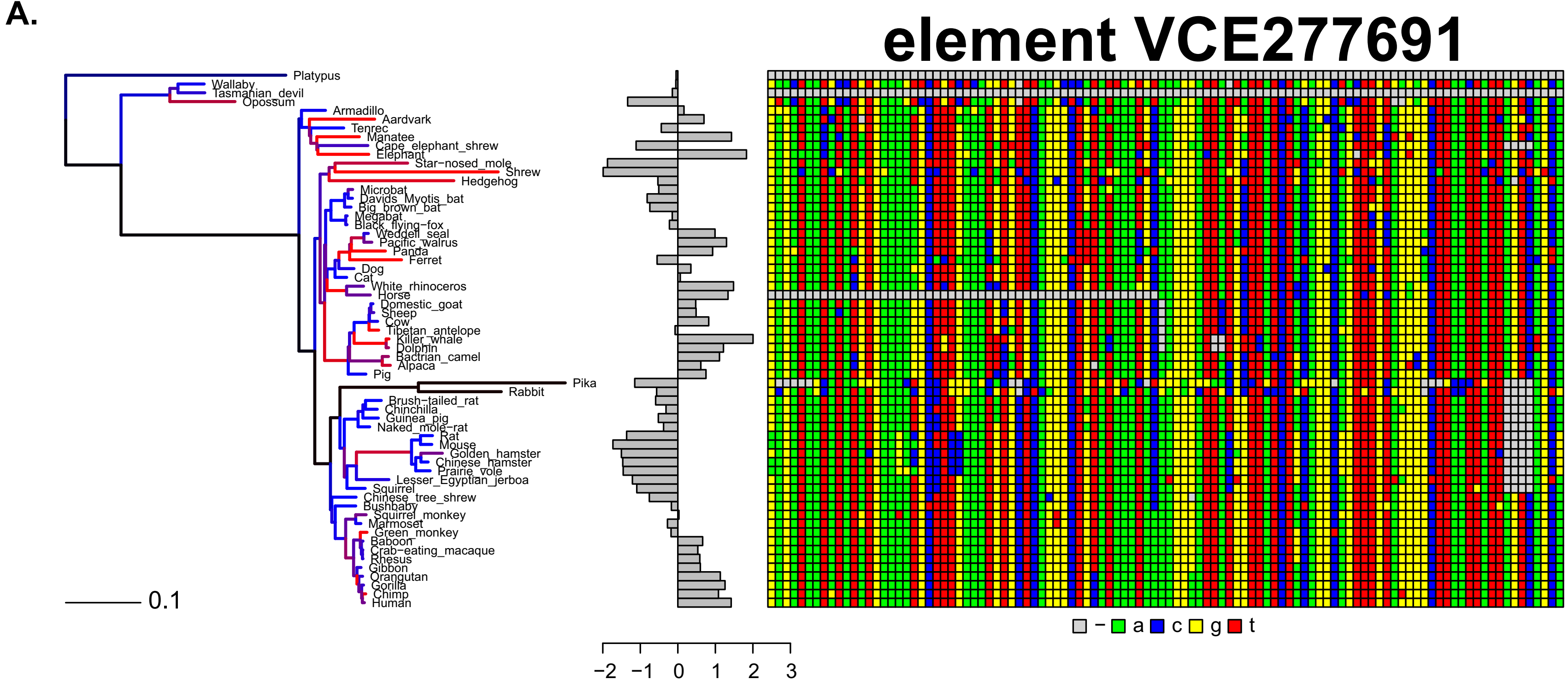
$$r_2 > 1 \quad if\ Z_s = 2\ \text{accelerated}$$

- estimate $log\,(\beta_3\ /\beta_2\ )$
- **can reveal both positive and negative associations of molecular and phenotypic rates**

Gemmell et al. 2024. *PLoS Comp. Biol.* (

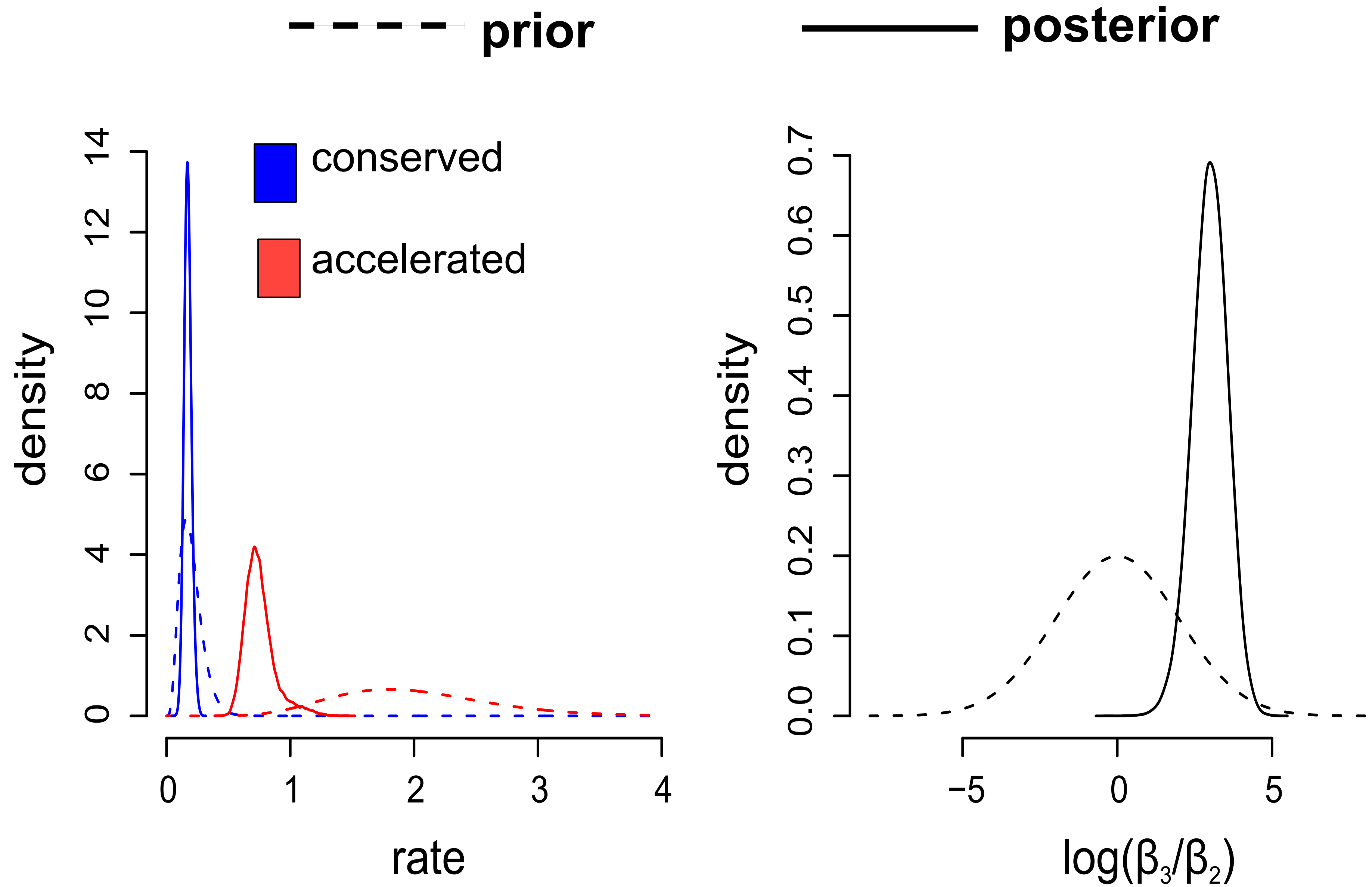
Linking sequence evolution and continuous trait data

Alignment of conserved noncoding element VCE277691



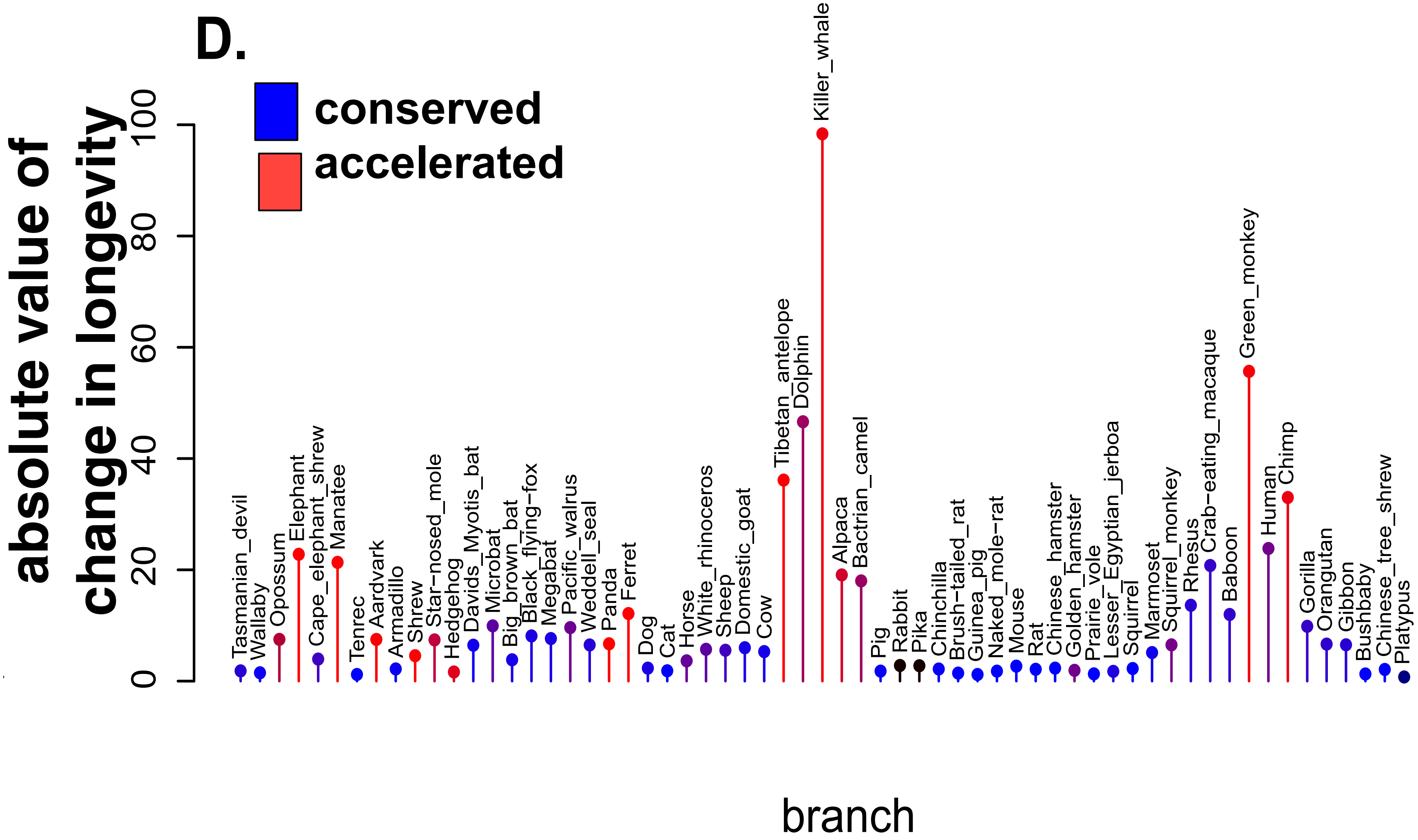
Gemmell et al. 2024. *PLoS Comp. Biol*. (4): e1011995.

Informative posterior means for molecular and phenotypic rates: CNEE VCE277691 as an example



Gemmell et al. 2024. *PLoS Comp. Biol*. (4): e1011995.

