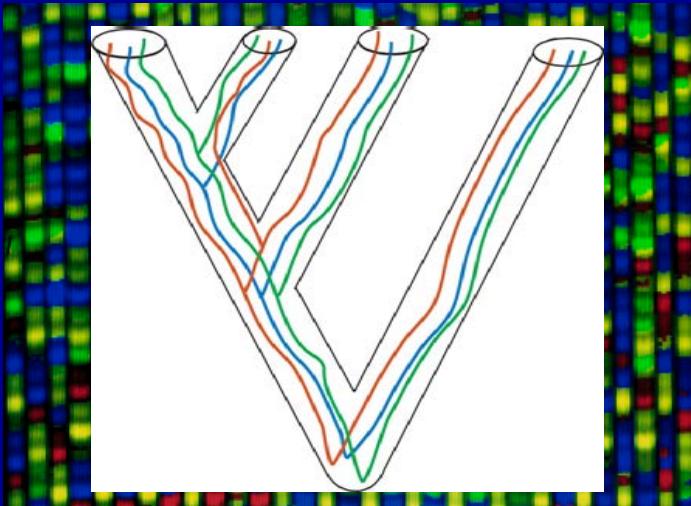
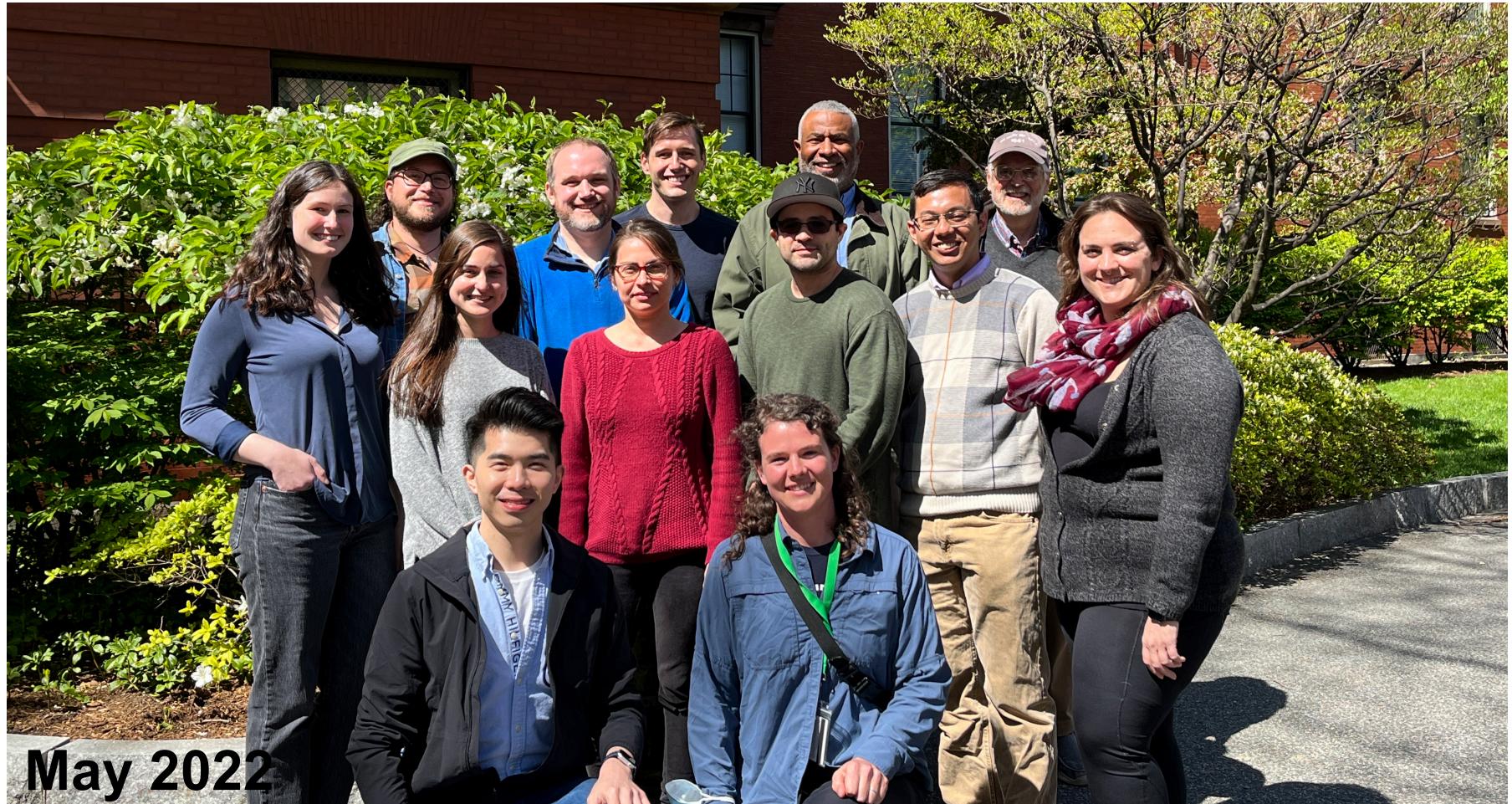
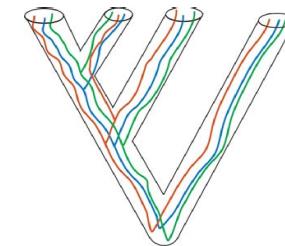
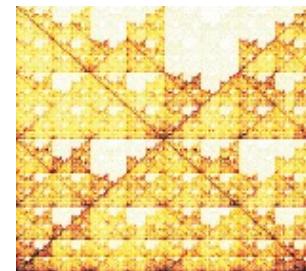
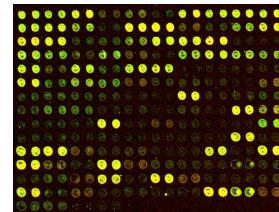
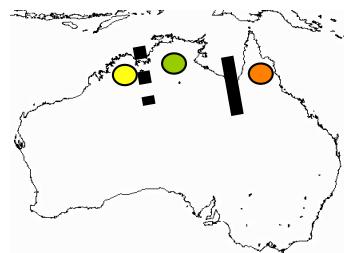


The continuum between phylogeography and phylogenetics



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!

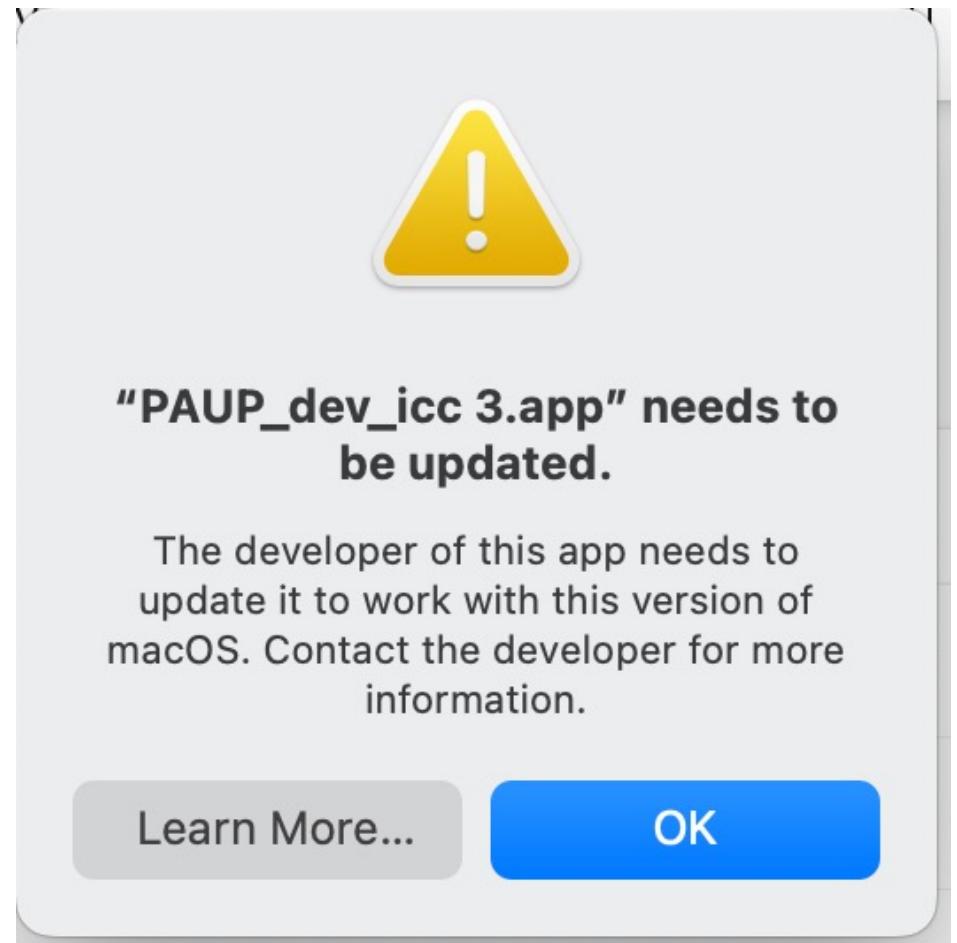
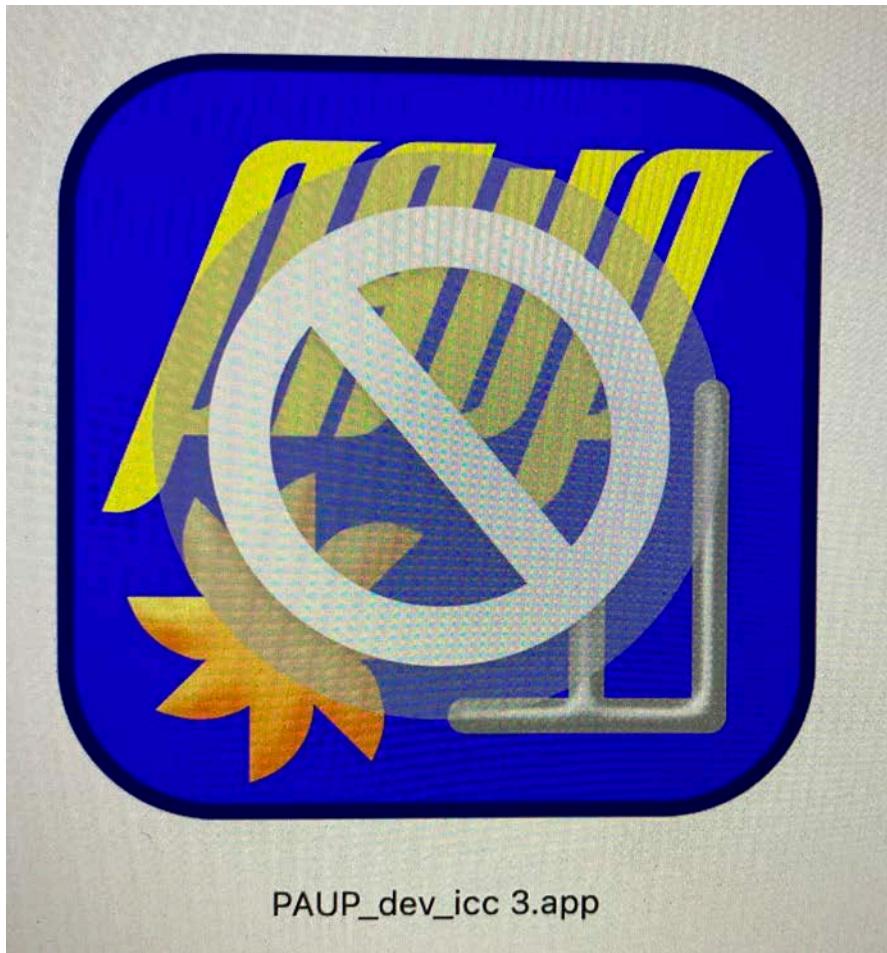


@ScottVEdwards1



**Adventure
Cycling
Association**

Oh no! End of the Paup* GUI era?

