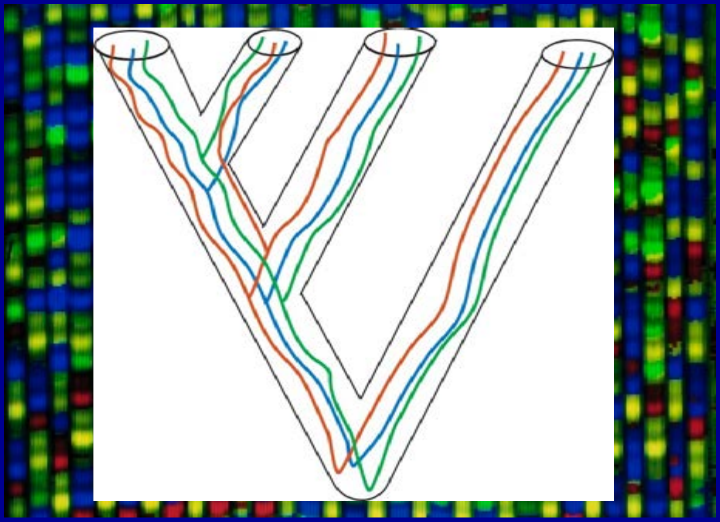
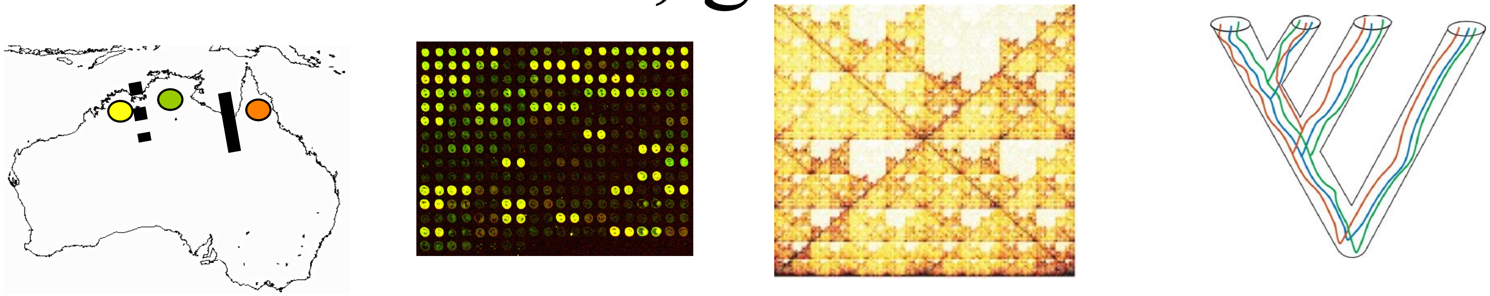
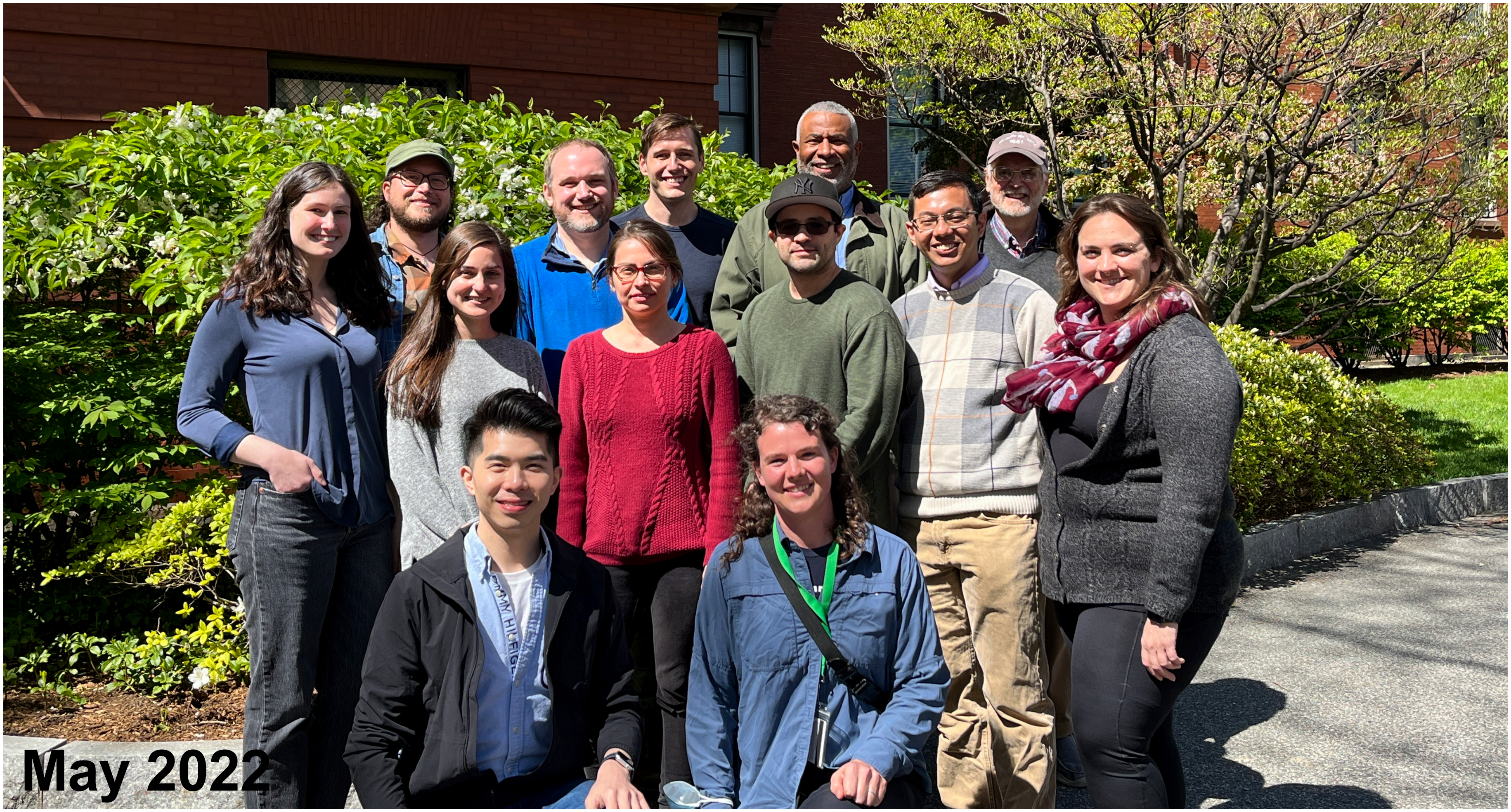
The continuum between phylogeography and phylogenetics



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





May 2022

I am a scientist, too!

Sunset Beach, OR

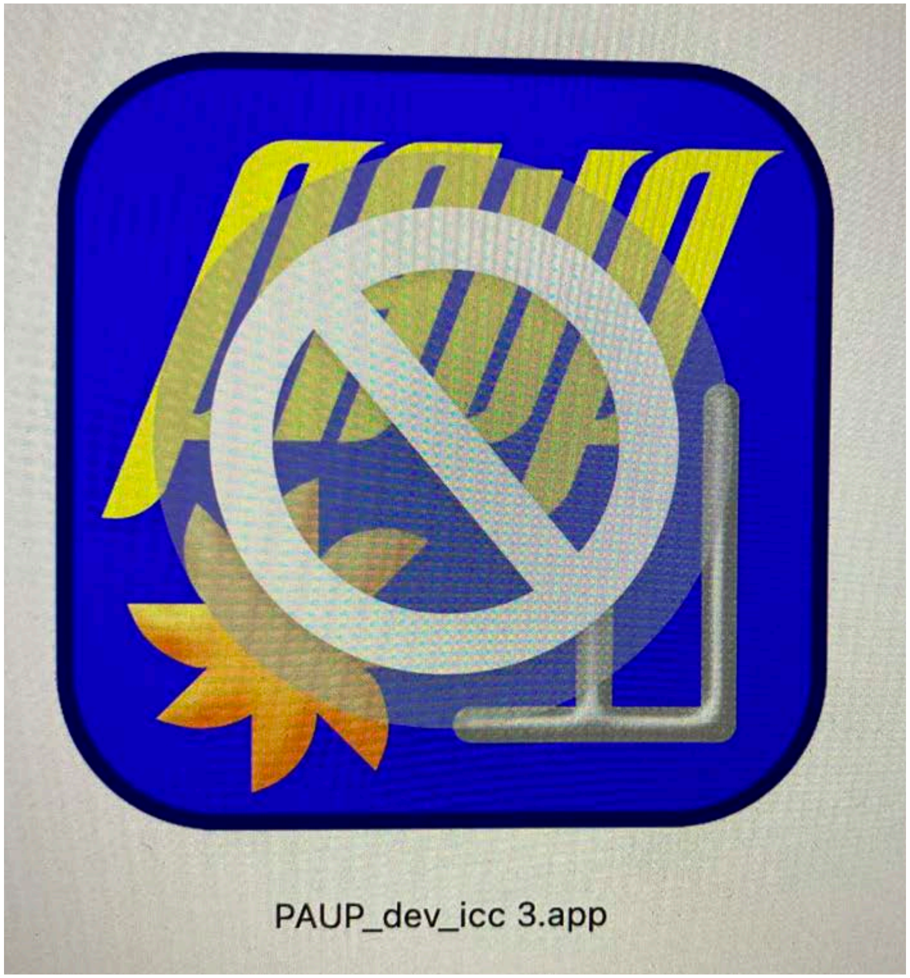
Plum Island, MA

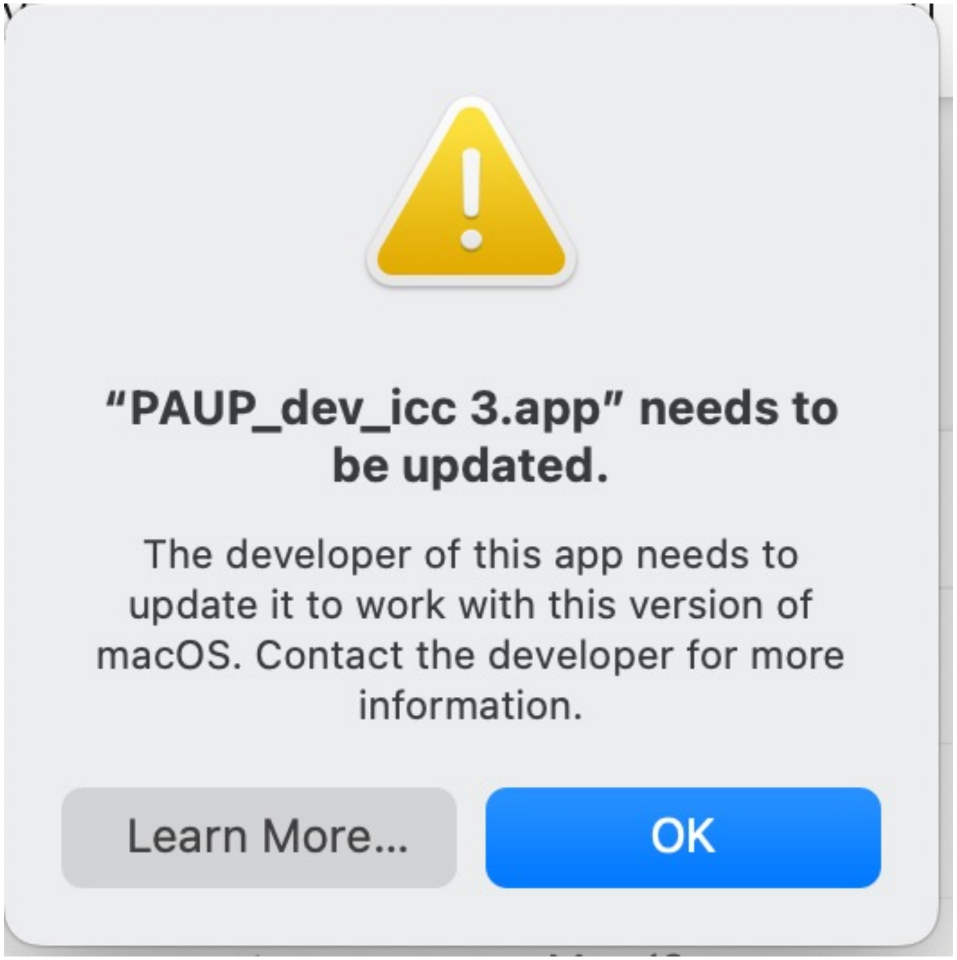
76 days
3848 miles

@ScottVEdwards1

Adventure Cycling Association

Oh no! End of the Paup* GUI era?





"PAUP_dev_icc 3.app" needs to be updated.

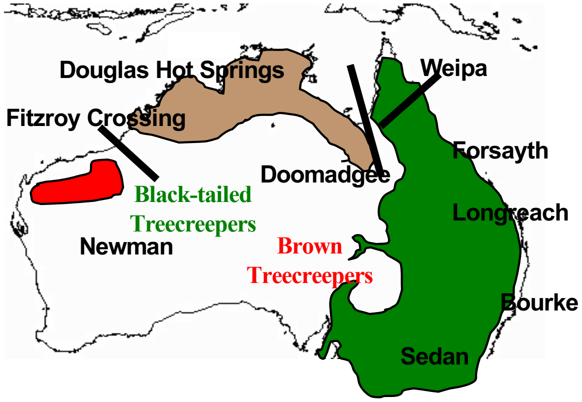
The developer of this app needs to update it to work with this version of macOS. Contact the developer for more information.

A 1980s conundrum in evolutionary biology

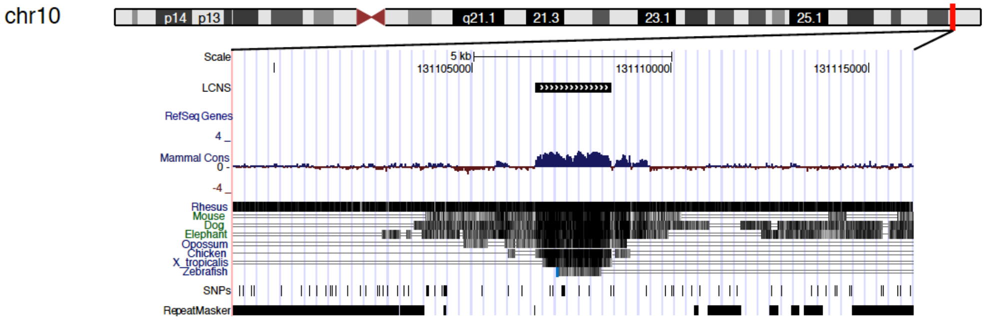
Oct. 2021, Seattle

Talk overview

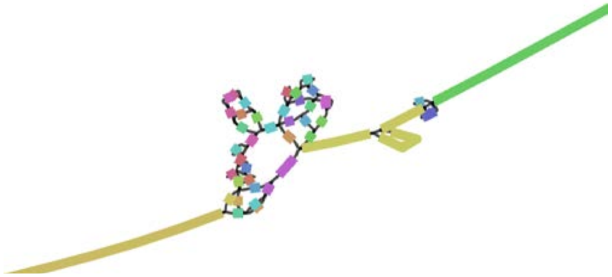
Part I: Reticulation and the emerging continuum between phylogeography and phylogenetics



Part II: PhyloG2P - Macroevolution and the origin of phenotypic traits



Part III: Pangenomes: the future of evolutionary genomics



Phylogeographic case studies

from the wilds of Australia

Babblers

Grassfinches

Treecreepers

PhyloG2P

Paleognaths

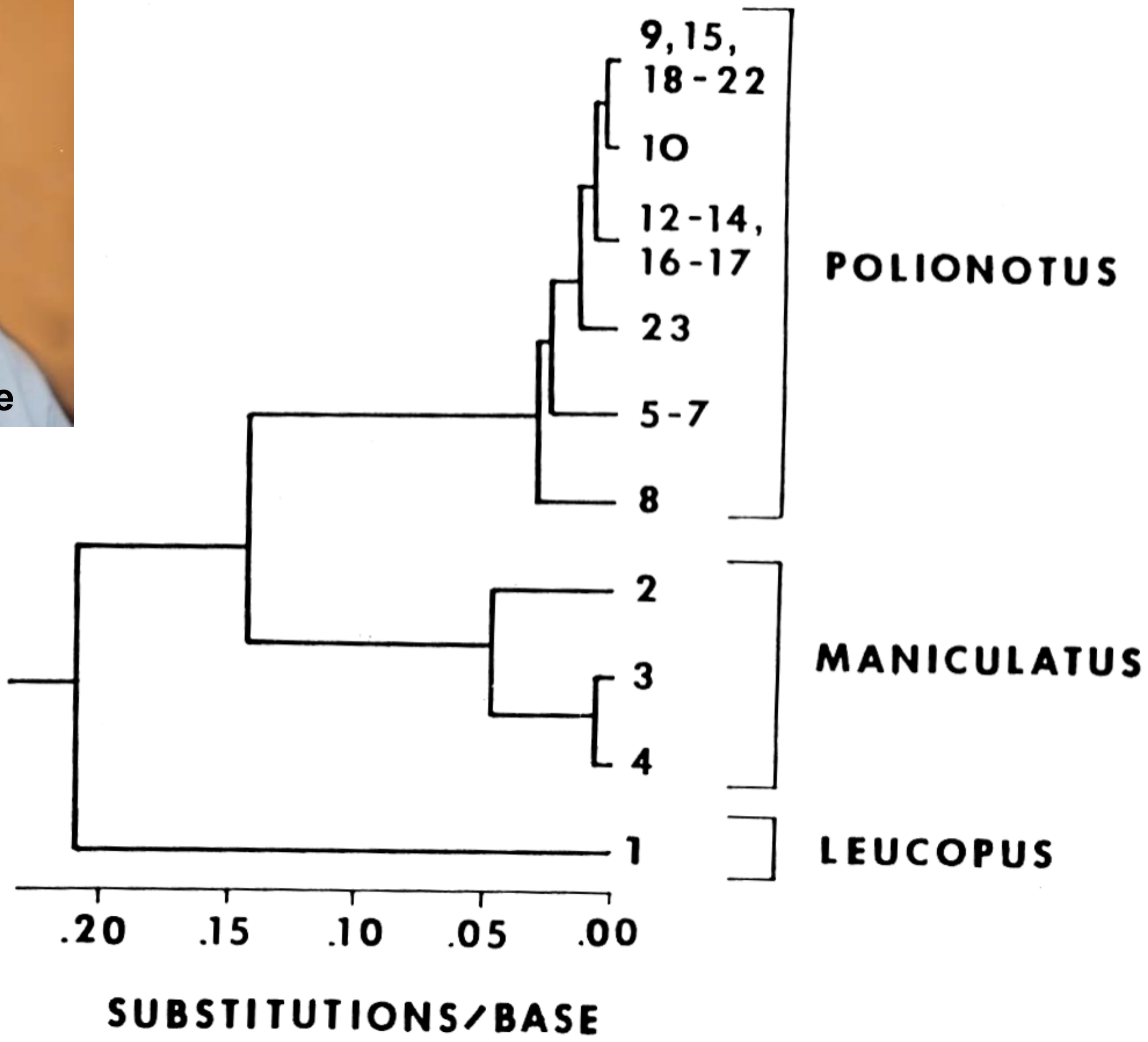
Honeyeaters

Pangenomes

Scrub Jays

The first ‘gene tree’, 1979

J. C. AVISE, R. A. LANSMAN AND R. O. SHADE

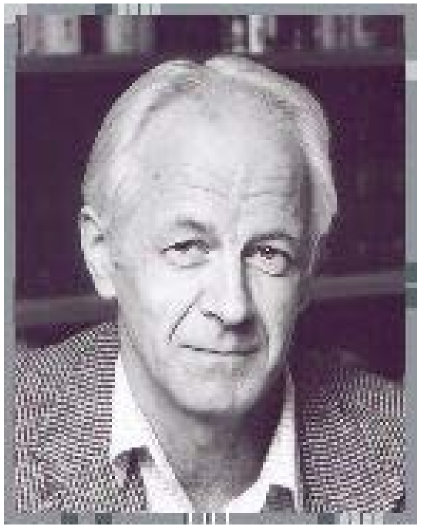


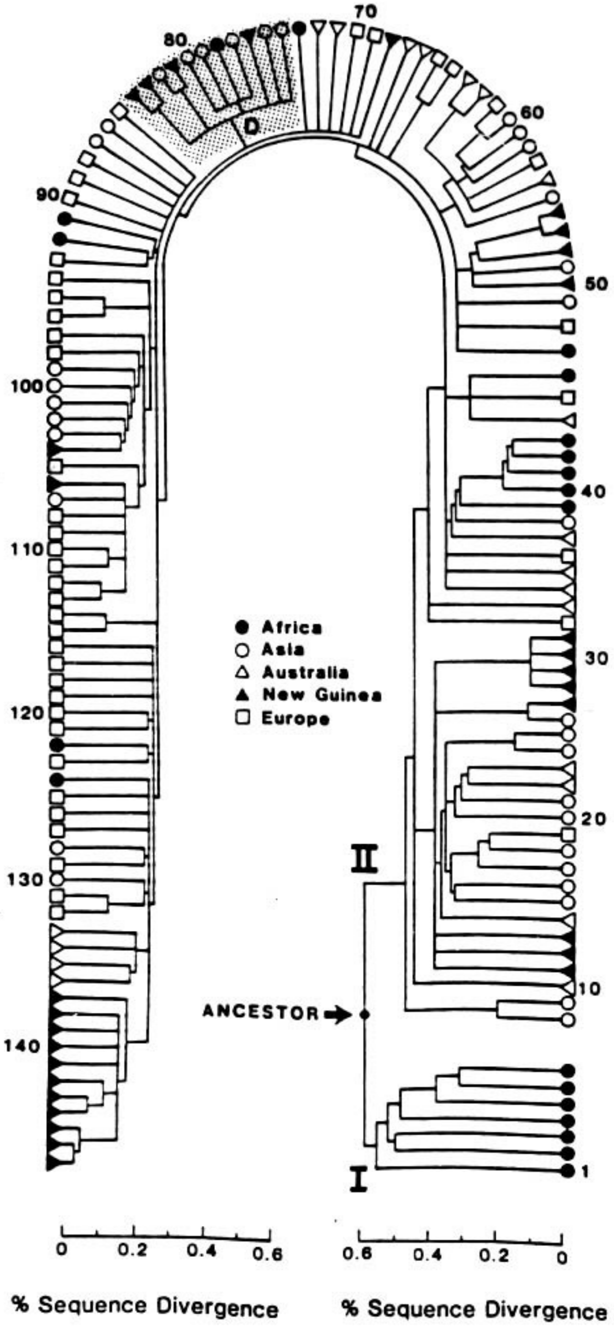
Reprinted from Nature, Vol. 325, No. 6099, pp. 31-36, 1 January 1987

Mitochondrial DNA and human evolution

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

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

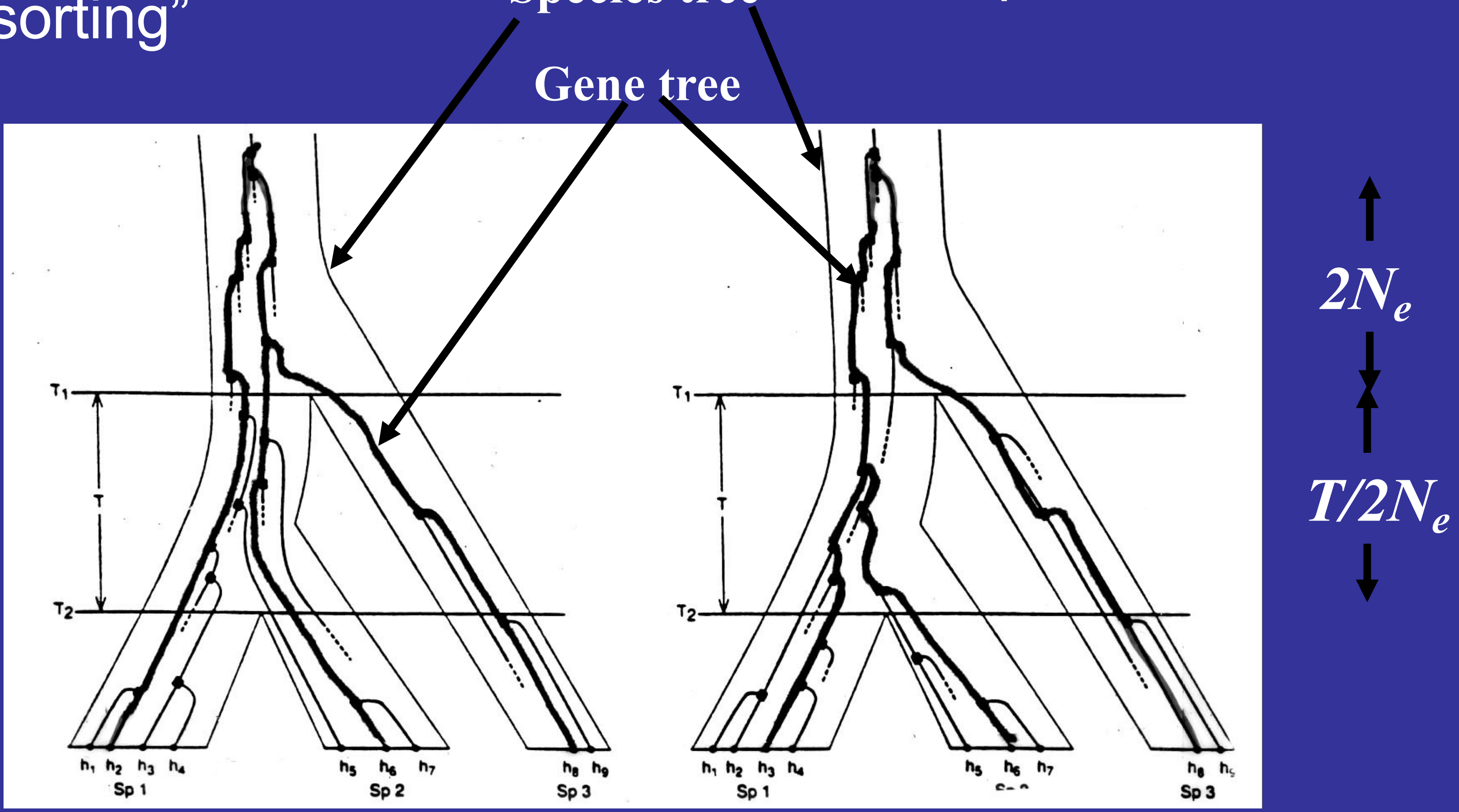




Stochastic gene tree variation during rapid radiations

"incomplete lineage sorting"

"deep coalescence"



Species tree

Gene tree

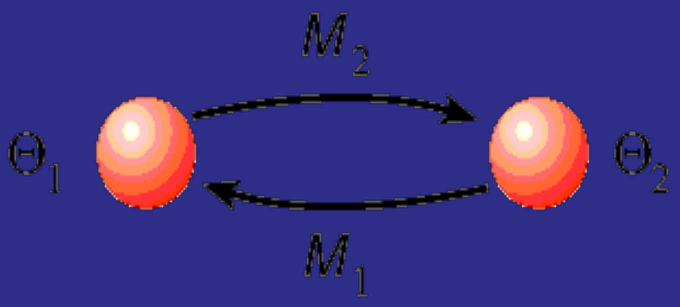
$2N_e$

$T/2N_e$

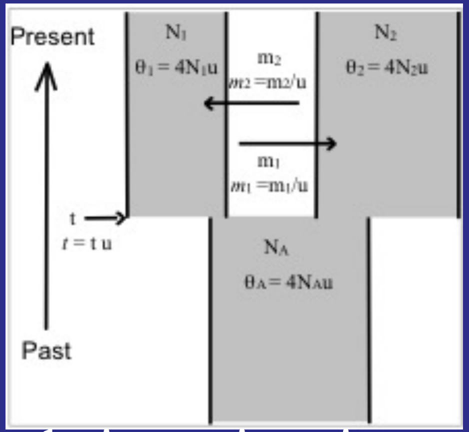
Hudson RR.1990. *Oxford Surv. Evol. Biol.* 7: 1-45