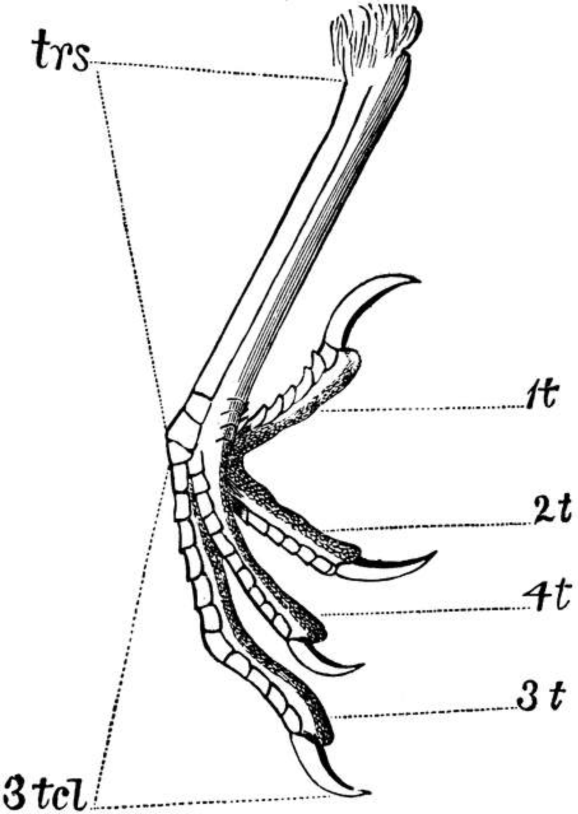
Molecular accelerations are associated with bigger changes in longevity (CNEE VCE277691)



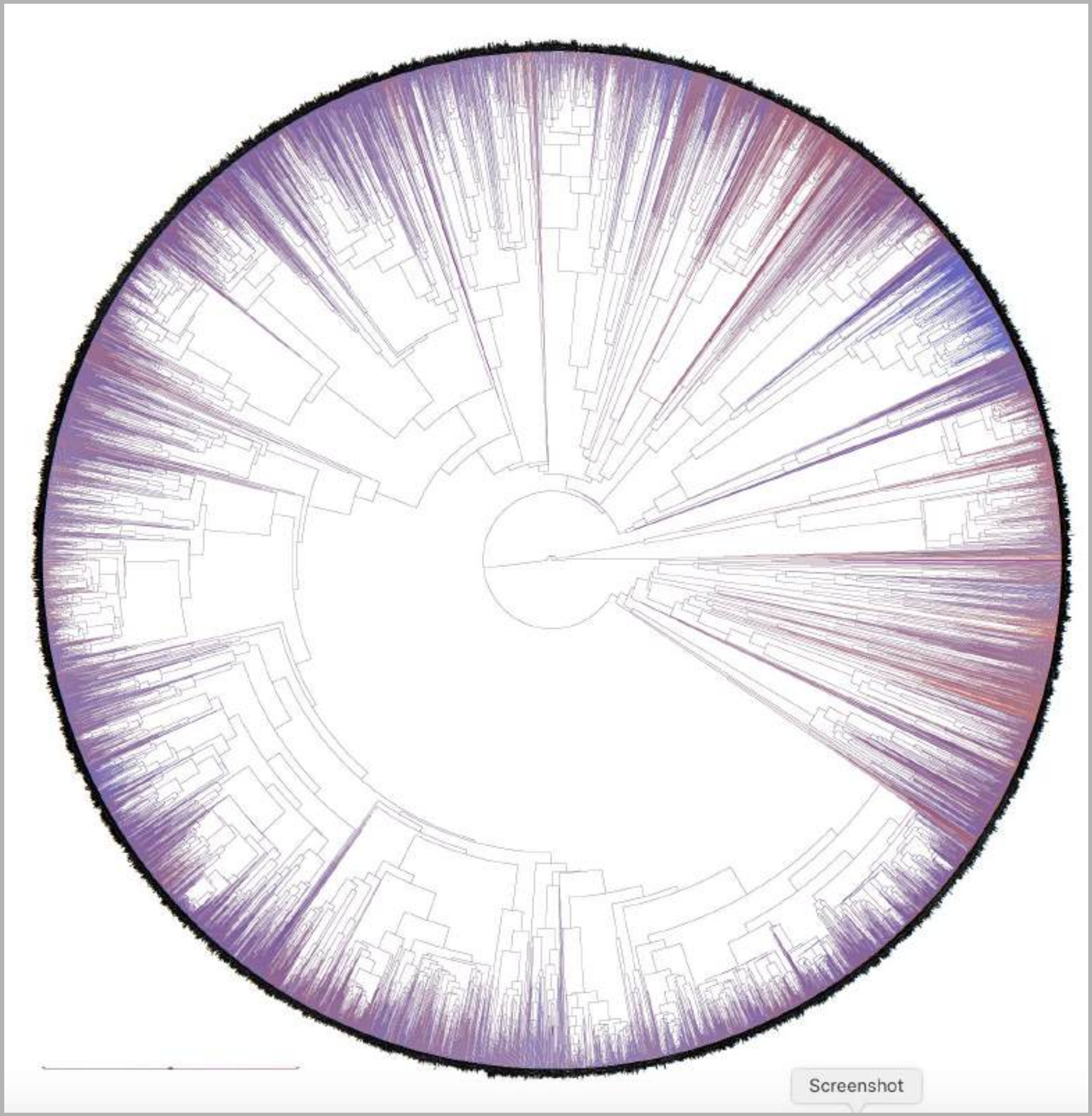
Gemmell et al. 2024. *PLoS Comp. Biol*. (4): e1011995.