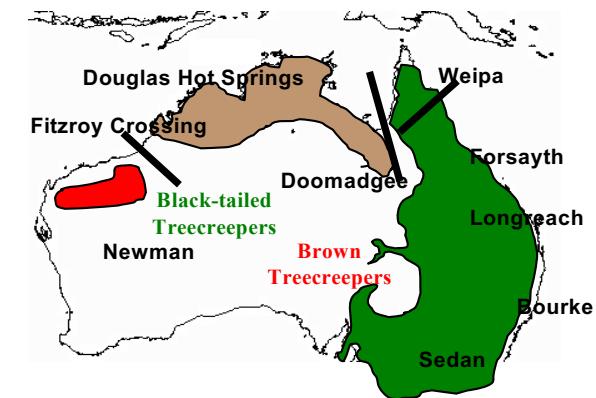


A 1980s conundrum in evolutionary biology

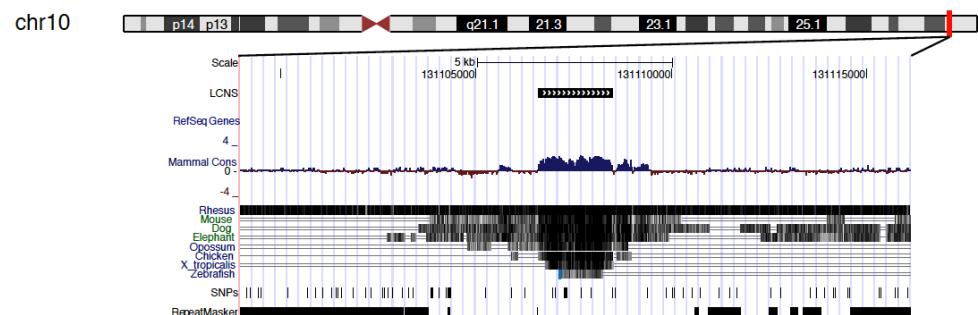


Talk overview

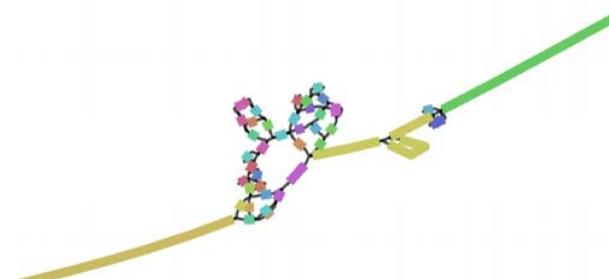
Part I: Reticulation and the emerging continuum between phylogeography and phylogenetics



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



Honeyeaters



Paleognaths



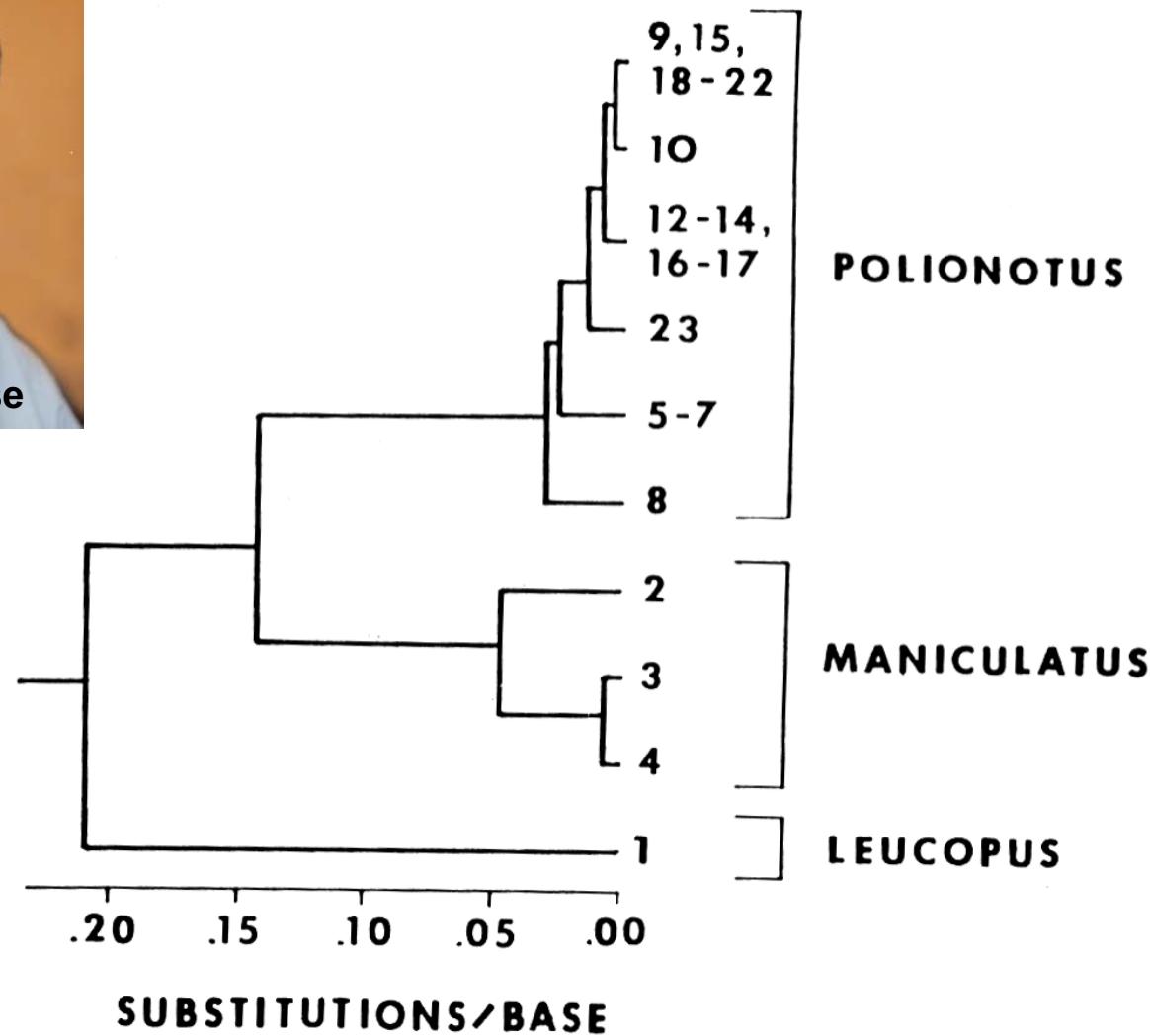
Scrub Jays

PhyloG2P

The first ‘gene tree’, 1979



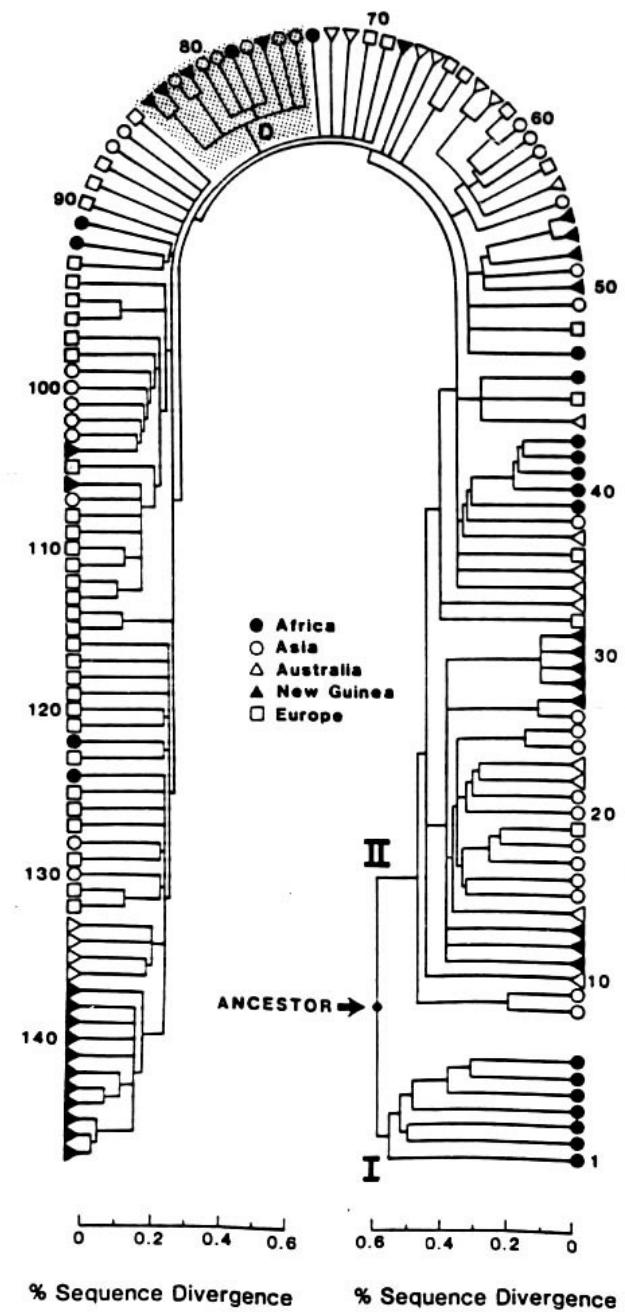
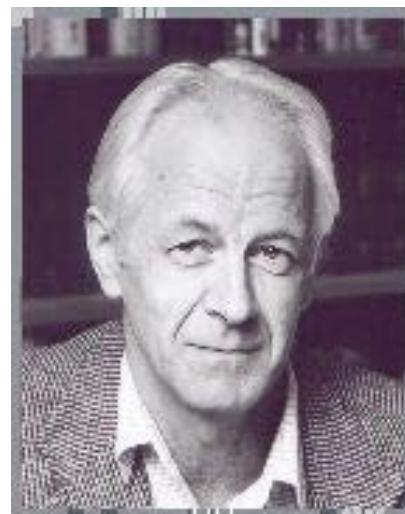
J. C. AVISE, R. A. LANSMAN AND R. O. SHADE



Mitochondrial DNA and human evolution

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

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



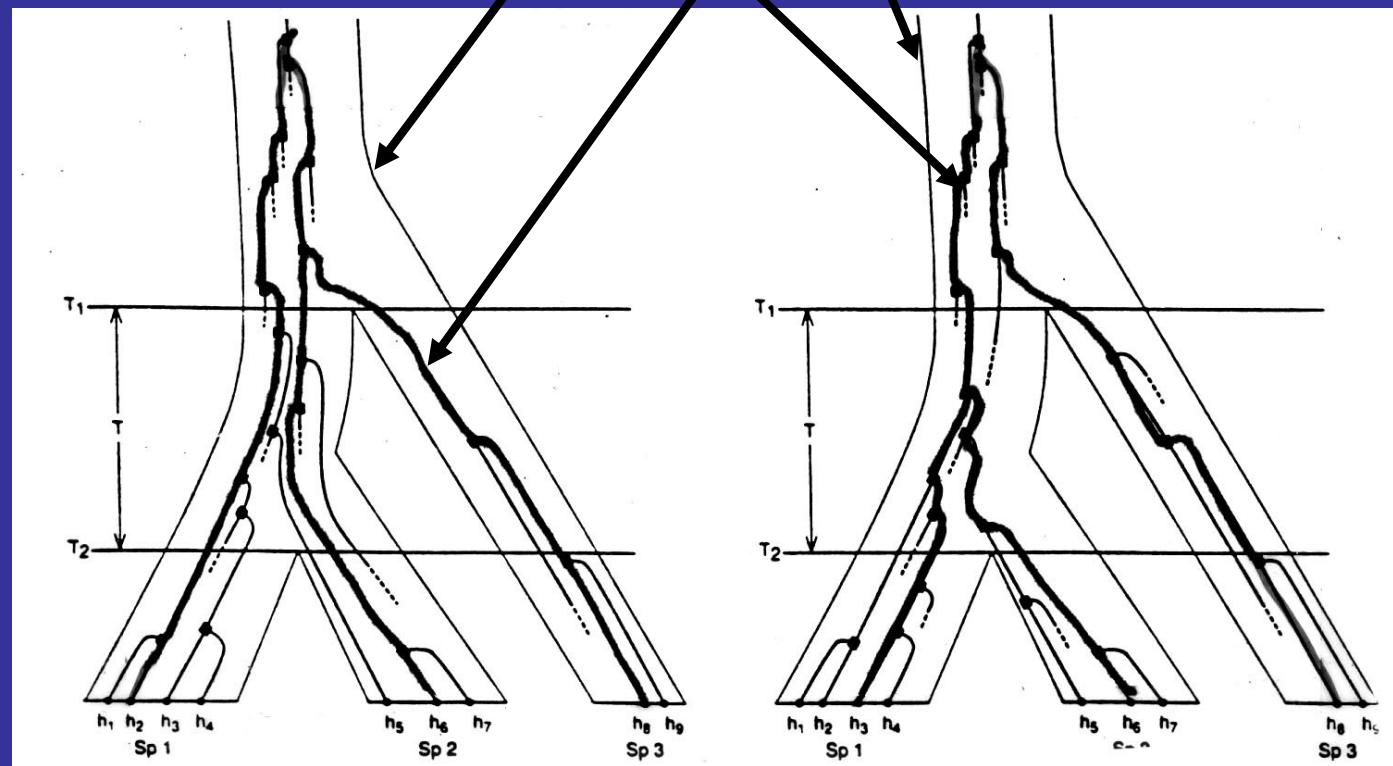
Stochastic gene tree variation during rapid radiations

“incomplete lineage
sorting”

Species tree

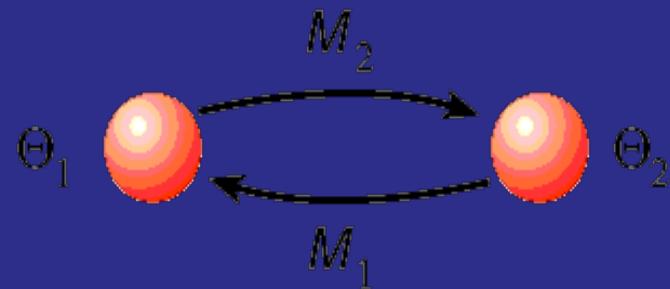
Gene tree

“deep coalescence”