Multilocus models in phylogeography

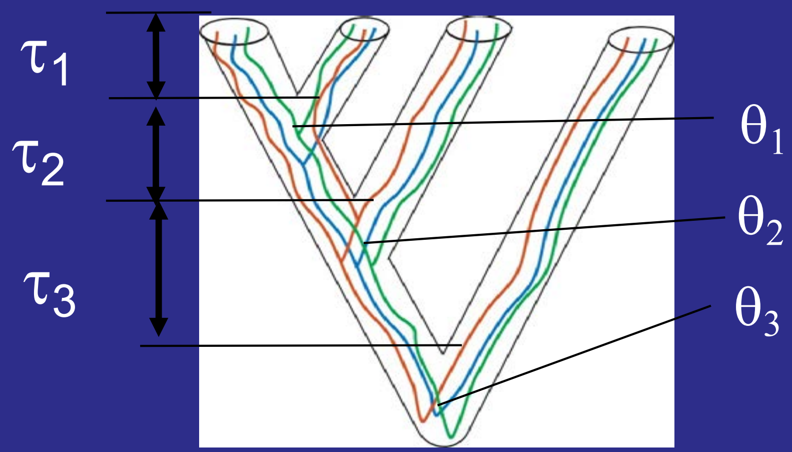
Population size: $\theta = 4N\mu$ Divergence time: $\tau=\mu t$ Gene flow: $M=m/\mu$



equilibrium migration model
MIGRATE: Beerli 2006 *Bioinformatics*



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

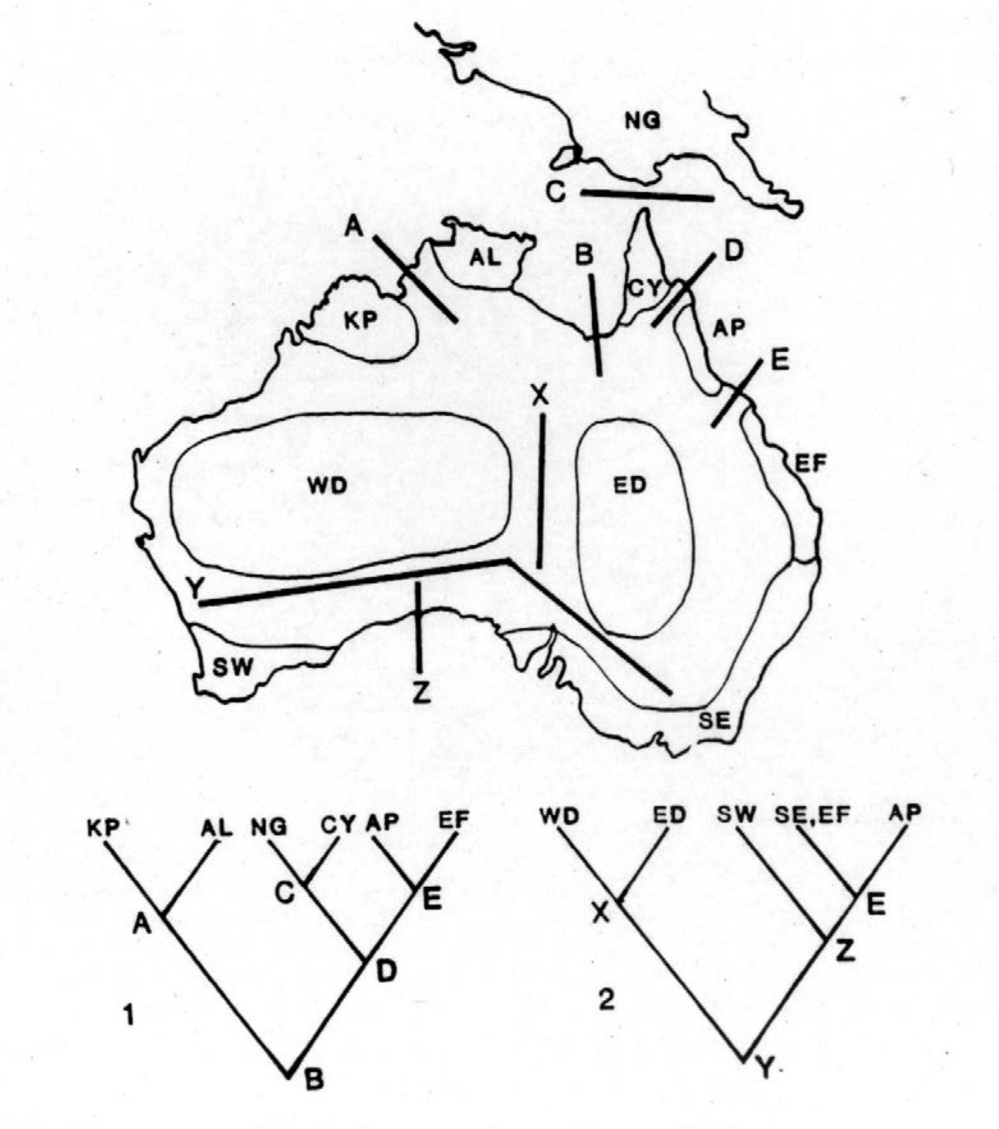


pure isolation
(phylogeny)
BEST: Liu and Pearl. 2007. *Syst. Biol*
BEAST*: Heled and Drummond. 2010. *Mol. Biol. Evol.*

~~Warm~~ welcome in the outback…

Stinking hot

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

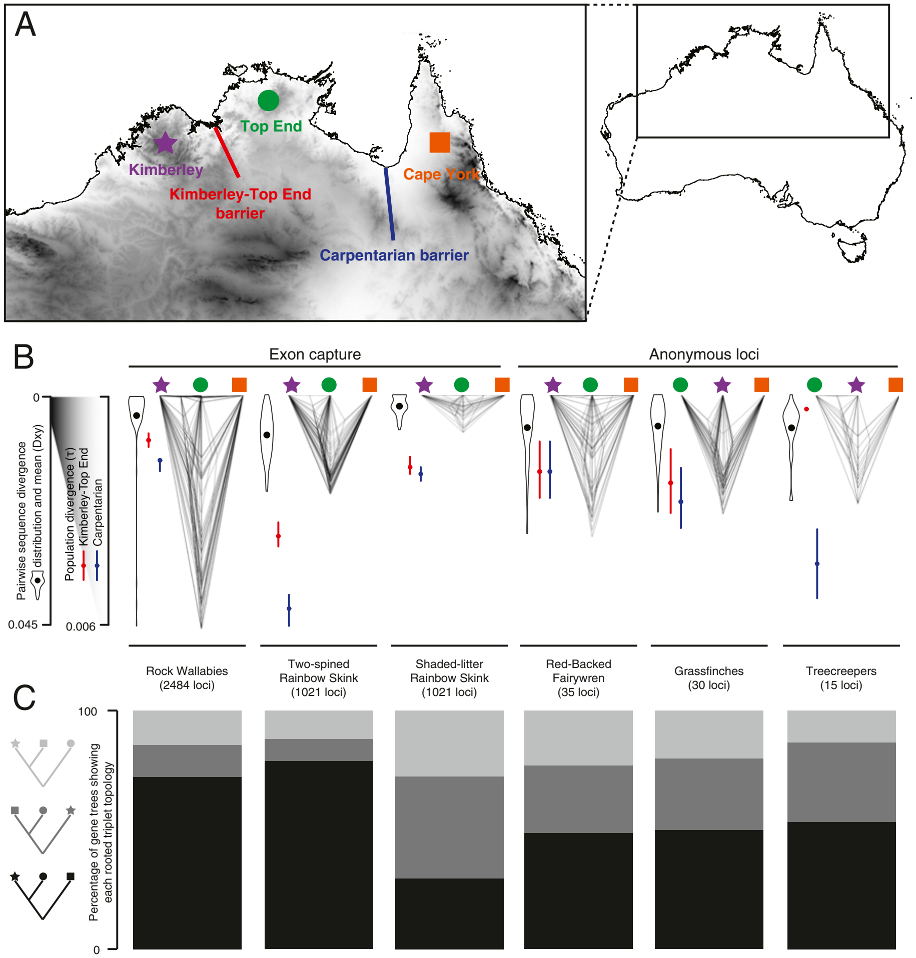


Cracraft 1986. *Evolution*

COLLOQUIUM PAPER

Reticulation, divergence, and the phylogeography–phylogenetics continuum

Scott V. Edwards[a,1], Sally Potter[b,c], C. Jonathan Schmitt[a], Jason G. Bragg[b,c], and Craig Moritz[b,c]

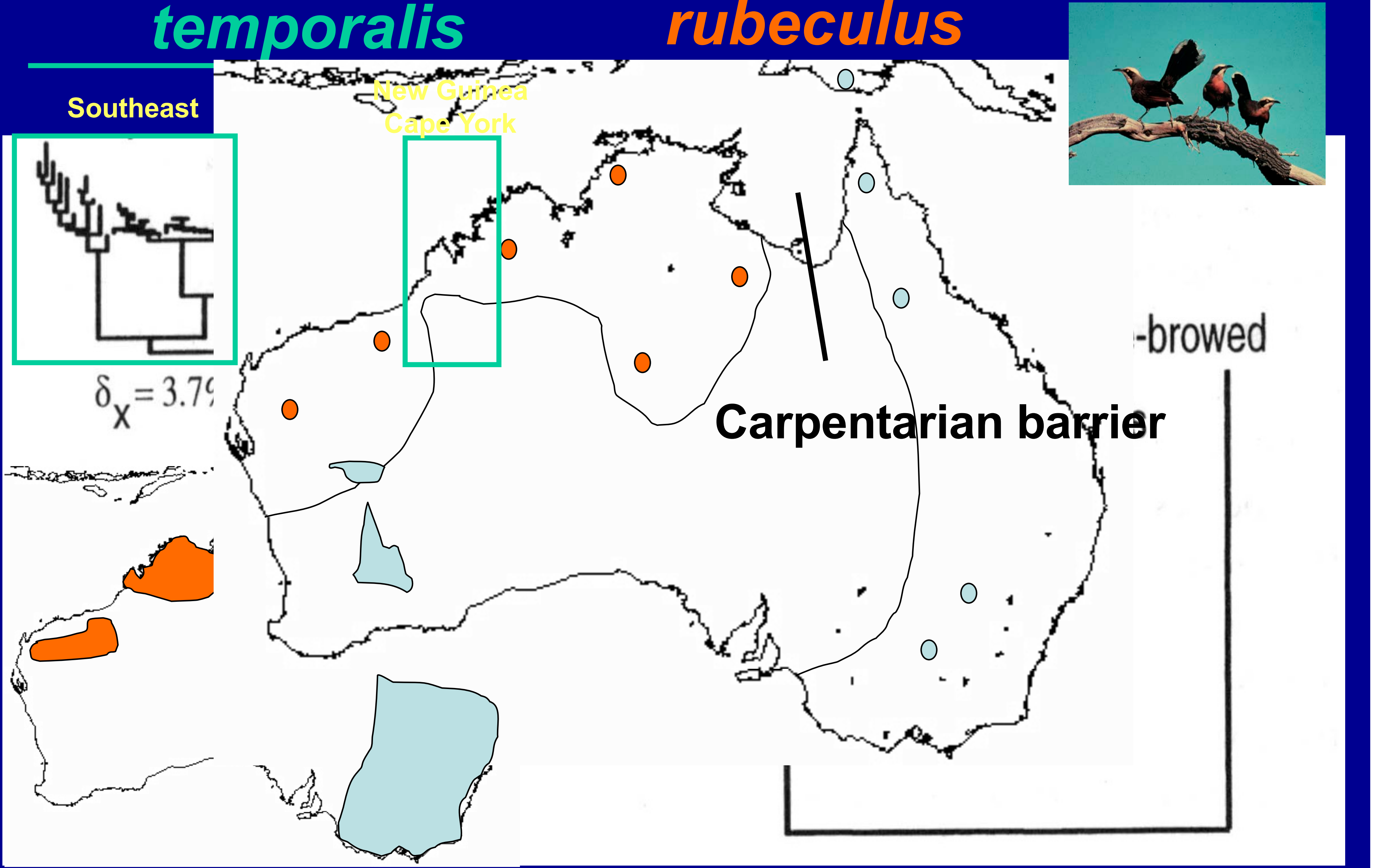


Edwards et al. 2016. *PNAS* 113: 8025-8032

Mitochondrial gene tree for grey-crowned babblers

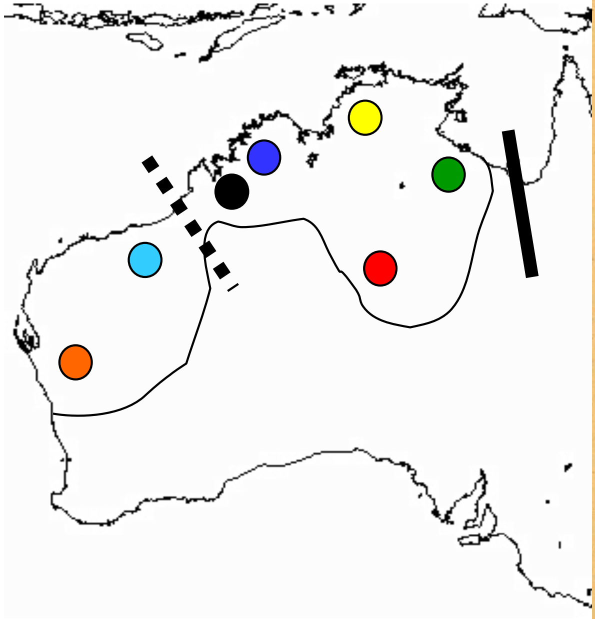
Sampling localities for grey-crowned babblers

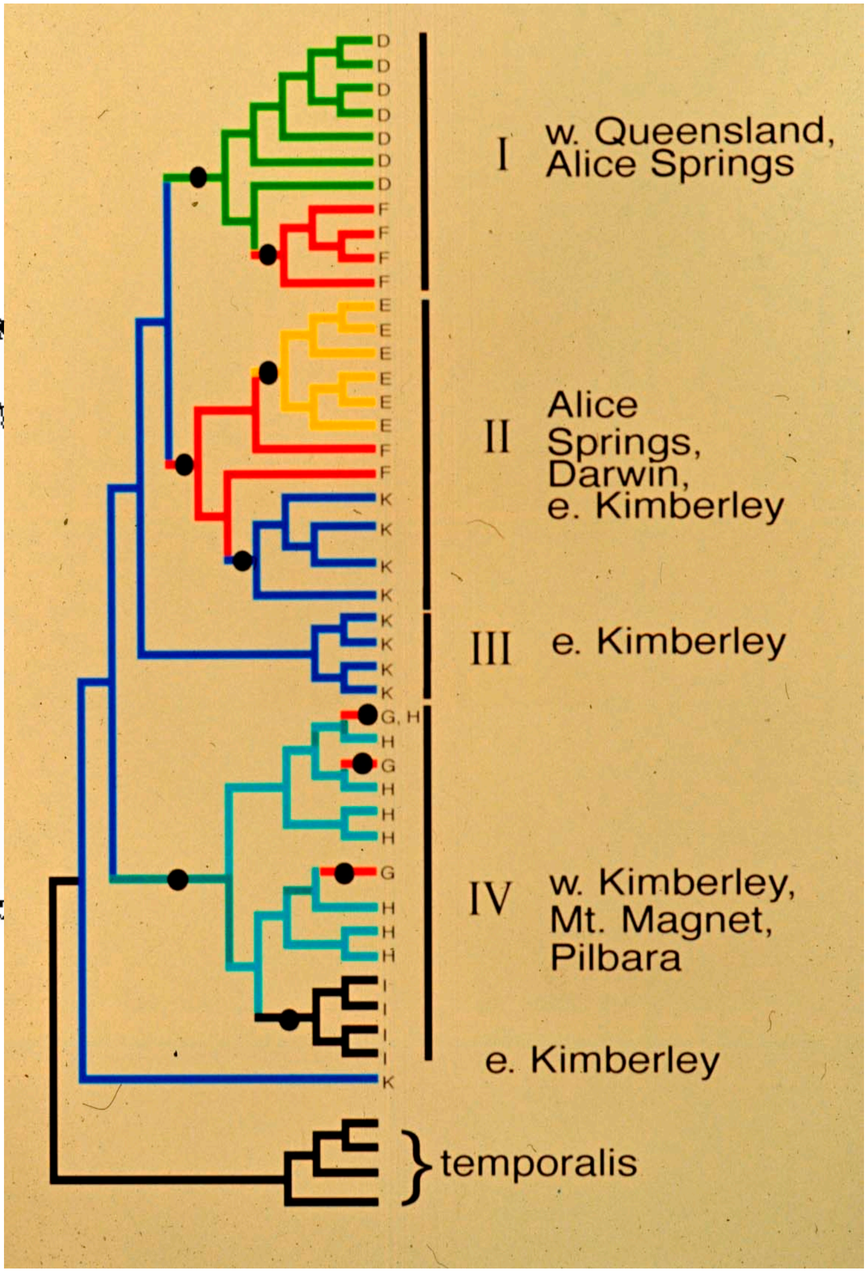
temporalis *rubeculus*



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

Gene flow erodes population monophyly





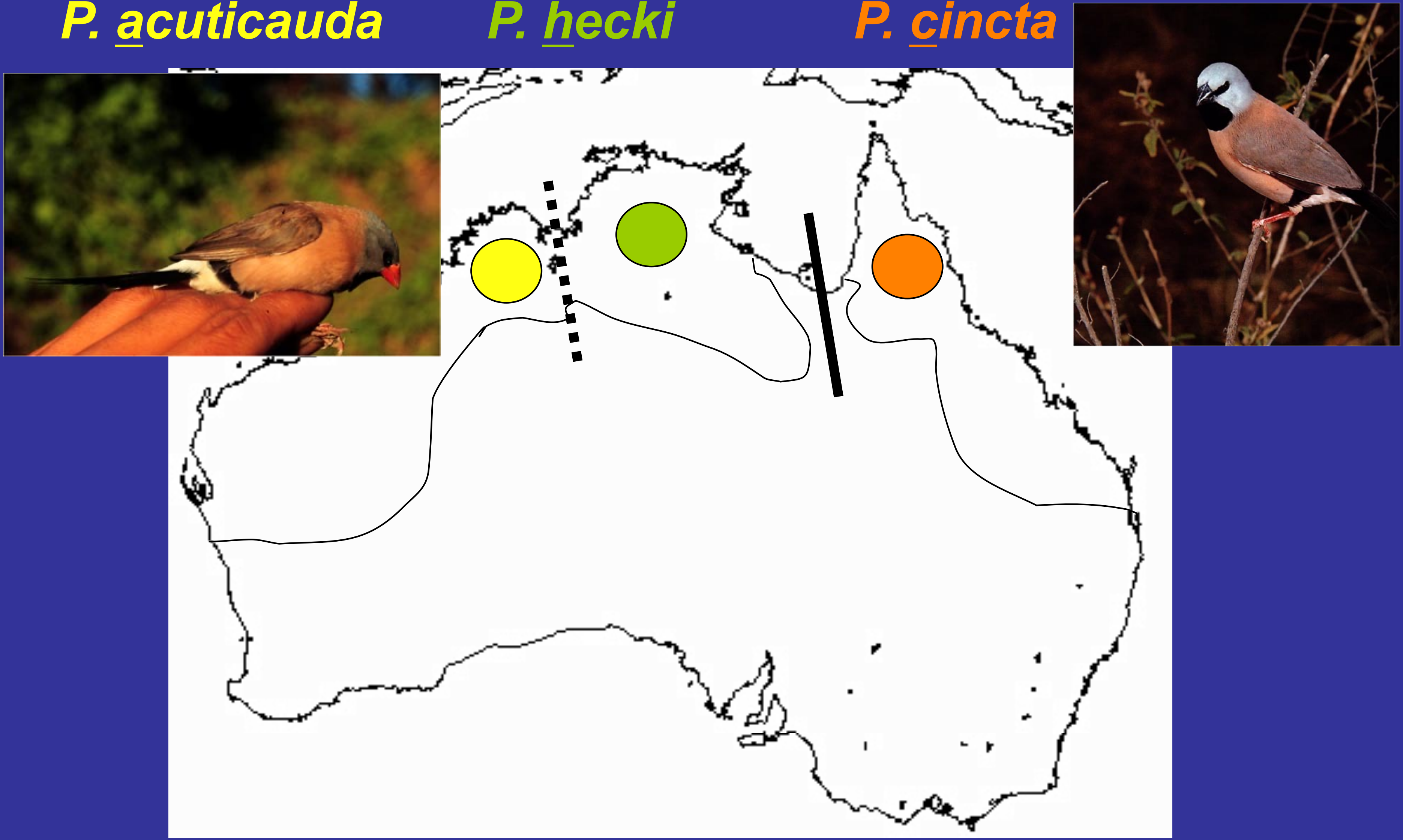
● **Migration event reconstructed by parsimony**

Bayesian approach: Advantages for estimating gene flow

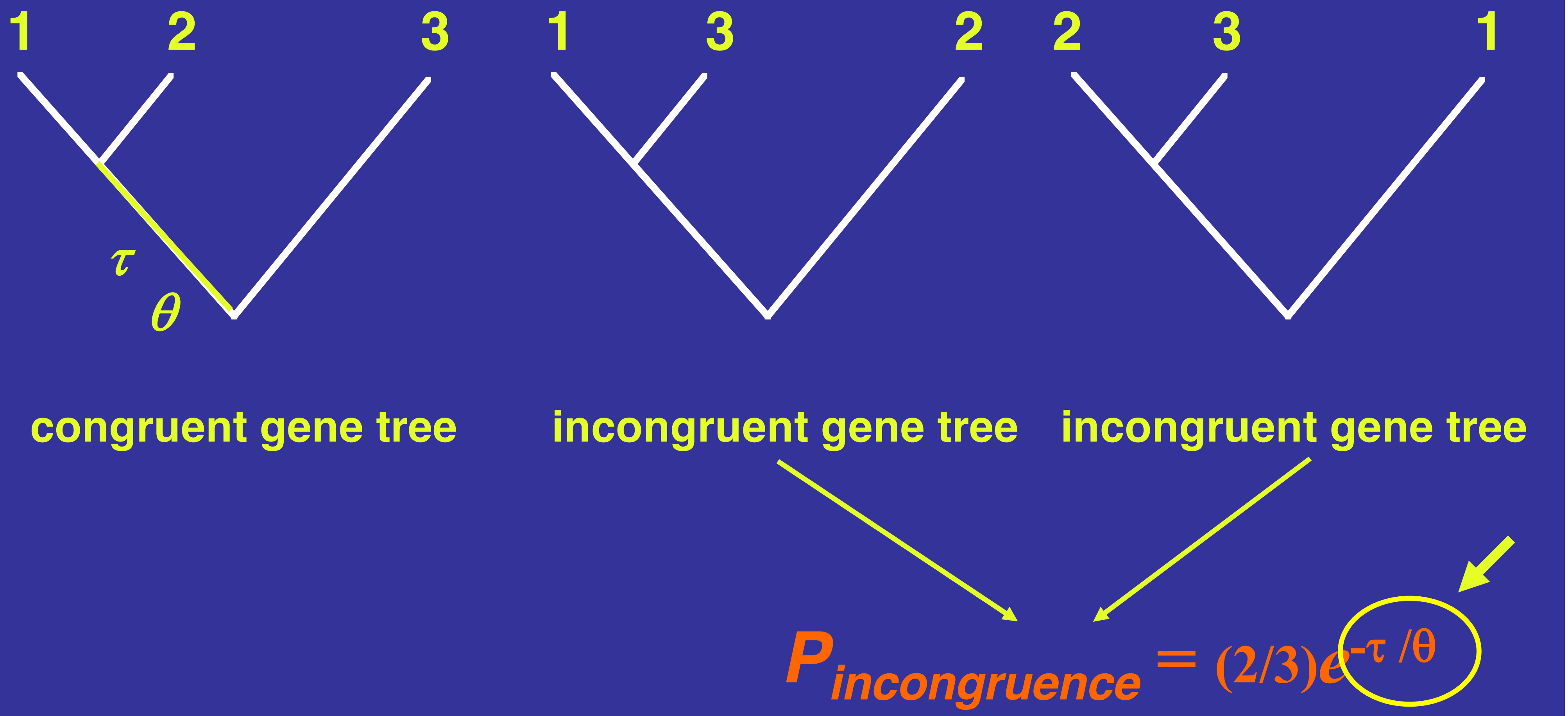
- Parsimony approach
  - Assumes complete certainty of gene tree
  - Statistical testing of different hypotheses of gene flow is cumbersome
  - Confidence limits reflect only uncertainty in *s*
- Bayesian approach (migrate-n)
  - Integrates the two sources of uncertainty
    - Accommodates uncertainty 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

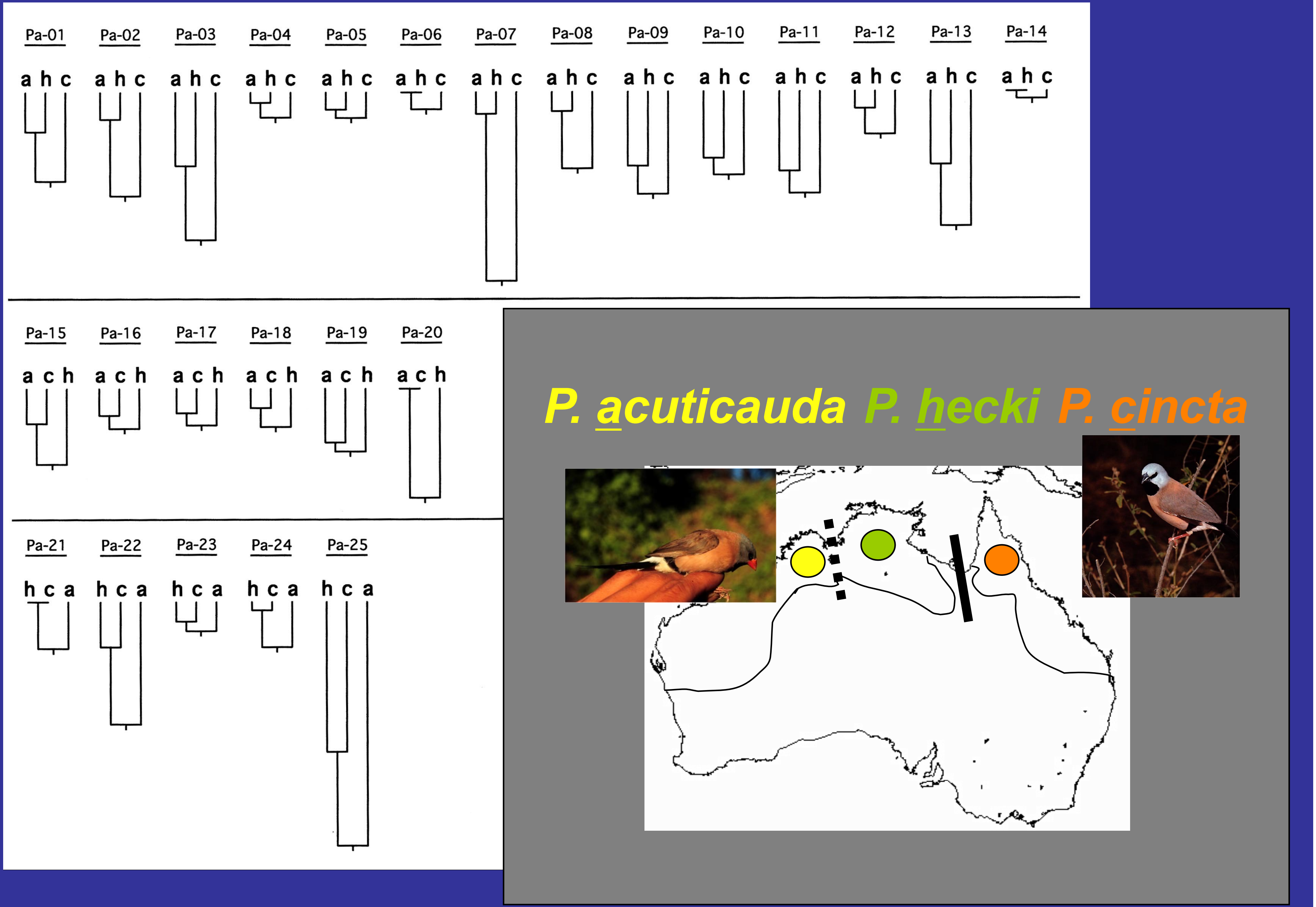
incongruent gene tree

incongruent gene tree

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

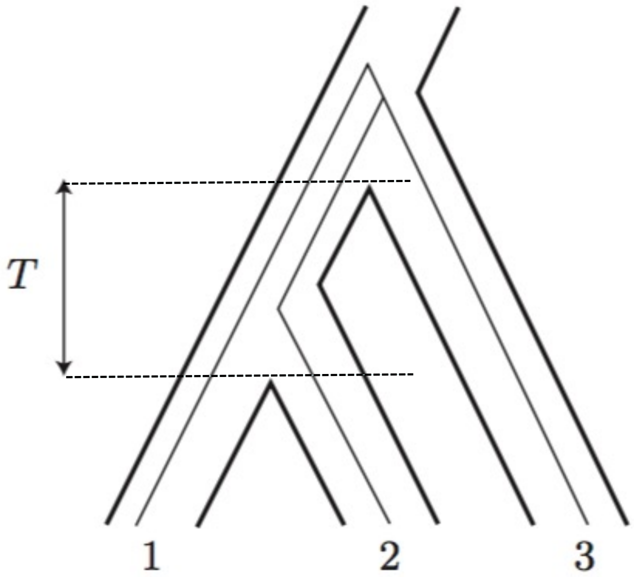
Pamilo and Nei 1988. Mol. Biol. Evol. 5: 568-583
Liu, et al. 2010. BMC Evolutionary Biology 10:302

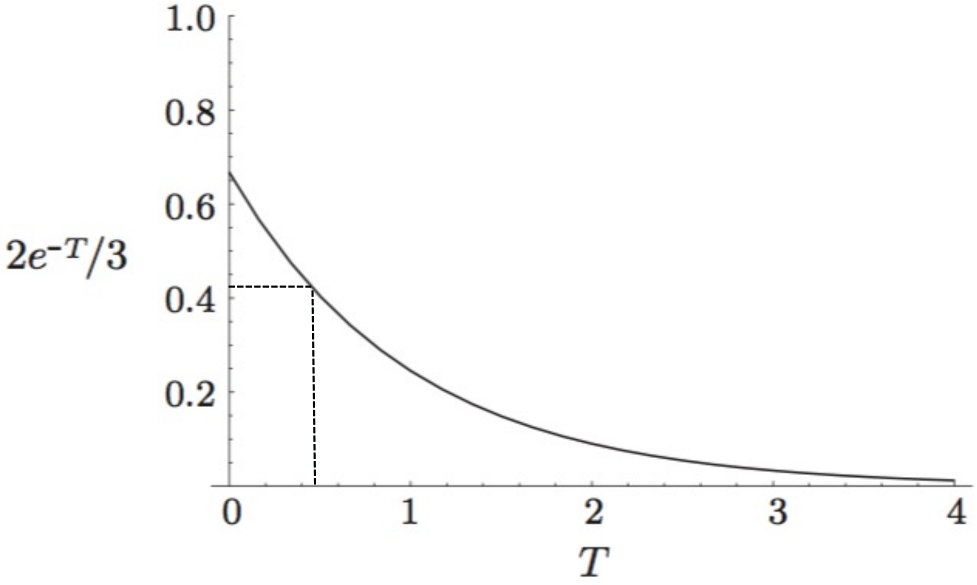
30 gene trees from Australian finches



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

Probability of discordance between gene tree and species tree





T = internode length = $2\mu t/4N\mu$ = t/2N generations long

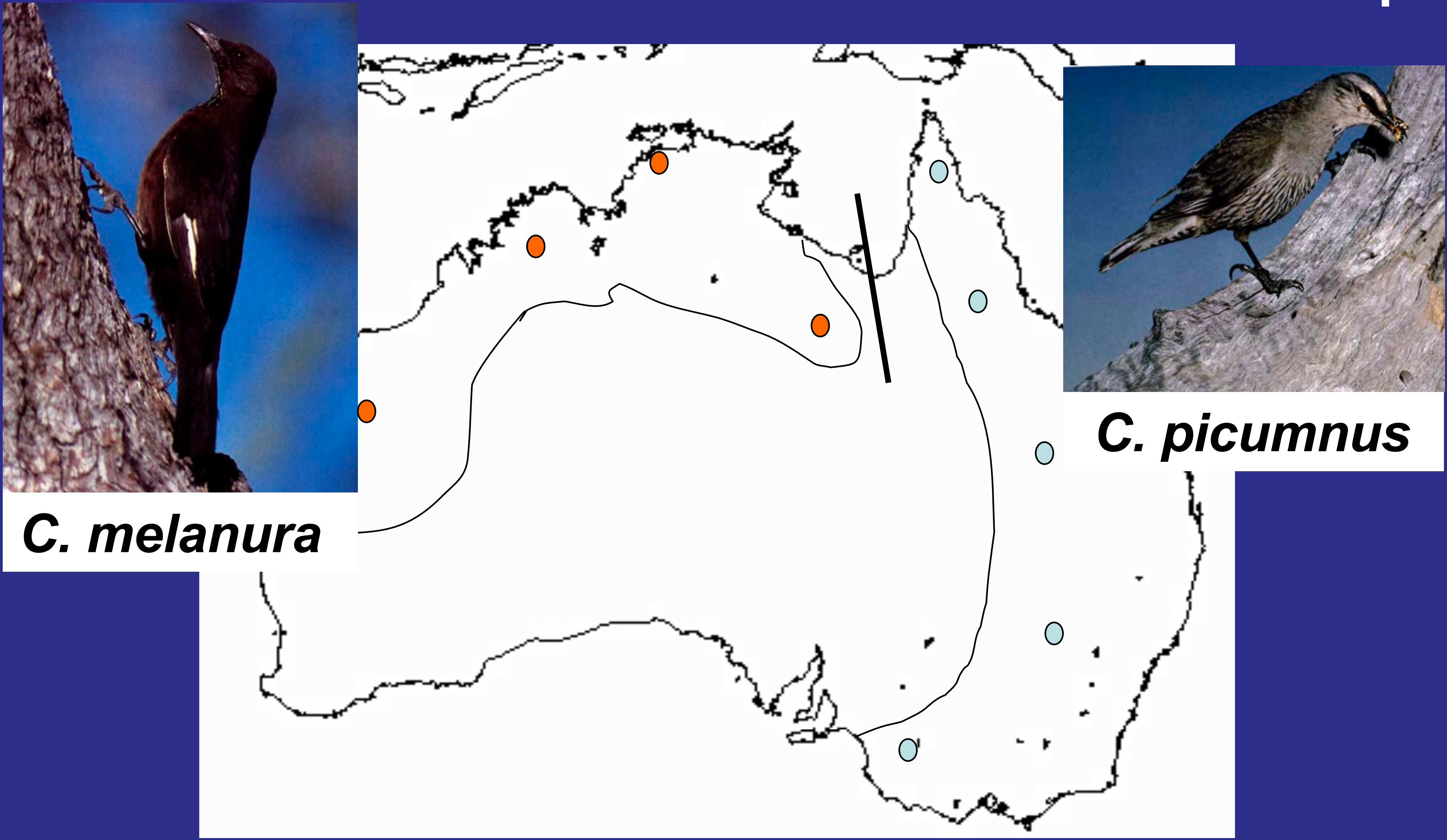
$$P\{\text{discordant}\} \quad = \quad \frac{2}{3}e^{-T}$$