Tarsus length available for all 10,800 species of bird

Subir Shakya



Shakya et al. 2025. *BMC Biol.*

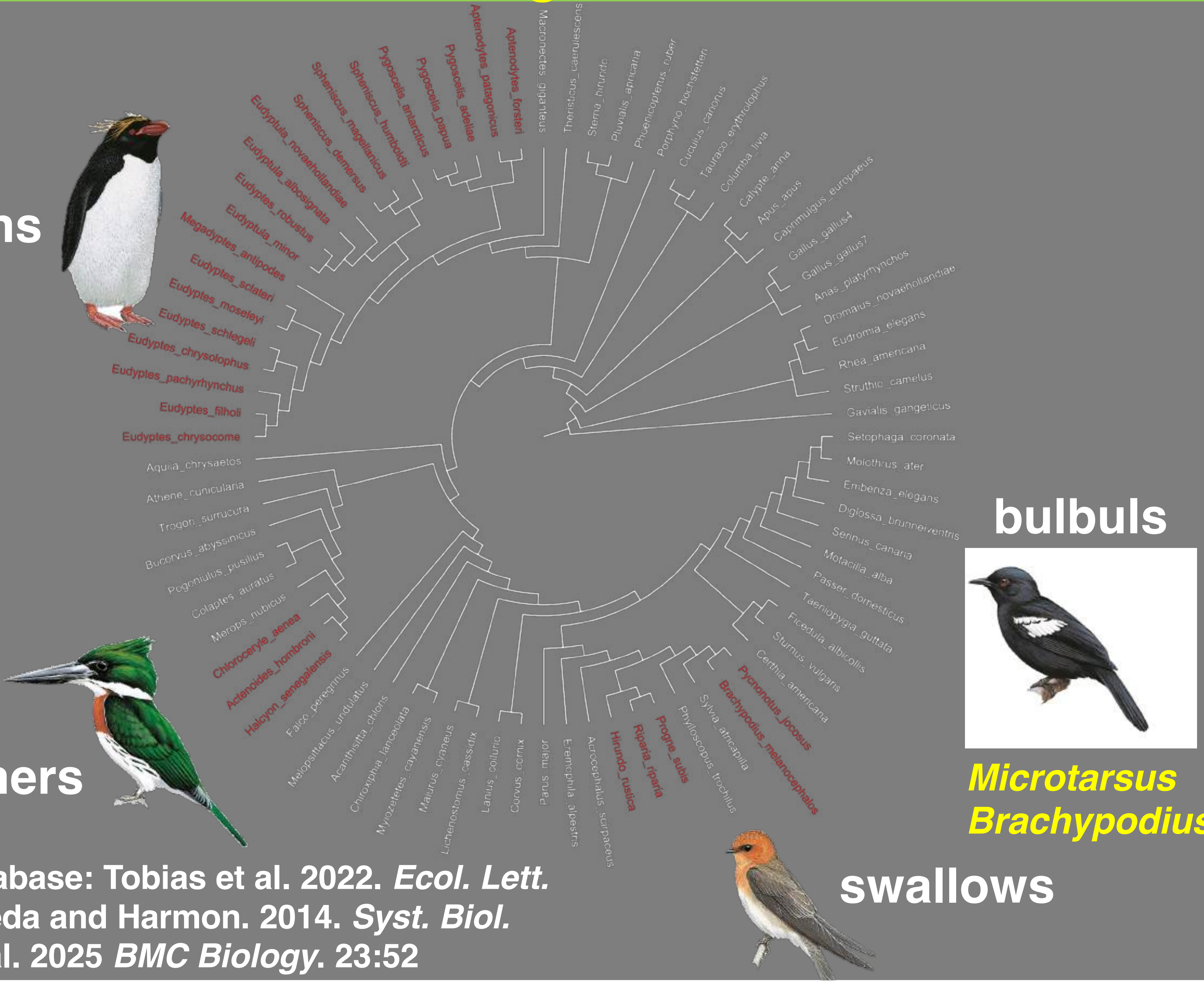


short **long**

Elliot Coues *Key to North American Birds* 1884

Avonet database: Tobias et al. 2022. *Ecol. Lett.*

Four groups of birds exhibit shifts to shorter tarsus length



penguins

kingfishers

bulbuls

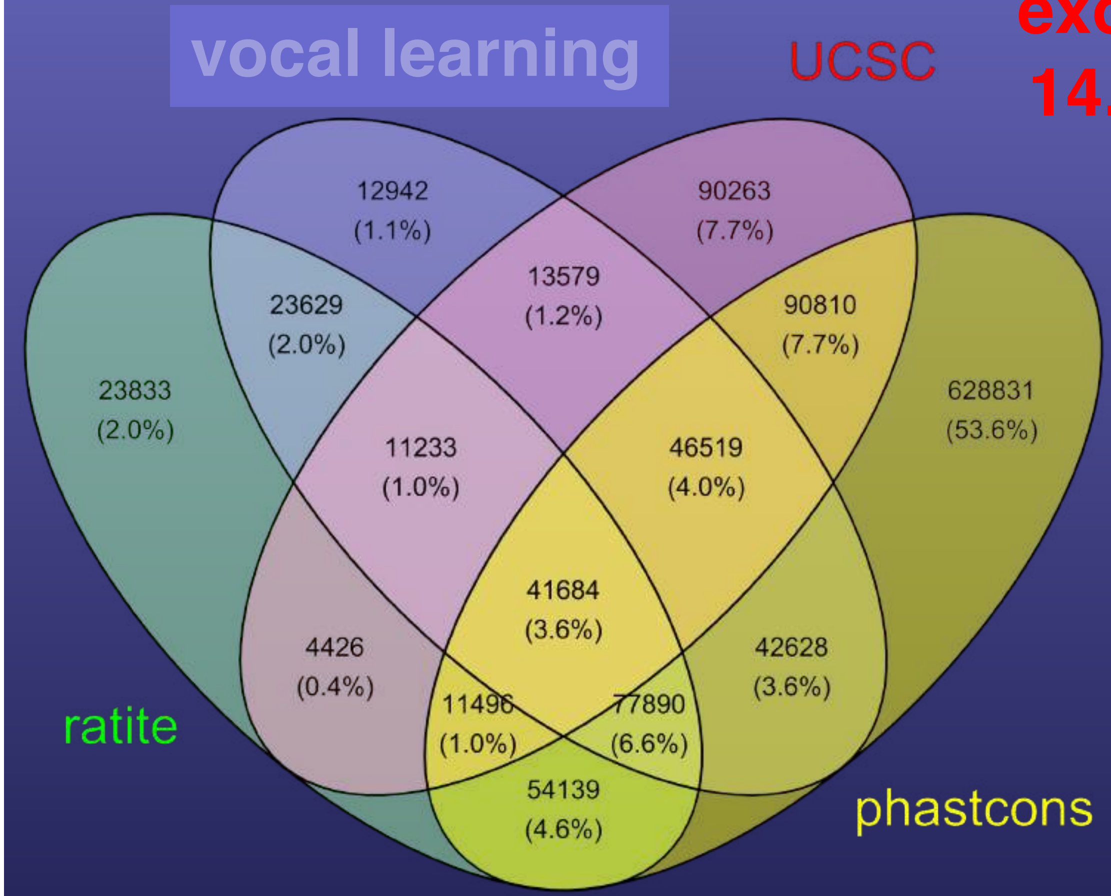
Microtarsus
Brachypodius

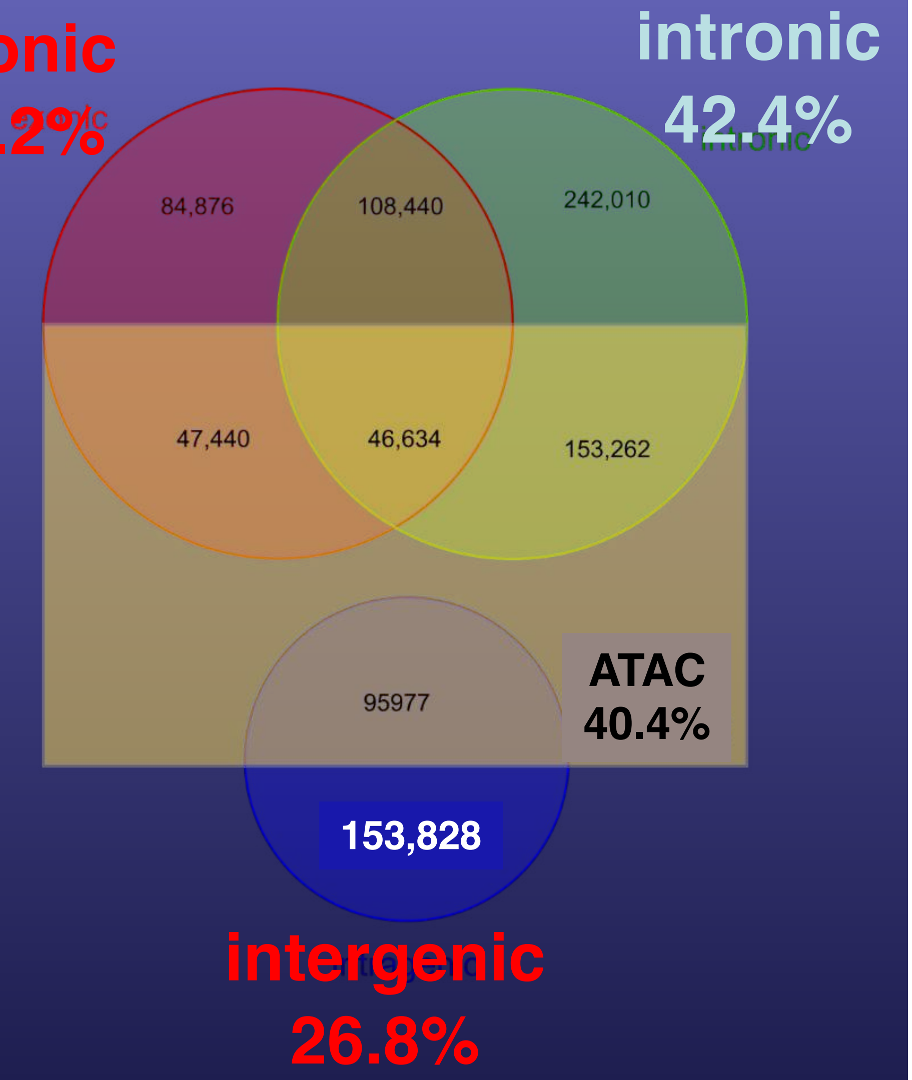
swallows

Avonet database: Tobias et al. 2022. *Ecol. Lett.*
Bayou: Uyeda and Harmon. 2014. *Syst. Biol.*
Shakya et al. 2025 *BMC Biology*. 23:52

A large collection of ~930,000 conserved elements

Four data sets combined



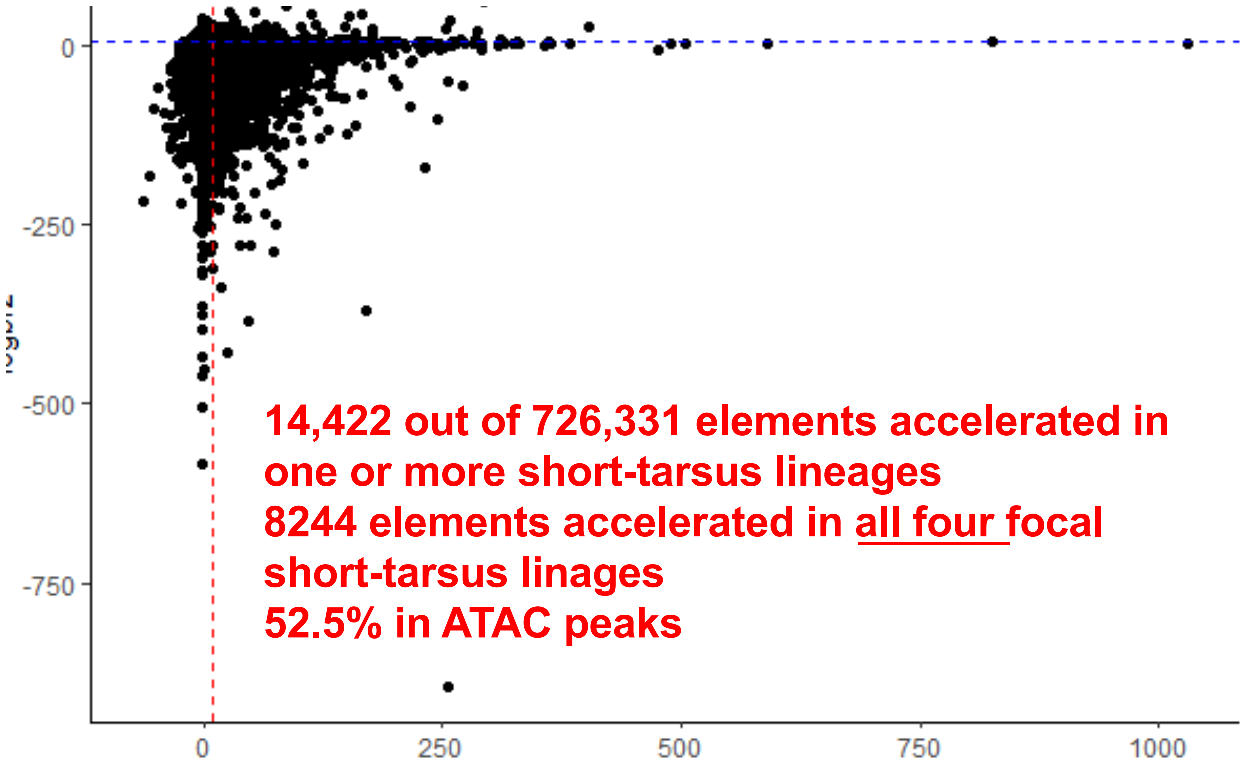


Sackton et al. 2019. *Science*
Cahill et al. 2021. *Genome Res.*
UCSC Genome Browser
Shakya et al. 2025 *BMC Biology*. 23:52

PhyloAcc identifies ~14,000 elements accelerated in short-tarsus lineages

347 elements where the target-lineage model is better than the unrestricted model

52.7% in ATAC peaks



14,422 out of 726,331 elements accelerated in one or more short-tarsus lineages

8244 elements accelerated in <u>all four</u> focal short-tarsus linages

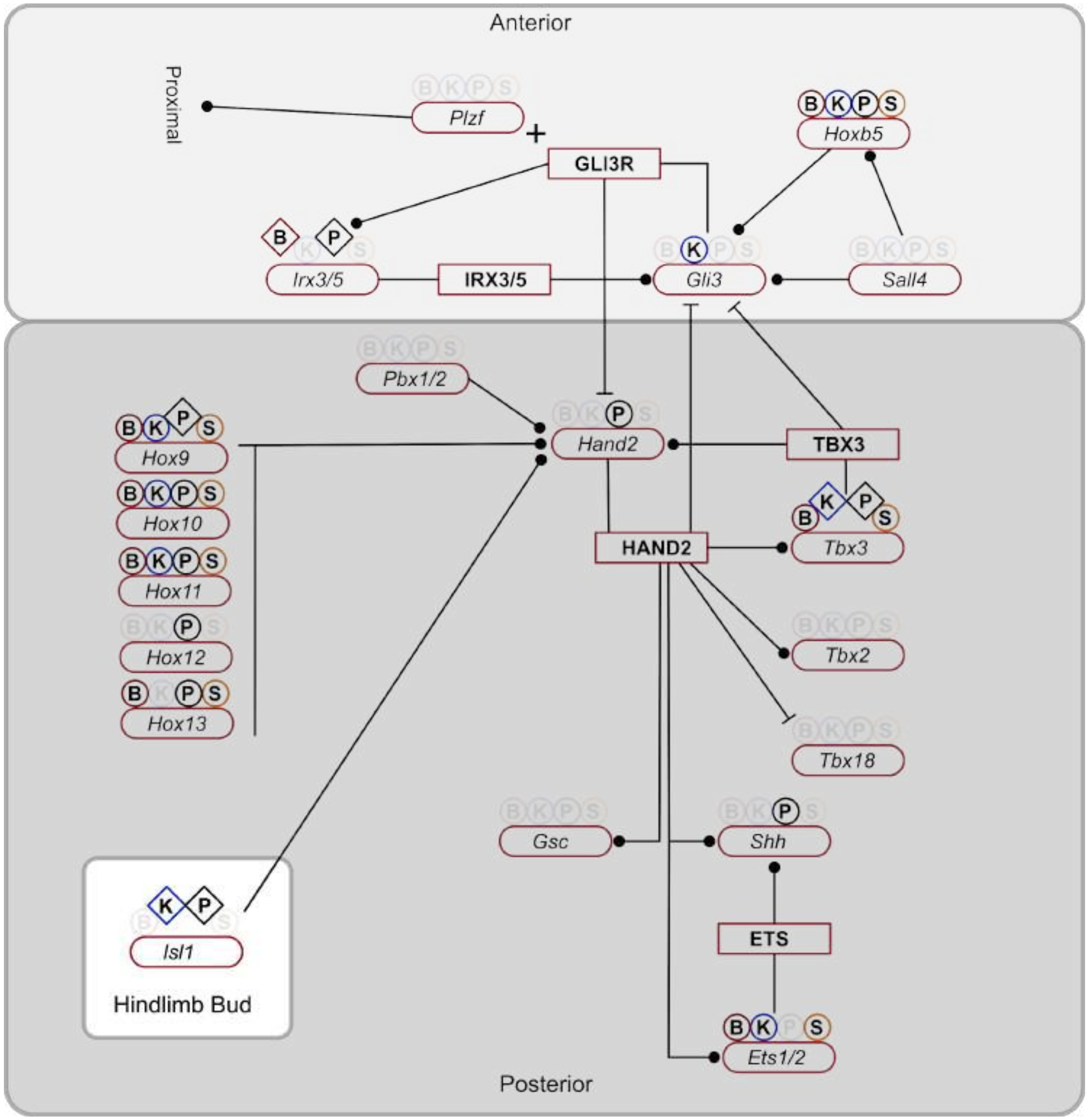
52.5% in ATAC peaks

Bayes Factor 2: target lineage model vs. unrestricted model

Bayes Factor 1: no acceleration vs. acceleration in targeted lineages

Shakya et al. 2025. *BMC Biol.*

Accelerated conserved elements near genes involved in limb bud development, field positioning, AER formation and axis patterning



Shakya et al. 2025. *BMC Biol.*

Limb Field Pre-patterning

B Bulbul

K Kingfisher

P Penguin

S Swallow

◇ Exclusively accelerated in target taxa

Pathway from Zuniga and Zeller, 2020

Surface scanning comparative phenotypic data for avian morphology in the Museum of Comparative Zoology

sternum
Tundra Swan
Cygnus columbianus– MCZ 343048

humerus
Northern Screamer
Chauna chavaria– MCZ 340307

tibiotarsus
Brant Goose
Branta bernicla– MCZ 336993

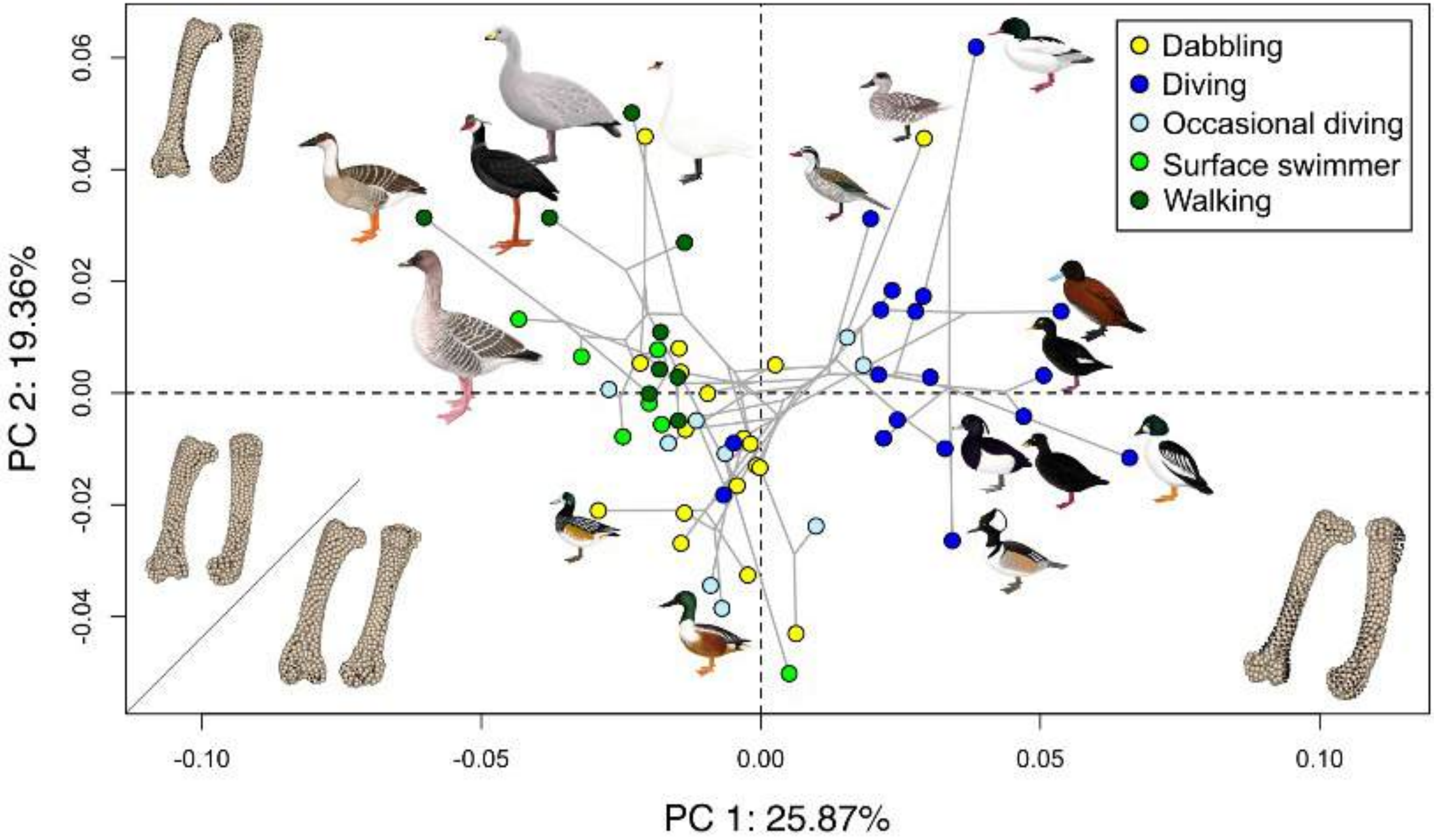
femur
Velvet Scoter
Melanitta fusca– MCZ 348703