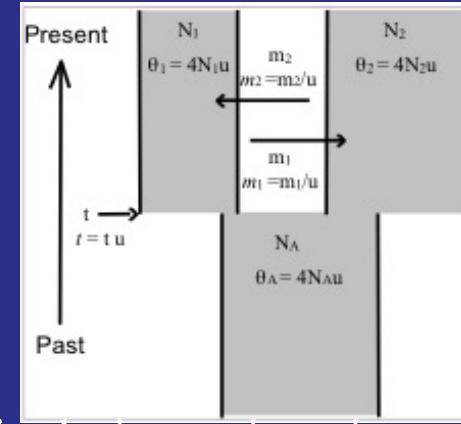


Multilocus models in phylogeography

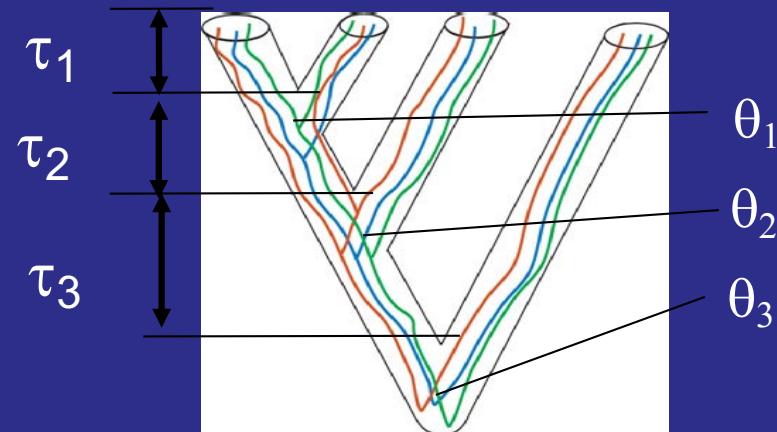
Population size: $\theta = 4N\mu$ Divergence time: $\tau = \mu t$ Gene flow: $M = m/\mu$



MIGRATE: Beerli 2006 *Bioinformatics*



IM: Hey and Nielsen 2004 *Genetics*



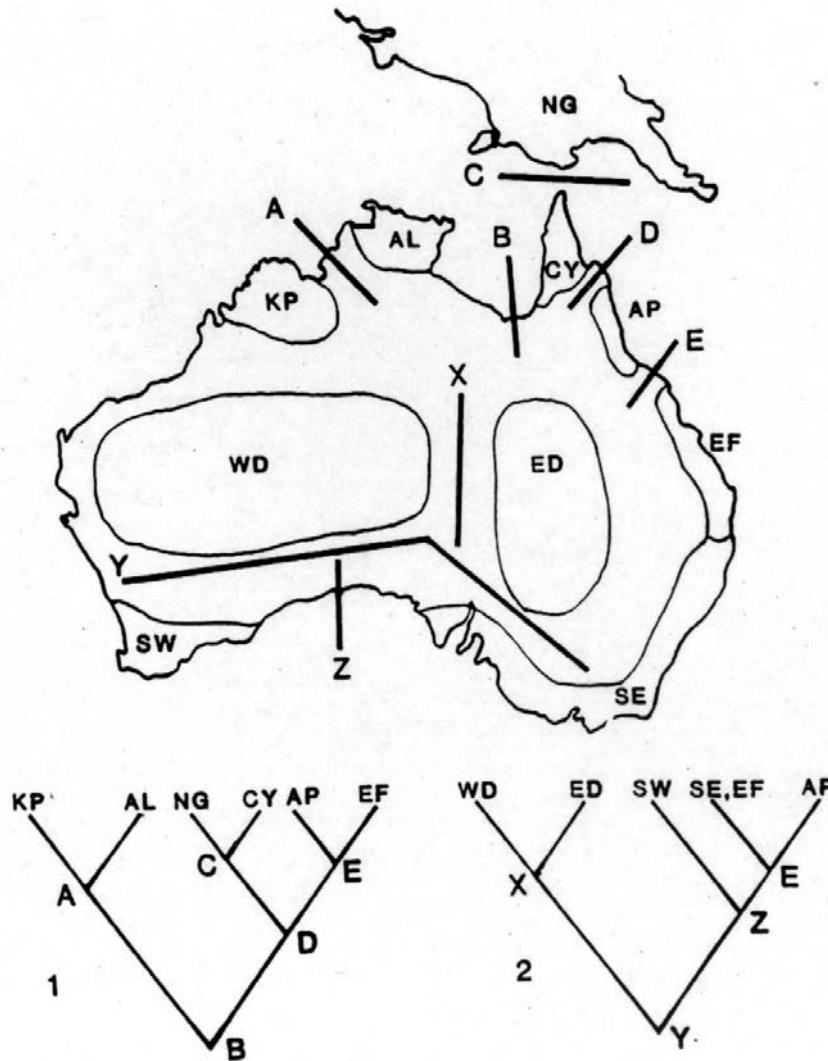
BEST: Liu and Pearl. 2007. *Syst. Biol*

BEAST*: Heled and Drummond. 2010. *Mol. Biol. Evol.*

~~Warm welcome in the outback...~~



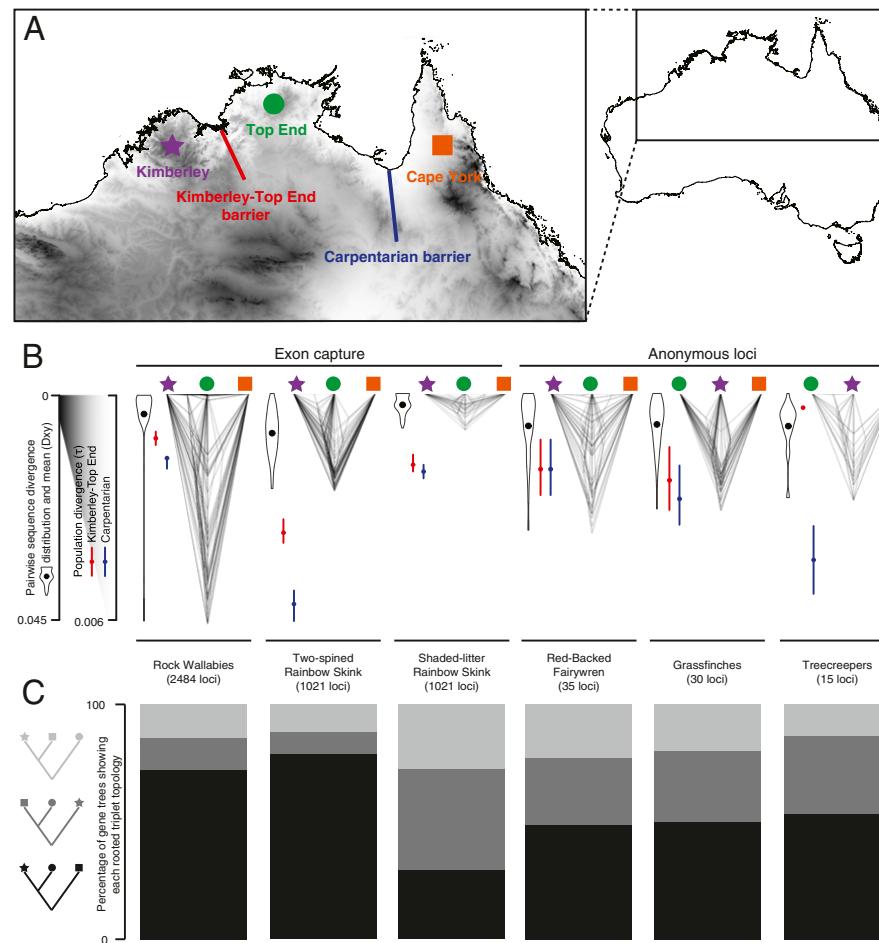
Carpentarian barrier (B) is deepest split in area cladograms of Australian biota



Cracraft 1986. *Evolution*

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}



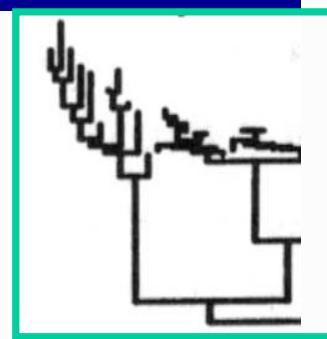
Edwards et al. 2016. *PNAS* 113: 8025–8032