= 12/28

Treecreepers (*Climacteris*)

Black-tailed treecreeper

Brown treecreeper



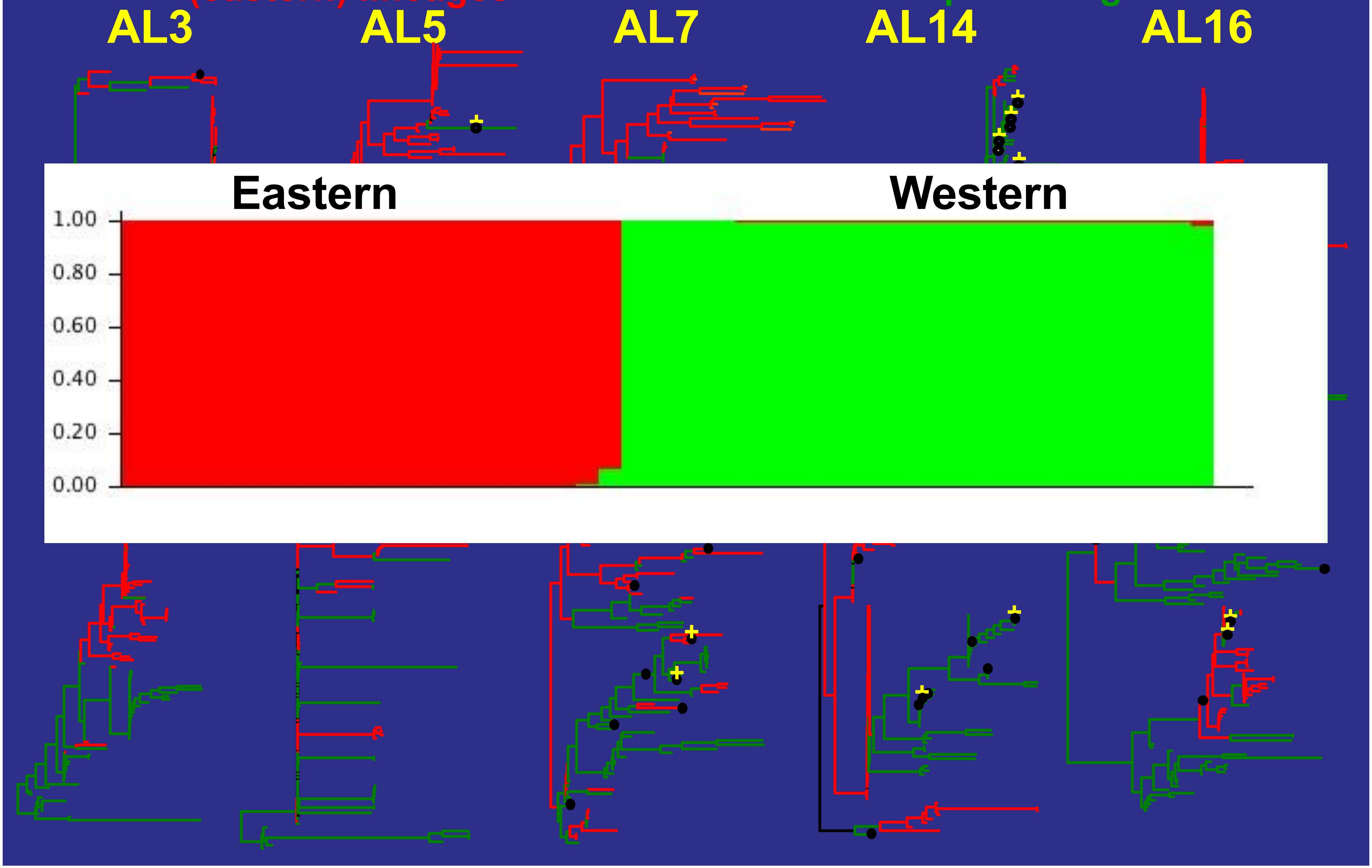
C. melanura

C. picumnus

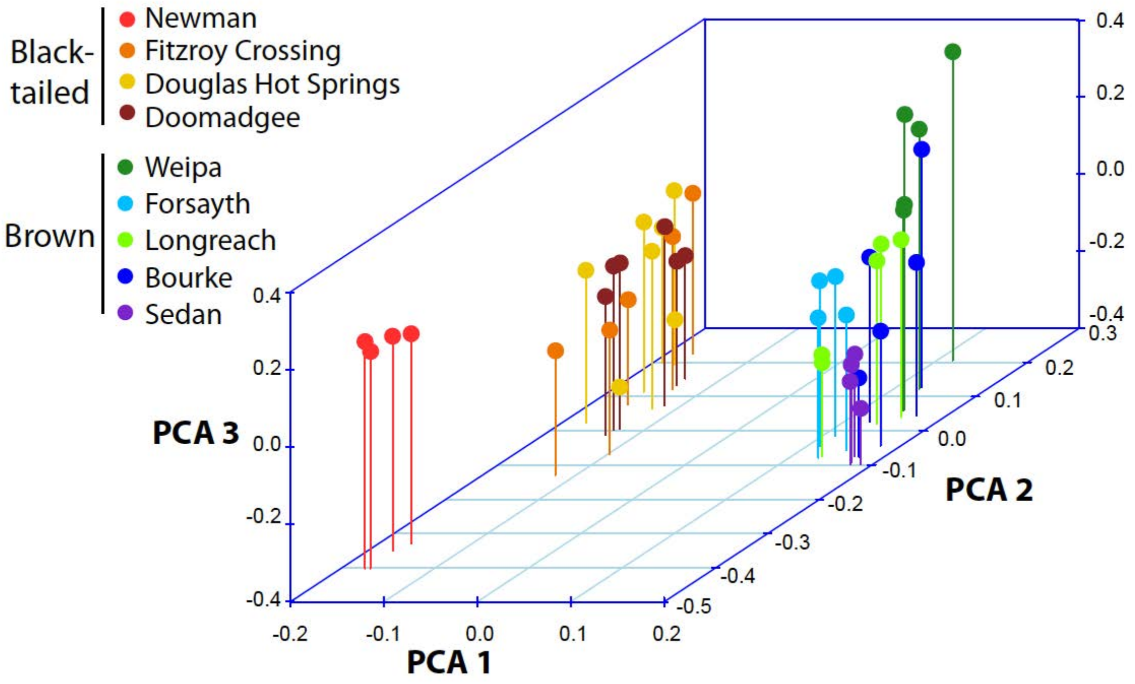
12 gene trees in Australian treecreepers

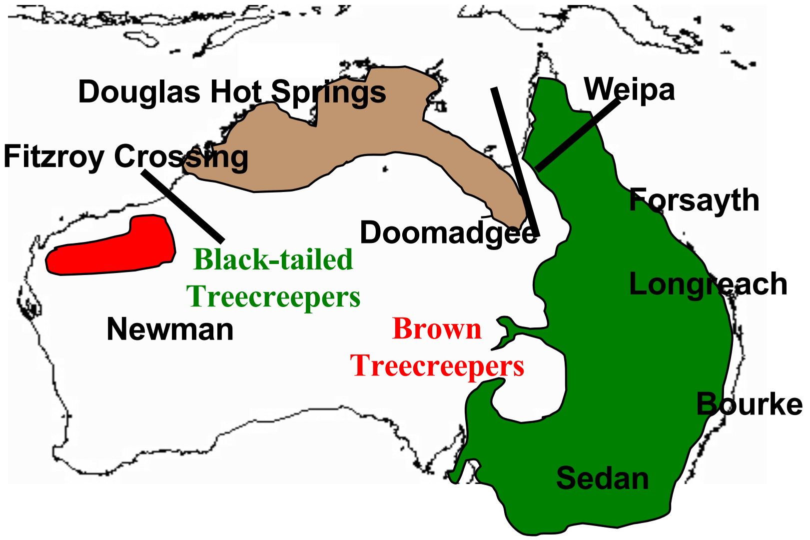
Brown Treecreeper (eastern) lineages

Black-tailed (western) Treecreeper lineages



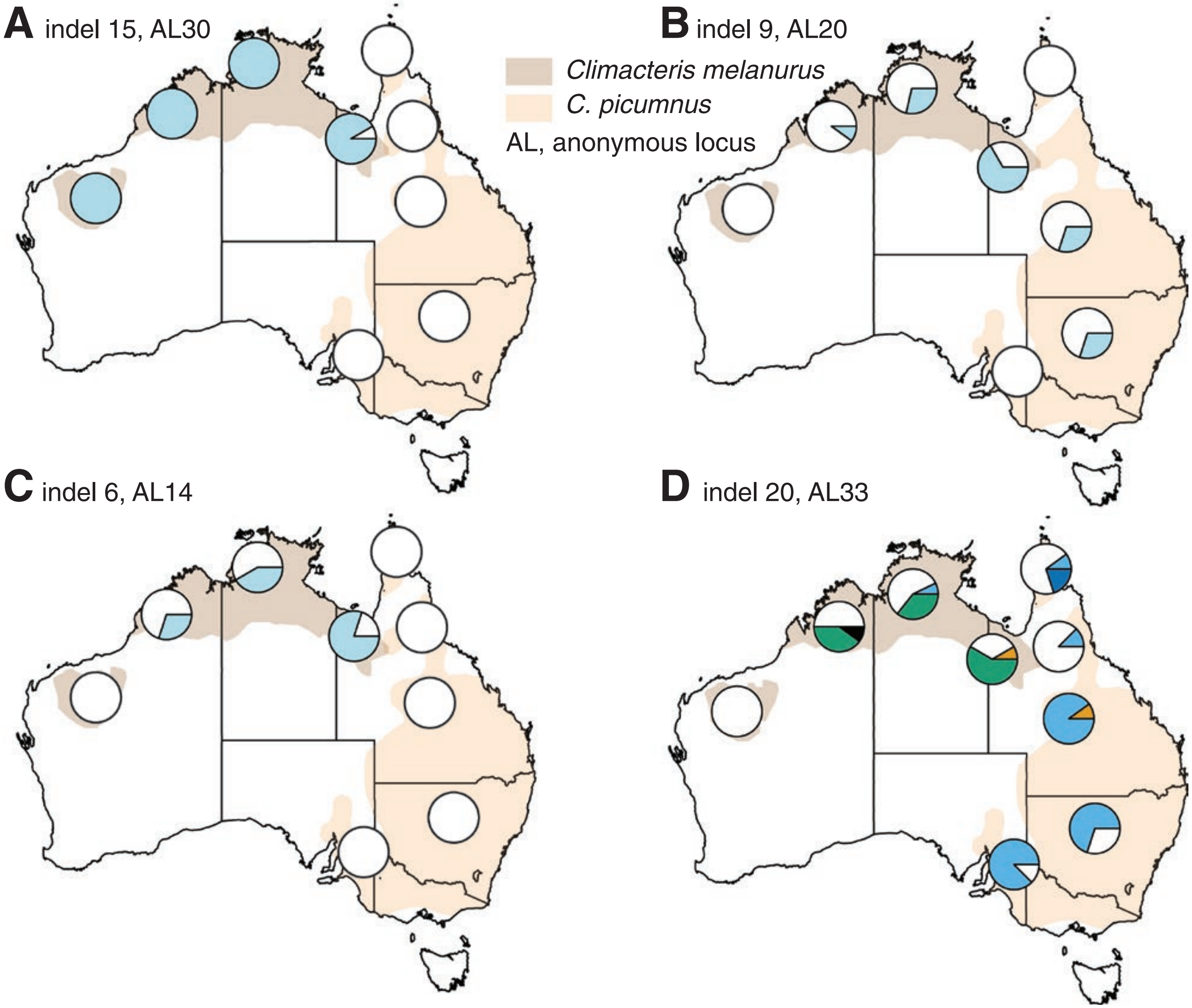
Correspondence between PCA and geography





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

Insertions-deletions provide significant phylogeographic signal

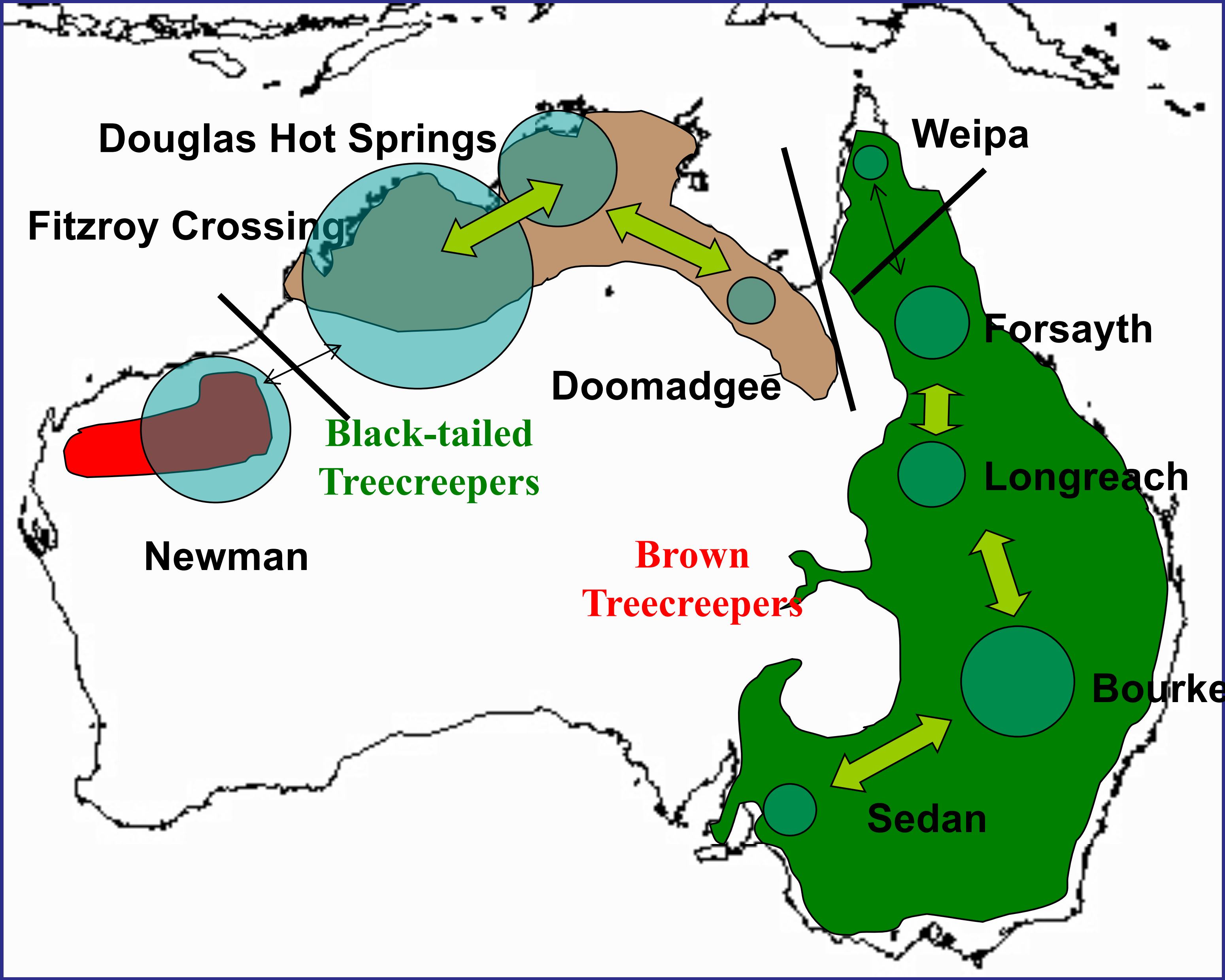


Edwards, Tonini, McInerney, Welch & Beerli. 2022. *Biol. J. Linn. Soc.*

Treecreeper populations are connected but variable in size (MIGRATE)

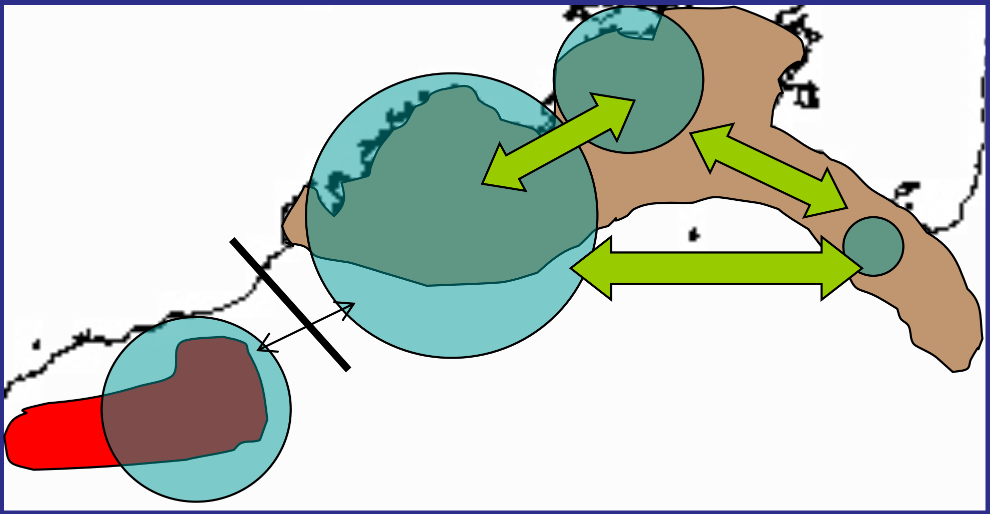
● **10,000 individuals**

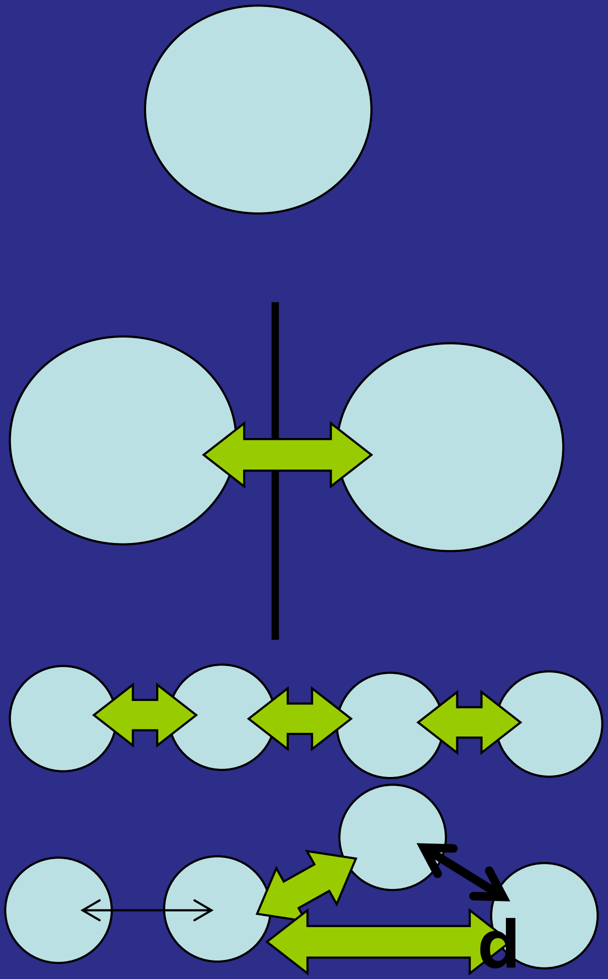
— **1-2 migrants per generation**



Rotzel, Edwards and Beerli, unpubl. data

Using Bayes Factors to evaluate phylogeographic models





$2(mL_1 – mL_2)$

-98.60

-83.14

-54.98

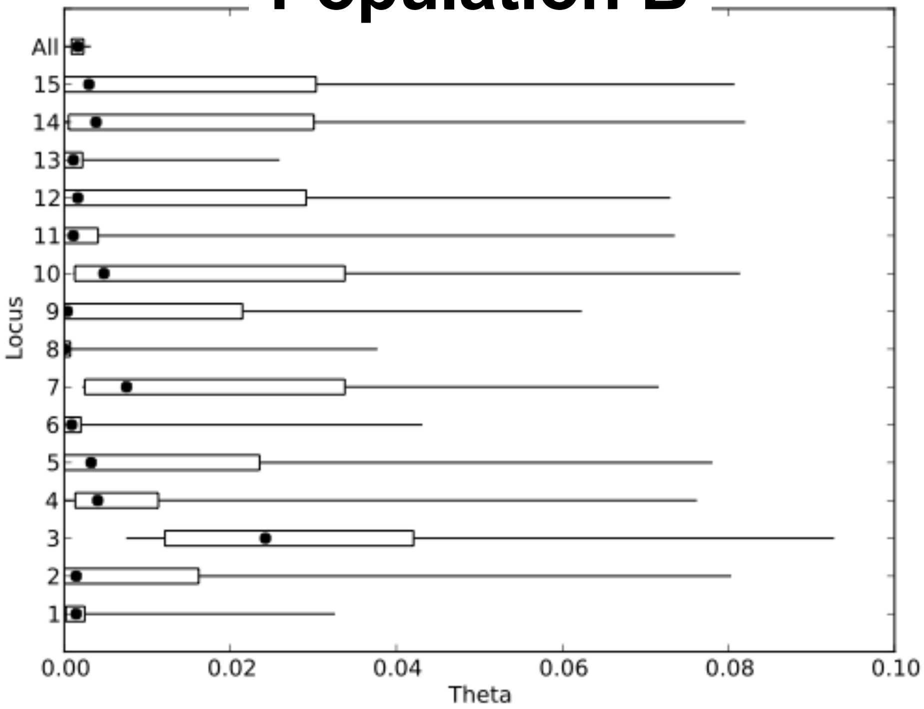
0.00

mL = marginal likelihood

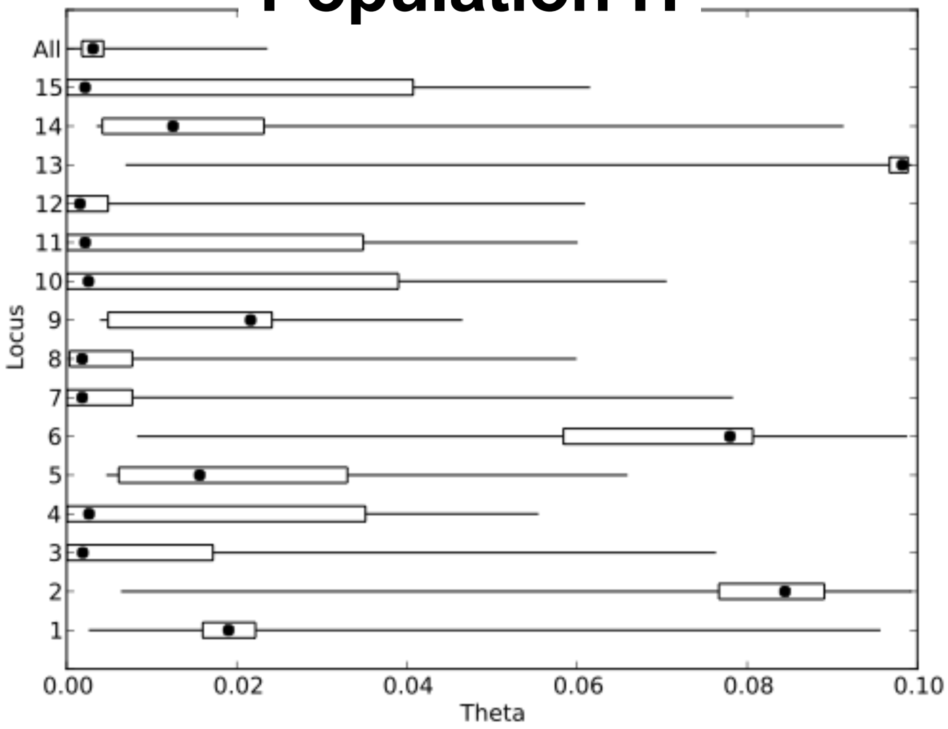
Edwards, Tonini, McInerney, Welch & Beerli. 2022. *Biol. J. Linn. Soc.*

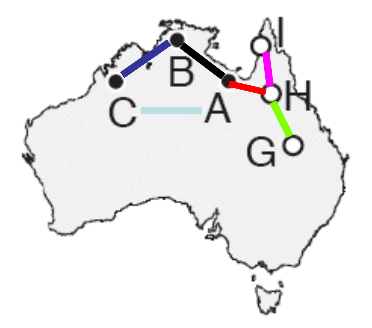
Higher precision estimates of demographic parameters (θ) with more loci