Genomic basis of skeletal variation in ducks (Anseriformes)

- correlations with foraging ecology
- loss of flight

Shakya et al. 2025. in revision, *Evolution*

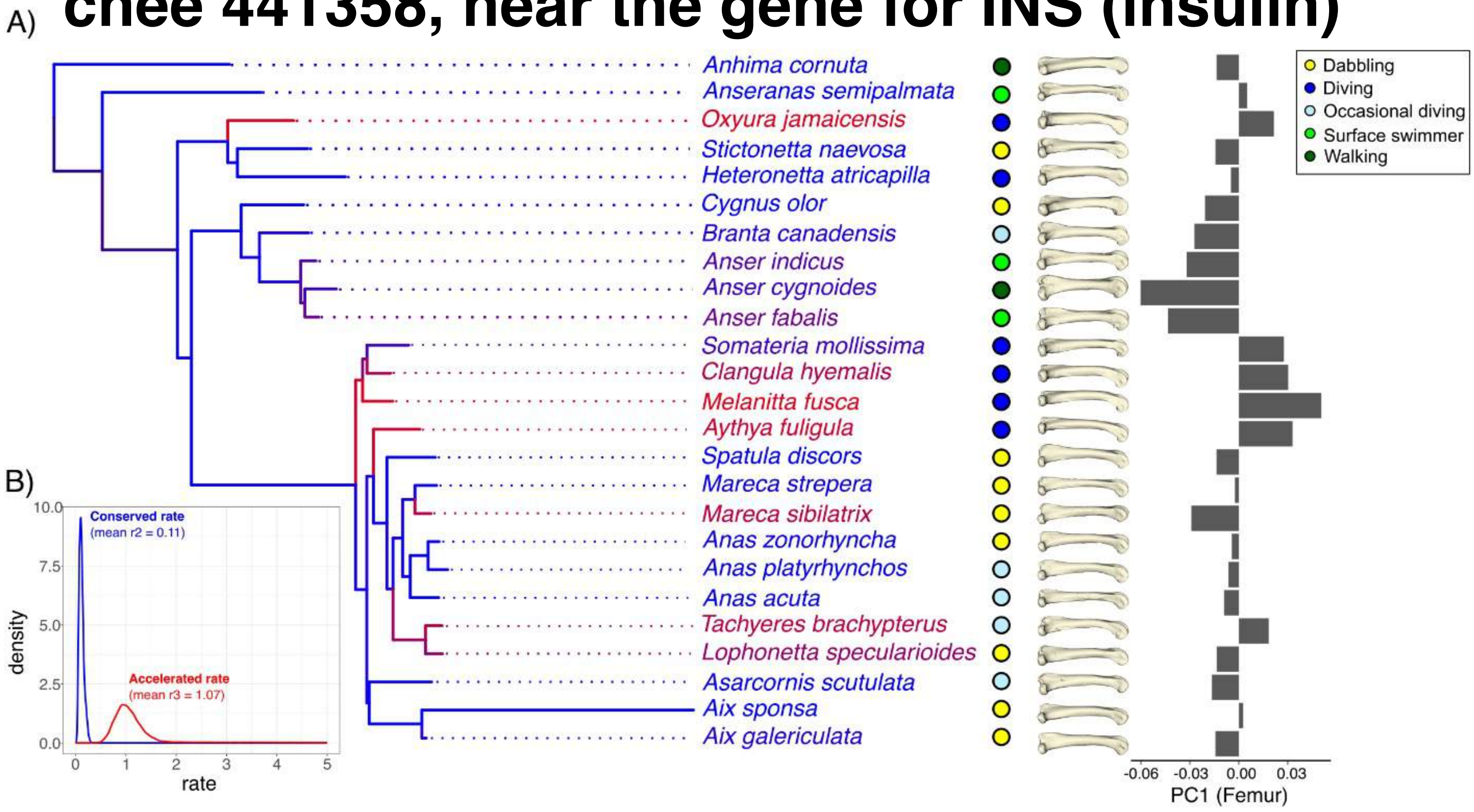
Femur shape in ducks correlates with foraging behavior



PhyloAcc-C identifies CNEEs associated with femoral shape

~877 k CNEEs analyzed

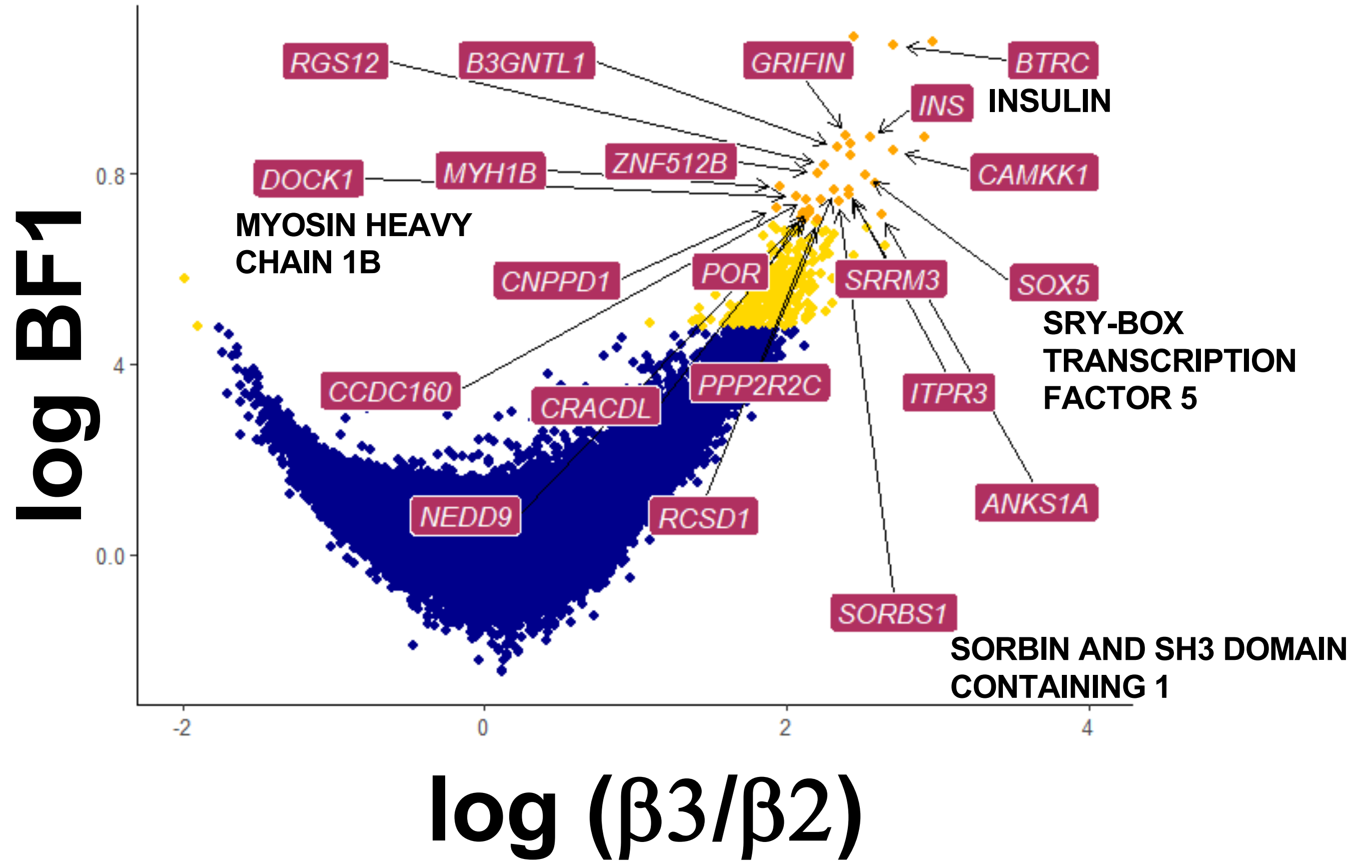
cnee 441358, near the gene for INS (insulin)