Mitochondrial genetic differences among eco-regional clades

temporalis

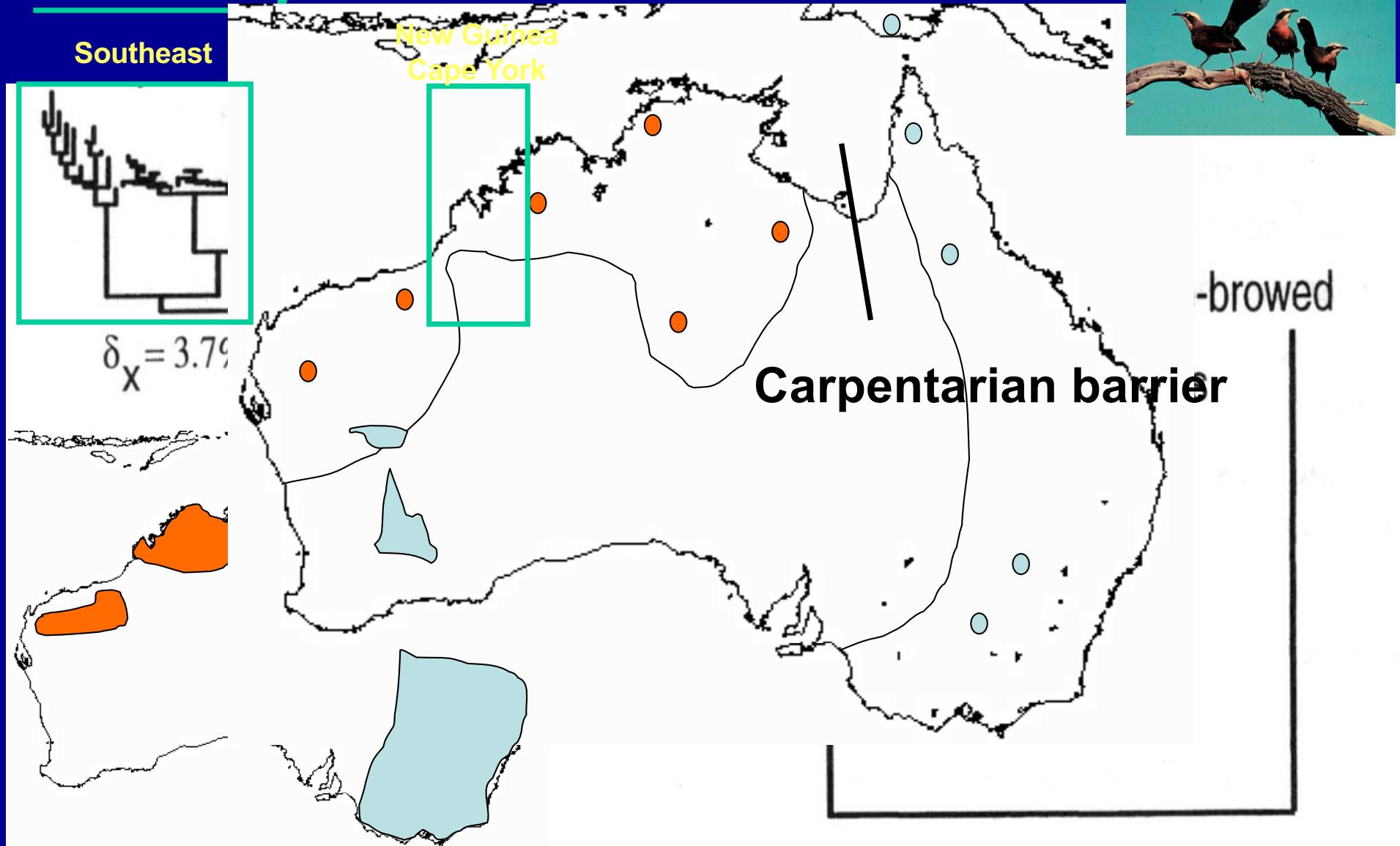
rubeculus

Southeast



$$\delta_X = 3.7\%$$

New Guinea
Cape York

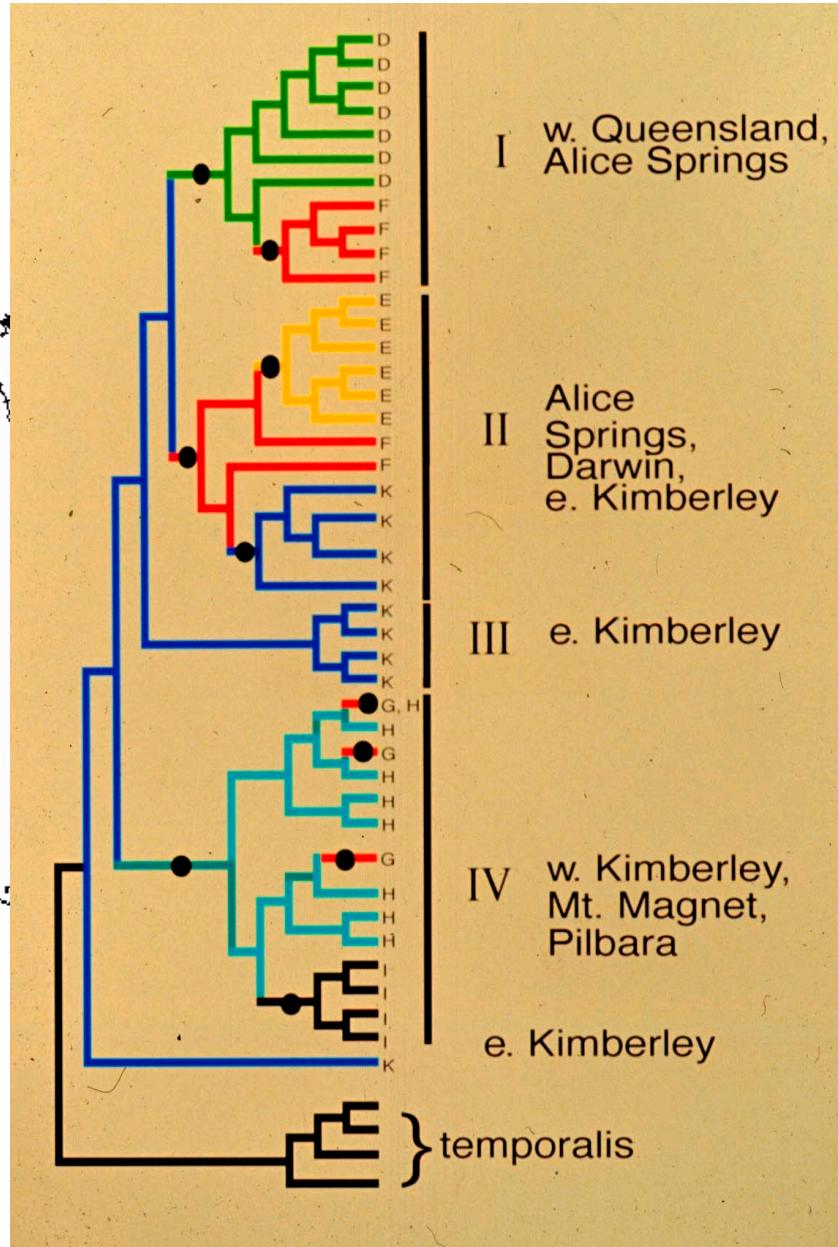
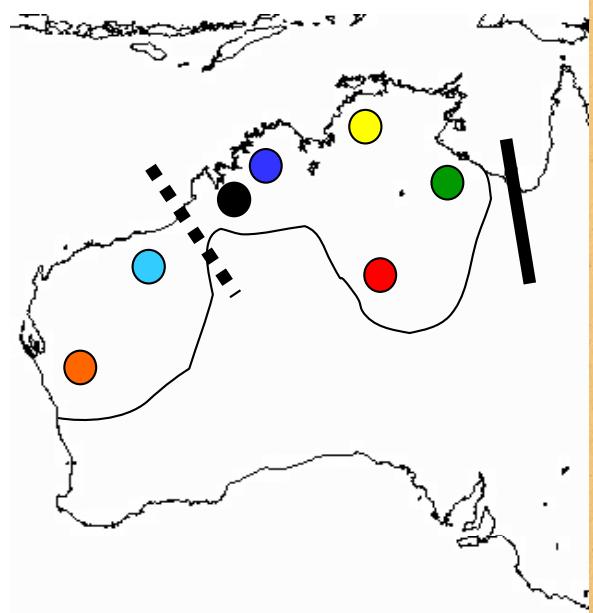


-browed

Carpentarian barrier

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

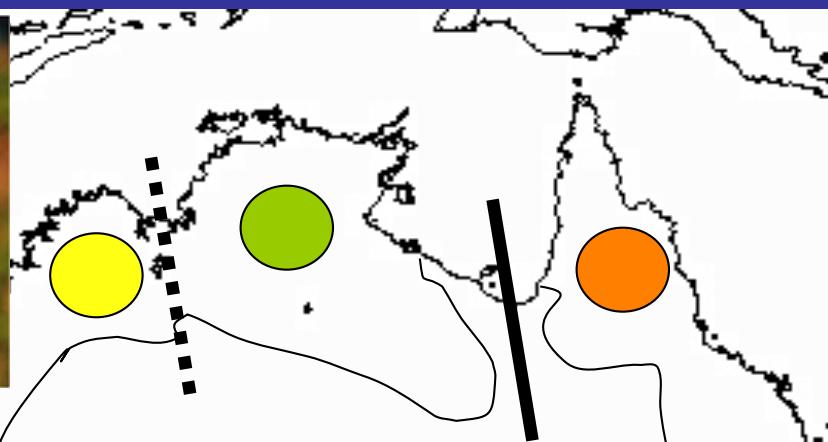
- Parsimony approach
 - Assumes complete certainty of gene tree
 - Statistical testing of different hypotheses of gene flow is cumbersome
 - Confidence limits reflect only uncertainty in s
- Bayesian approach (migrate-n)
 - Integrates the two sources of uncertainty
 - Accommodates uncertainty in trees by integrating over all trees
 - Accommodates the range of Nm given these trees
 - Hypothesis testing easier