Population B

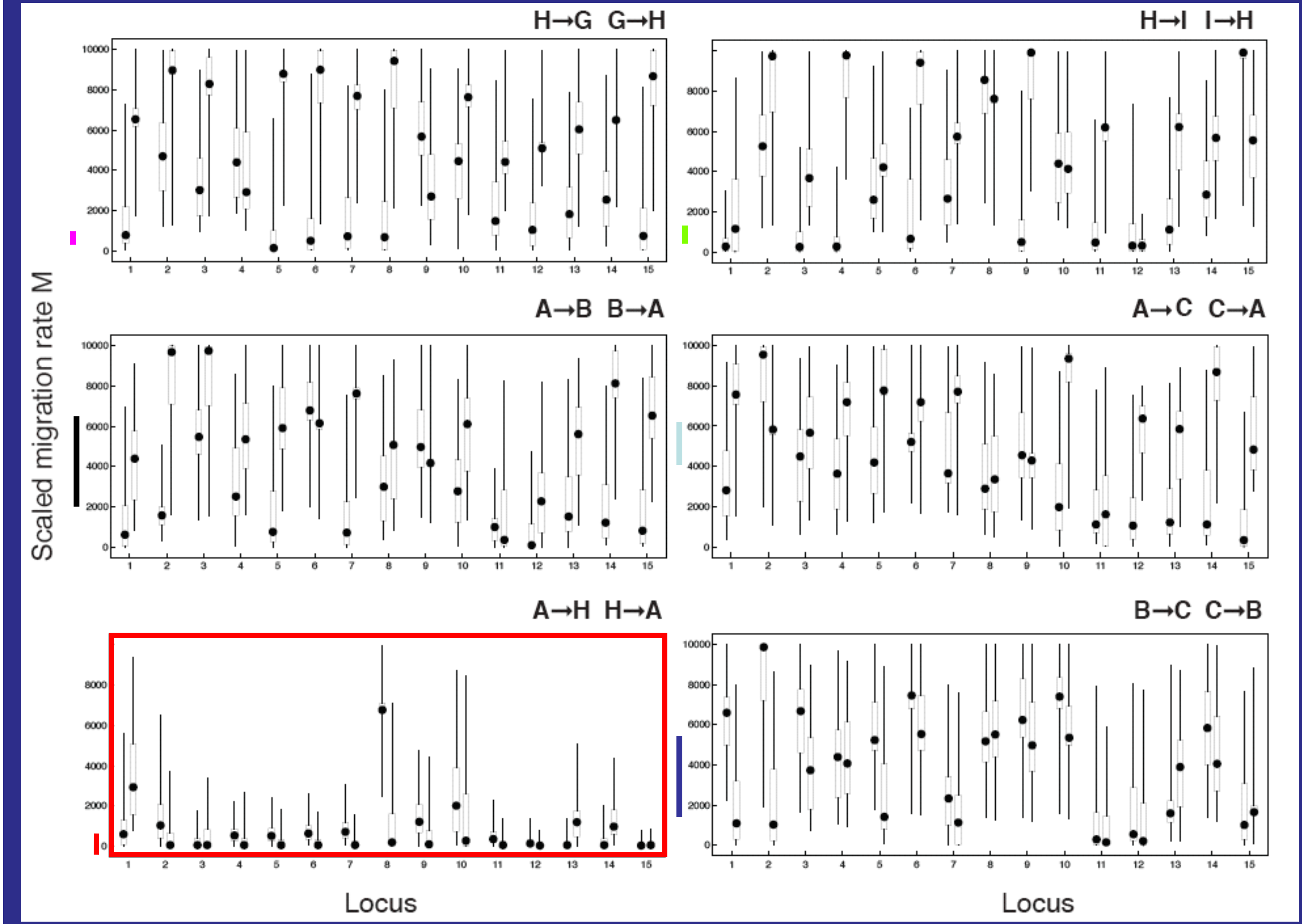


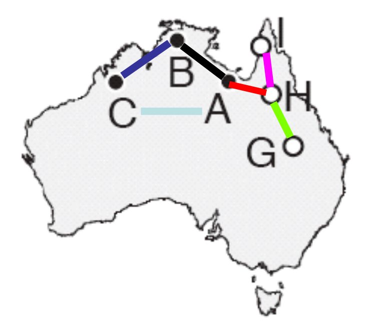
Population H



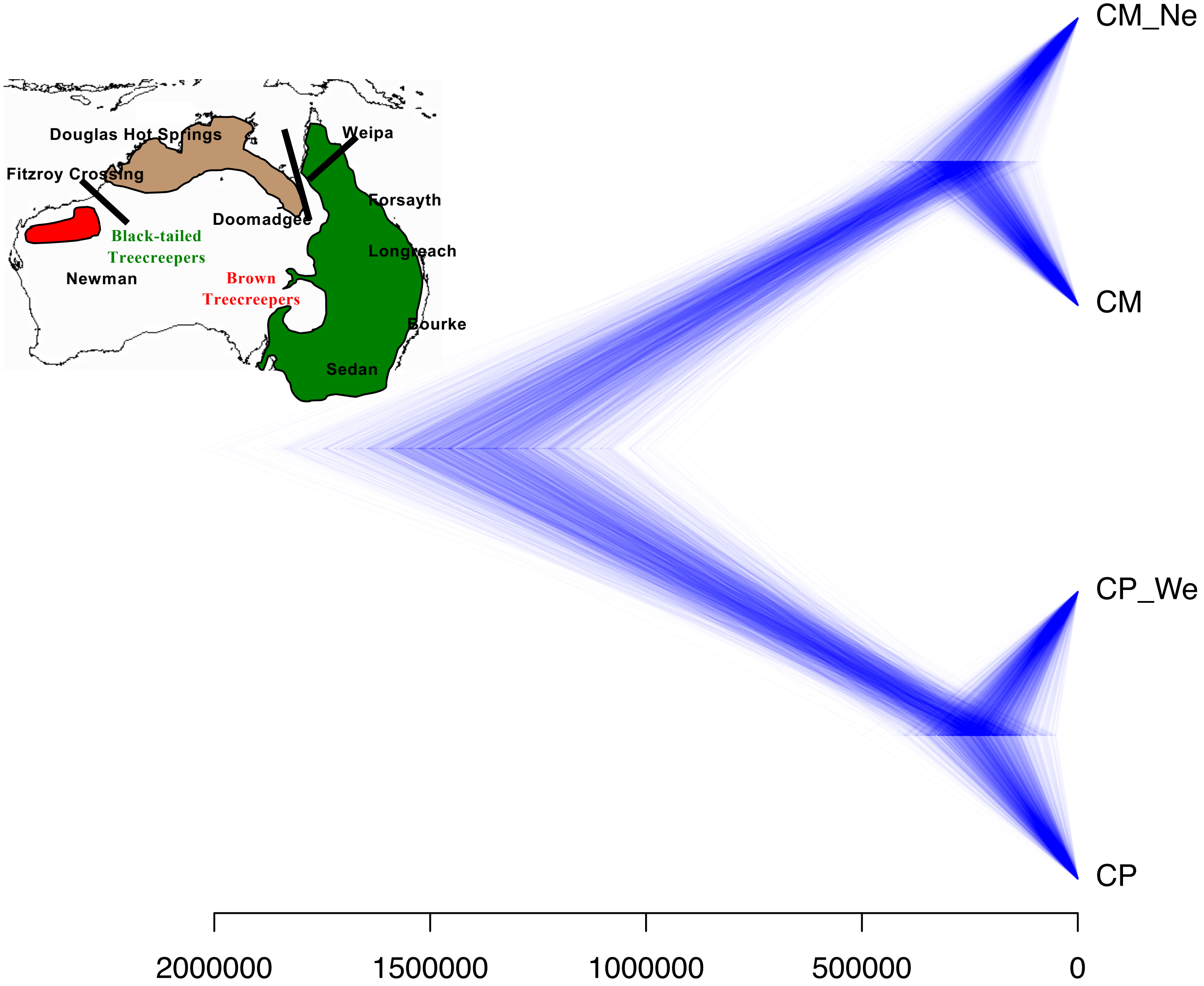


Multilocus estimates of migration rate





Phylogeny and divergence times



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

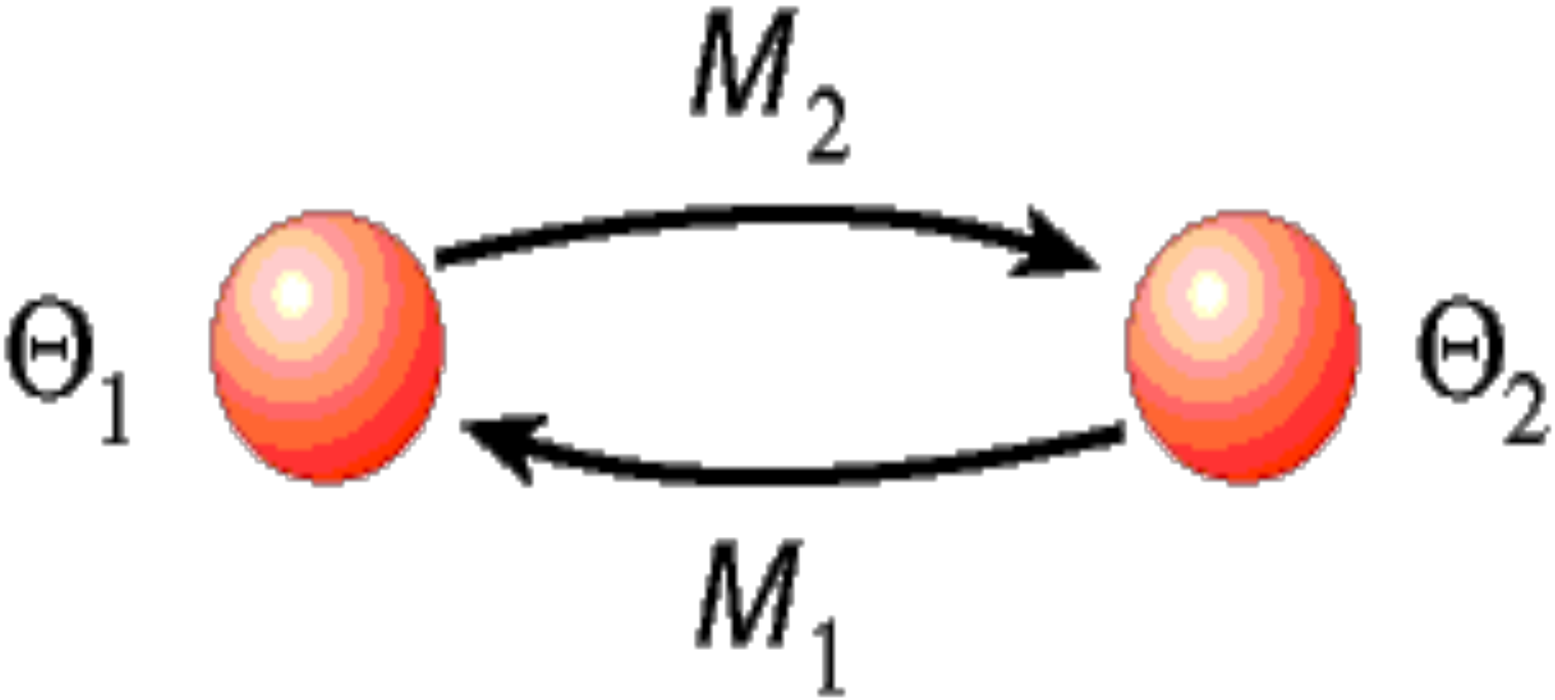
Model selection using Migrate

1

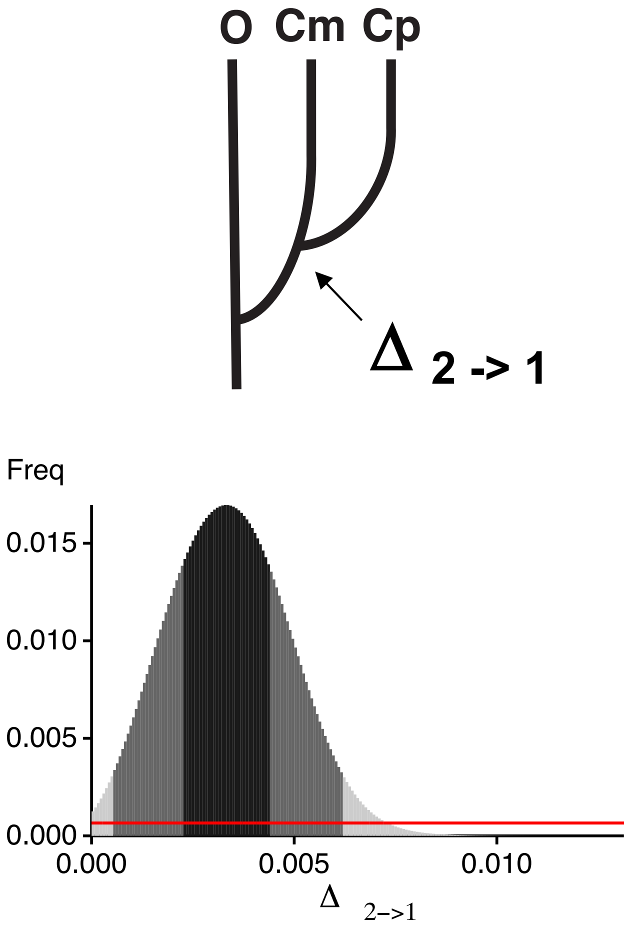
- 0 0 0
0 - 0 d
0 0 - d
d 0 0 -

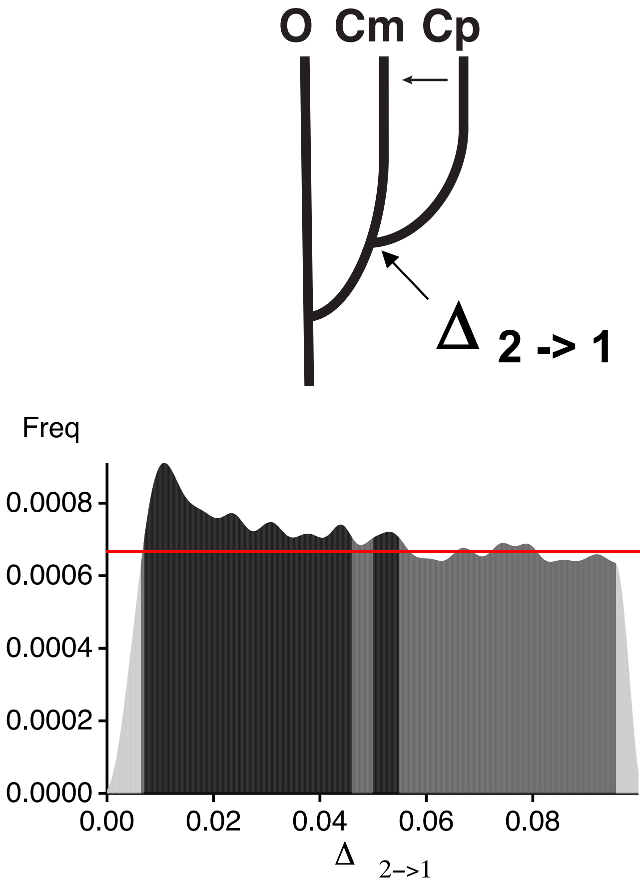
2 **Best model!**

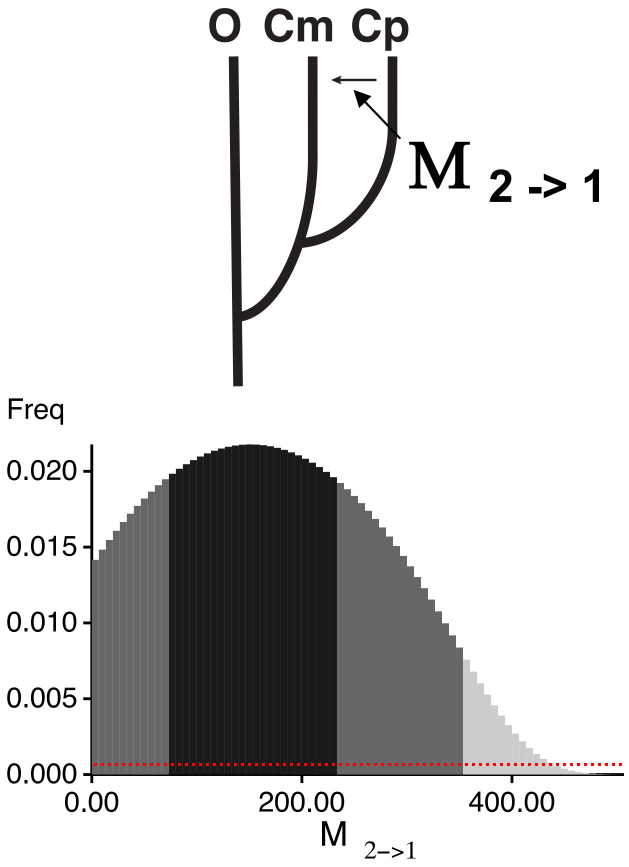
- 0 0
0 - d
d 0 -



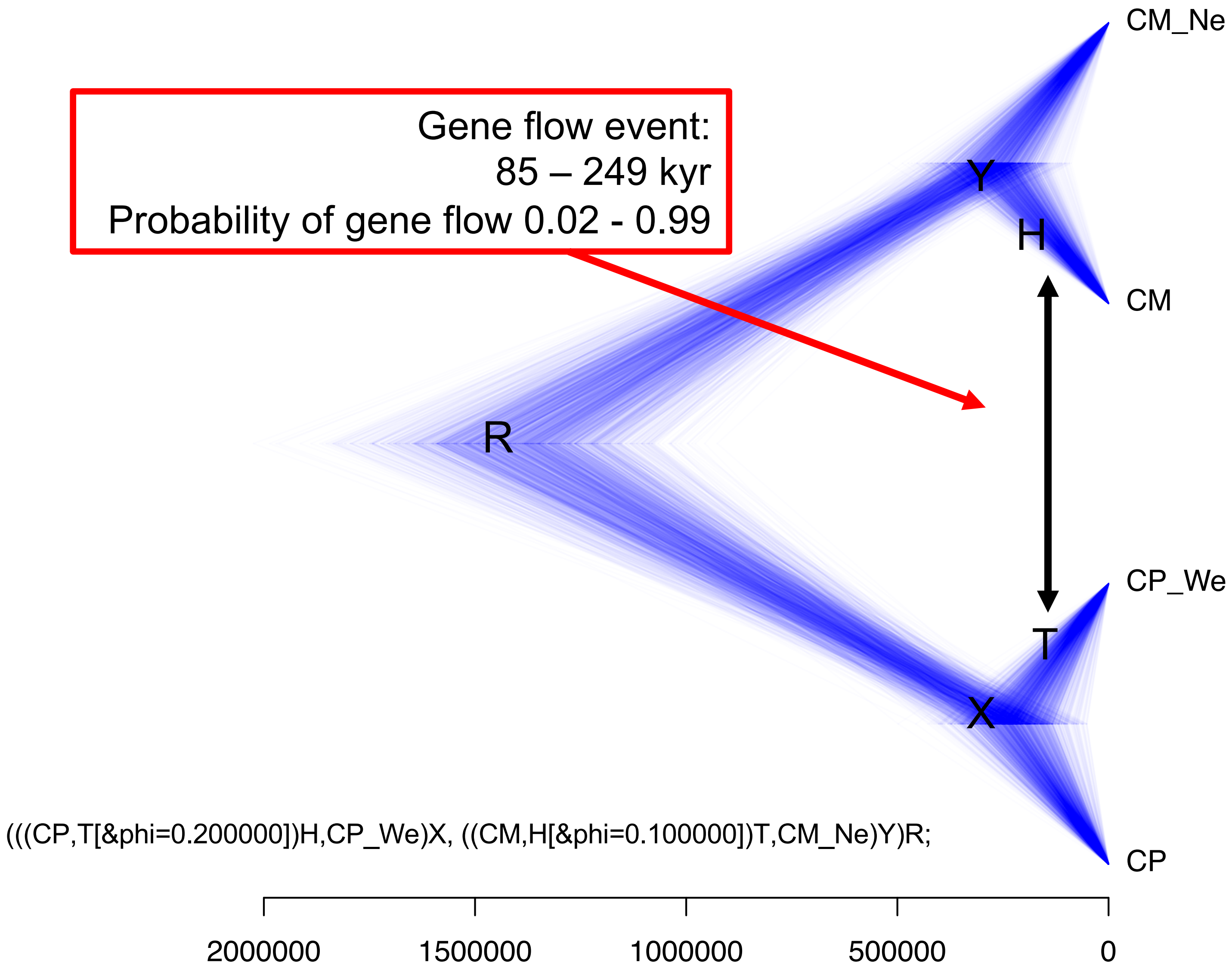
Challenges estimating isolation-migration parameters







Phylogeny and gene flow with bpp

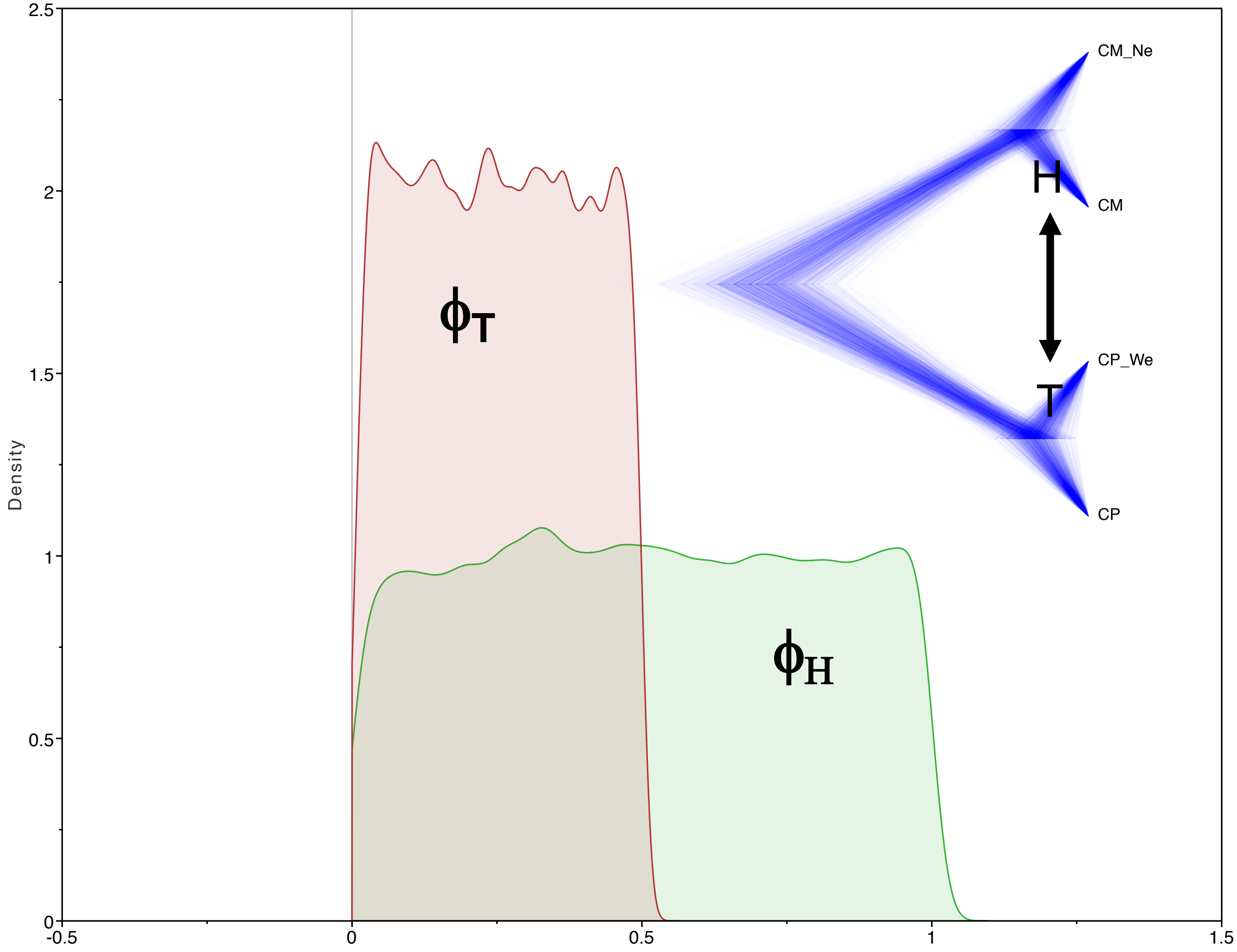


Gene flow event:
85 – 249 kyr
Probability of gene flow 0.02 - 0.99

(((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

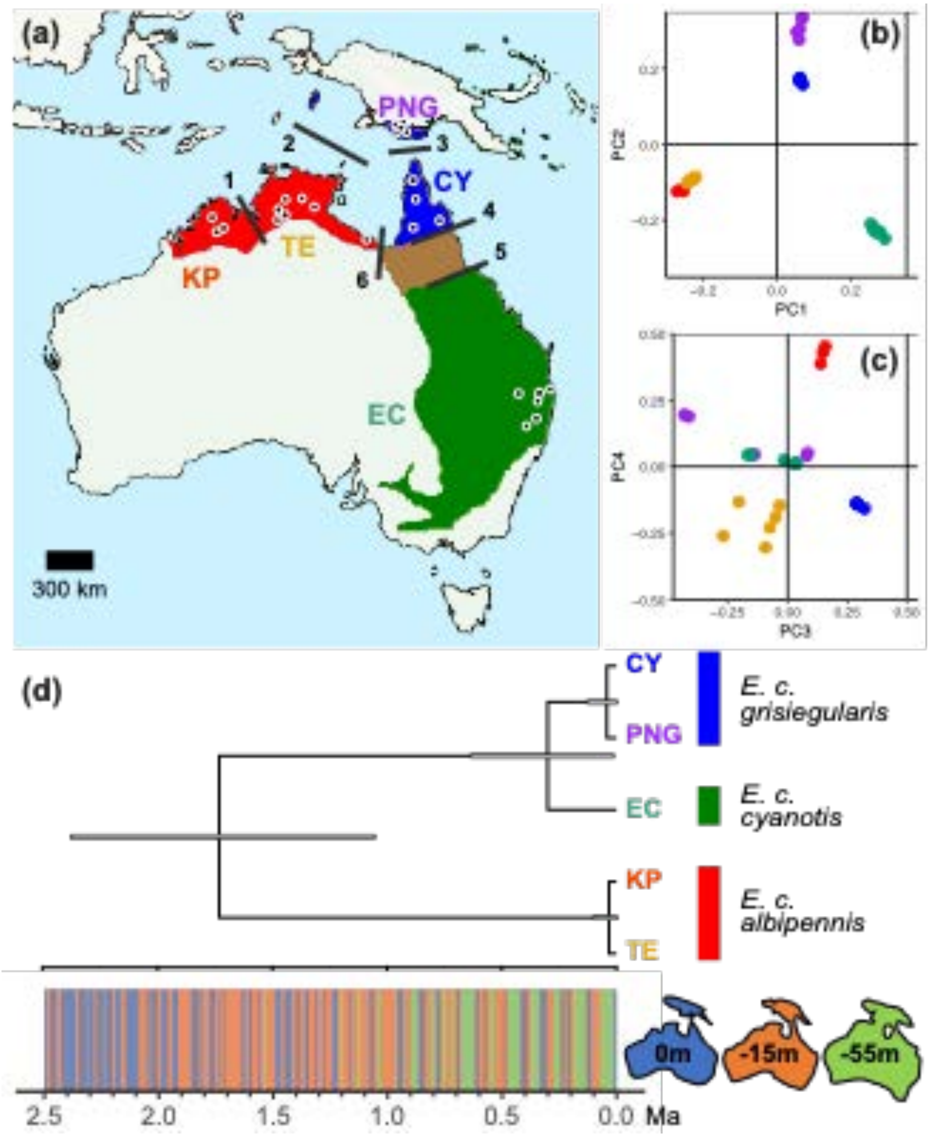
BPP: flat posterior distributions of gene flow probability



Probability of introgression ϕ

Whole-genome phylogeography of a widespread Australian honeyeater

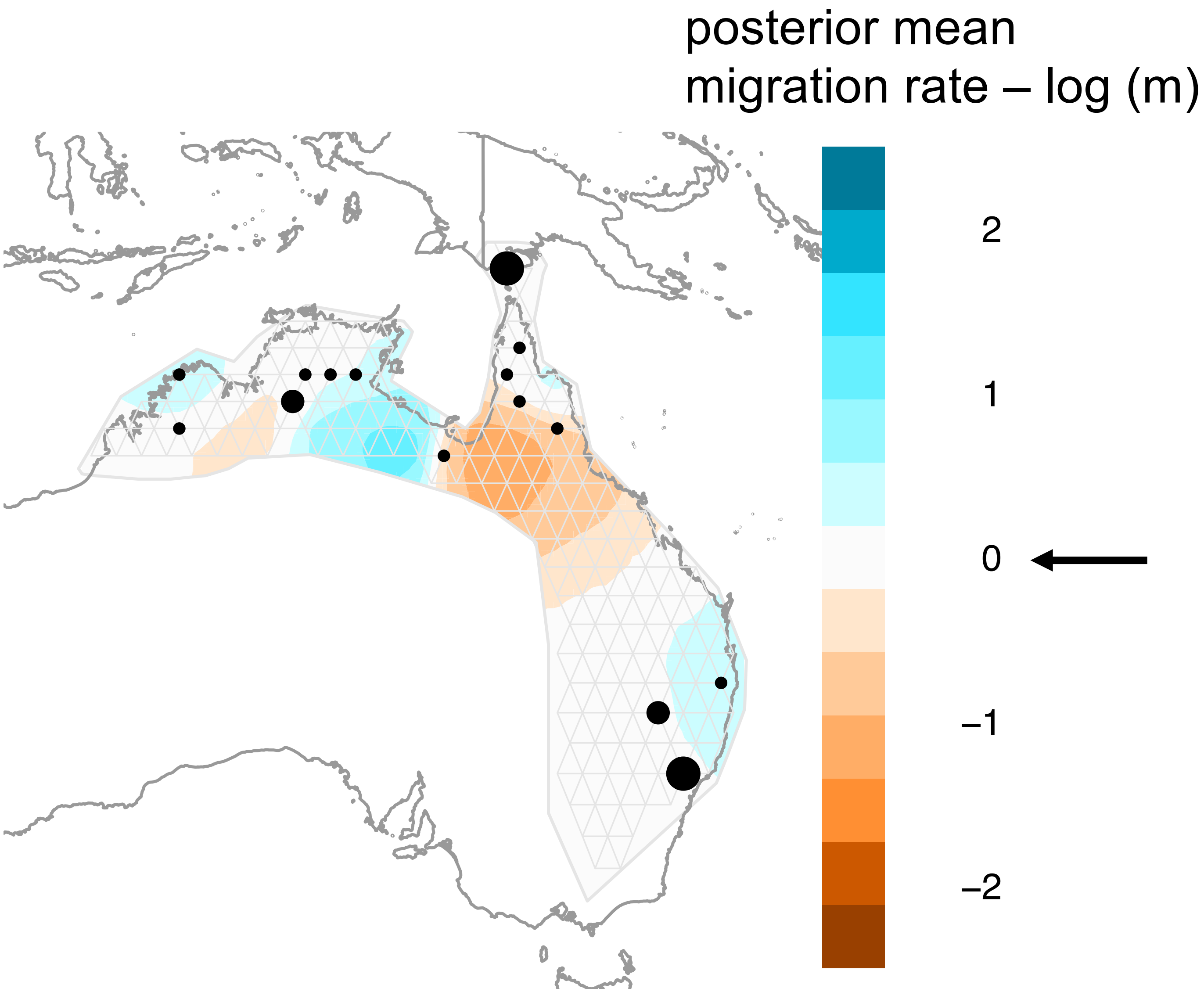
Blue-faced honeyeater
Entomyzon cyanotis