Shakya et al. 2025. in revision, *Evolution*

PhyloAcc-C identifies multiple CNEEs associated with rates of femoral evolution

21 CNEEs occur within 10 kb of a gene of relevance to femoral development



INSULIN

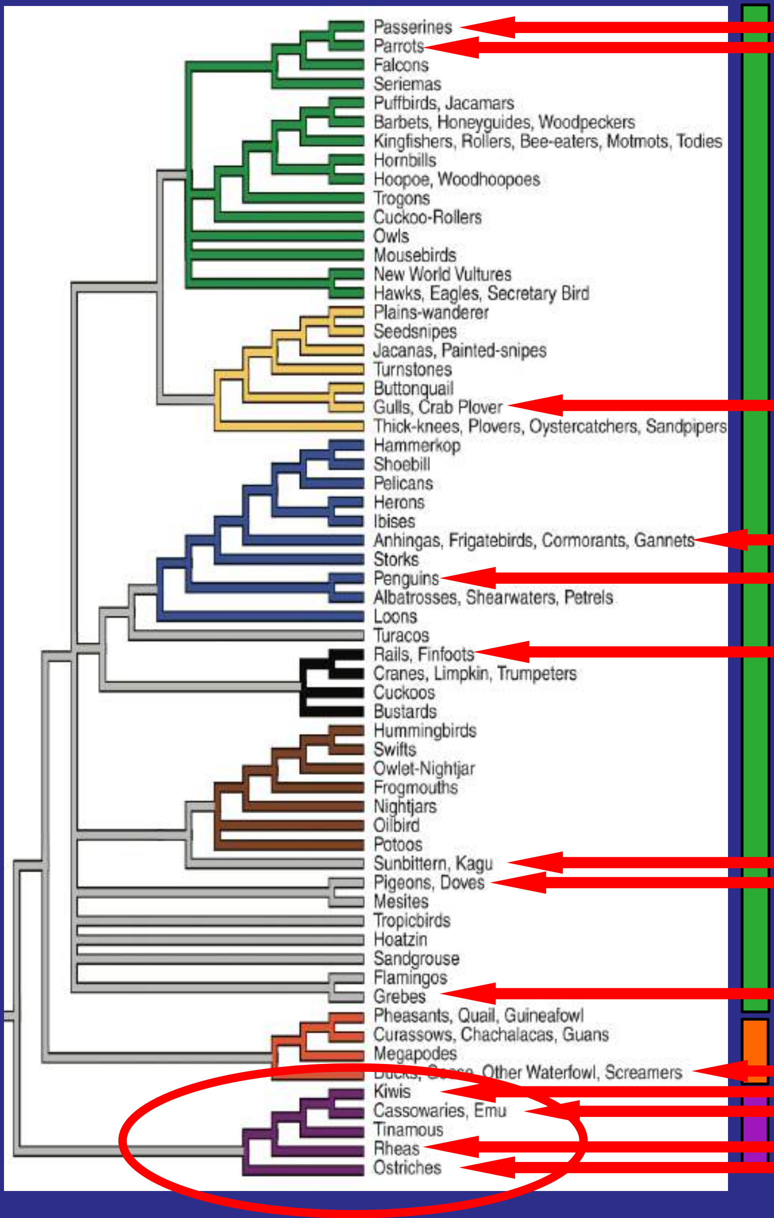
MYOSIN HEAVY CHAIN 1B

SRY-BOX TRANSCRIPTION FACTOR 5

SORBIN AND SH3 DOMAIN CONTAINING 1

Shakya et al. 2025. in revision, *Evolution*

Convergent loss of flight in birds: a complex trait



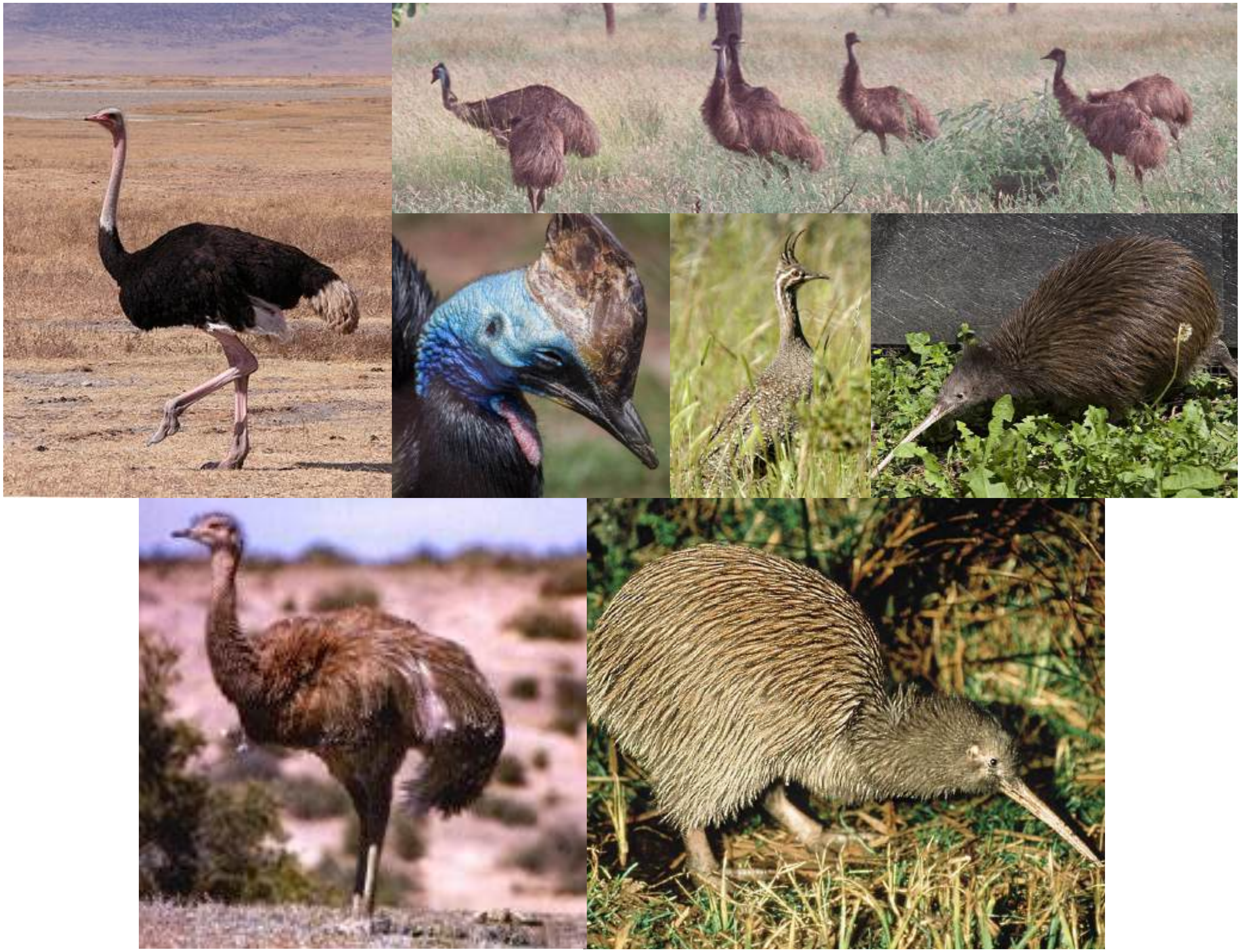
Neoaves ~9,500 species

Galloanserae ~450 species

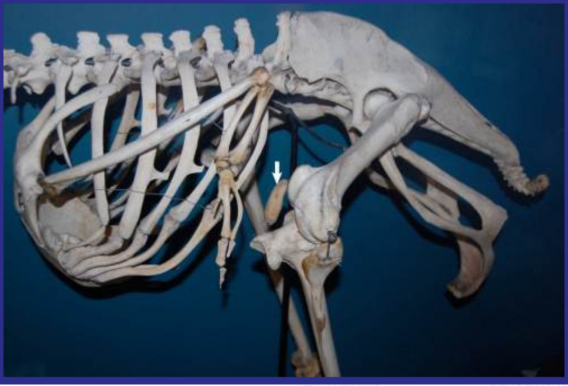
Paleognathae ~60 species

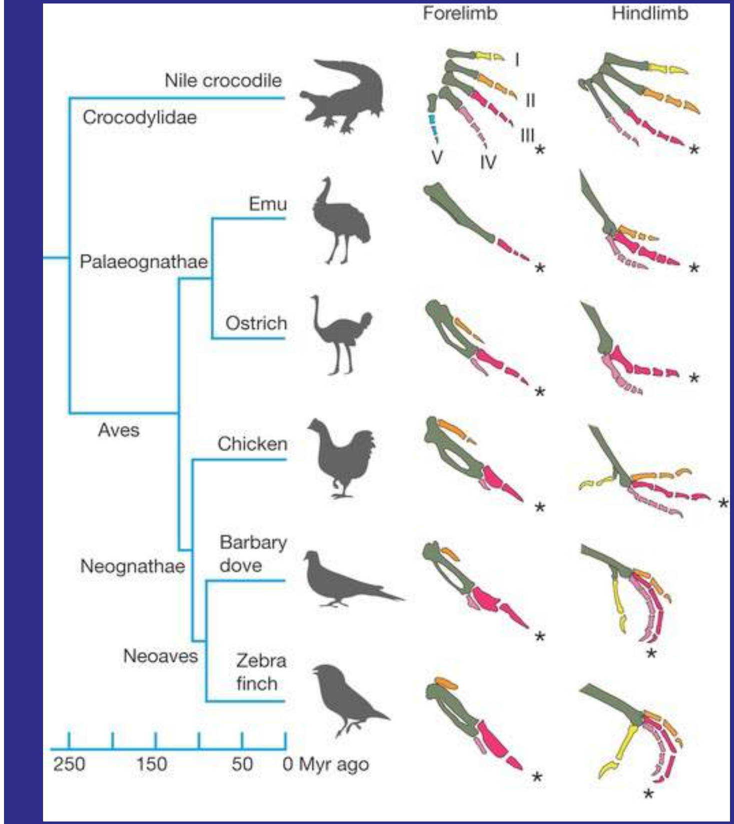
Hackett et al. 2008, *Science*

CNEEs and the convergent evolution of flightlessness in Palaeognathae



Skeletal modifications for flightlessness





Little-spotted kiwi sternum

Emu and ostrich keelless sterna

De Bakker et al. 2013. *Nature* 500, 445–448.

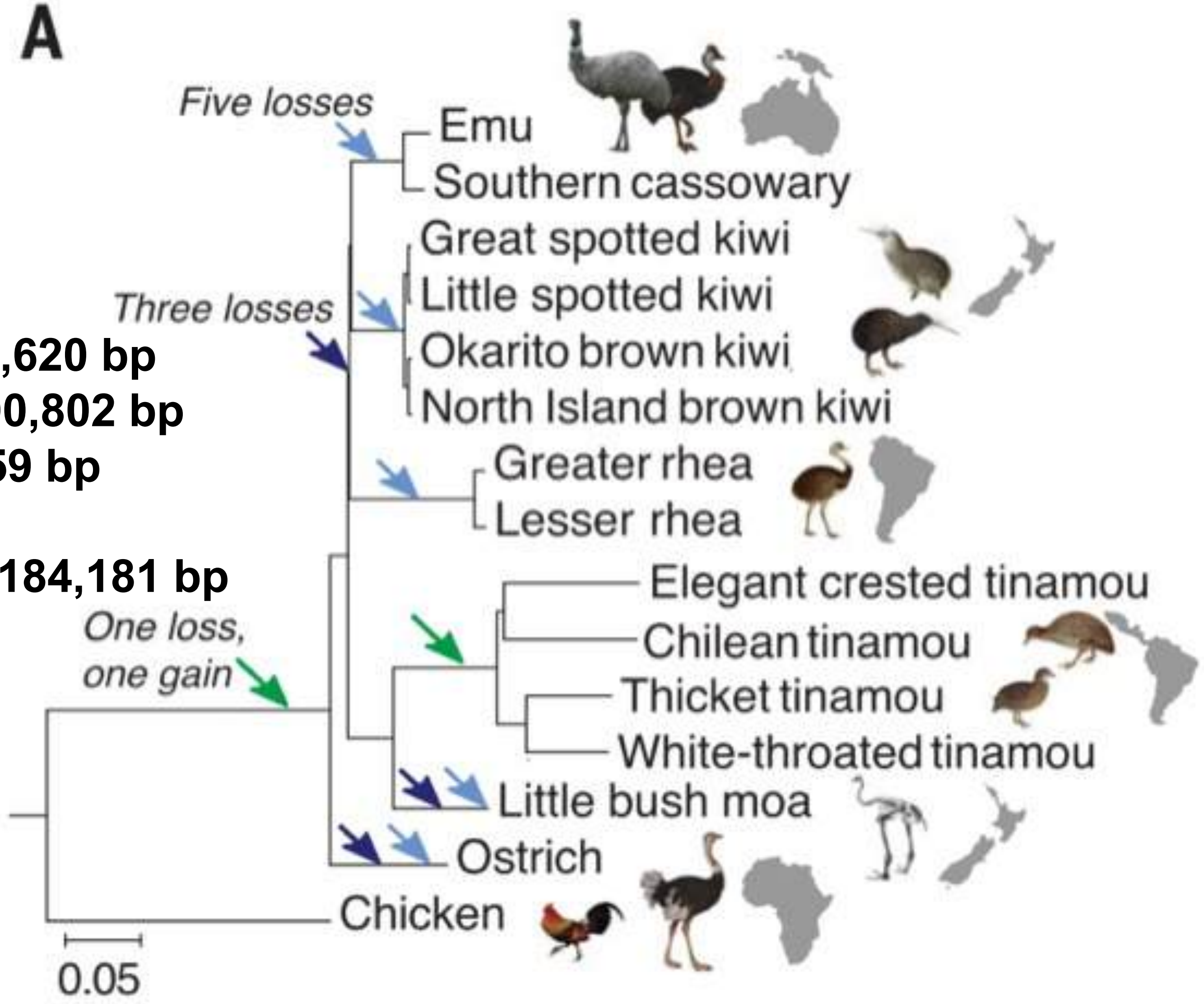
11 new paleognath genomes

Image (all CC): David Cook; Quartl; Jim, the Photographer, Tim Sackton

Coalescent* analyses place tinamous inside flightless radiation, suggesting convergent loss of flight

12,676 CNEEs - 4,794,620 bp
5,016 introns - 27,890,802 bp
3,158 UCEs - 8,498,759 bp

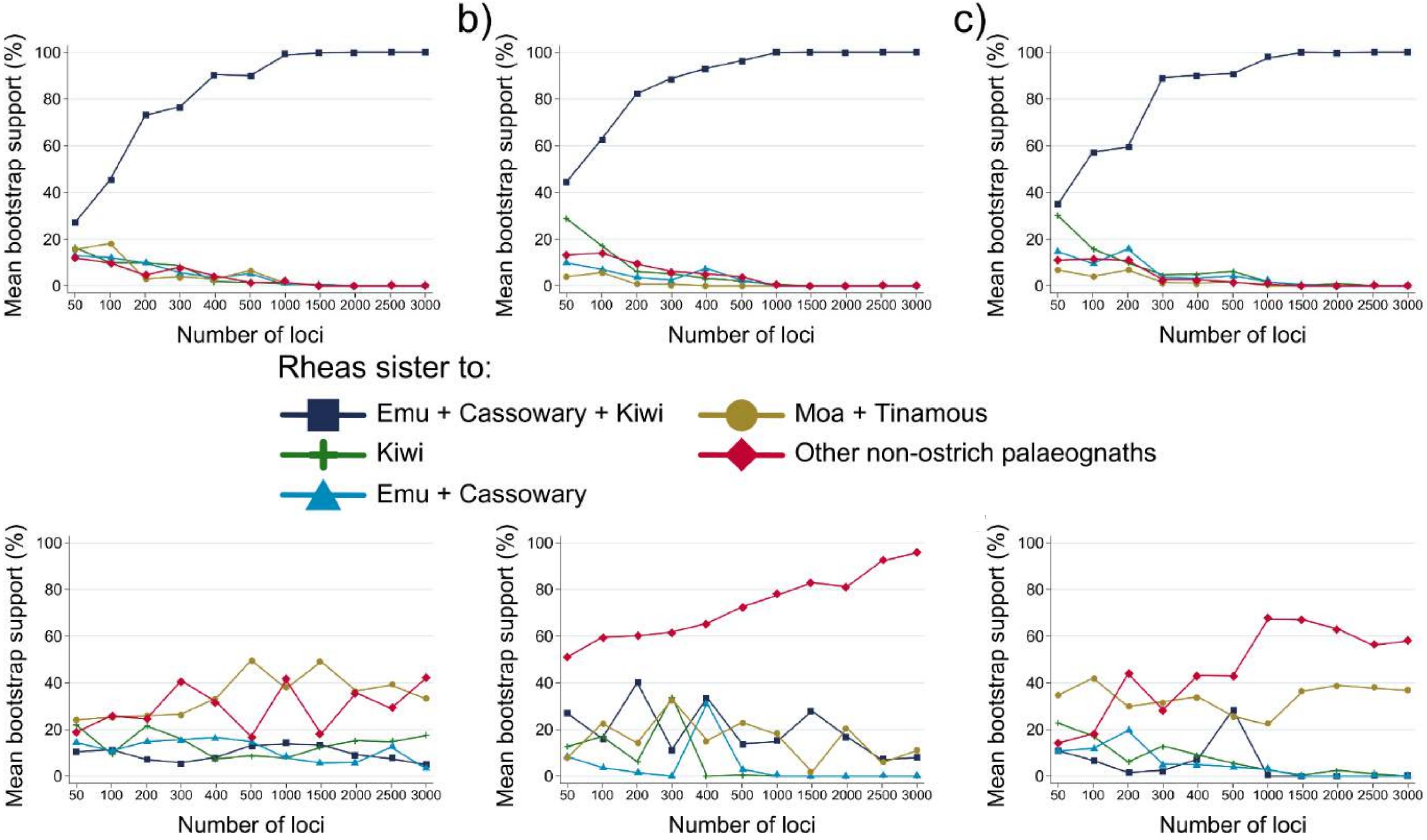
Total: 20,850 loci; 41,184,181 bp



MP-EST: Liu et al. 2010. *BMC Evol. Bic
Sackton et al. 2019. *Science* 364: 74-78