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

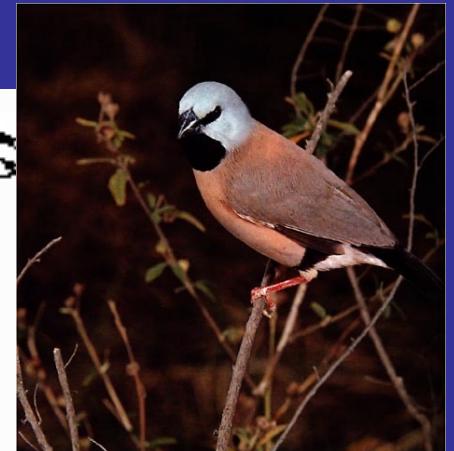
P. acuticauda



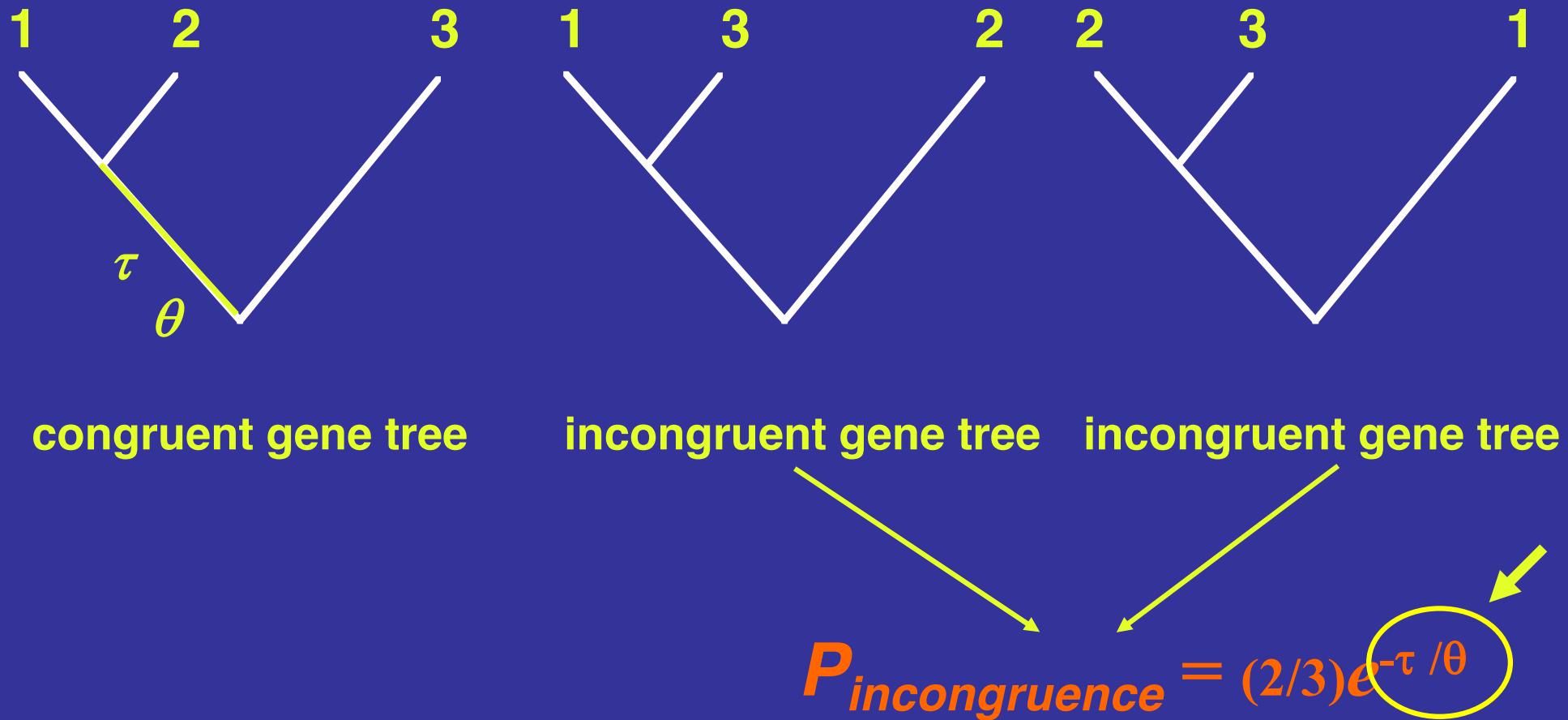
P. hecki



P. cincta

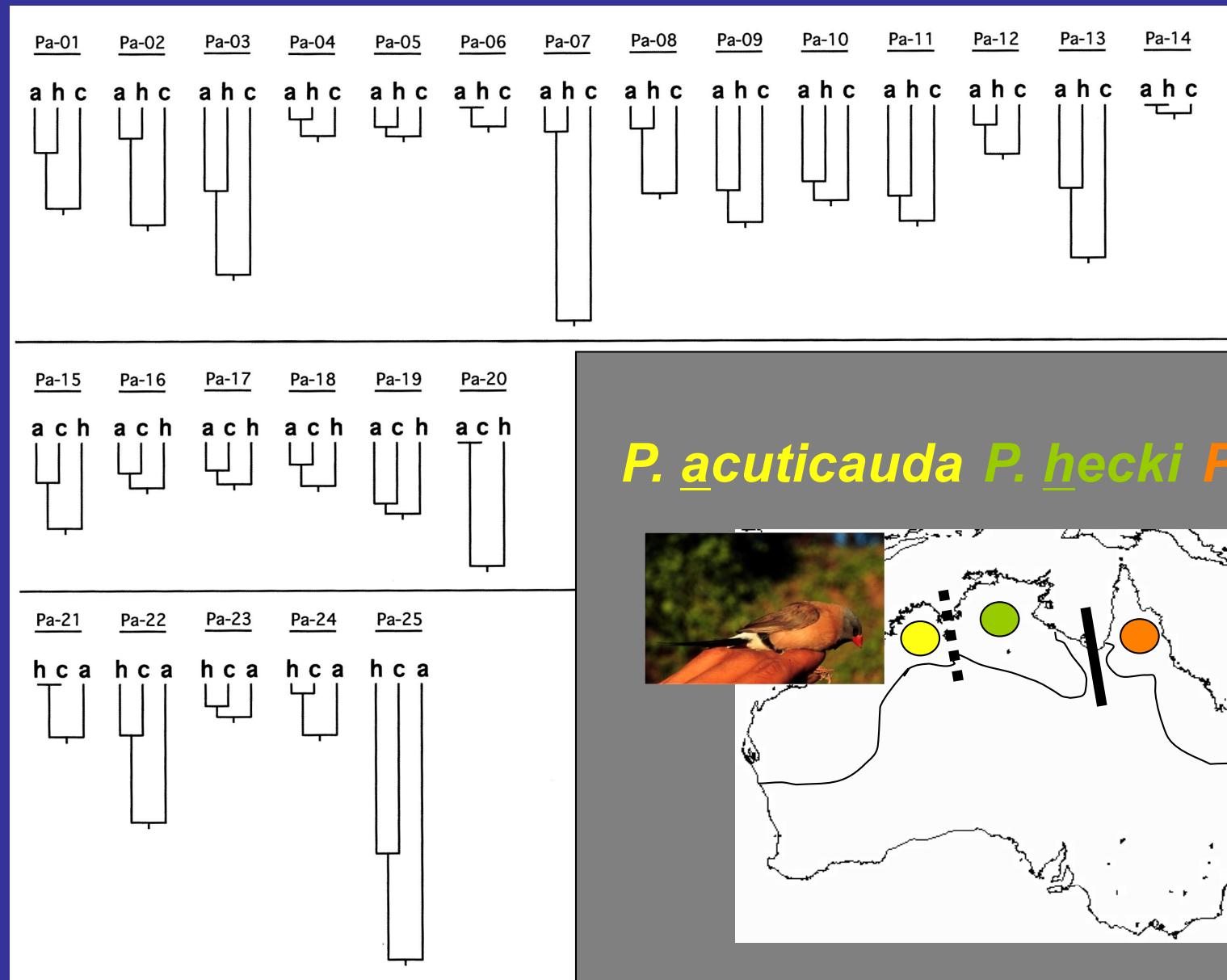


Probability of gene trees in a rooted 3-tip species tree

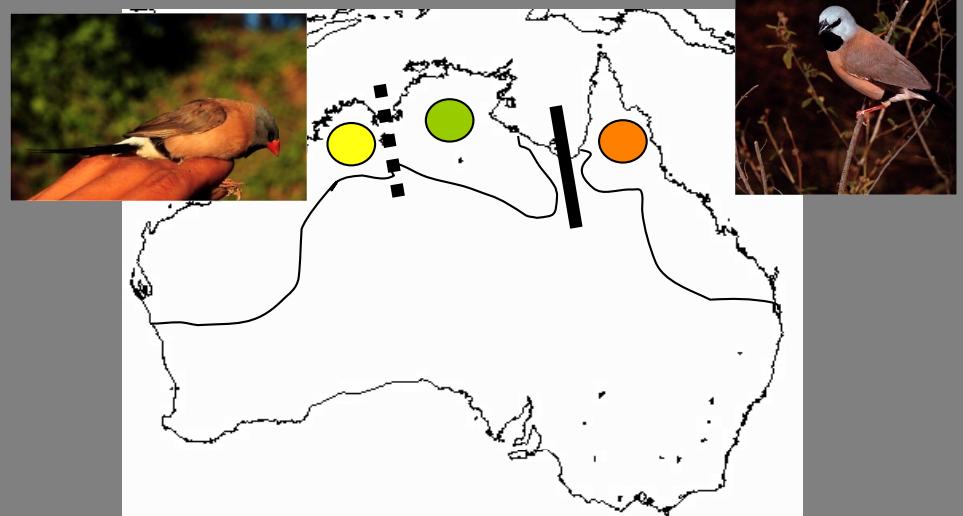


Pamilo and Nei 1988. Mol. Biol. Evol. 5: 568-583
Liu, et al. 2010. BMC Evolutionary Biology 10:302

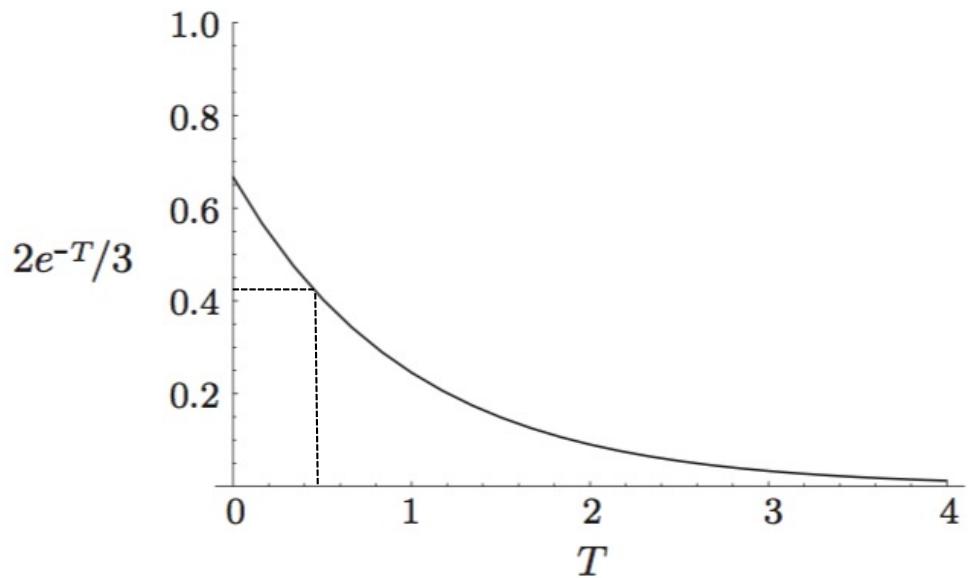
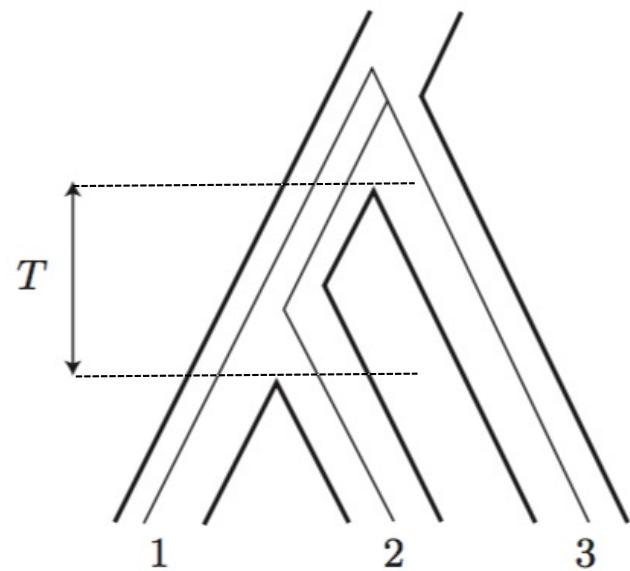
30 gene trees from Australian finches



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



Probability of discordance between gene tree and species tree



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

$$\begin{aligned} P\{\text{discordant}\} &= \frac{2}{3}e^{-T} \\ &= 12/28 \end{aligned}$$

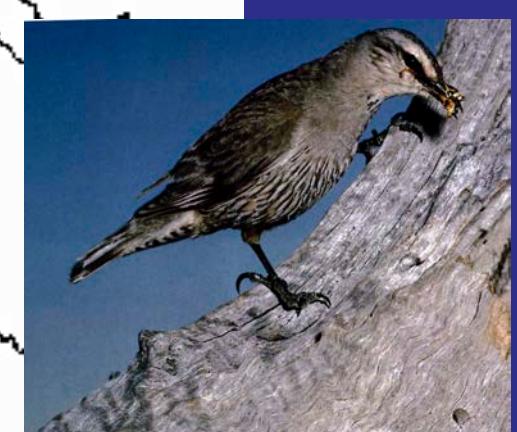
Treecreepers (*Climacteris*)

Black-tailed treecreeper

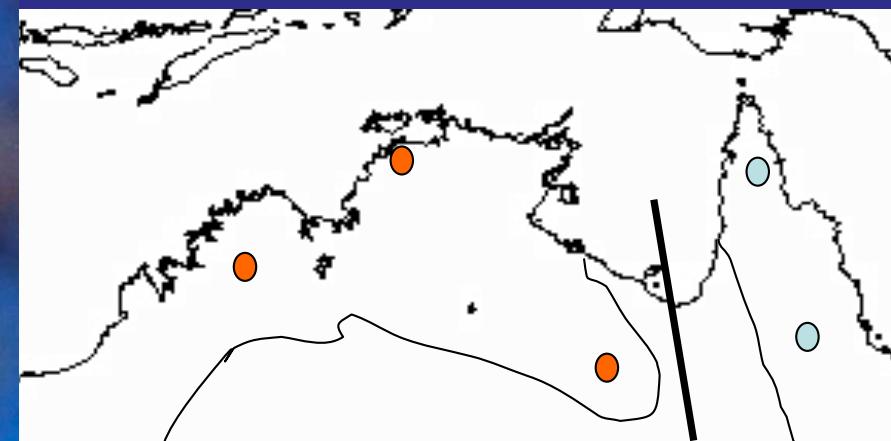


C. melanura

Brown treecreeper



C. picumnus



12 gene trees in Australian treecreepers

Brown Treecreeper
(eastern) lineages

Black-tailed (western)
Treecreeper lineages

AL3

AL5

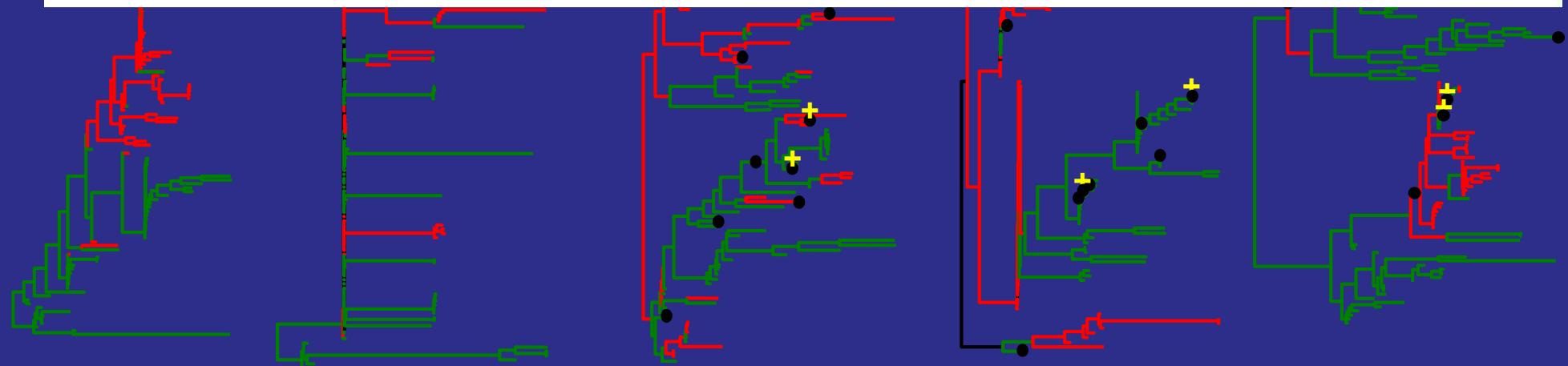
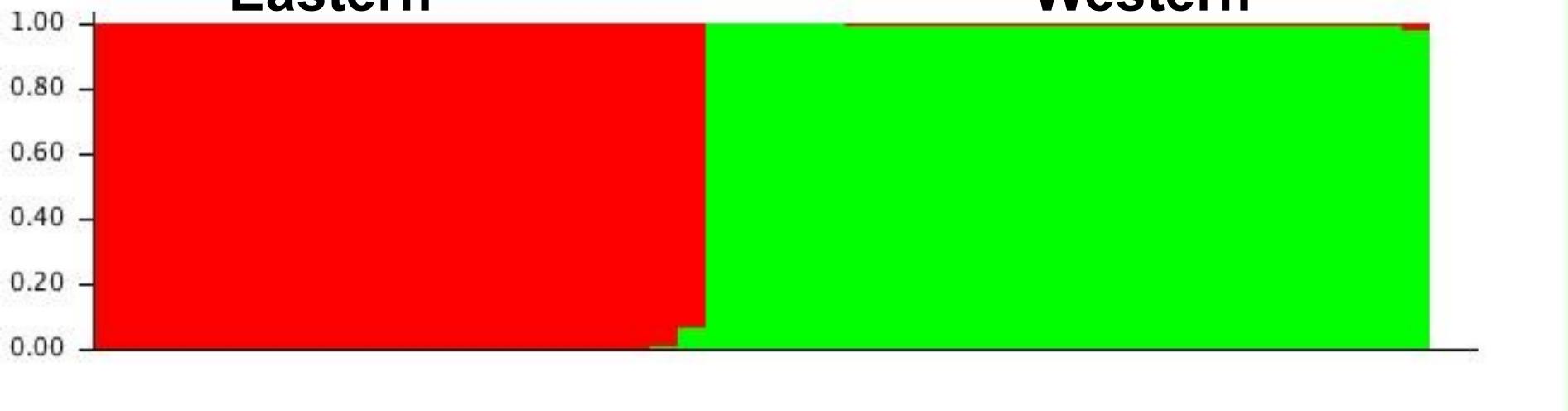
AL7