- **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

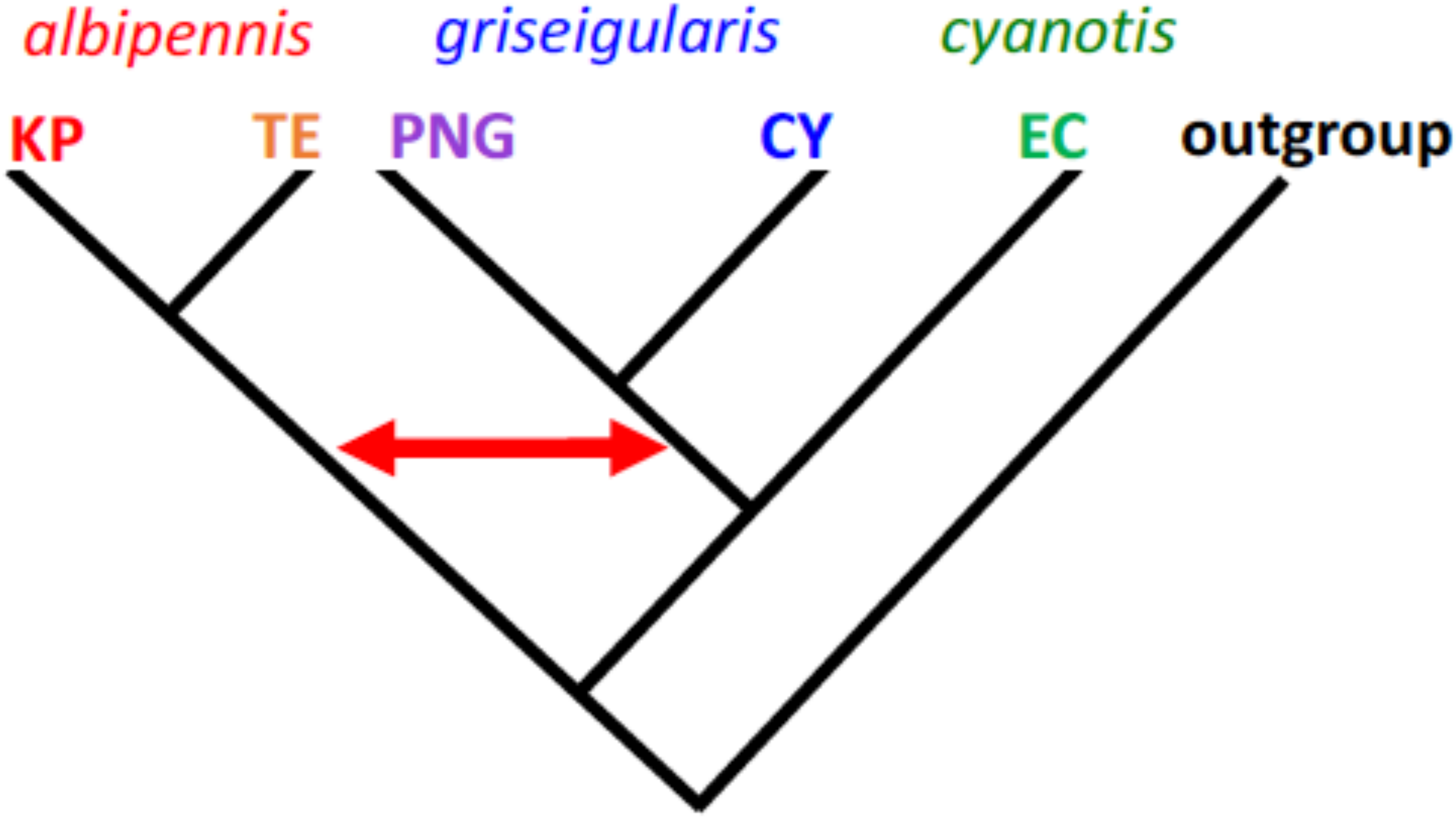
$$D = \frac{Num.\ ABBAs - Num.\ BABAs}{Num\ SNPs}$$

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

A **G** **A** **G**

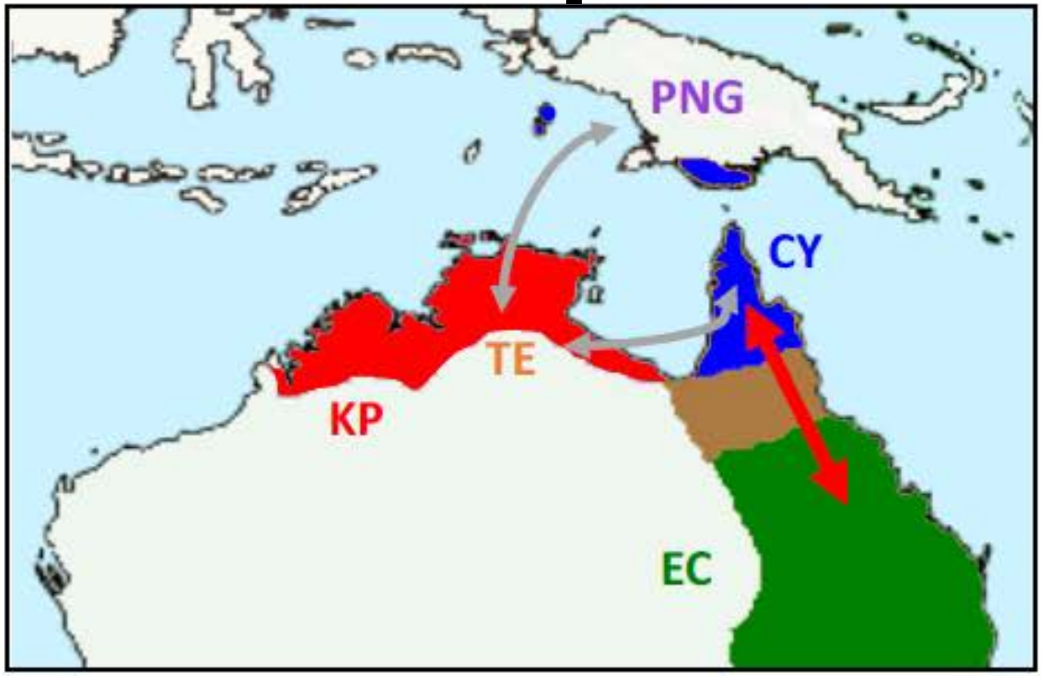
ABBA-BABA tests reveal signals of past introgression

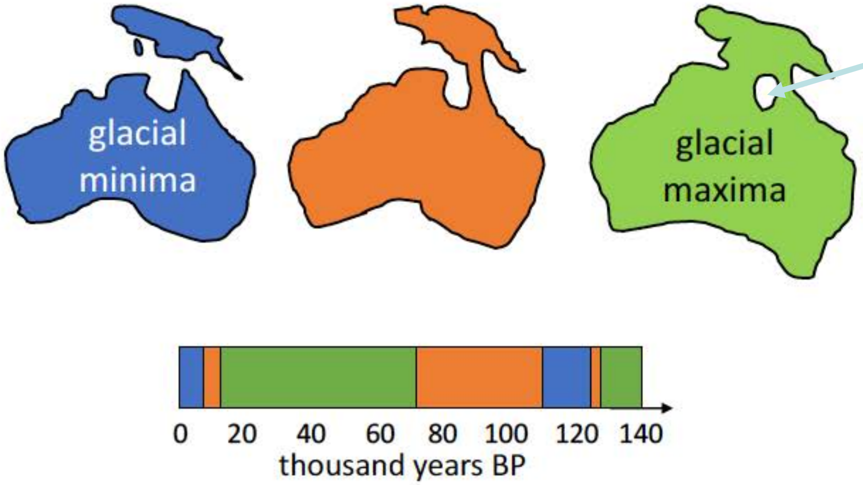
(Dsuite, Malinsky et al. 2021. *Mol. Ecol. Res.*)



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

Phylogeographic shadows of Lake Carpentaria





Pleistocene Lake Carpenta

Trends in Ecology & Evolution

CellPress REVIEWS

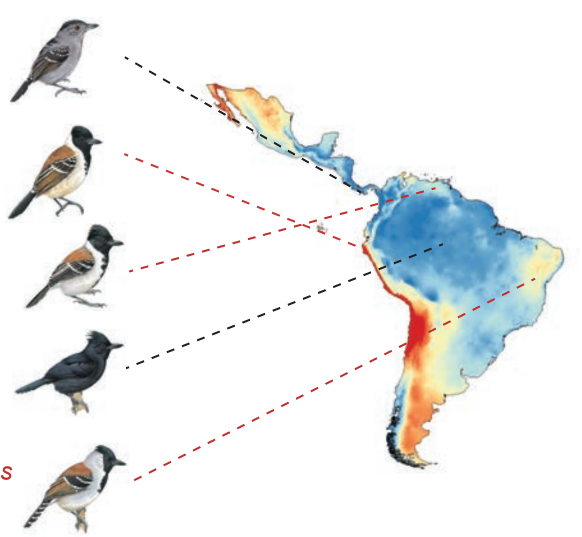
Opinion

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]

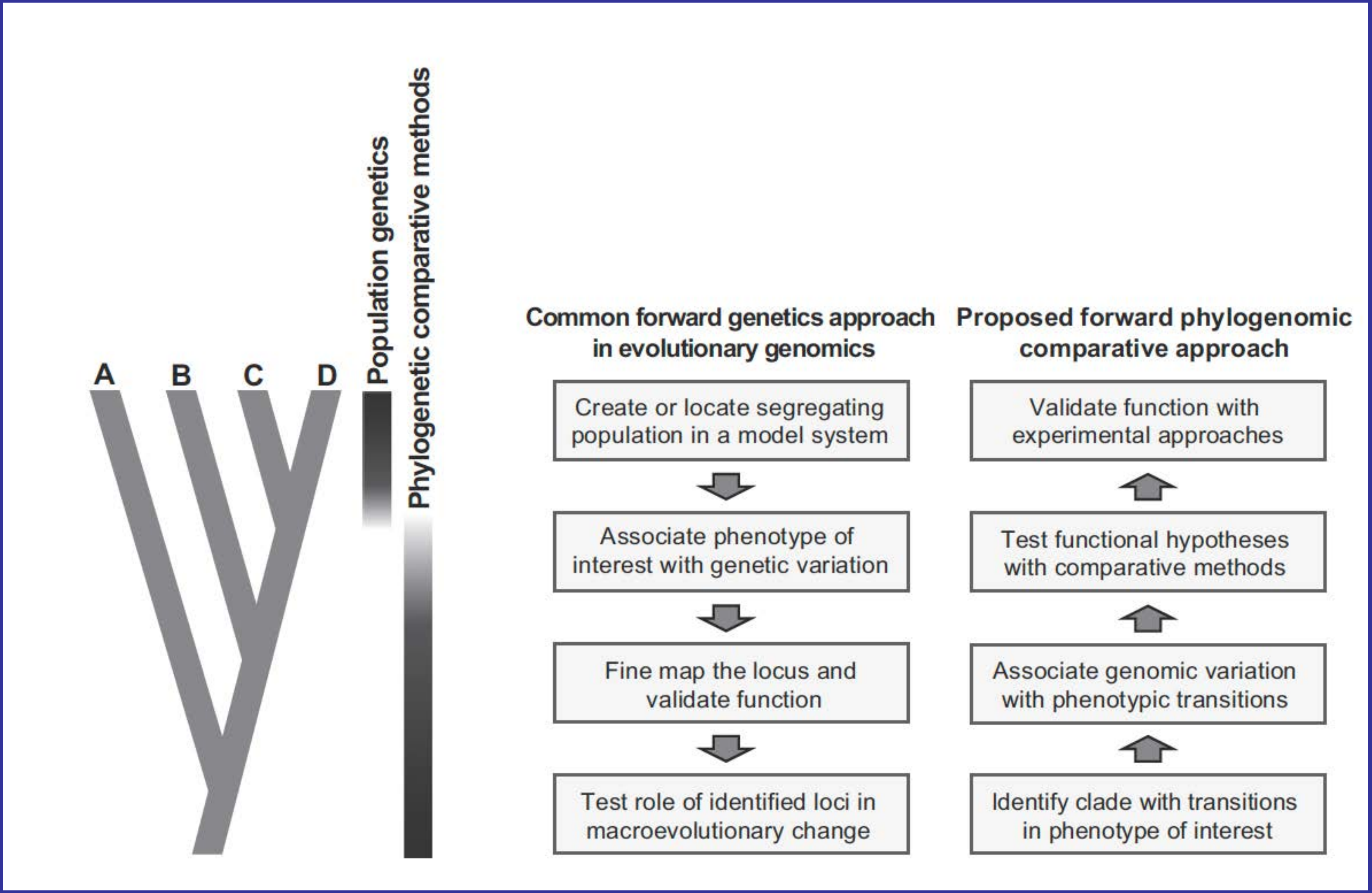
Smith, et al. 2020. *TREE* 35: P415-525

- 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

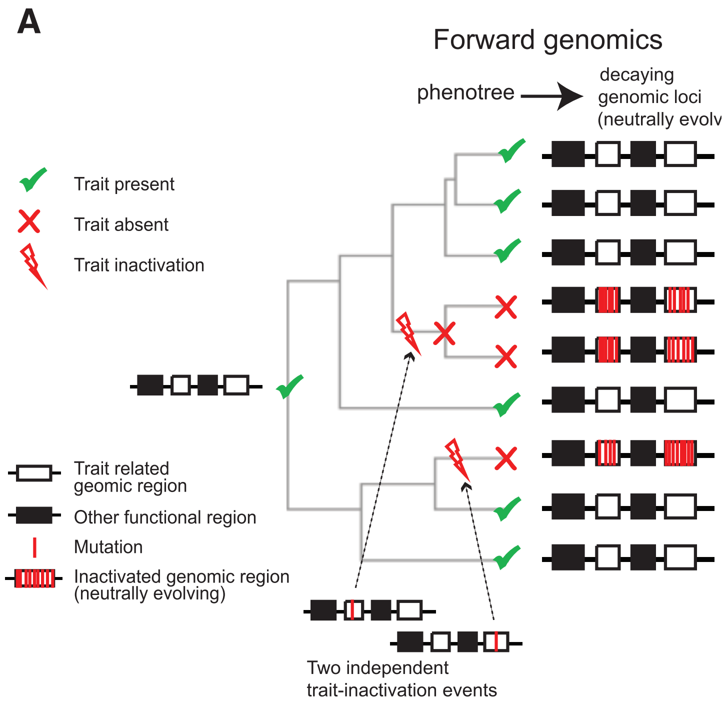
- Validate function with experimental approaches
- Test functional hypotheses with comparative methods
- Associate genomic variation with phenotypic transitions
- Identify clade with transitions in phenotype of interest

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

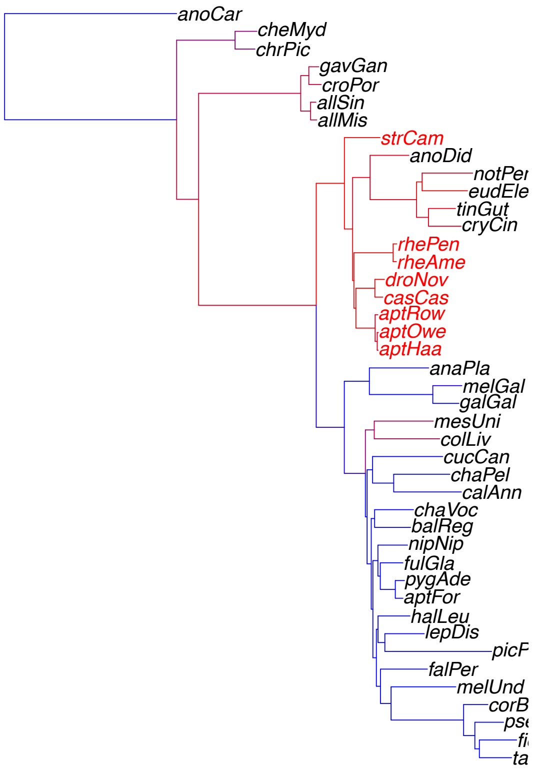
Genomic signatures of trait associations

Deletion or inactivation of conserved noncoding elements

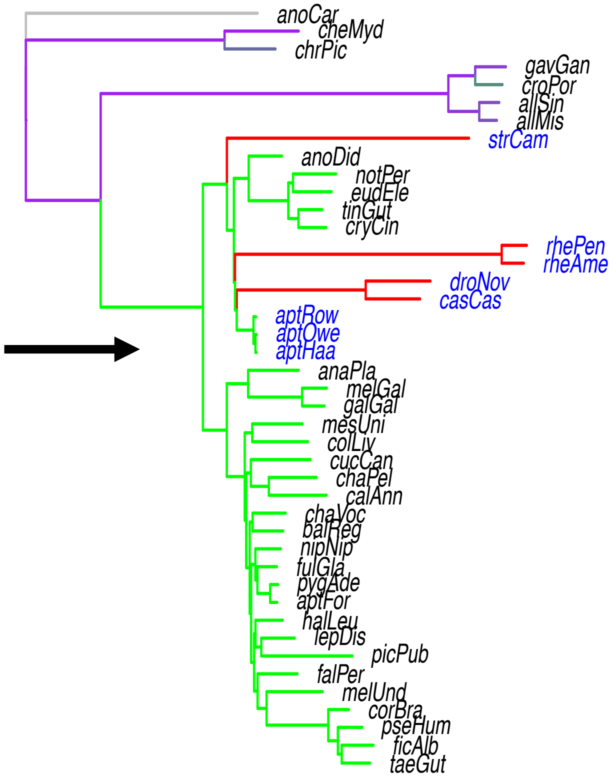
Acceleration of clade-wide conserved noncoding elements



Genome-wide neutral tree



Single locus accelerated tree



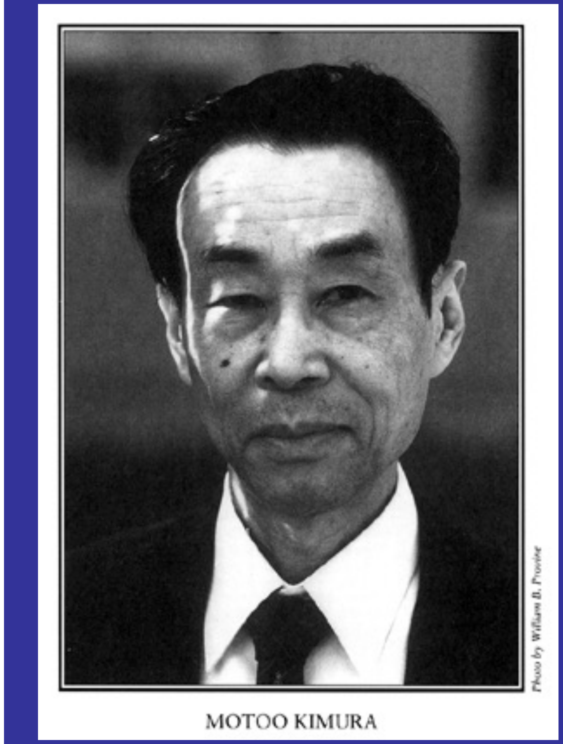
Marcovitz et al. 2016

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



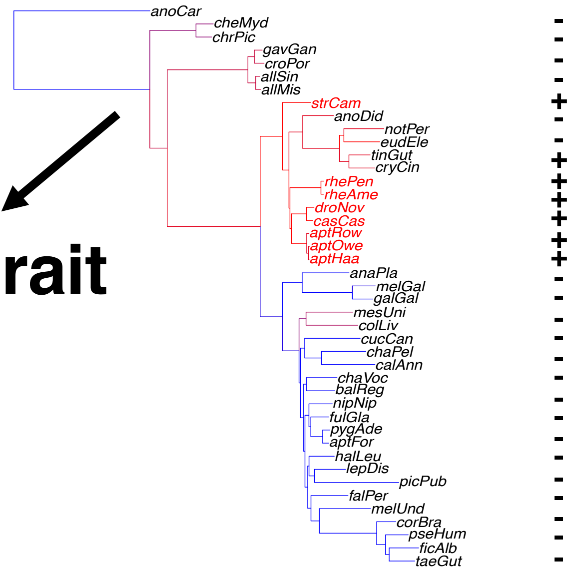
Motoo Kimura (1924-1994)

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

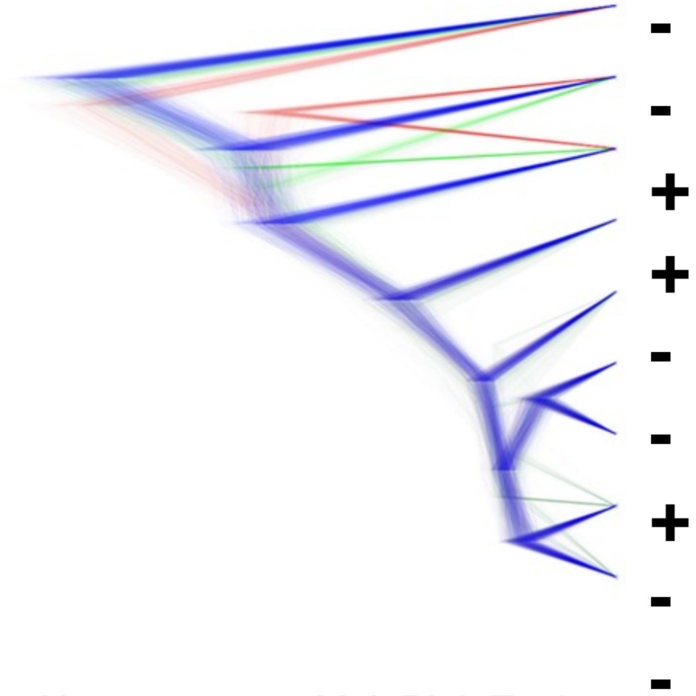
Functionally less important parts of genomes will have a higher substitution rate than functionally important ones

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

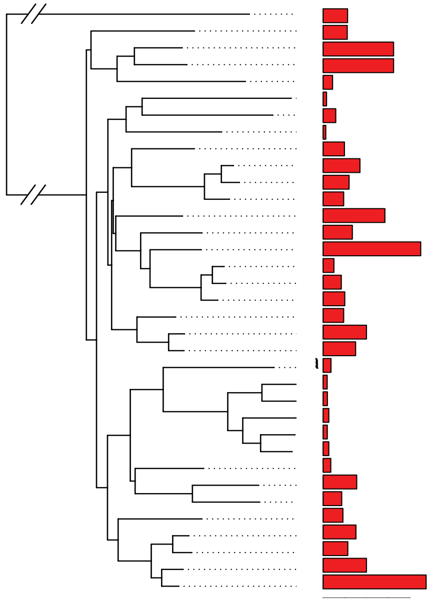
PhyloAcc binary trait



PhyloAcc-GT binary trait



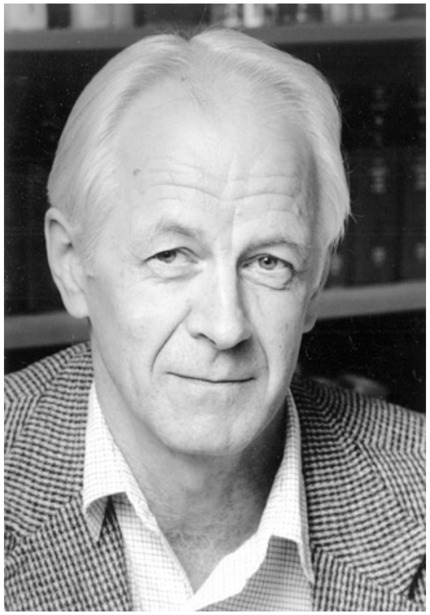
PhyloAcc-C continuous trait



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

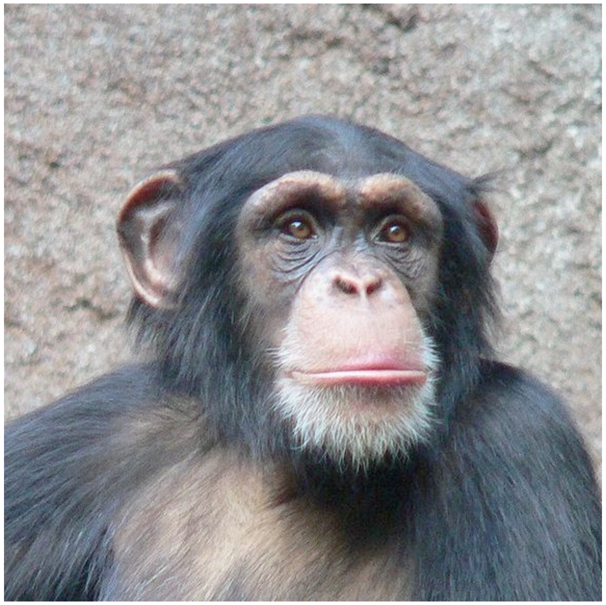
Yan et al. 2023. bioRxiv. https://doi.org/10.1101/2022.12.23.521765

Evolutionary change: genes or gene regulation?



Evolution at Two Levels in Humans and Chimpanzees

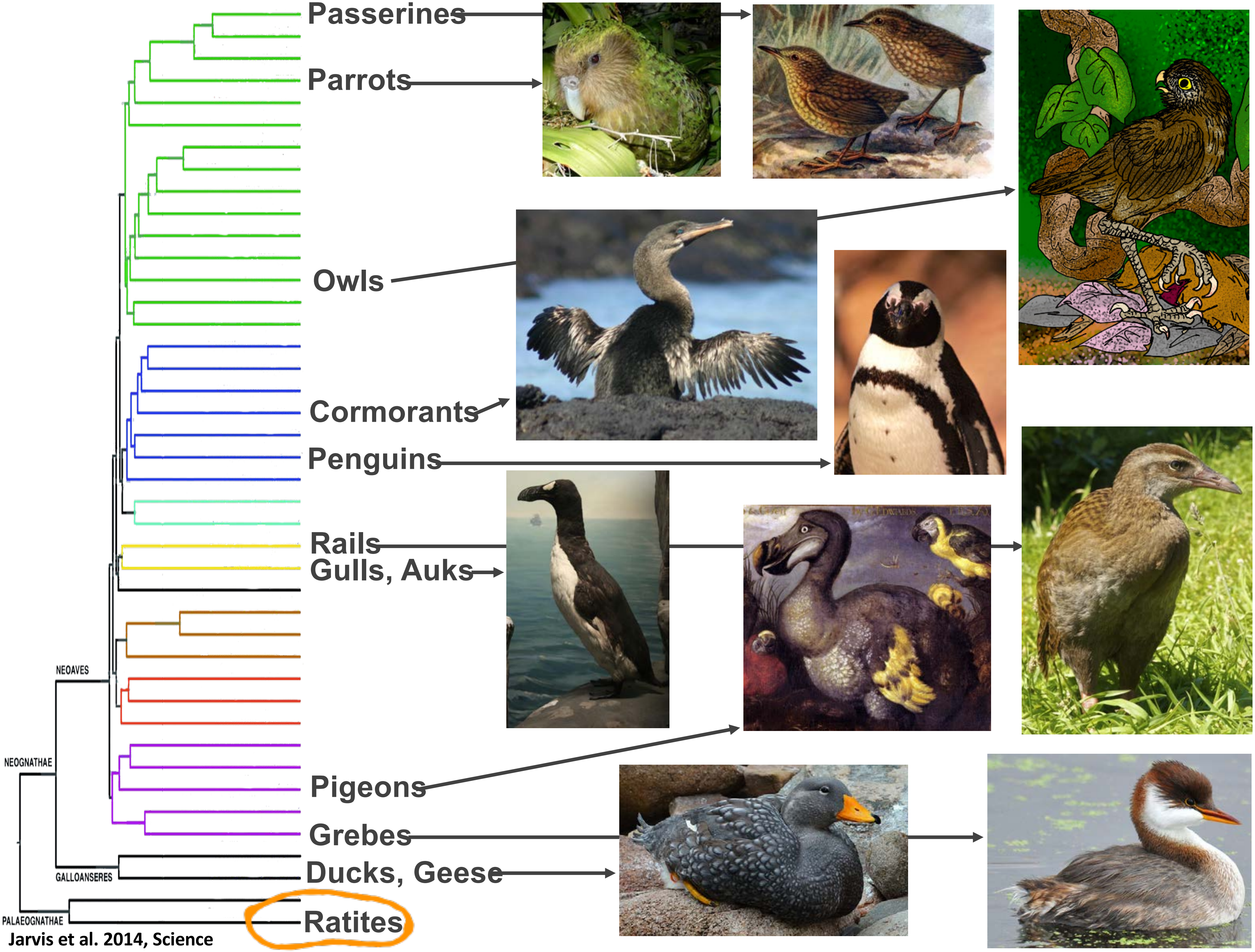
Their macromolecules are so alike that regulatory mutations may account for their biological differences.