Phylogenetic subsampling and consistent accumulation of phylogenetic signal using MP-EST

CNEEs Introns UCEs



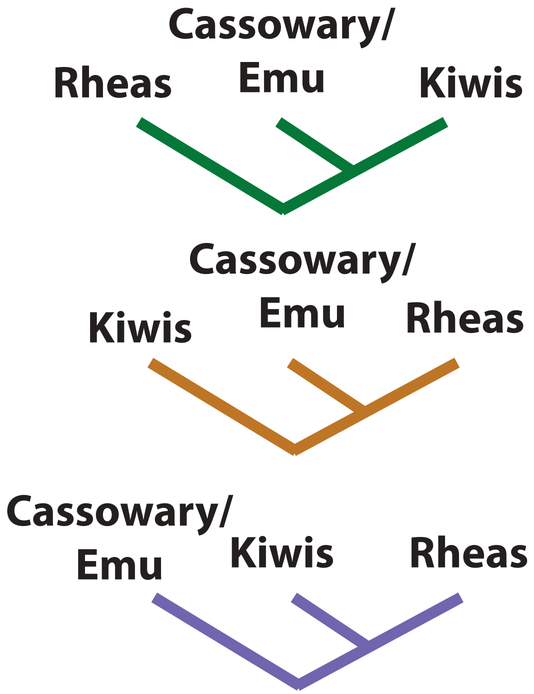
Cloutier et al. 2019. *Syst. Biol.*; Edwards, S. 2016. *Zool. Scripta*

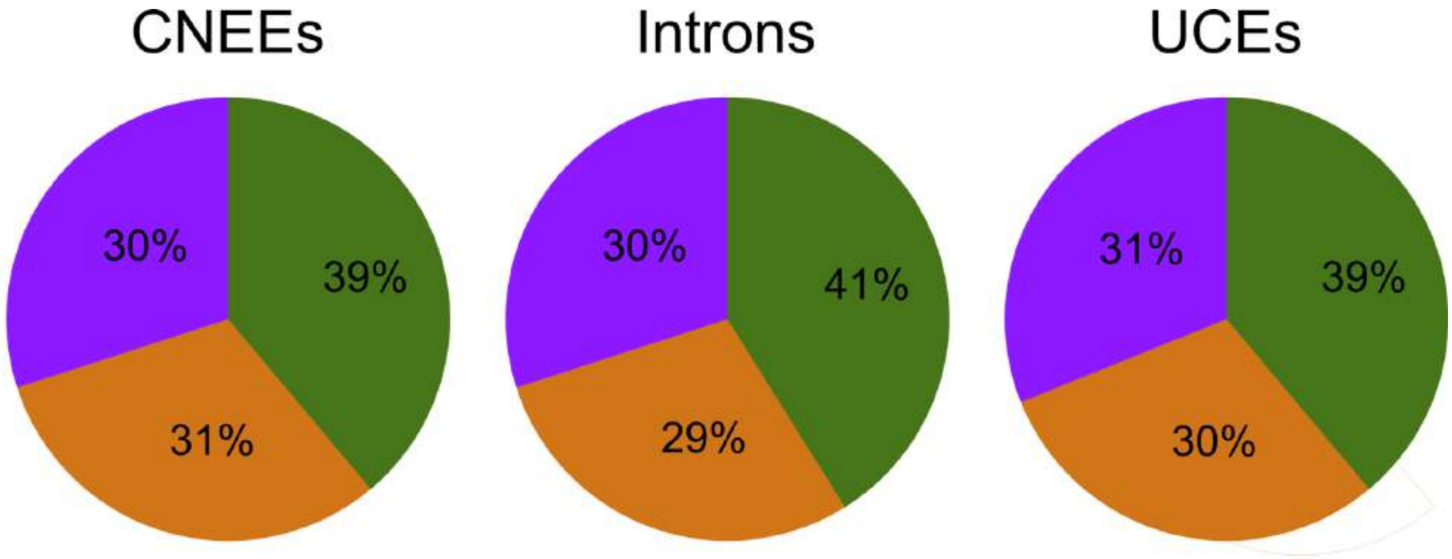
Gene tree distribution suggests a near polytomy at base of ratites