AL14

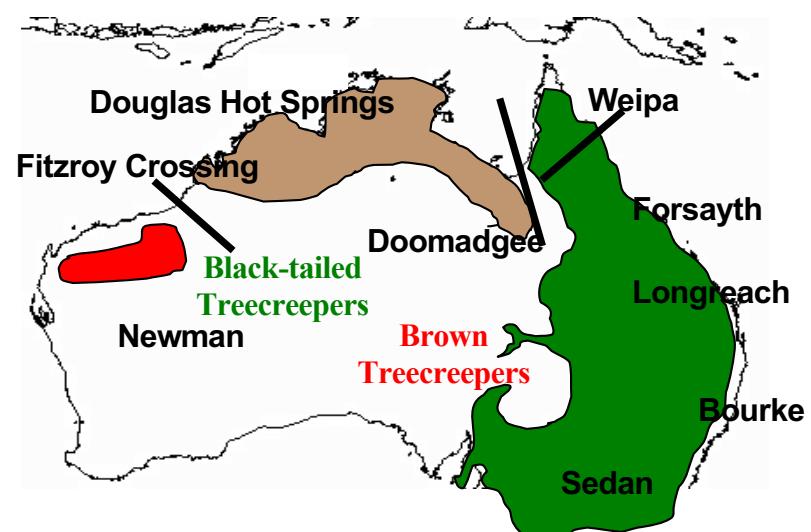
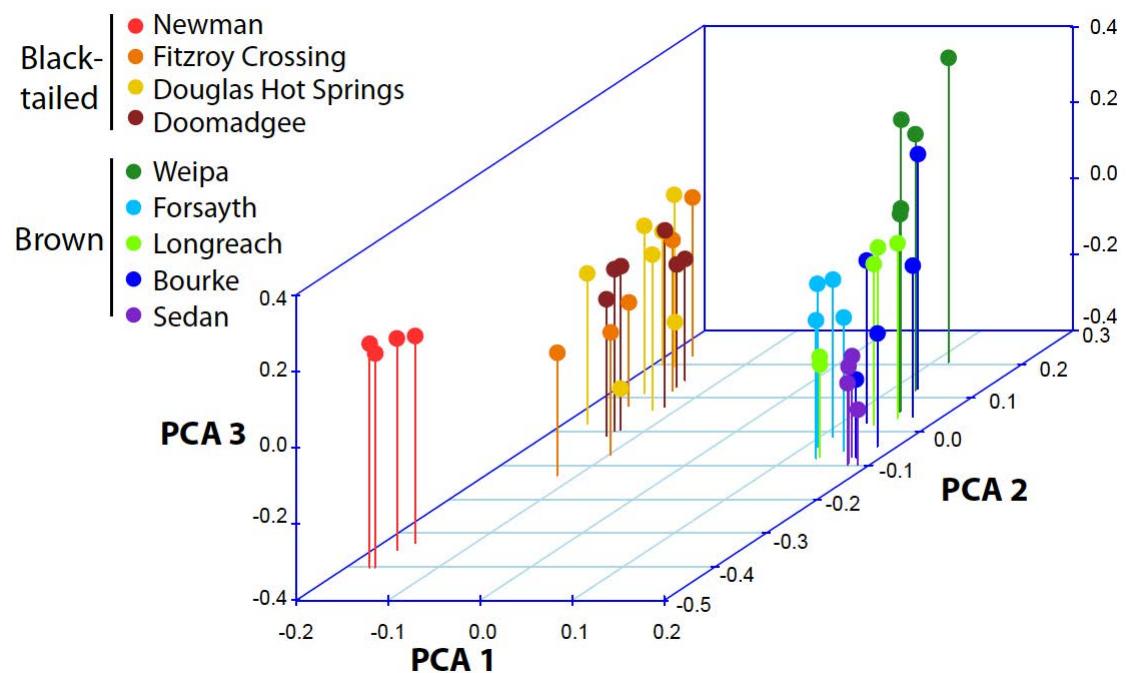
AL16

Eastern

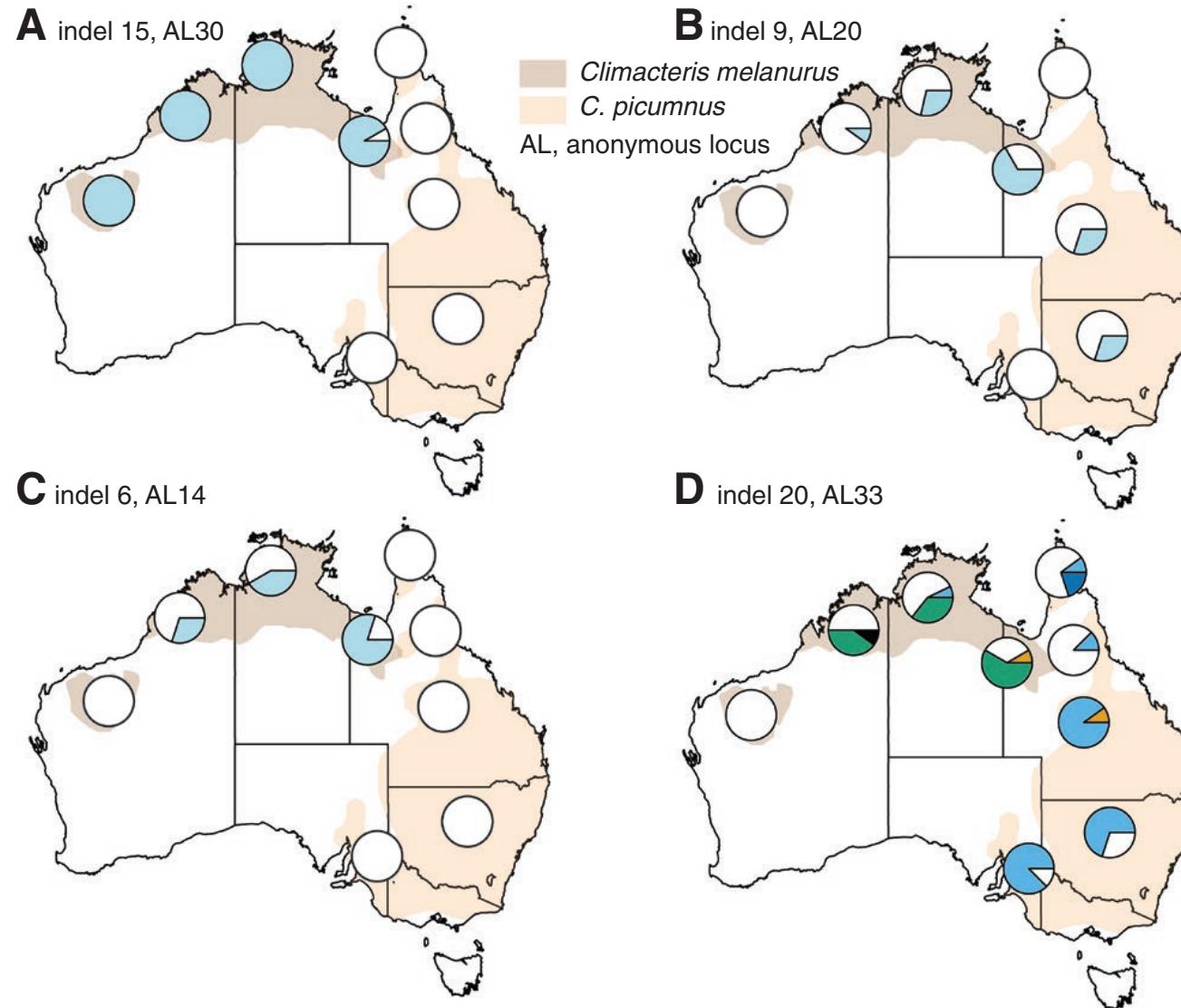
Western



Correspondence between PCA and geography



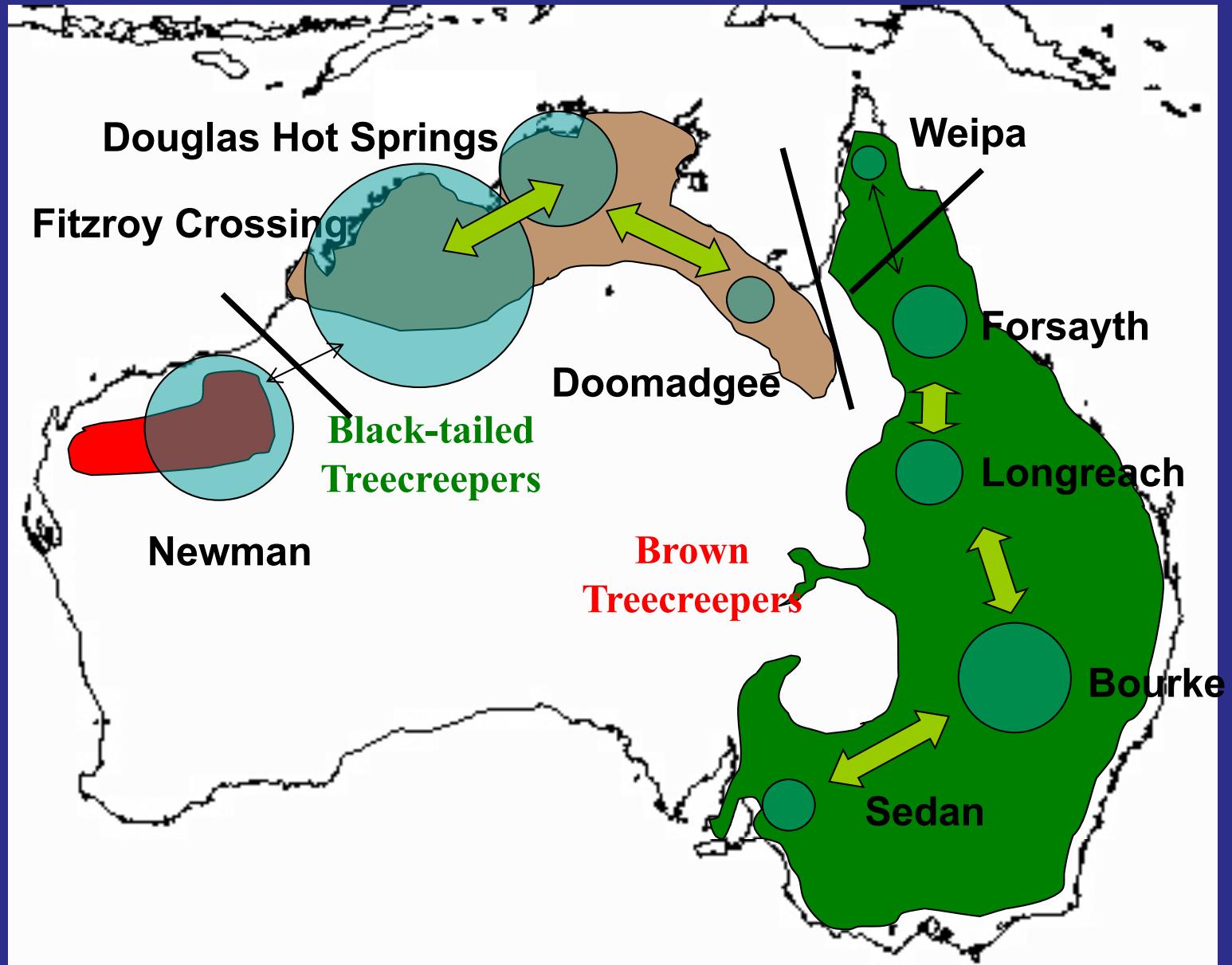
Insertions-deletions provide significant phylogeographic signal



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

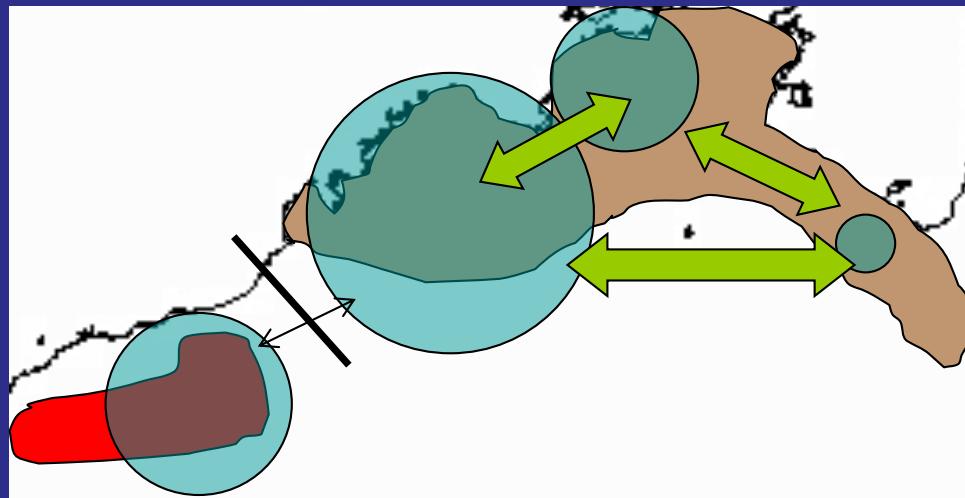
Treecreeper populations are connected but variable in size (MIGRATE)