Mary-Claire King and A. C. Wilson

SCIENCE

11 April 1975, Volume 188, Number 4184



Passerines

Parrots

Owls

Cormorants

Penguins

Rails

Gulls, Auks

Pigeons

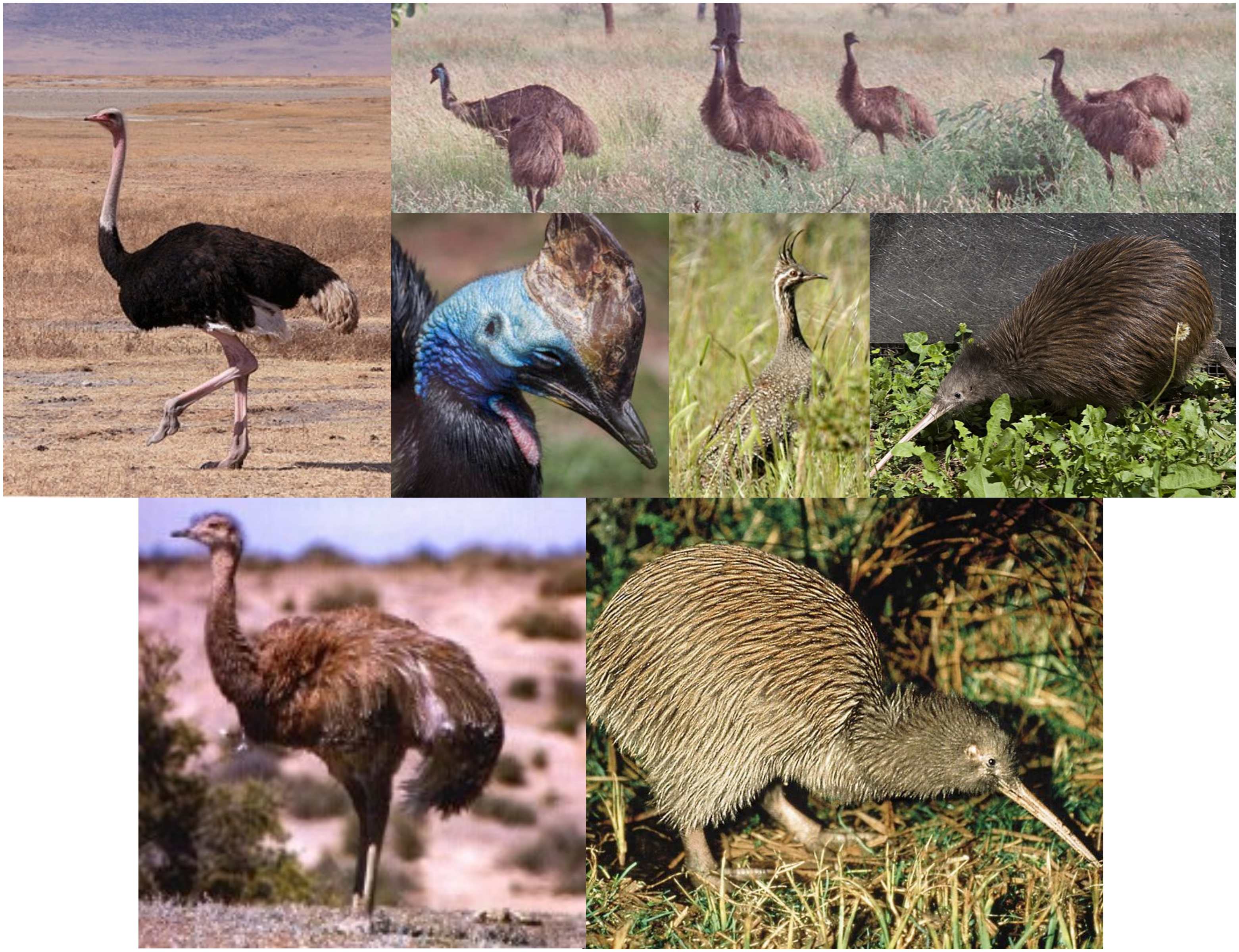
Grebes

Ducks, Geese

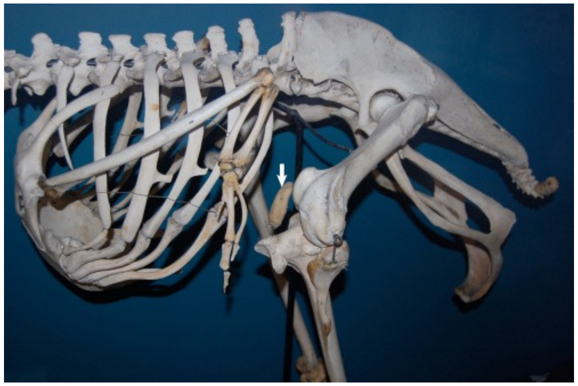
Ratites

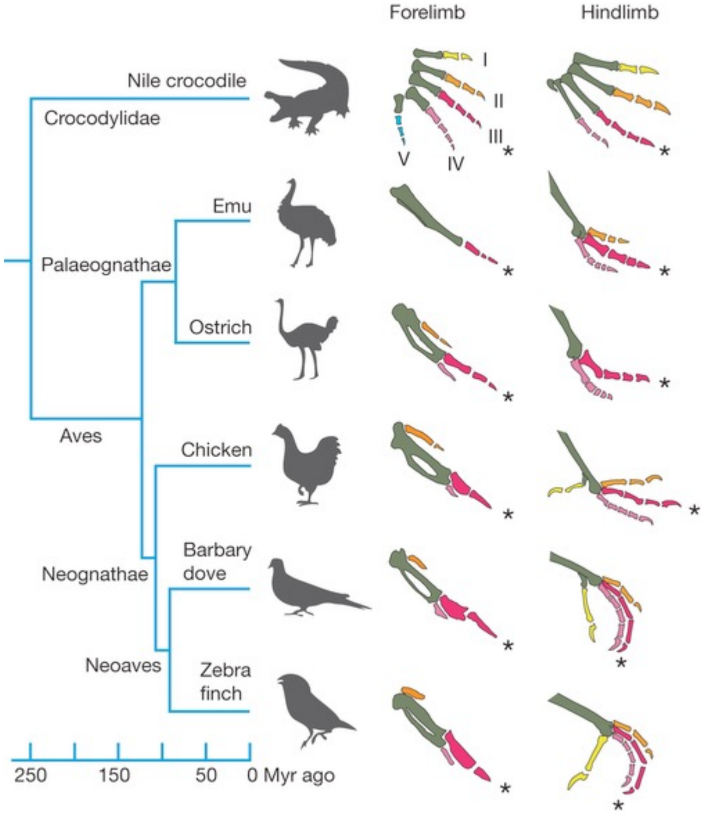
Jarvis et al. 2014, Science

CNEEs and the convergent evolution of flightlessness in Palaeognathae



Skeletal modifications for flightlessness

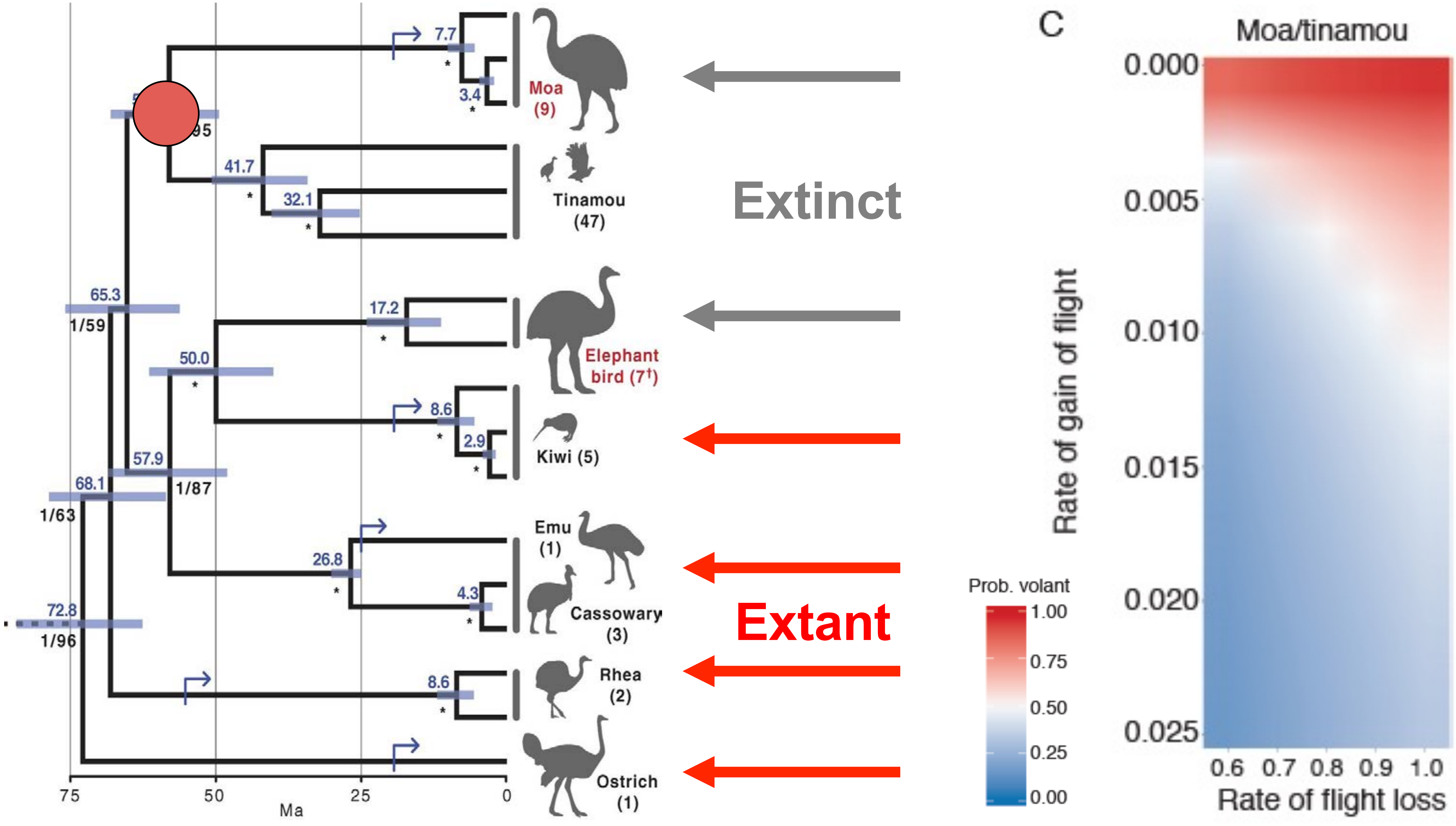




Little-spotted kiwi sternum

Emu and ostrich keelless sterna

Convergent losses of flight allow comparative genomics to identify genomic regions for flightlessness



Extinct

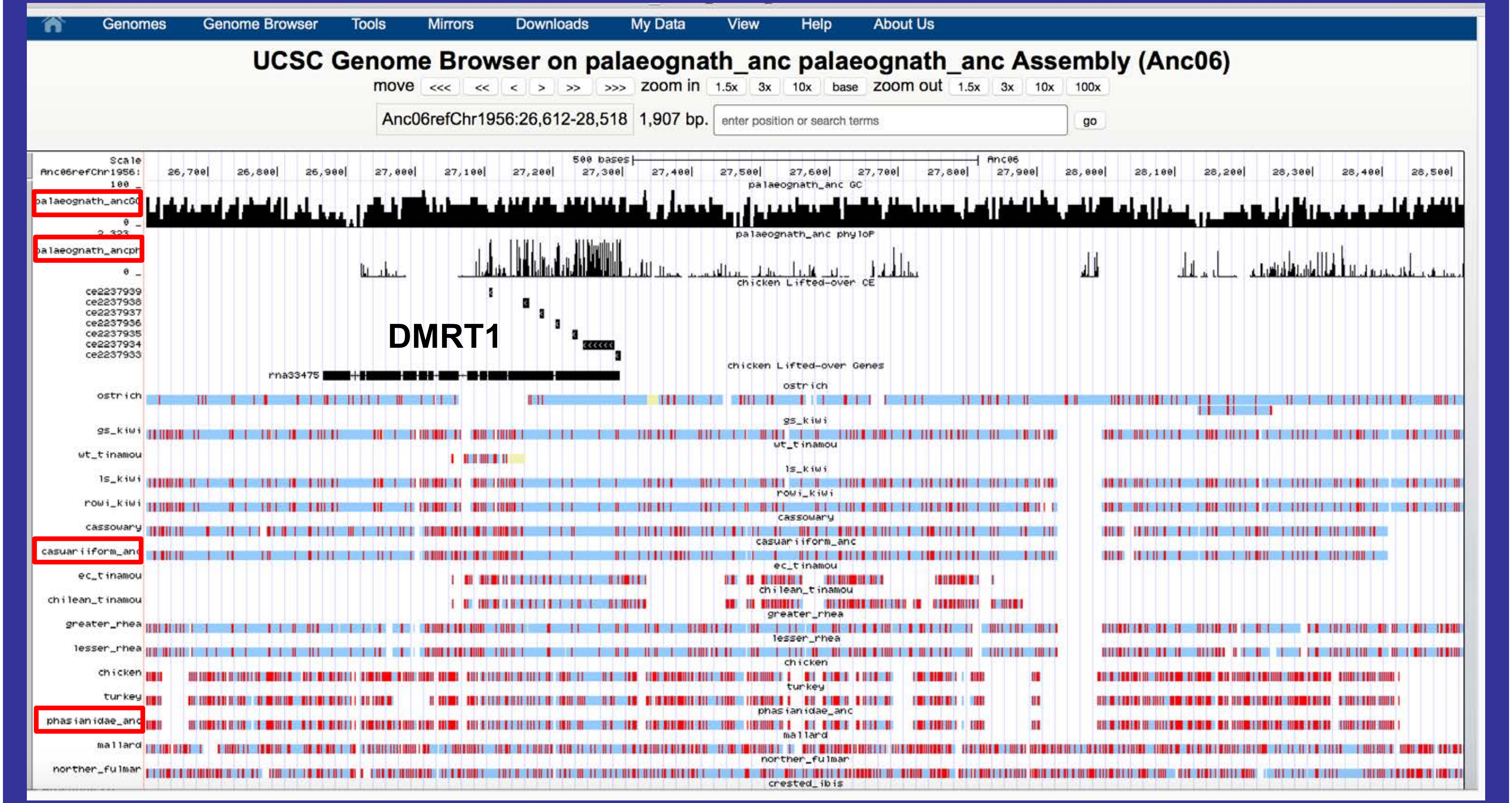
Extant

Mitchell et al 2014

11 new paleognath genomes

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

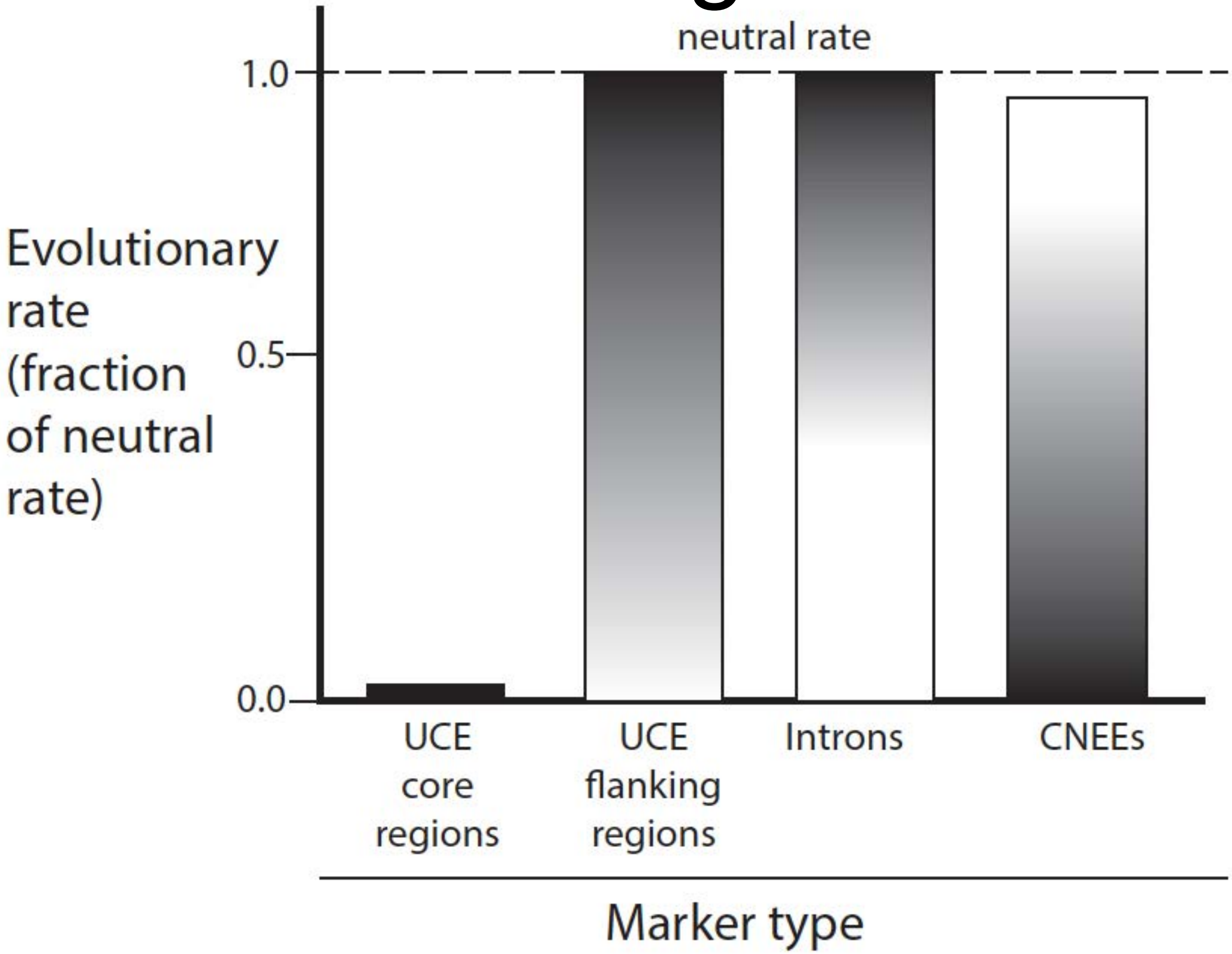
42-species whole genome alignment for birds using ProgressiveCactus



UCSC Genome Browser on palaeognath_anc palaeognath_anc Assembly (Anc06)

Anc06refChr1956:26,612-28,518 1,907 bp.

Relative rates of different noncoding markers

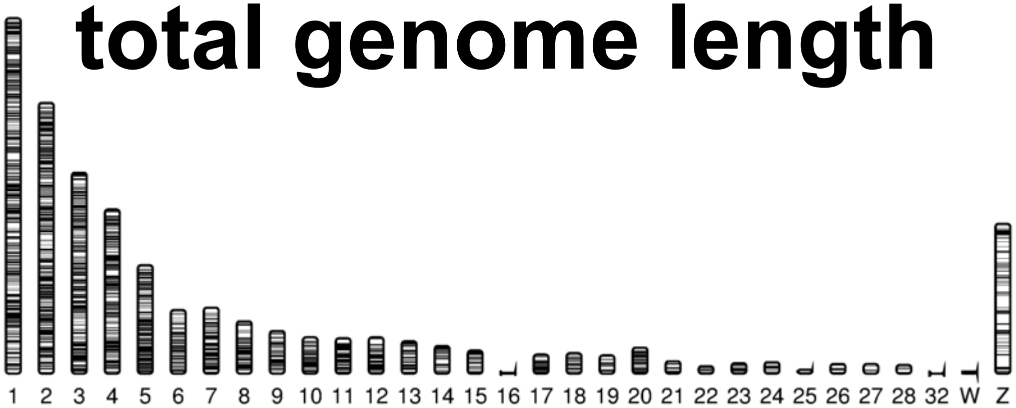


Edwards et al. 2017. *Syst. Biol*. 66: 1028

Phylogenomic markers cover *c*. 3% of total genome length

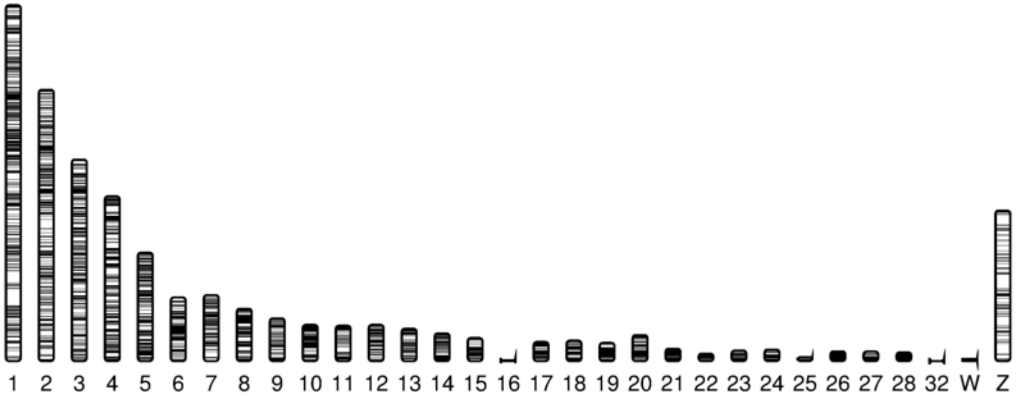
12,676

CNEEs



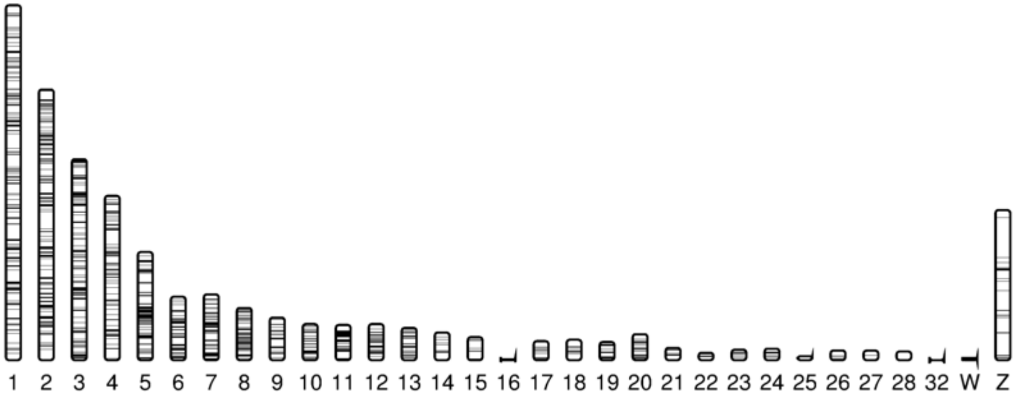
5,016

Introns



3,158

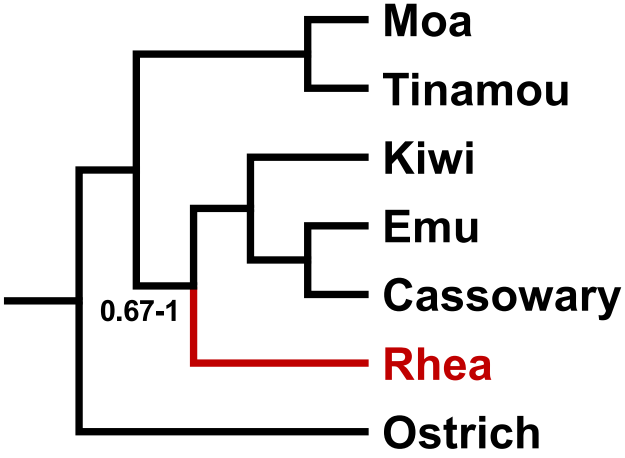
Ultraconserved elements (UCEs)



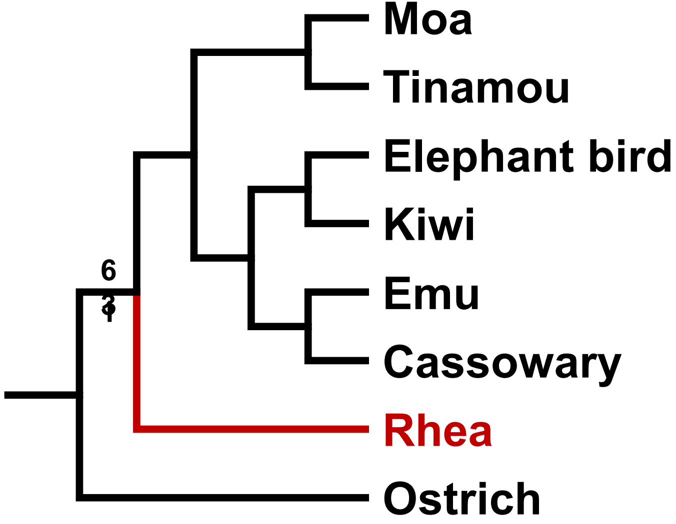
Cloutier et al. 2019. *Syst. Biol*. 10.1093/sysbio/s

Relationships of rheas unclear

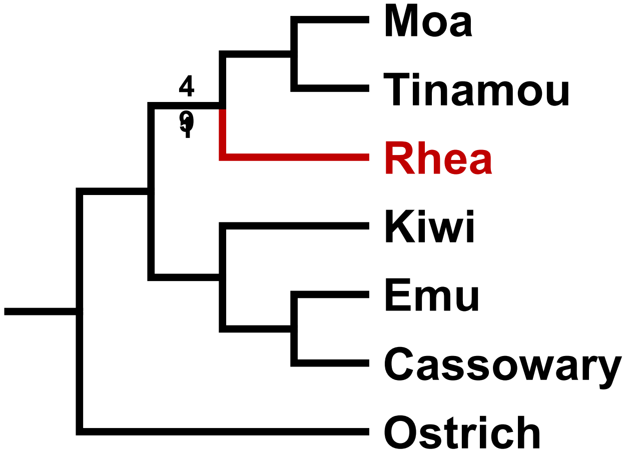
Haddrath & Baker (2012) 27 nuclear loci



Mitchell et al. (2014) mtDNA



Smith et al. (2013) 60 nuclear loci + mtDNA

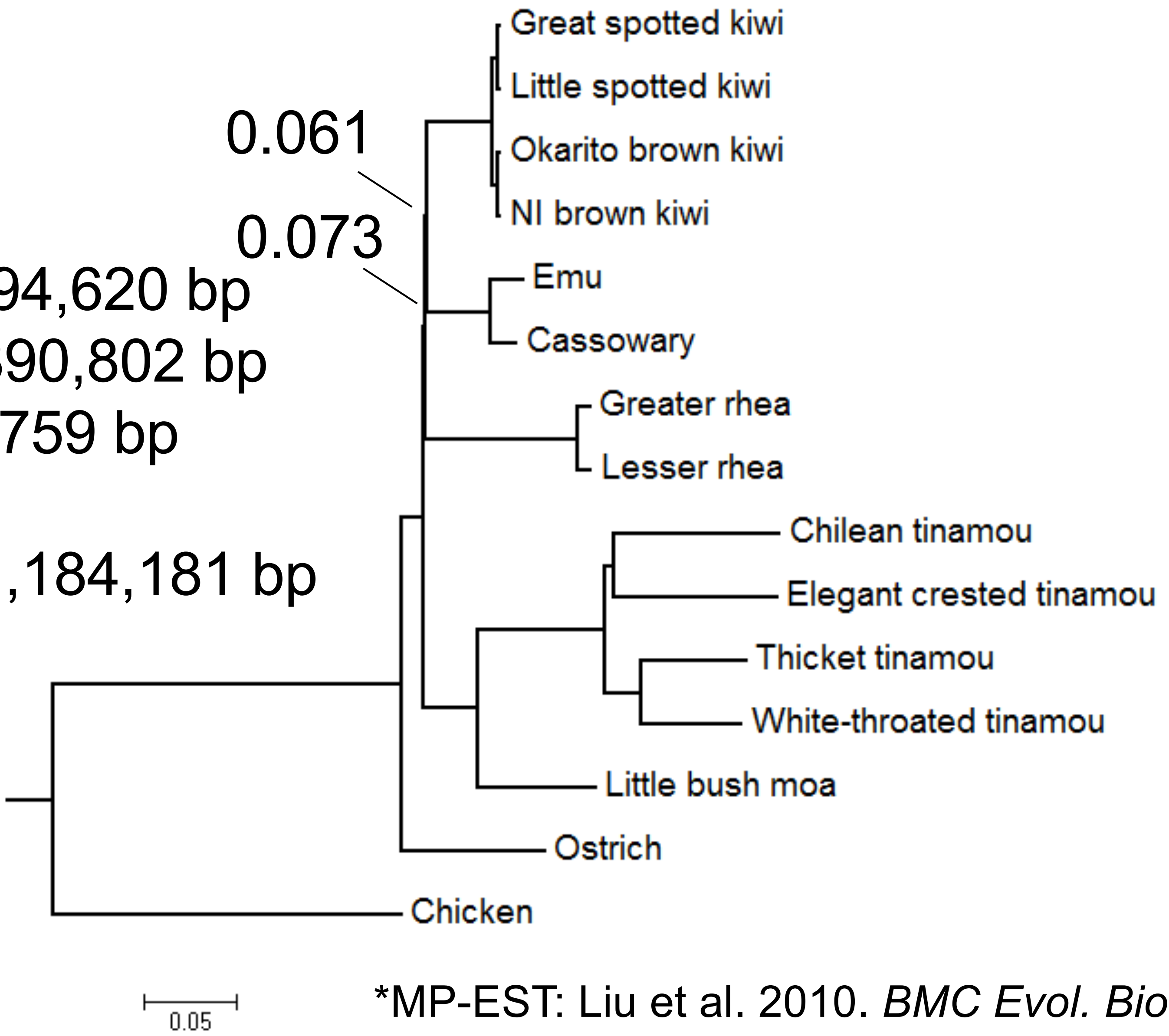


Coalescent* analyses resolve the position of rheas and reveal an ancient rapid radiation

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

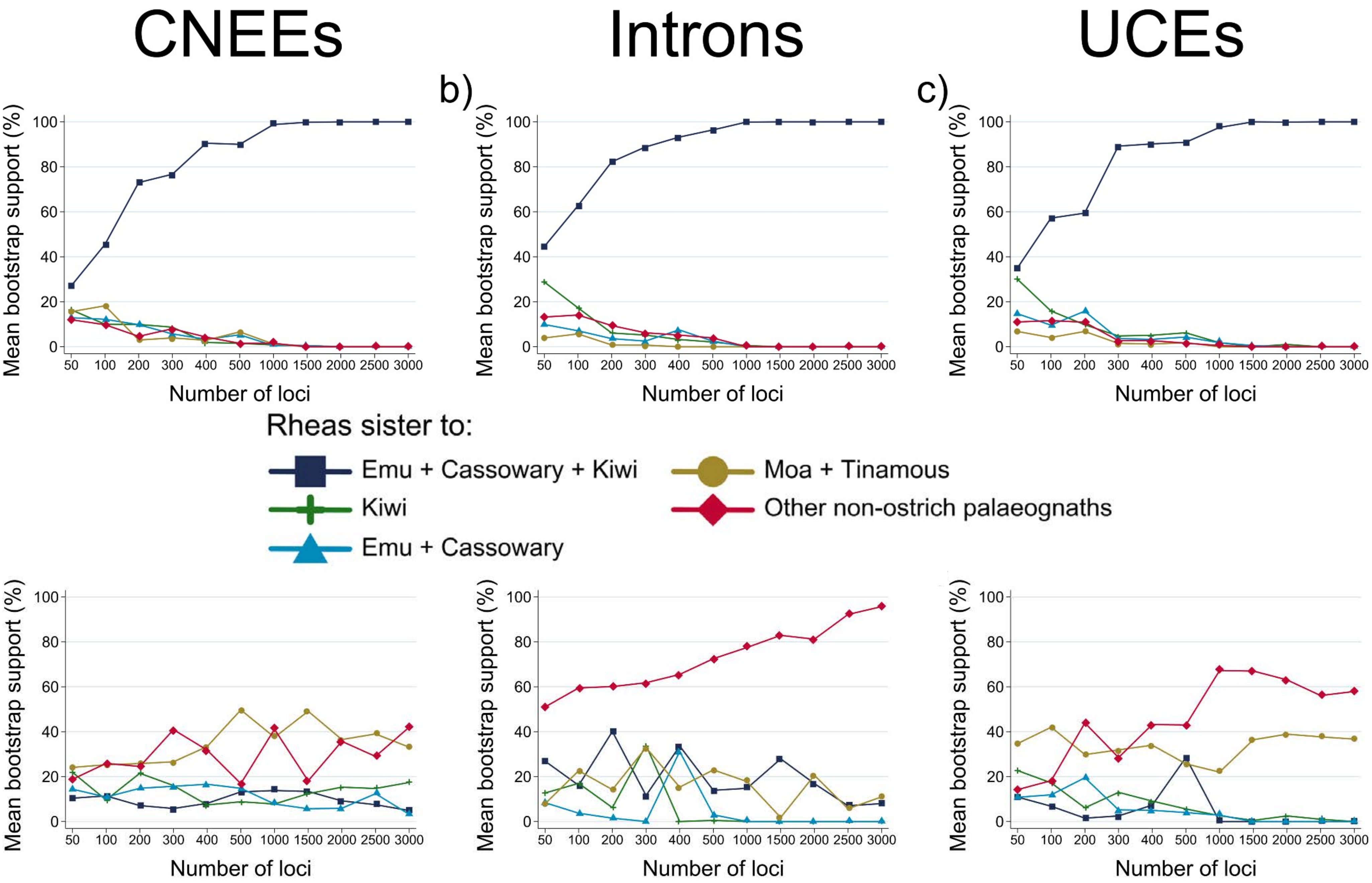
Total: 20,850 loci; 41,184,181 bp

Branch lengths in coalescent units



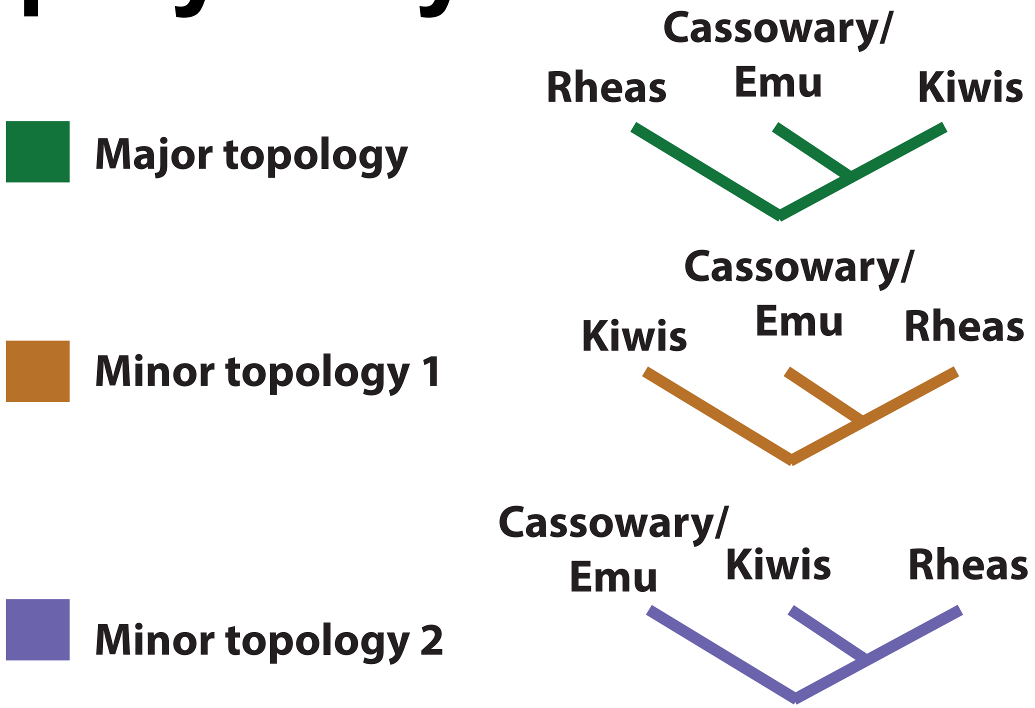
*MP-EST: Liu et al. 2010. *BMC Evol. Biol.*