🟩 **Major topology**

🟧 **Minor topology 1**

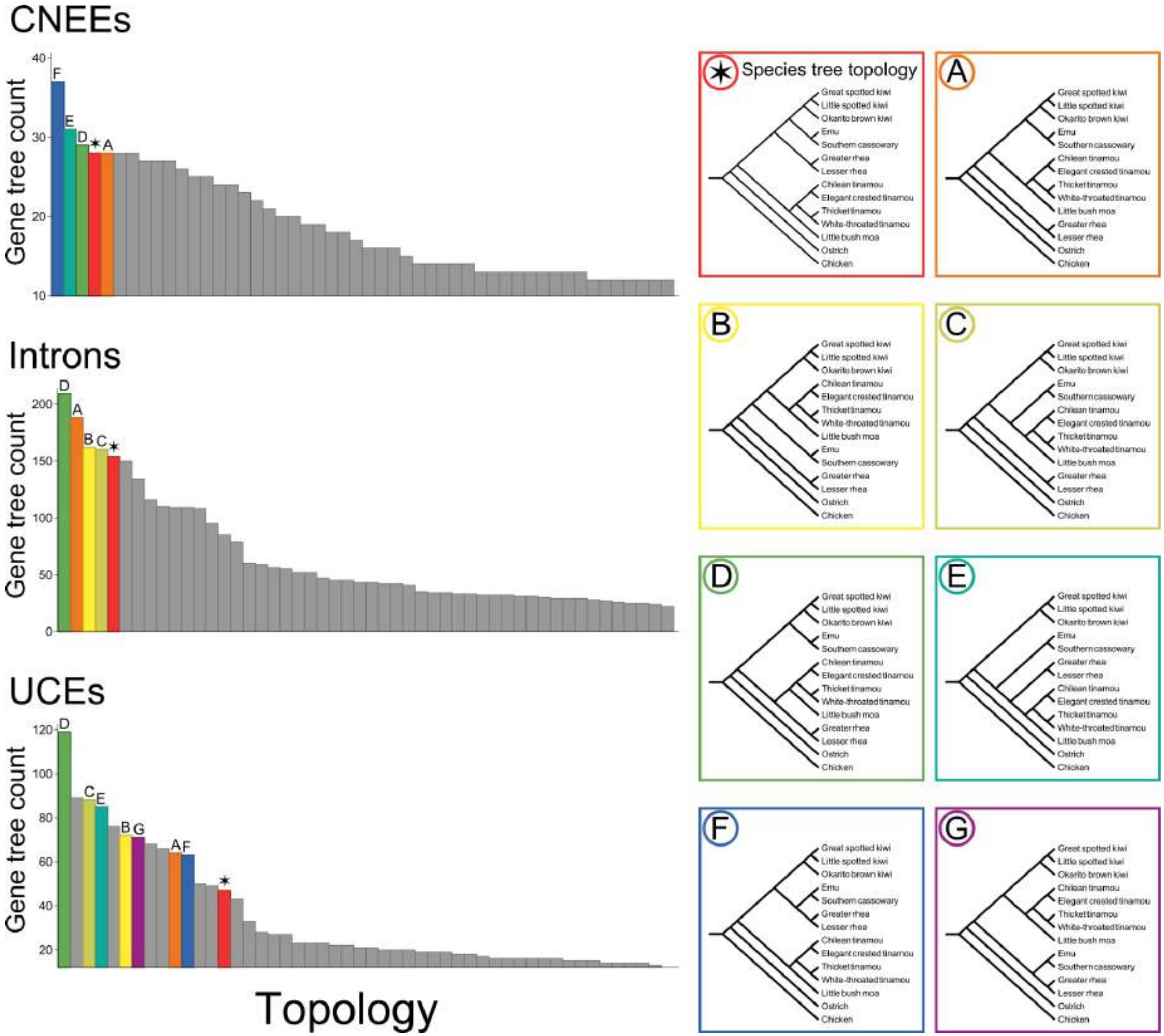
🟪 **Minor topology 2**





Cloutier et al. 2019. *Syst. Biol*. 10.1093/sysbio/syz019

Anomaly zone: most common gene tree does not match the species tree

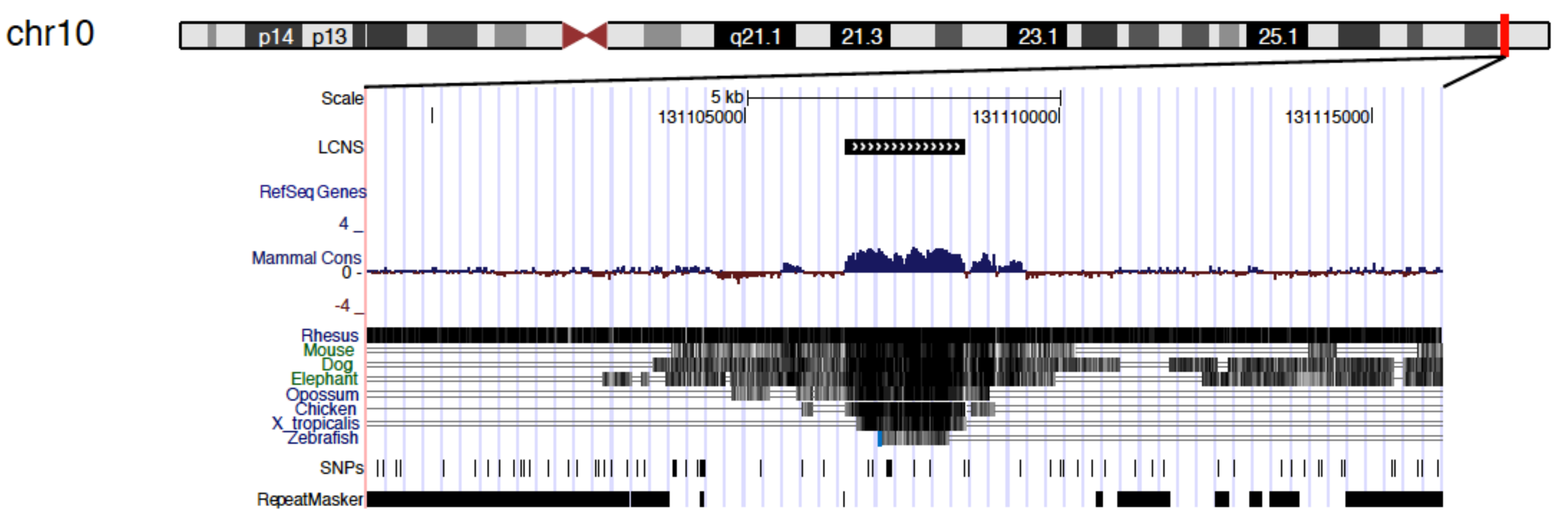


Cloutier et al. 2019. *Syst. Biol.* 10.1093/sysbio/syz019

CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements

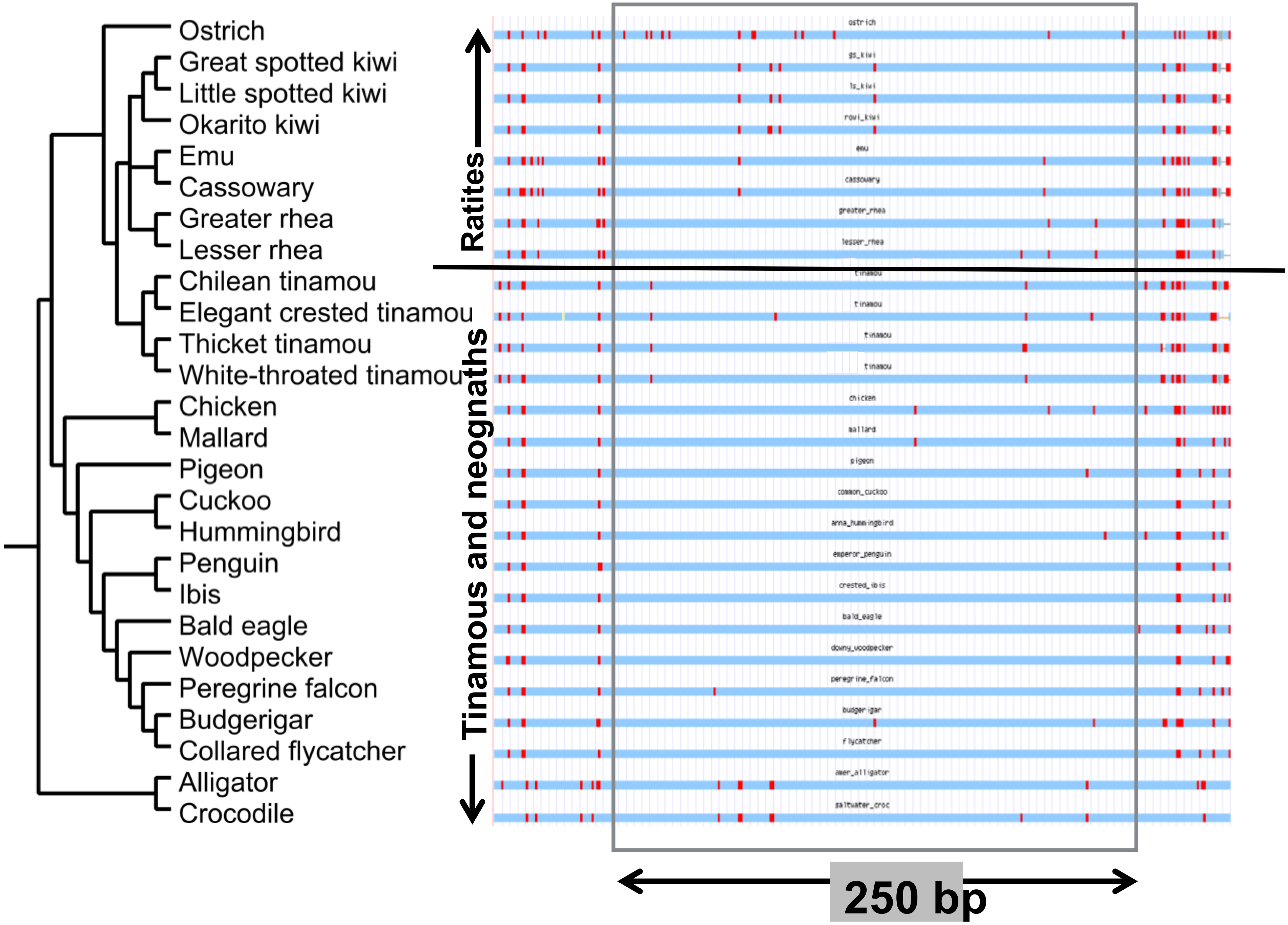
284,001 long (* > 50 bp) CNEEs in data set