● 10,000 individuals
— 1-2 migrants per generation



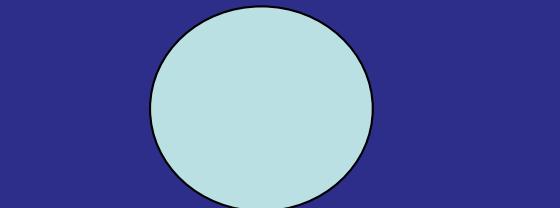
Rotzel, Edwards and Beerli, unpubl. data

Using Bayes Factors to evaluate phylogeographic models

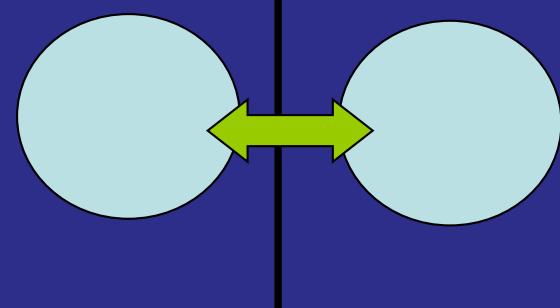


$$2(mL_1 - mL_2)$$

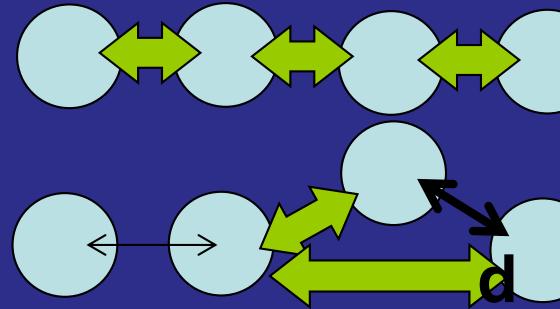
-98.60



-83.14



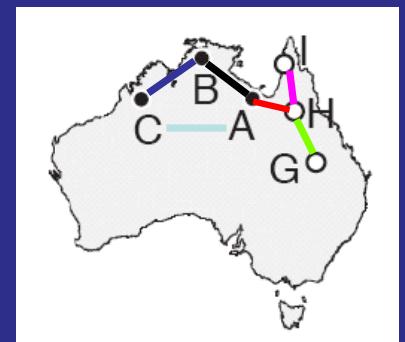
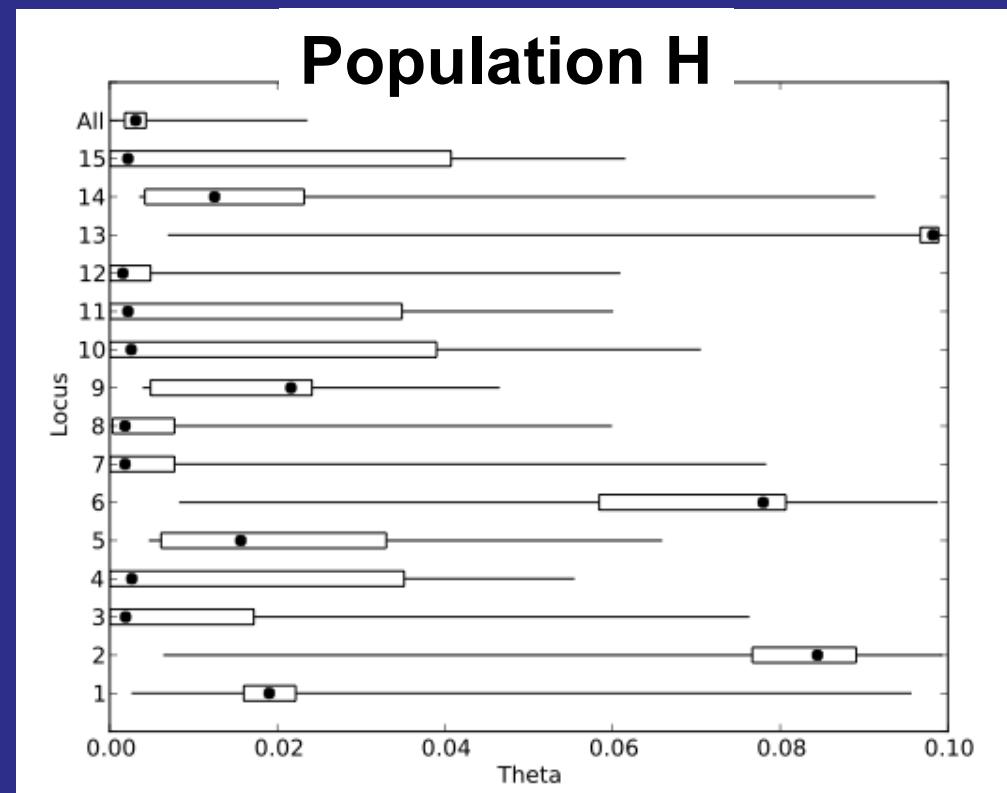
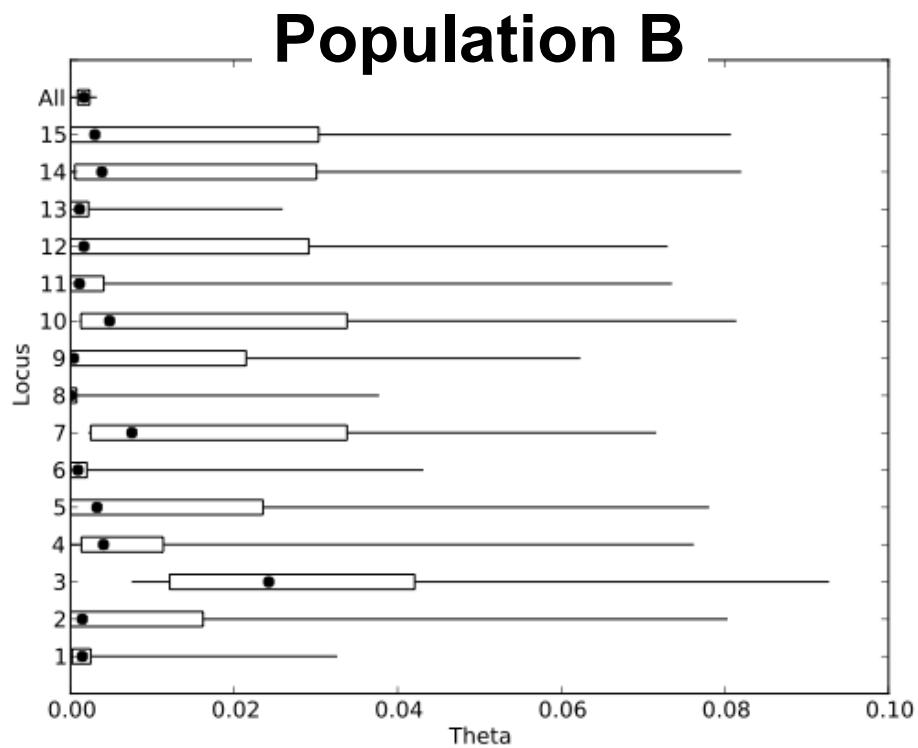
-54.98



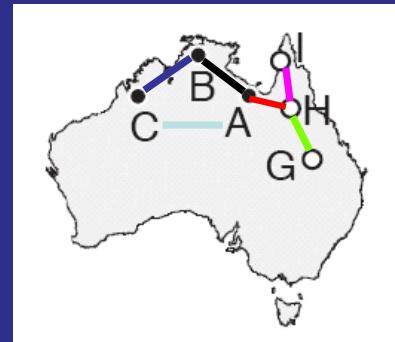
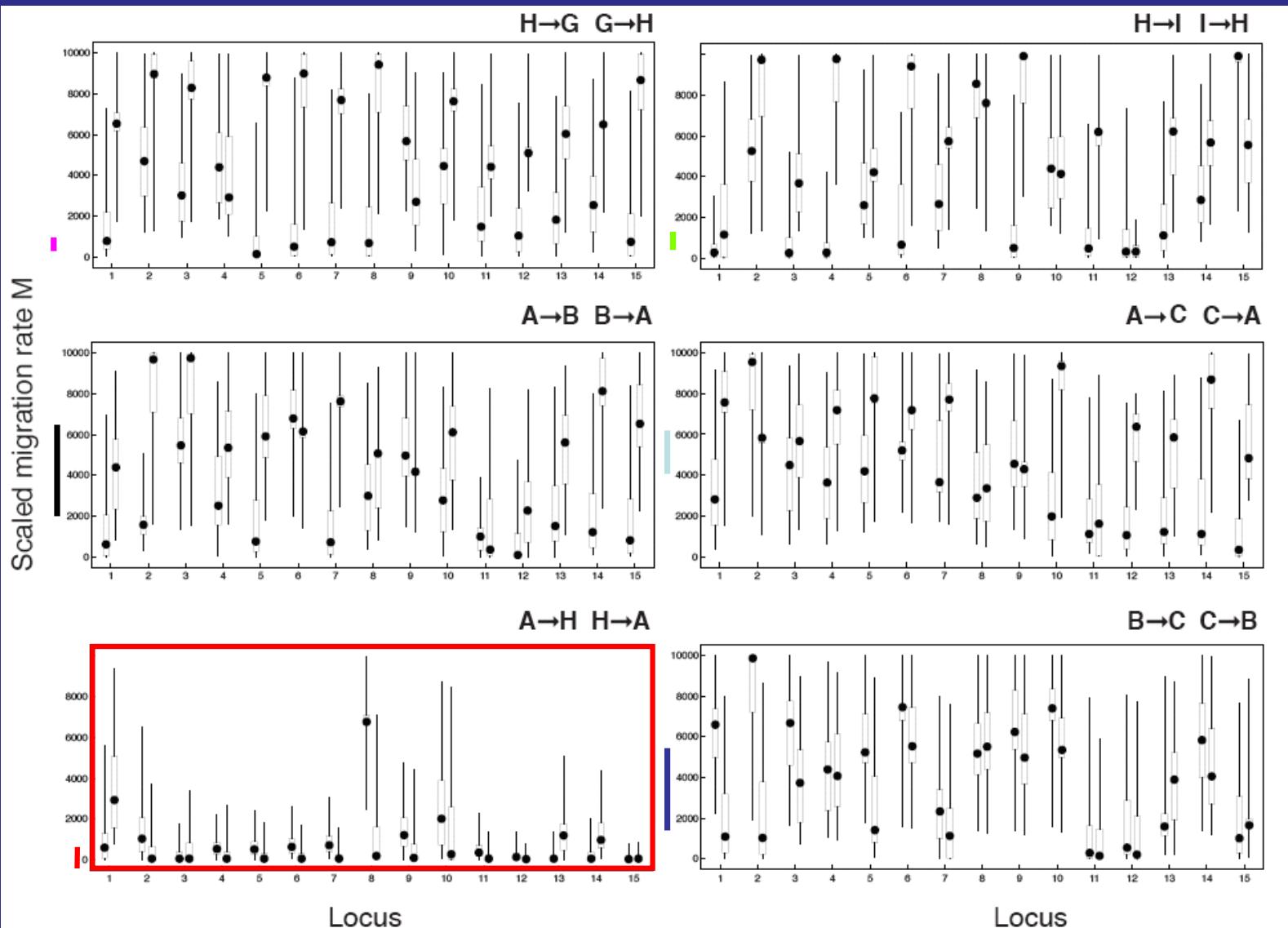
0.00

mL = marginal likelihood

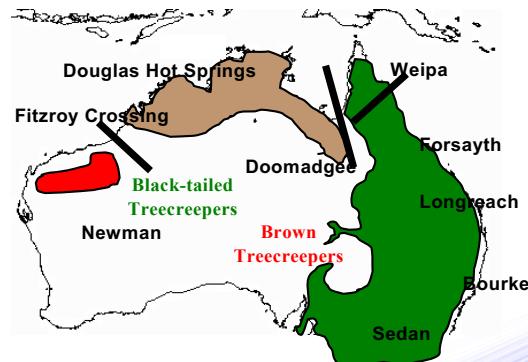
Higher precision estimates of demographic parameters (θ) with more loci