Consistent accumulation of phylogenetic signal using MP-EST



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

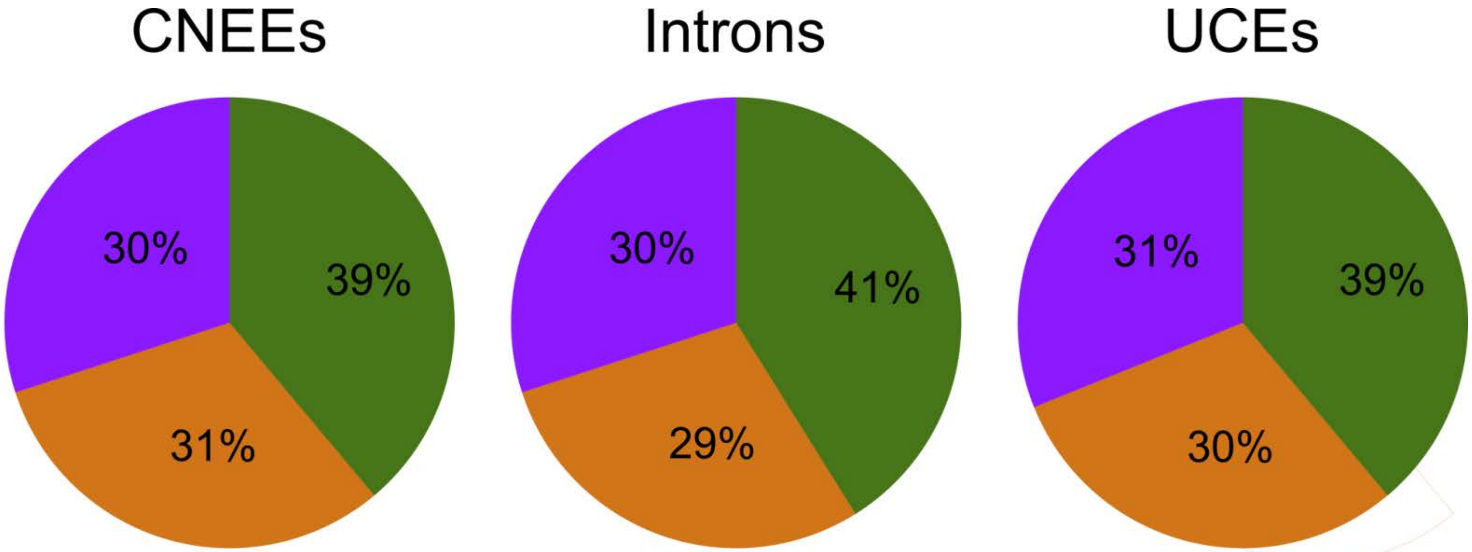
Gene tree distribution suggests a near polytomy at base of ratites



🟩 **Major topology**

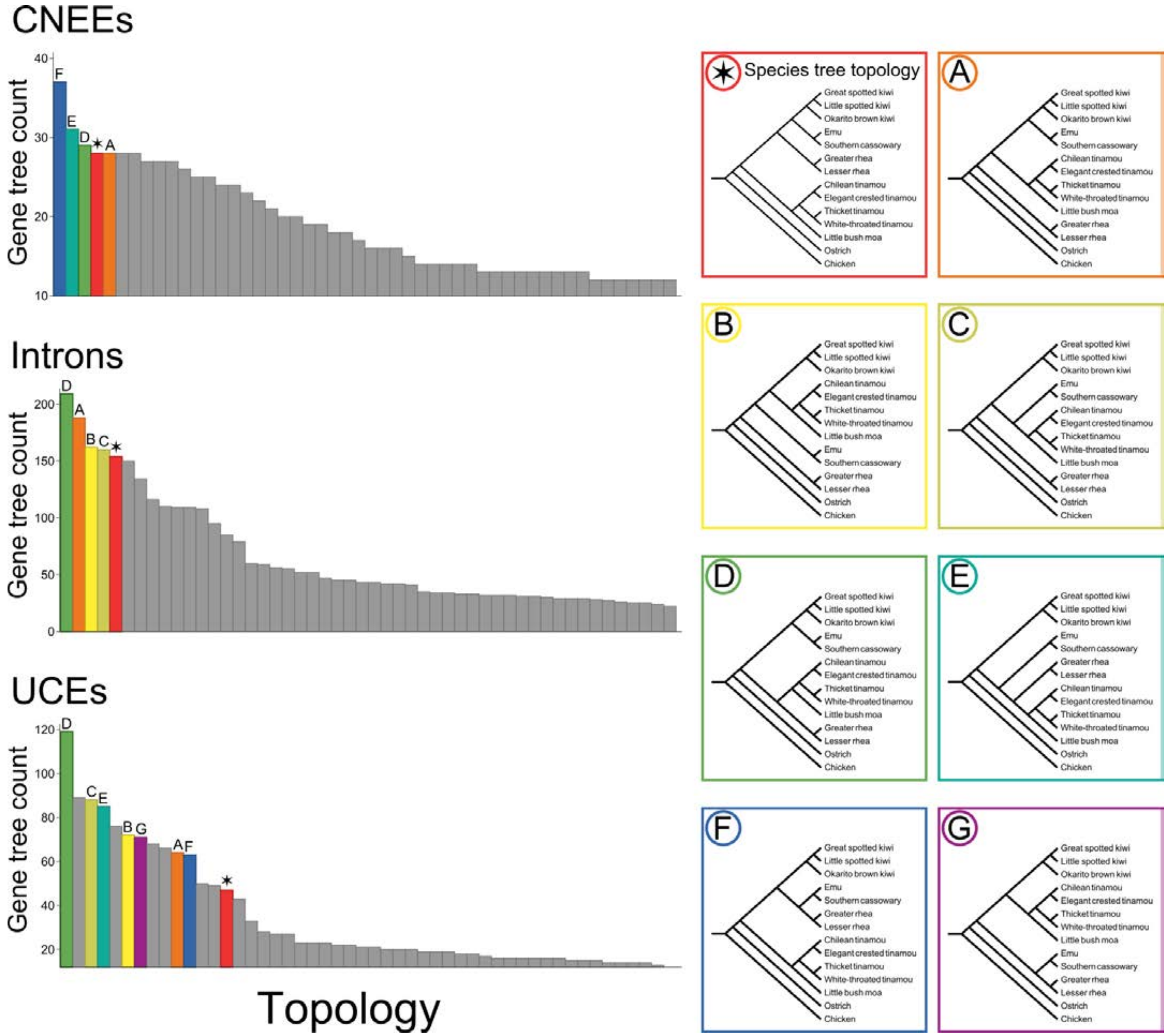
🟧 **Minor topology 1**

🟪 **Minor topology 2**



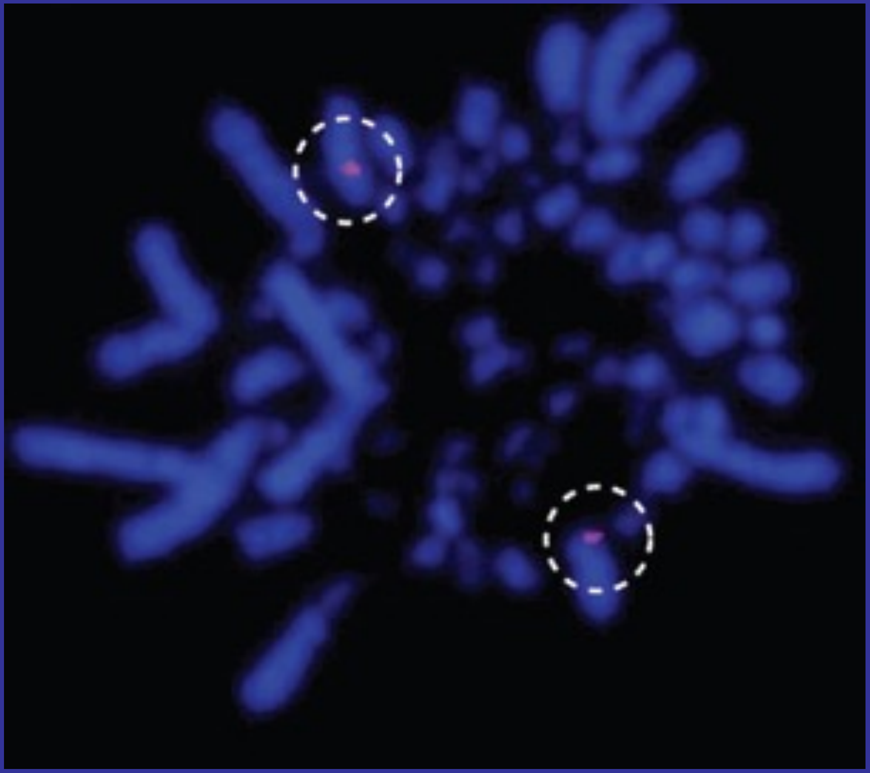
Cloutier et al. 2019. *Syst. Biol.* 10.1093/sysbio/syz019

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



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

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

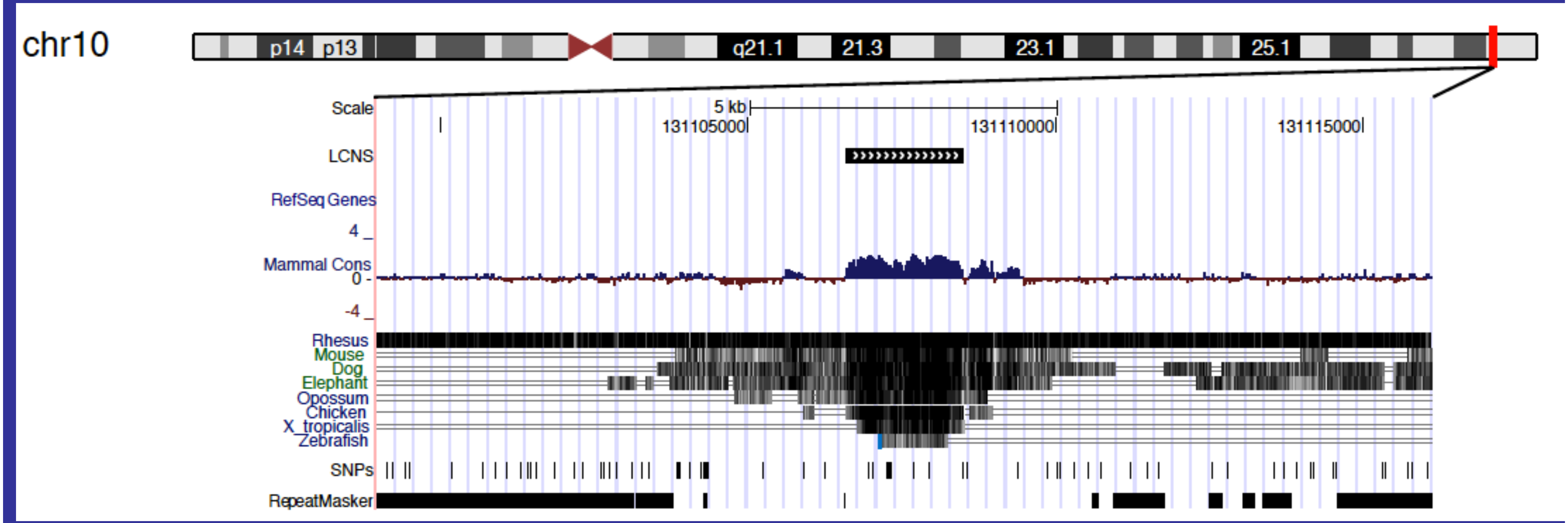


Karyotype of an Emu

CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements

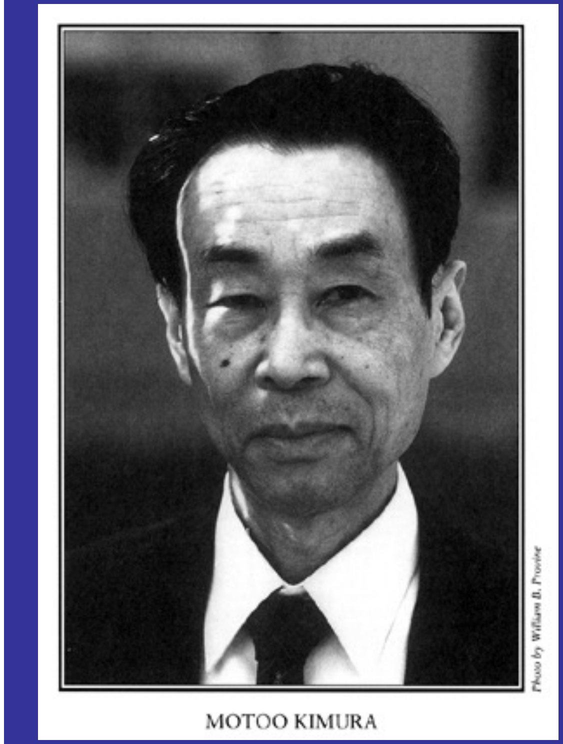
284,001 long (* > 50 bp) CNEEs in data set



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

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

Neutral Theory of Molecular Evolution

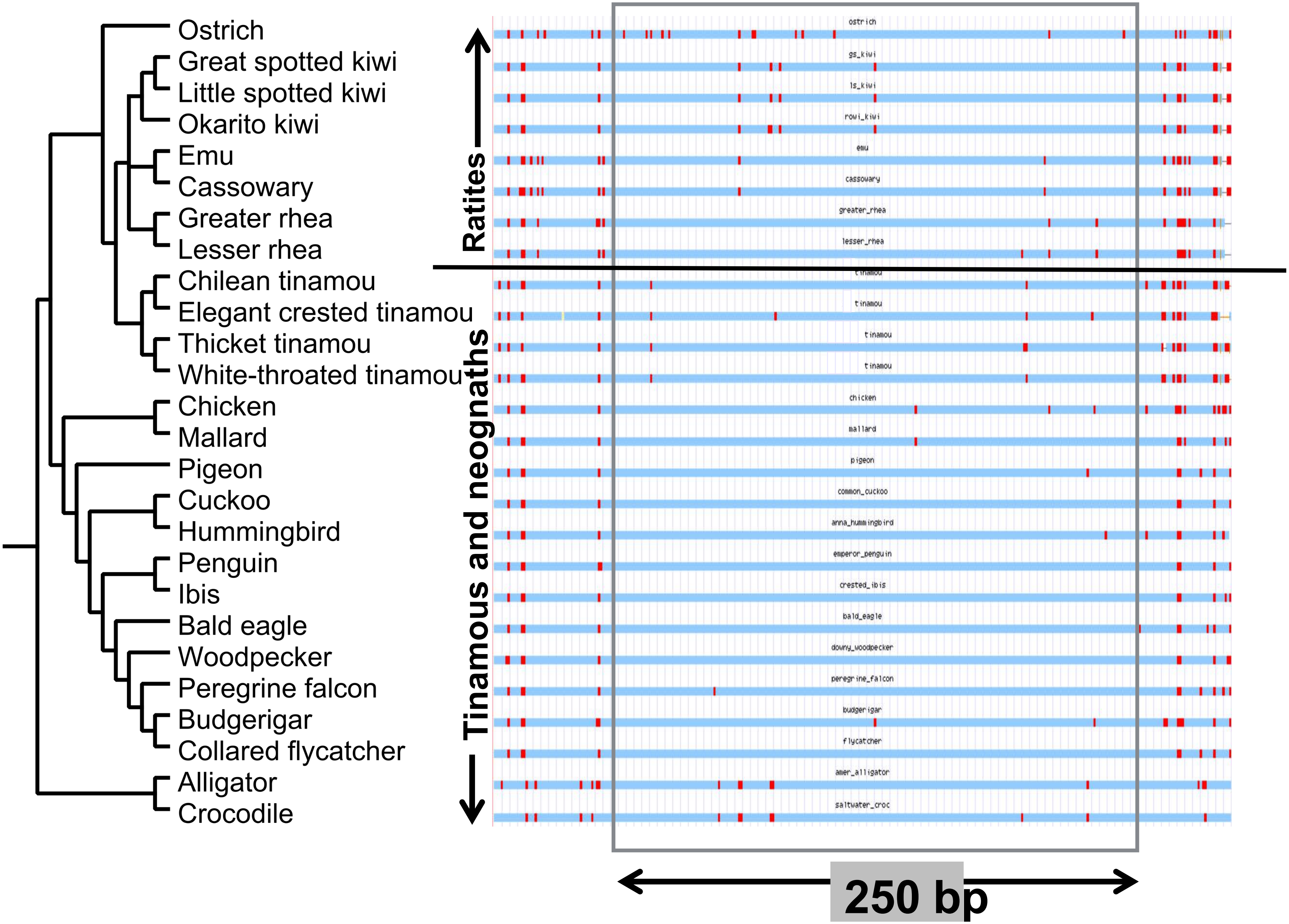


Motoo Kimura (1924-1994)

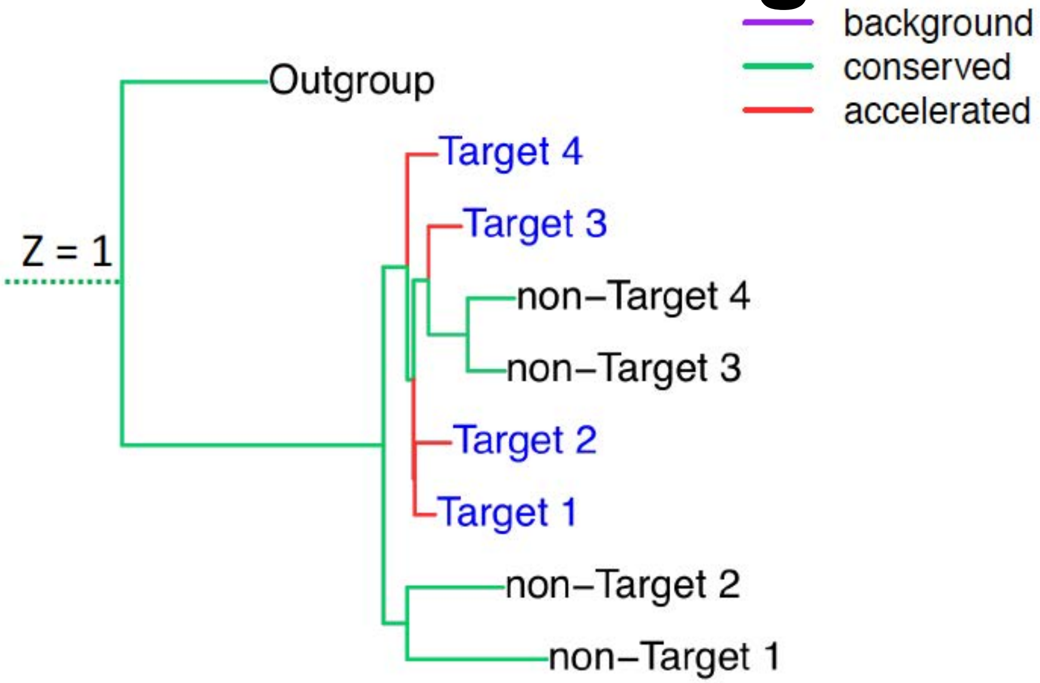
Functionally less important molecules or <u>parts thereof</u> will have a higher substitution rate than functionally important ones

Functionally less important <u>parts of genomes</u> will have a higher substitution rate than functionally important ones

Convergent loss of function of CNEEs in ratite lineages



Branch-specific Bayesian model of noncoding rate accelerations



for noncoding element *i*

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

α = probability of gain of conserved state

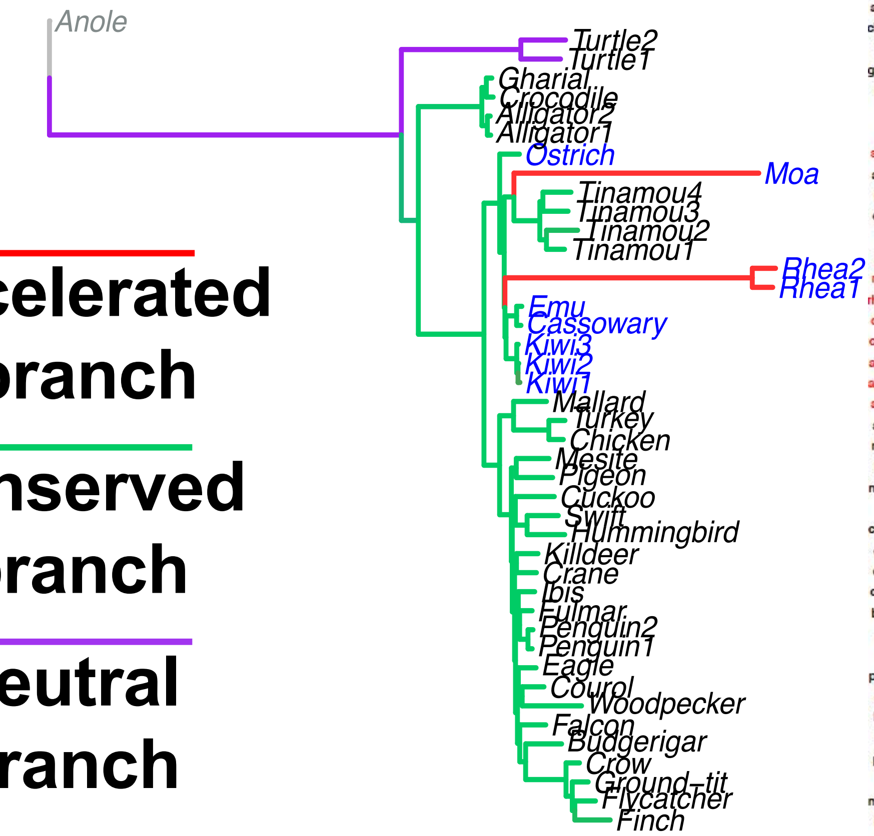
β = probability of loss of conserved state

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

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

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

A convergently accelerated CNEE detected with a novel Bayesian method

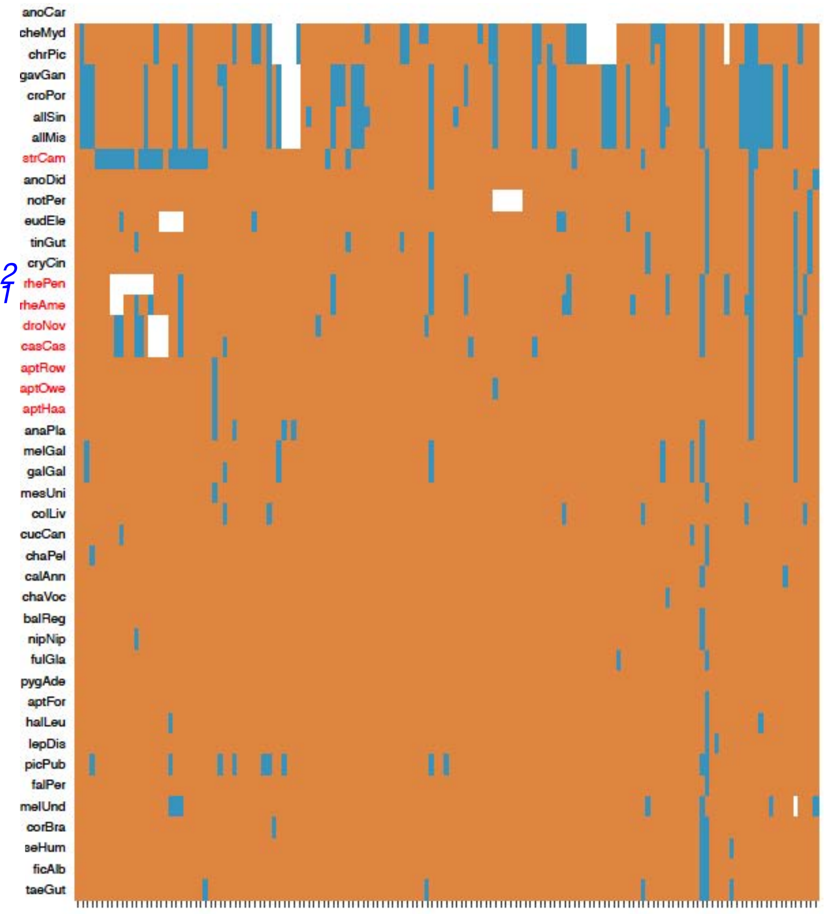


accelerated branch

conserved branch

neutral branch

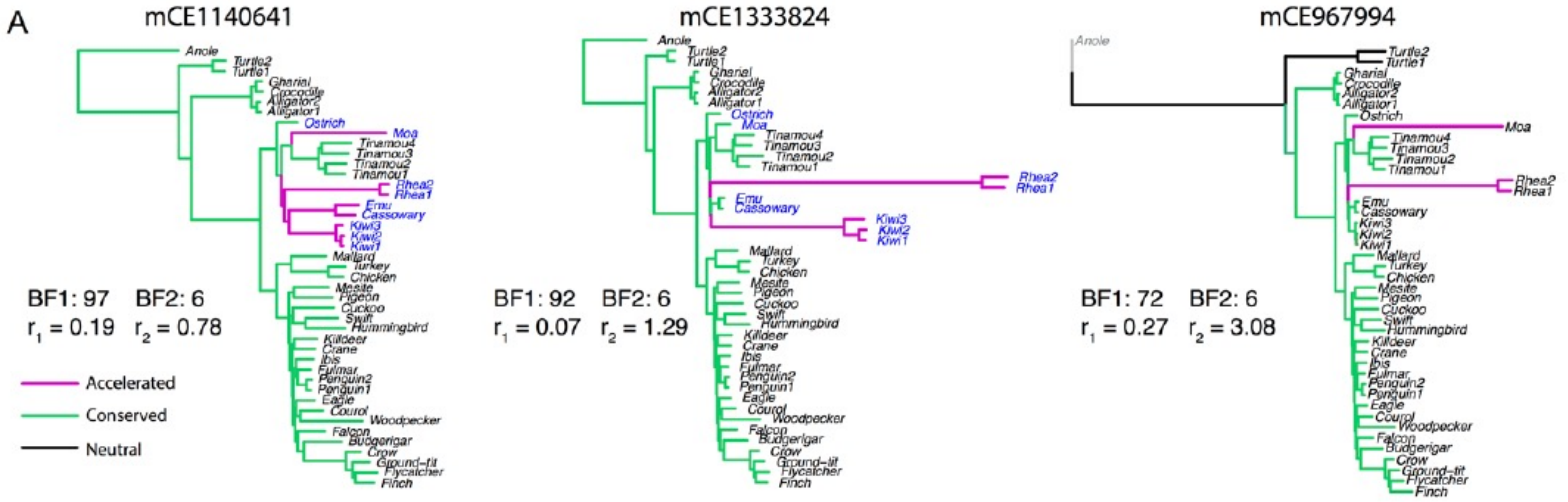
Branch lengths relative to conserved rate