View of a segment of human chromosome 10 using UCSC Genome Browser

Janes et al. (2011) *Genome Biol. Evol.* 3:102–113

Convergent loss of function of CNEEs in ratite lineages

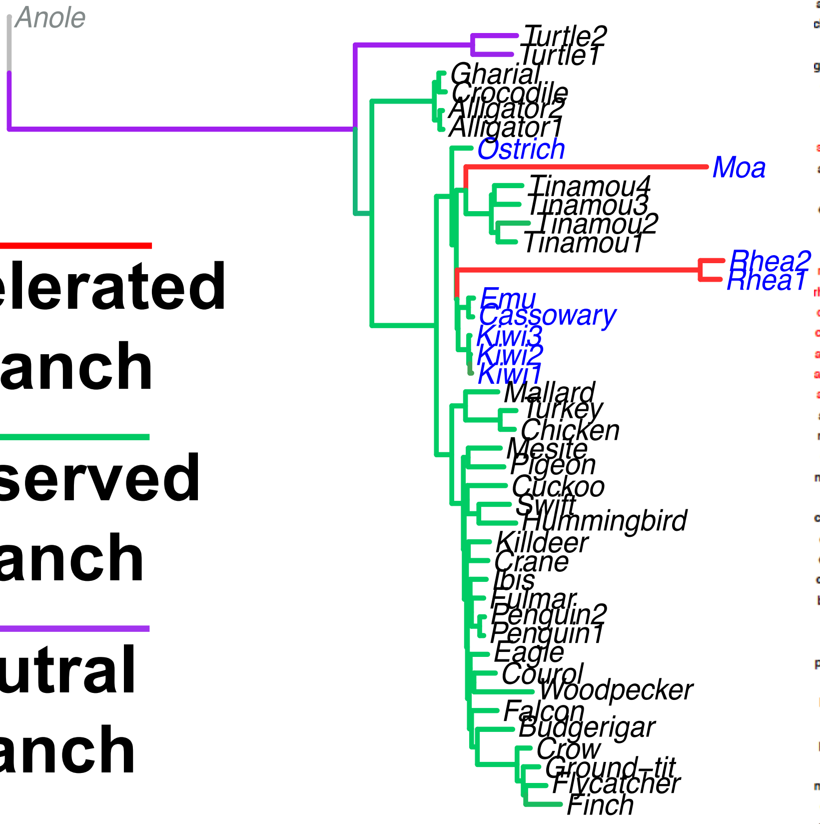


A convergently accelerated CNEE detected with a novel Bayesian method

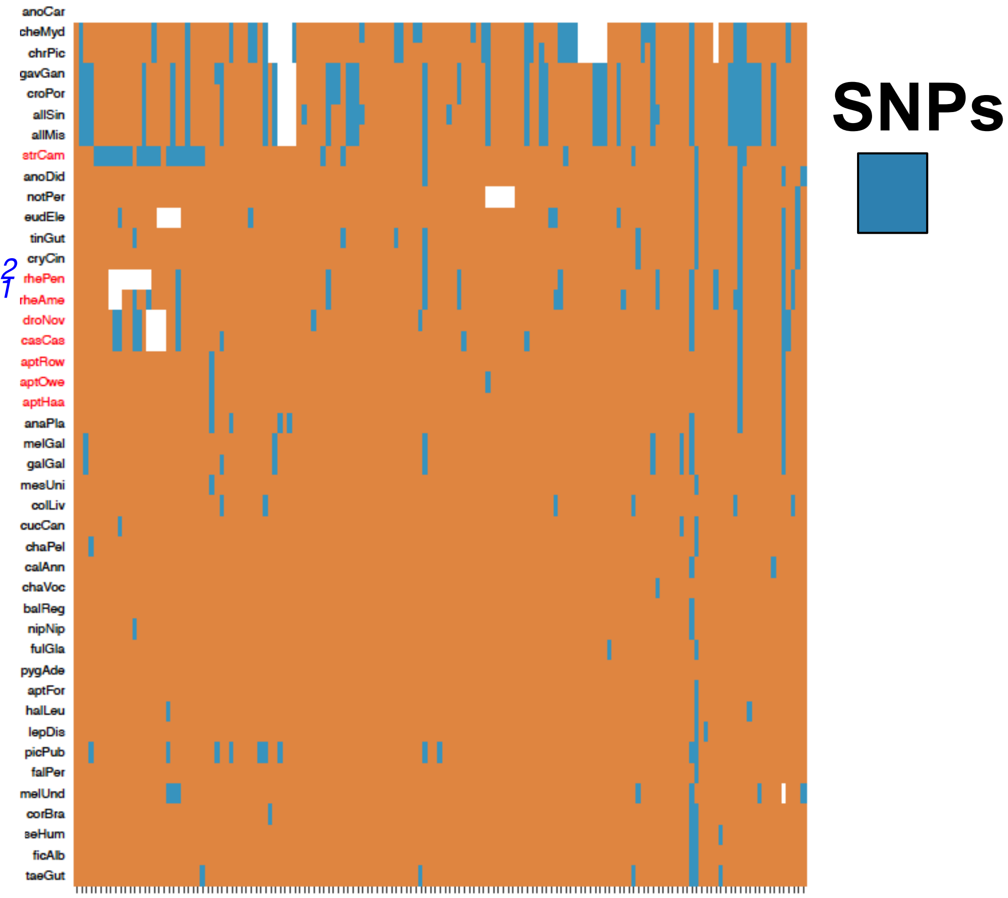
accelerated branch

conserved branch

neutral branch

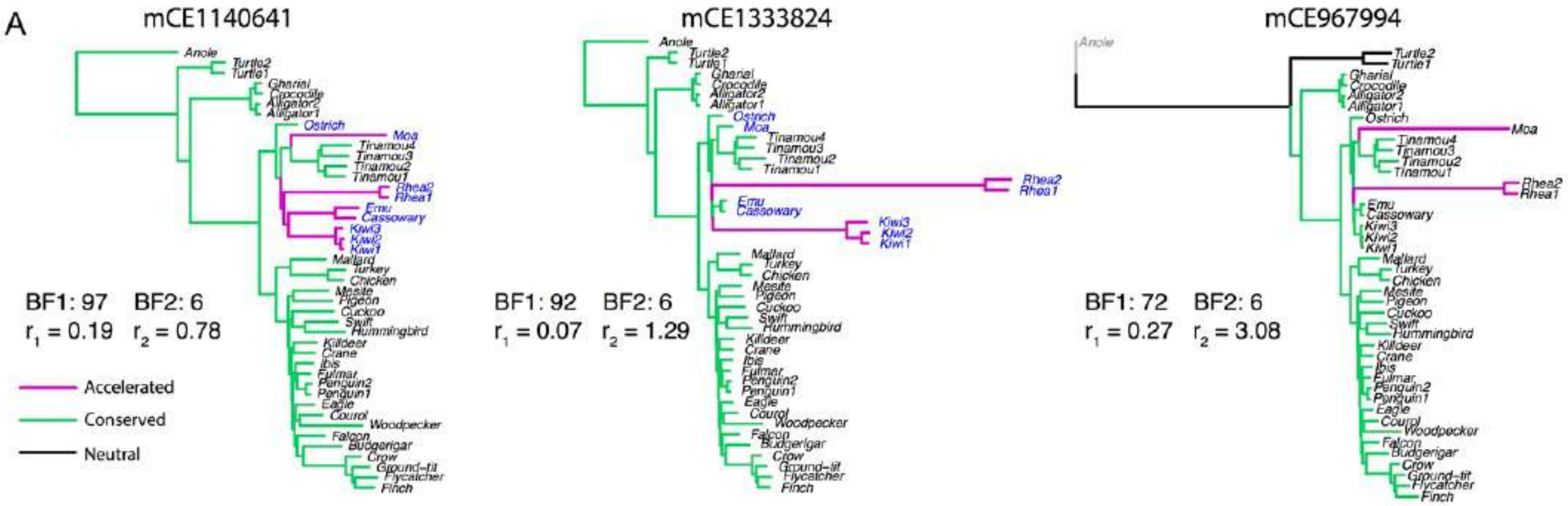


Branch lengths relative to conserved rate



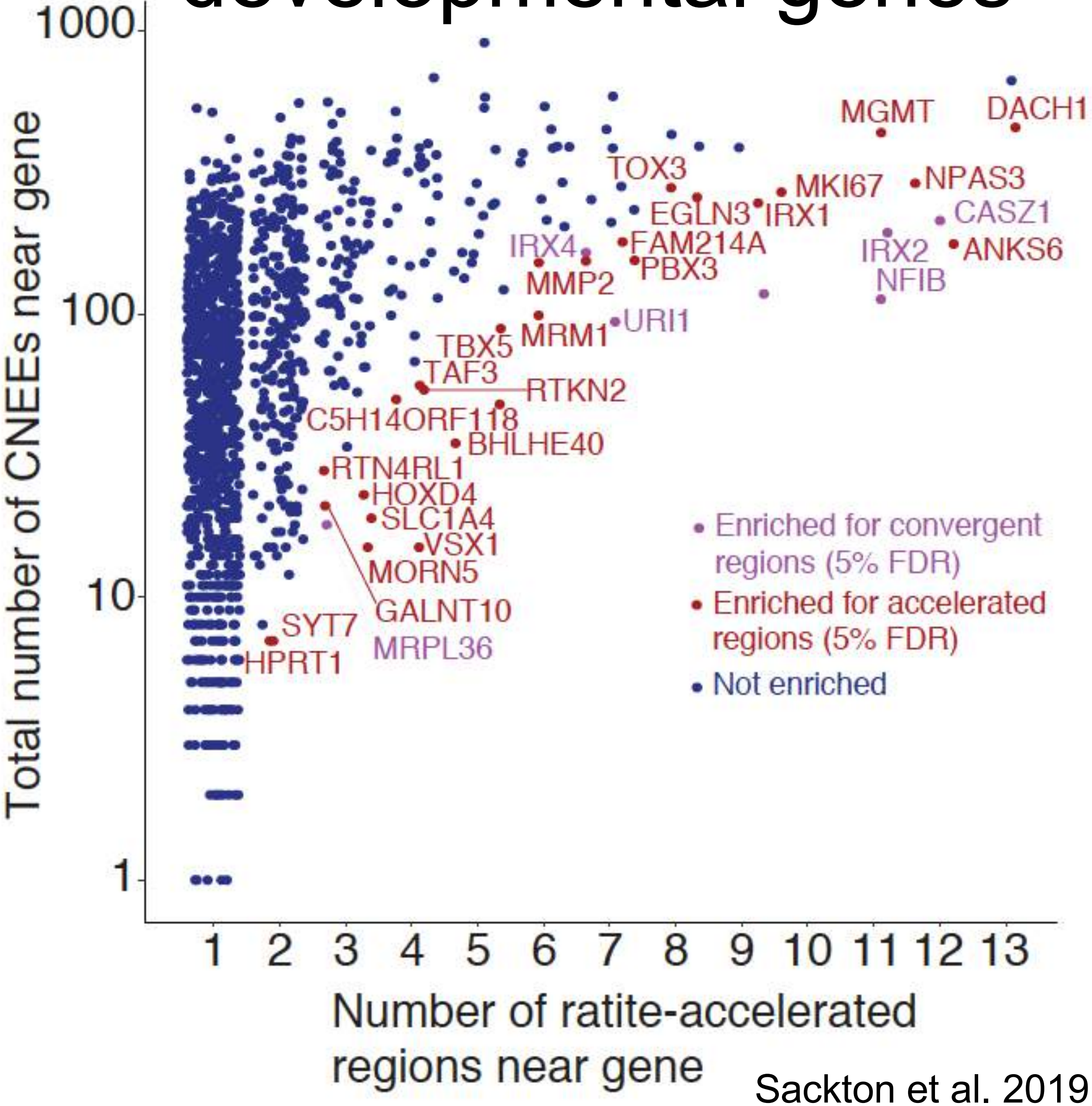
Hu, Z., et al. 2019. *Mol. Biol. Evol*. 36: 1086

2335 CNEEs (~0.8 %) evolve rapidly in ratites



Hu, Z., et al. 2019. *Mol. Biol. Evol*. 36: 1086

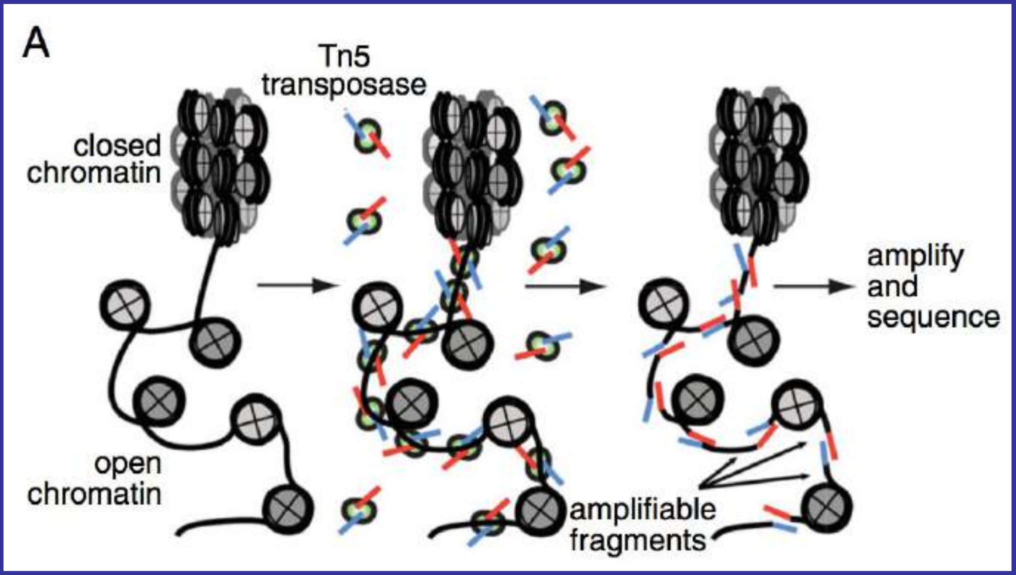
Rapid regulatory evolution near developmental genes



Sackton et al. 2019. *Science* 364: 74-78

Assay for Transposase-Accessible Chromatin

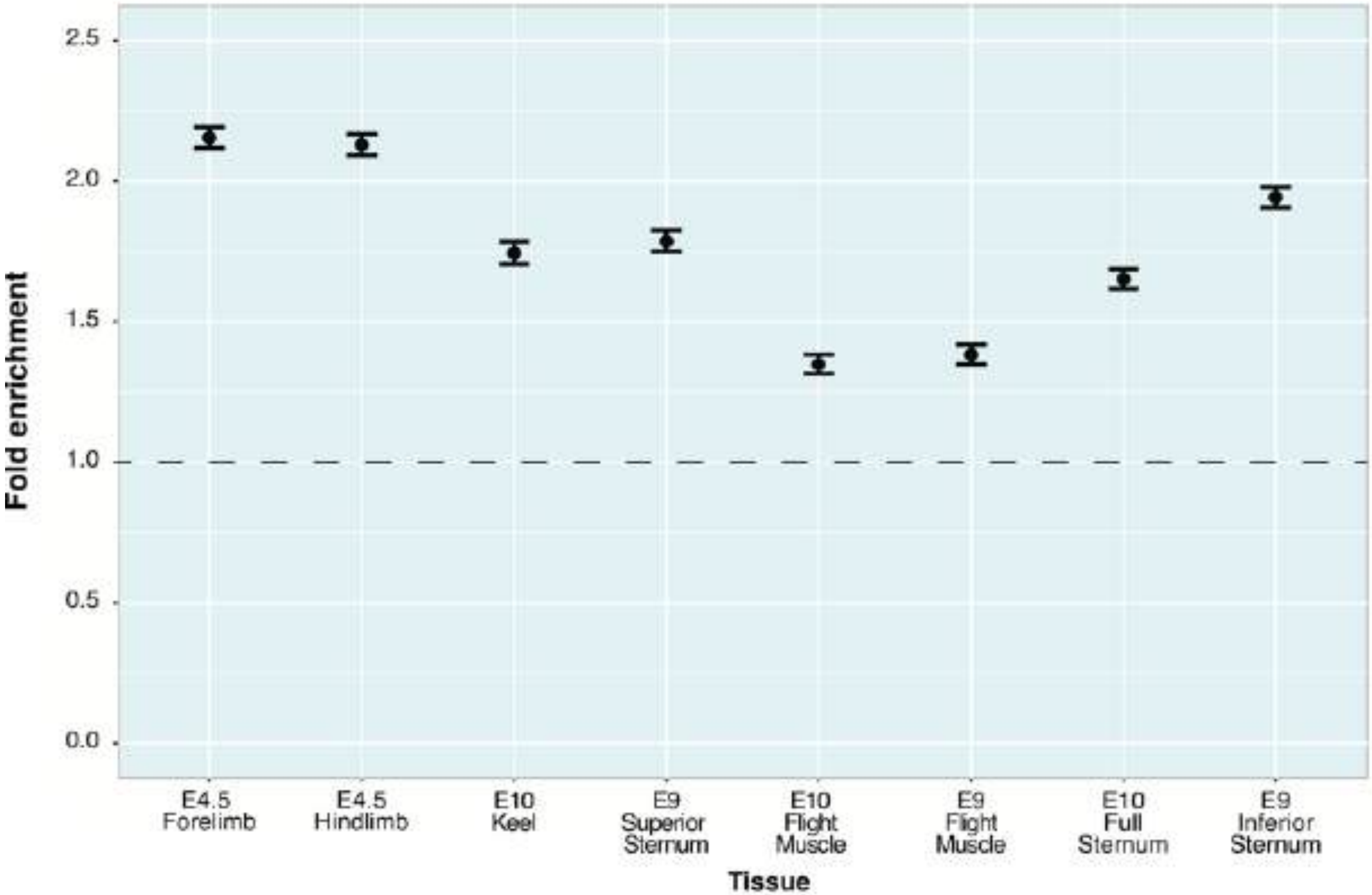
ATAC-Seq identifies DNA with open chromatin, accessible to transcription factors



Buenrostro et al. 2015. Curr Protoc.Biol. 2015; 109: 21.29.1–21.29.9.

Stage HH24-25 chickens and rheas

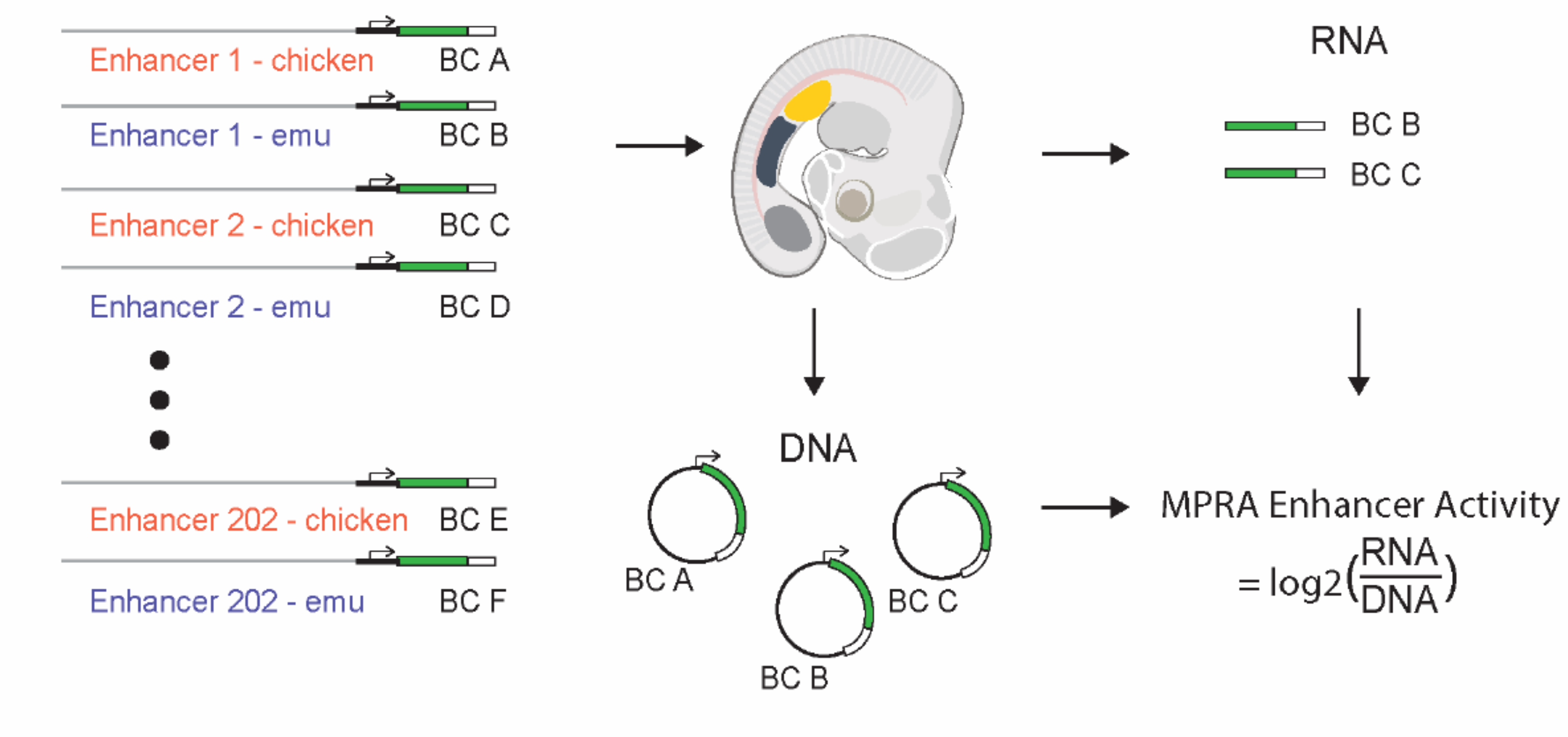
ATAC-seq: CNEEs in multiple flight-related tissues are active in early chick development



Sackton et al. 2019. *Science* 364: 74-78

Massively parallel reporter assays screen for enhancers active *in vivo* in fore- and hindlimb

HH16 → HH21



MPRA: injected at HH16, mRNA harvested at HH23