SNPs 🟦

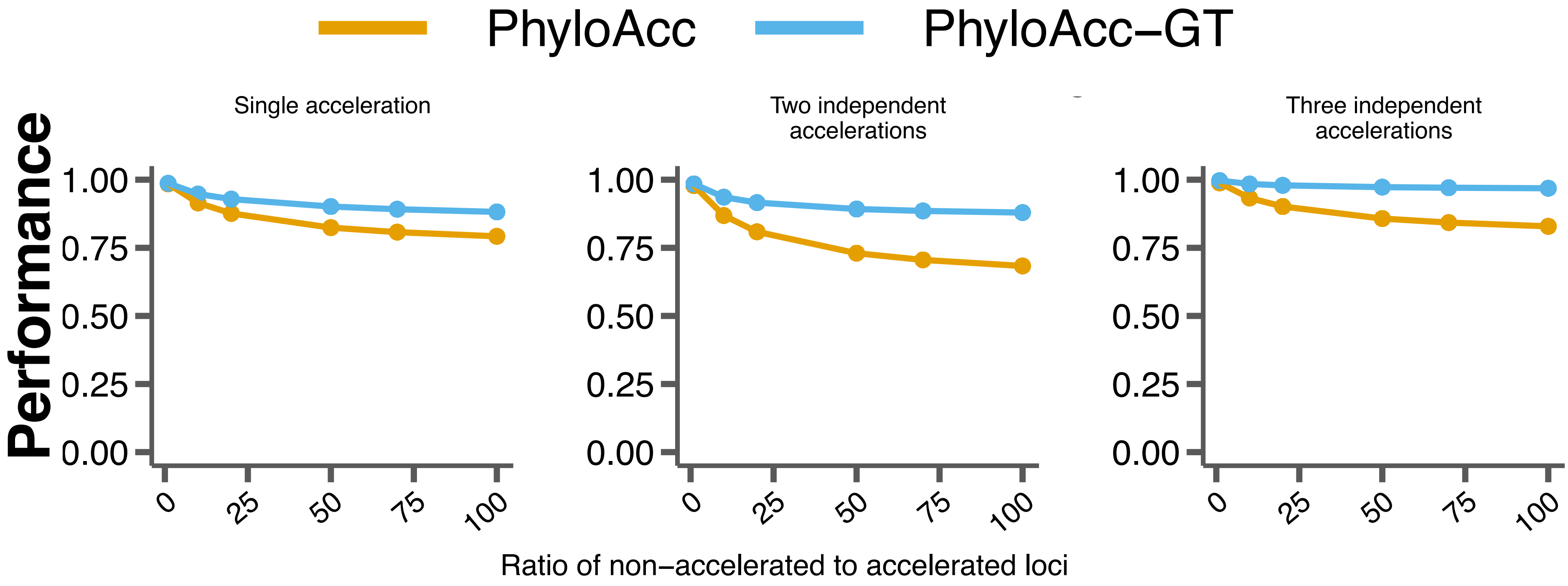
Hu, Z., et al. 2019. *Mol. Biol. Evol.* 36: 1086

Additional examples of convergently accelerated CNEEs



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

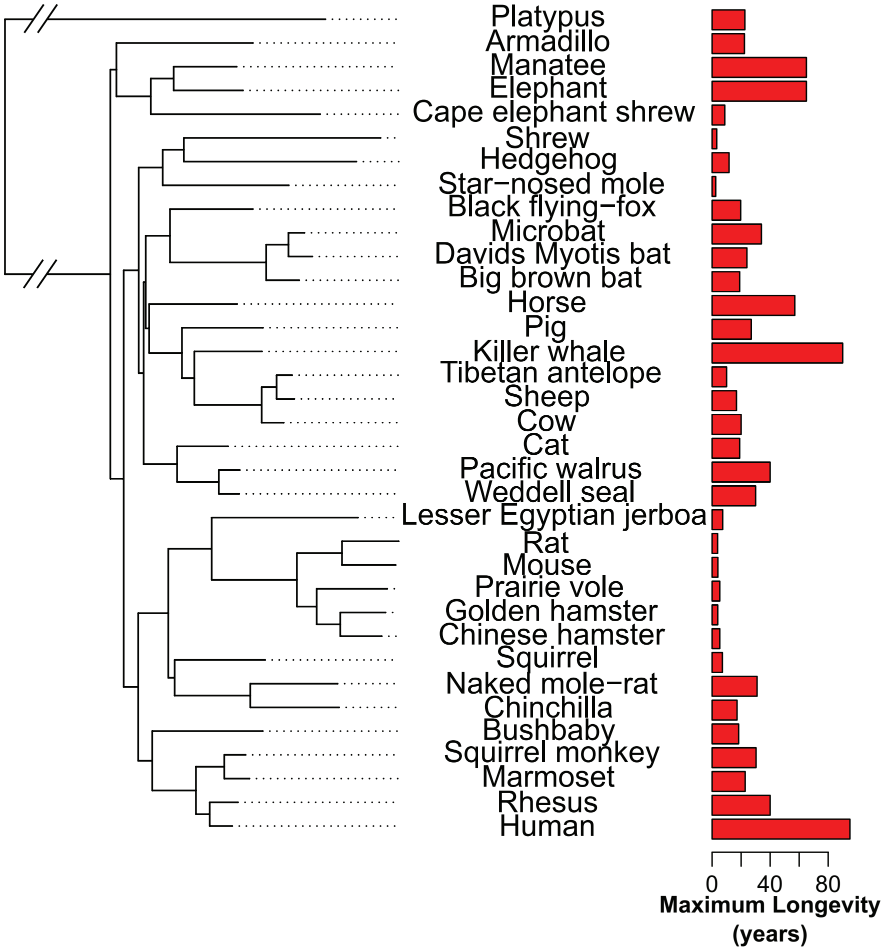
PhyloAcc-GT: Detecting accelerations with gene tree variation

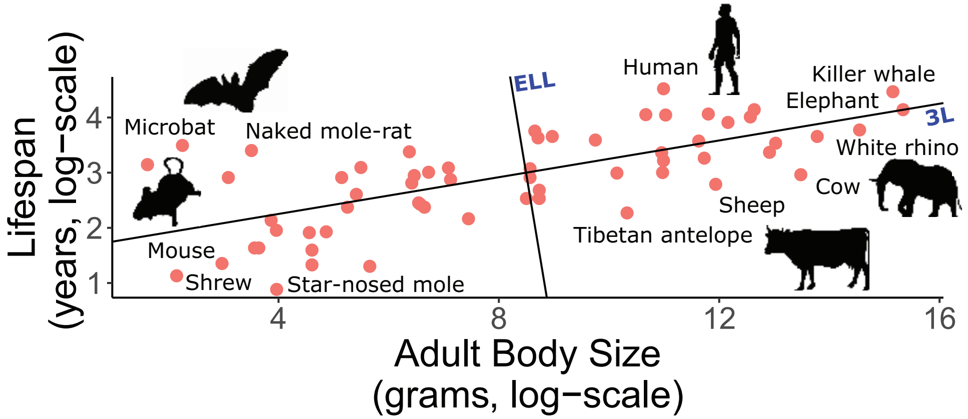


https://www.biorxiv.org/content/10.1101/2022.12.23.521765v1

PhyloAcc-C: Models connecting molecular rates and continuous traits

longevity in mammals



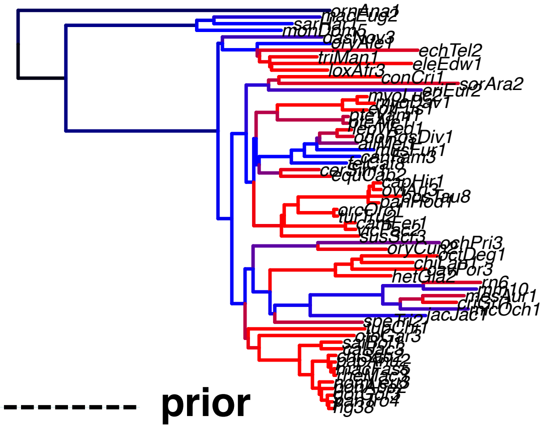


https://github.com/phyloacc/PhyloAcc-C

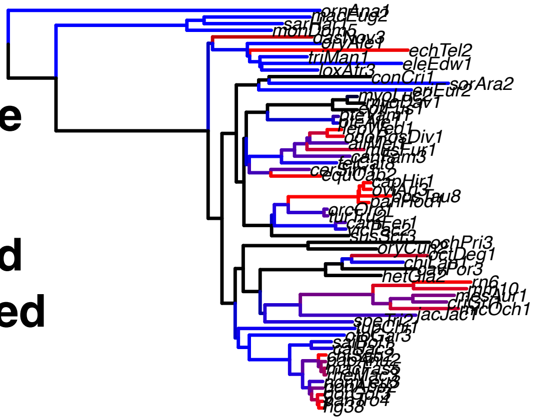
Kowalczyk et al. 2020. *eLife*.

Testing two noncoding elements for rate associations with longevity in mammals

posterior Z
VCE138247.fna



posterior Z
VCE160635.fna

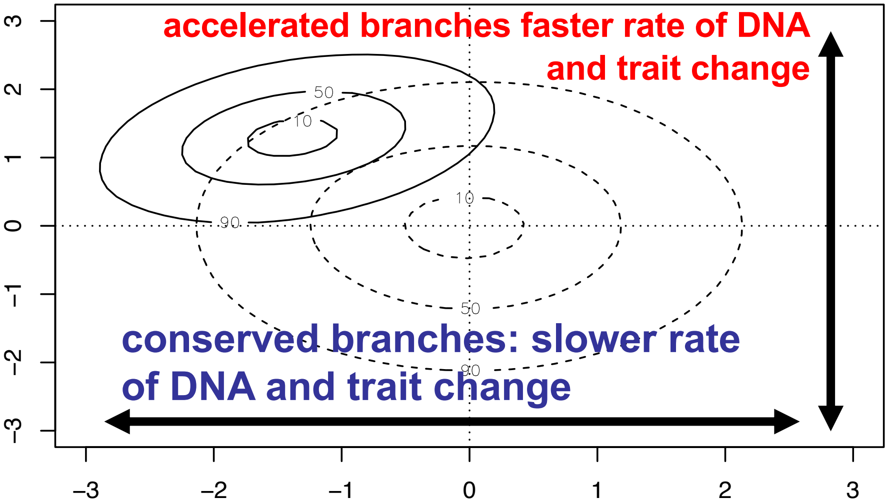


Conservation state

— neutral
— conserved
— accelerated

- - - - prior
— posterior

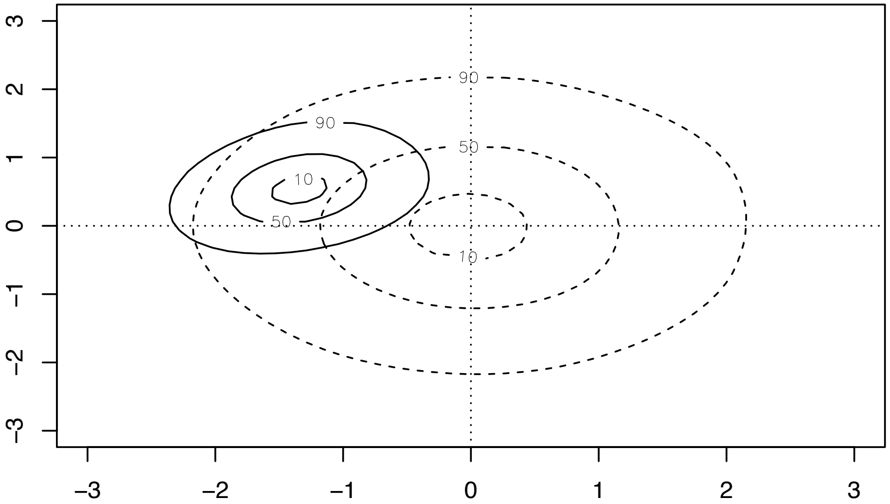
posterior beta
VCE138247.fna



accelerated branches faster rate of DNA and trait change

conserved branches: slower rate of DNA and trait change

posterior beta
VCE160635.fna

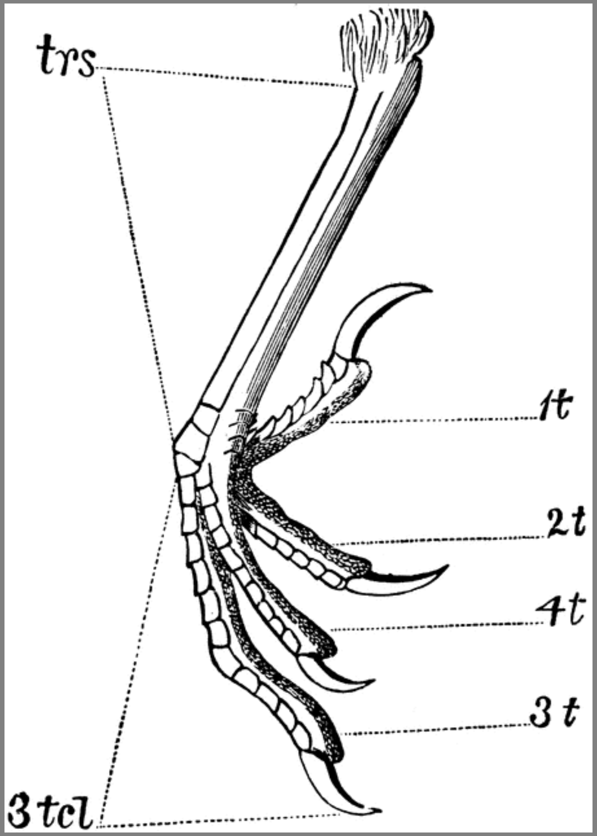


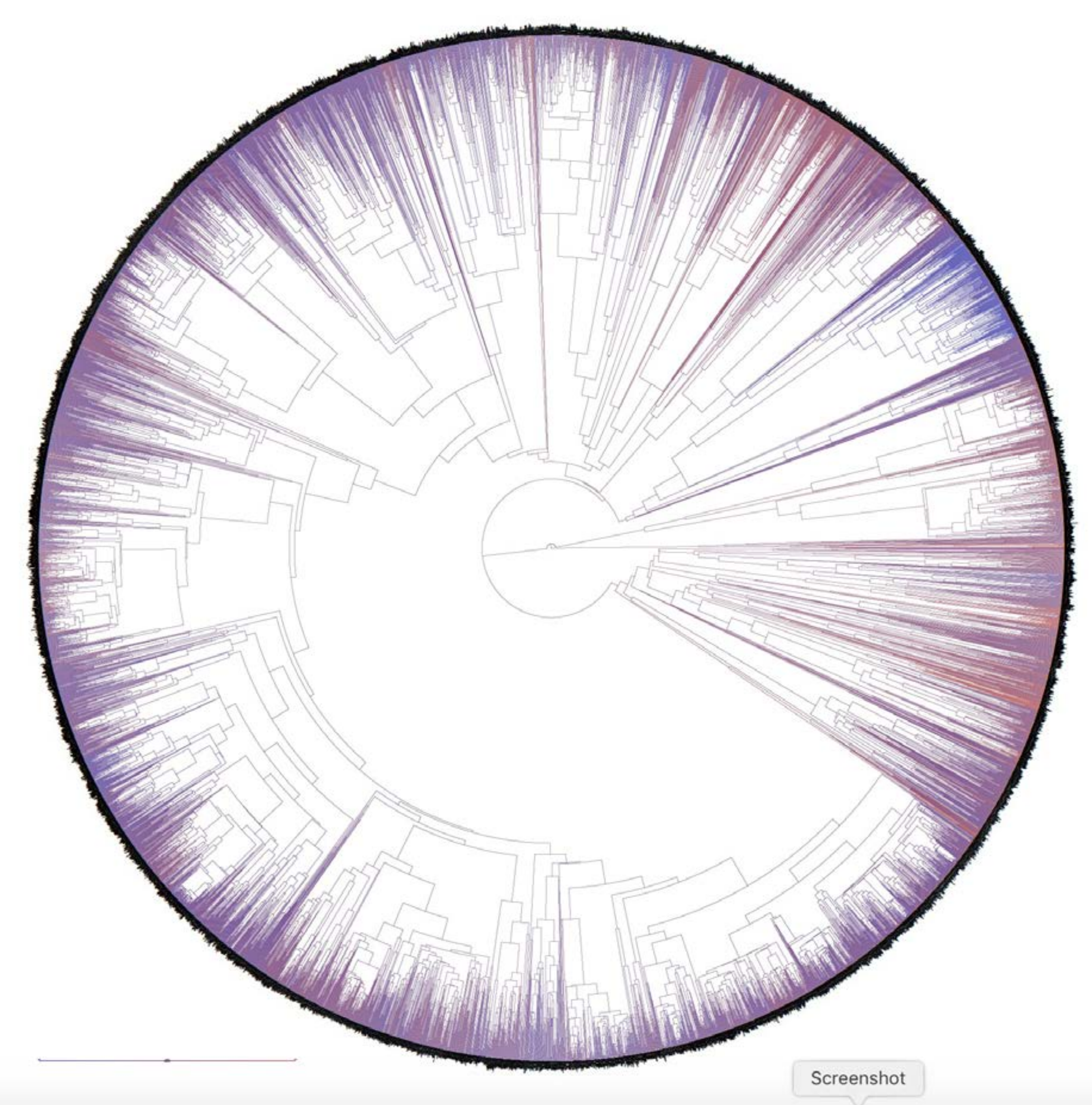
log β_2: multiplier for variance of trait change on accelerated branches

log β_3: multiplier for variance of trait change on conserved branches

log β_3: multiplier for variance of trait change on conserved branches

Tarsus length available for all 10,800 species of birds





short **long**

Elliot Coues *Key to North American Birds* 1884

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

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

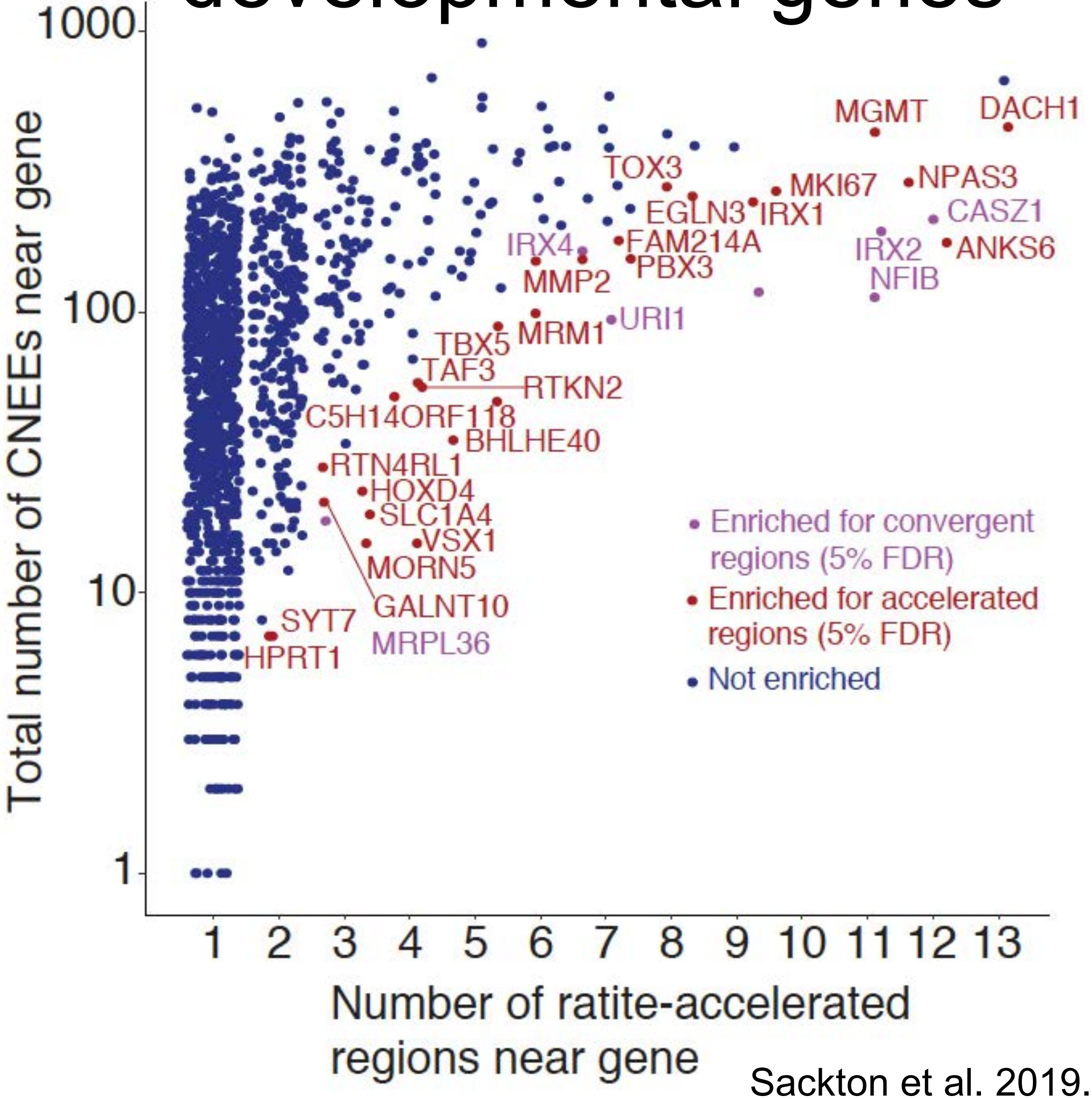
sternum
Tundra Swan
Cygnus columbianus– MCZ 343048

humerus
Northern Screamer
Chauna chavaria– MCZ 340307

tibiotarsus
Brant Goose
Branta bernicla– MCZ 336993

femur
Velvet Scoter
Melanitta fusca– MCZ 348703

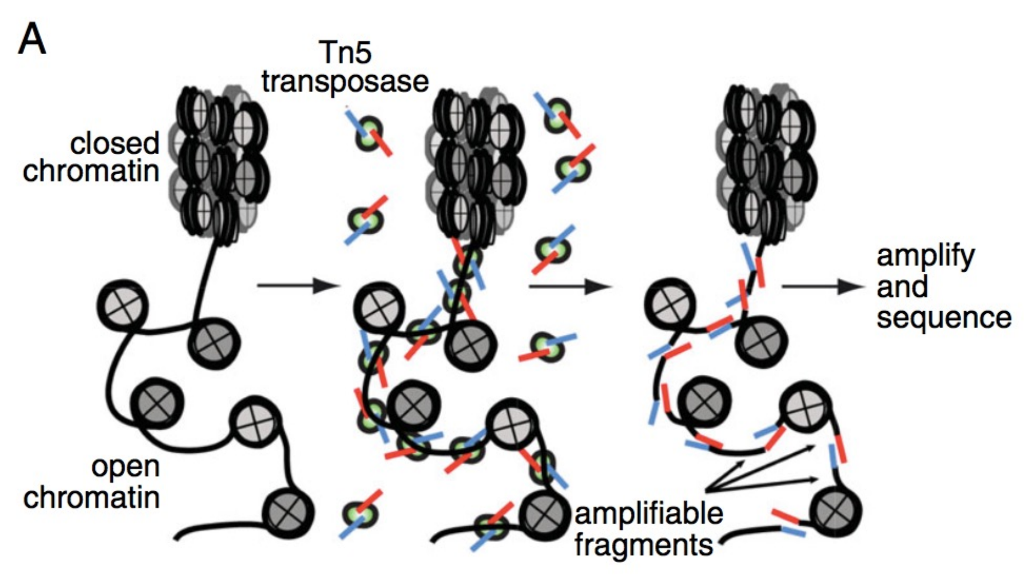
Rapid regulatory evolution near developmental genes



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

Assay for Transposase-Accessible Chromatin

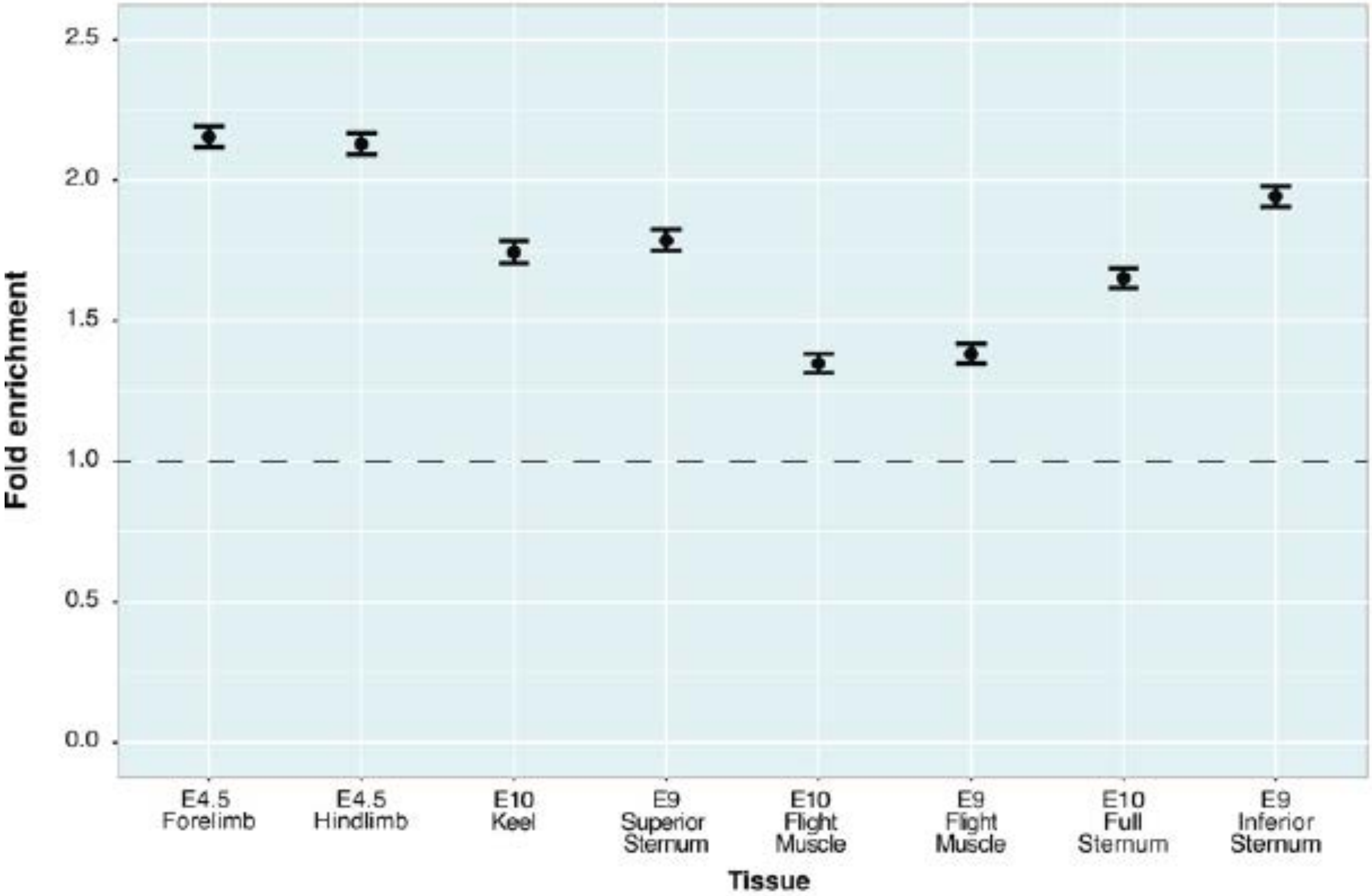
ATAC-Seq identifies DNA with open chromatin, accessible to transcription factors



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

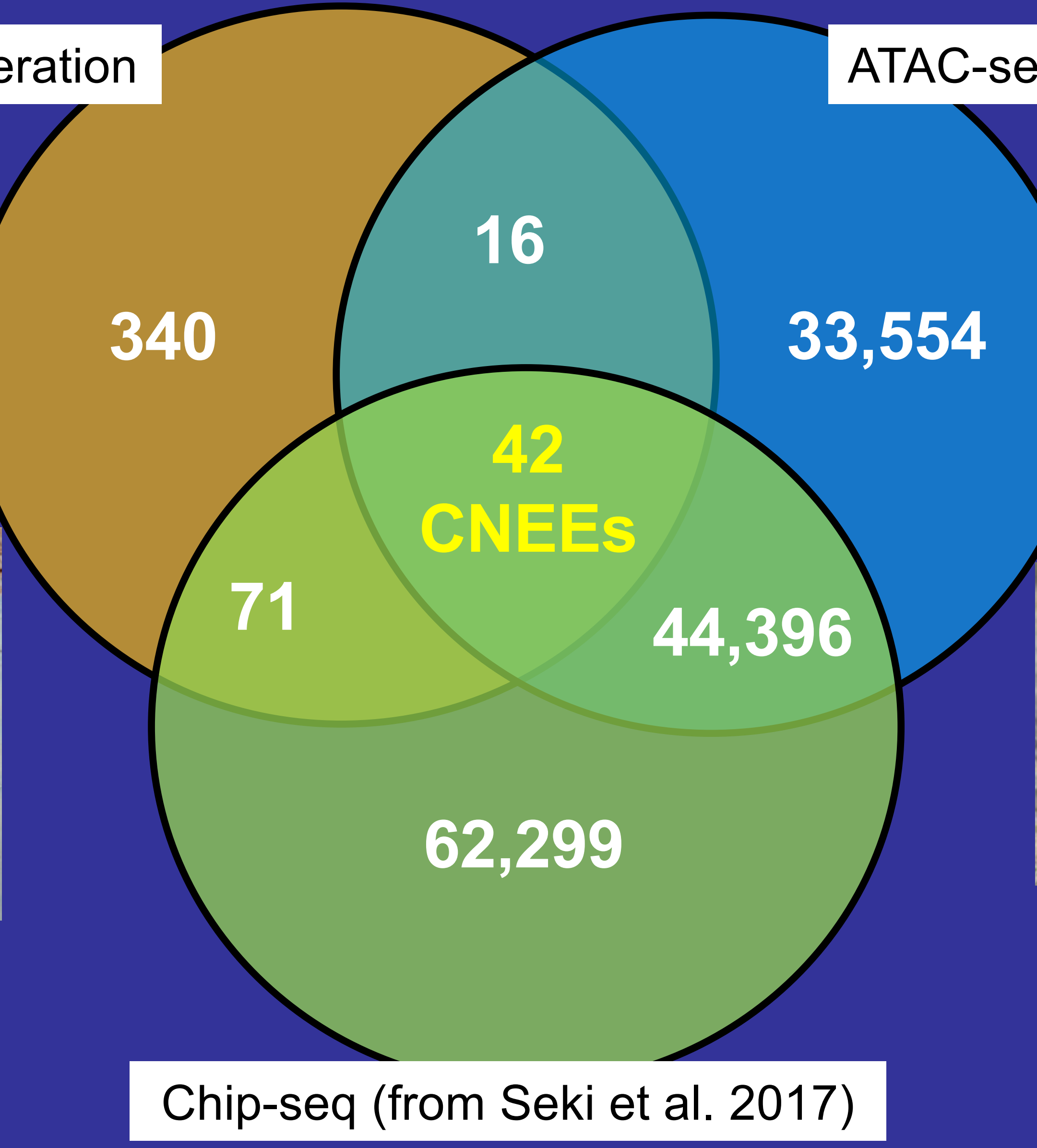
Stage HH24-25 chickens and rheas

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



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

Combined information from multiple sources suggests candidate enhancers for flightlessness phenotypes



Rate acceleration

ATAC-seq

Chip-seq (from Seki et al. 2017)

340

16

33,554

42 CNEEs

71

44,396

62,299