Multilocus estimates of migration rate



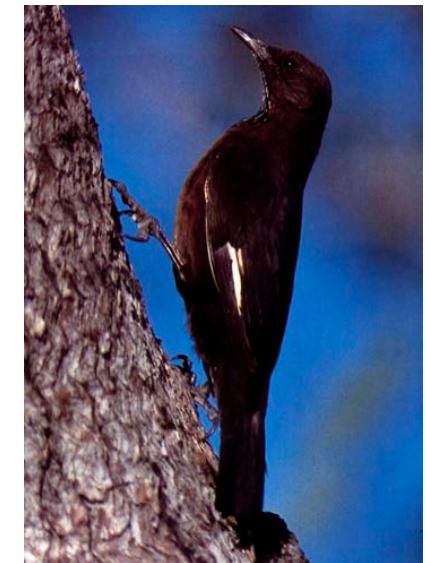
Phylogeny and divergence times



2000000 1500000 1000000 500000 0

CM_Ne

CM



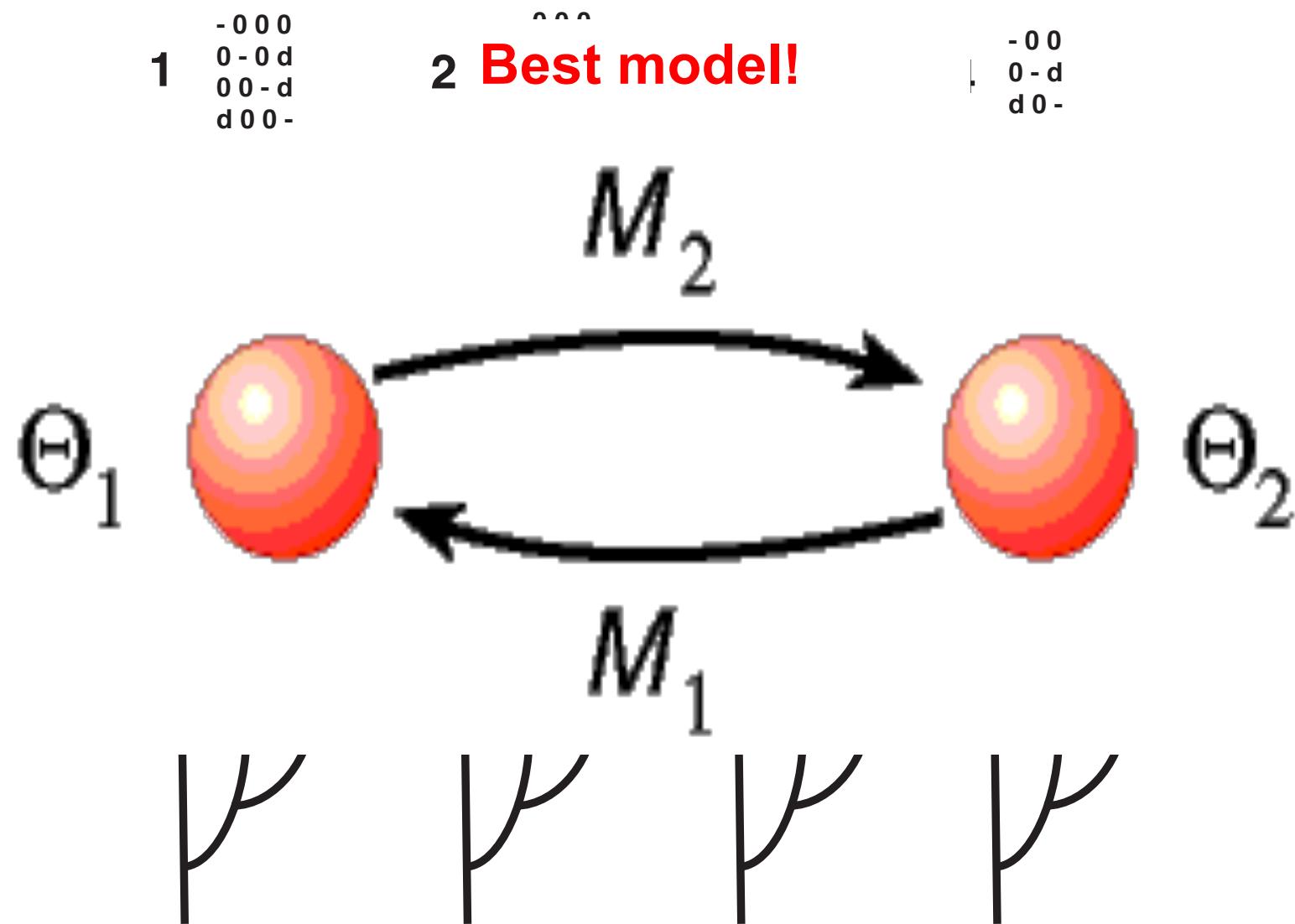
CP_We

CP

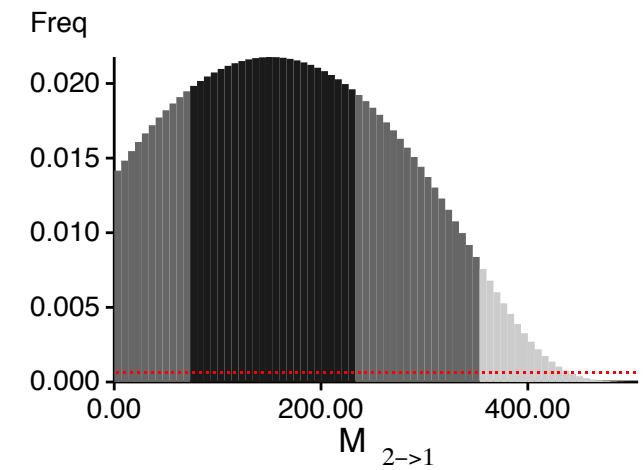
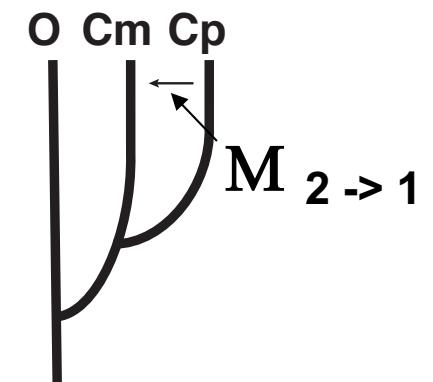
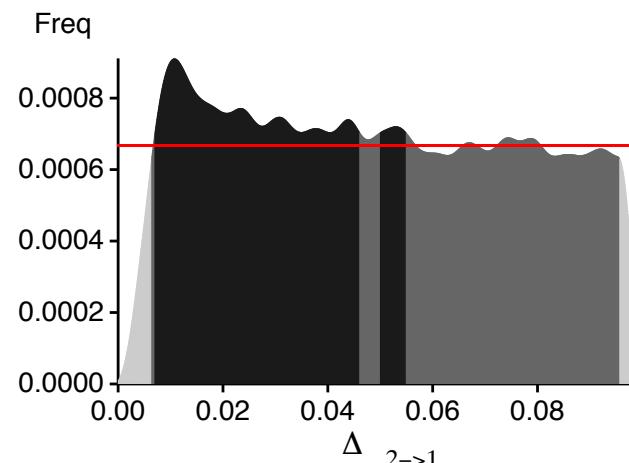
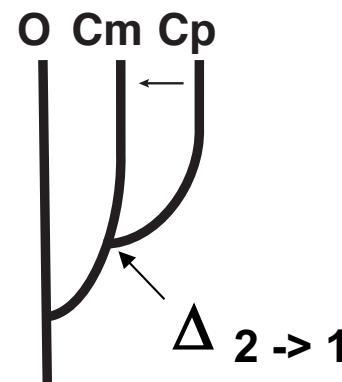
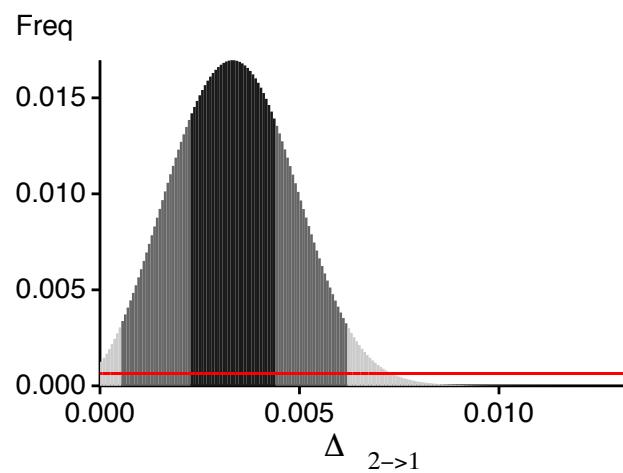
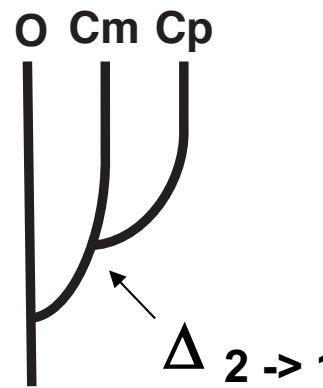


Divergence time (years), gen = 1 yr., $\mu = 2.2\text{e-}09$

Model selection using Migrate

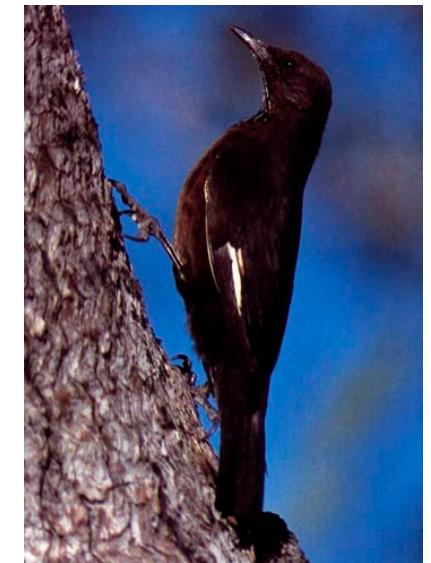
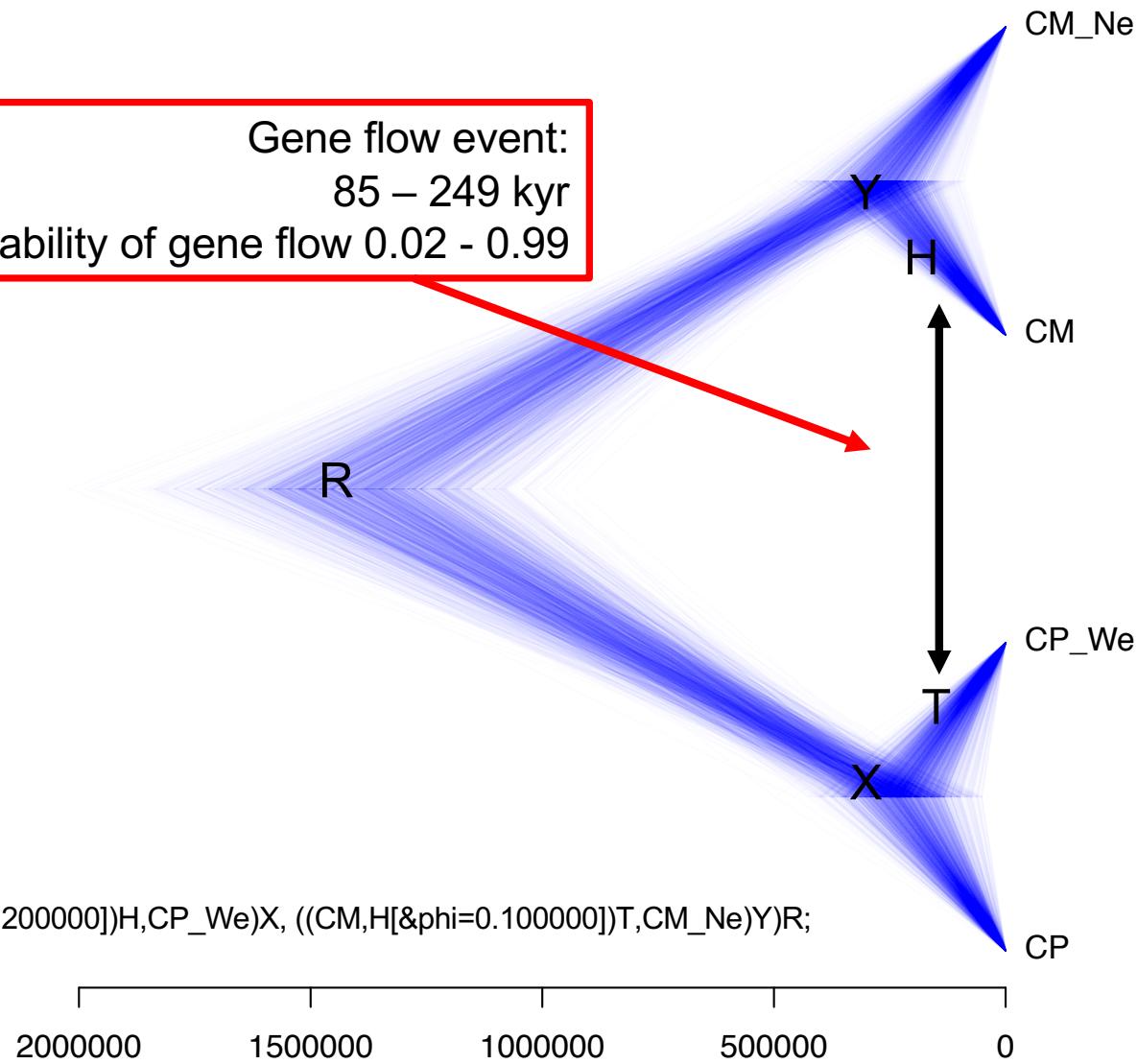


Challenges estimating isolation-migration parameters



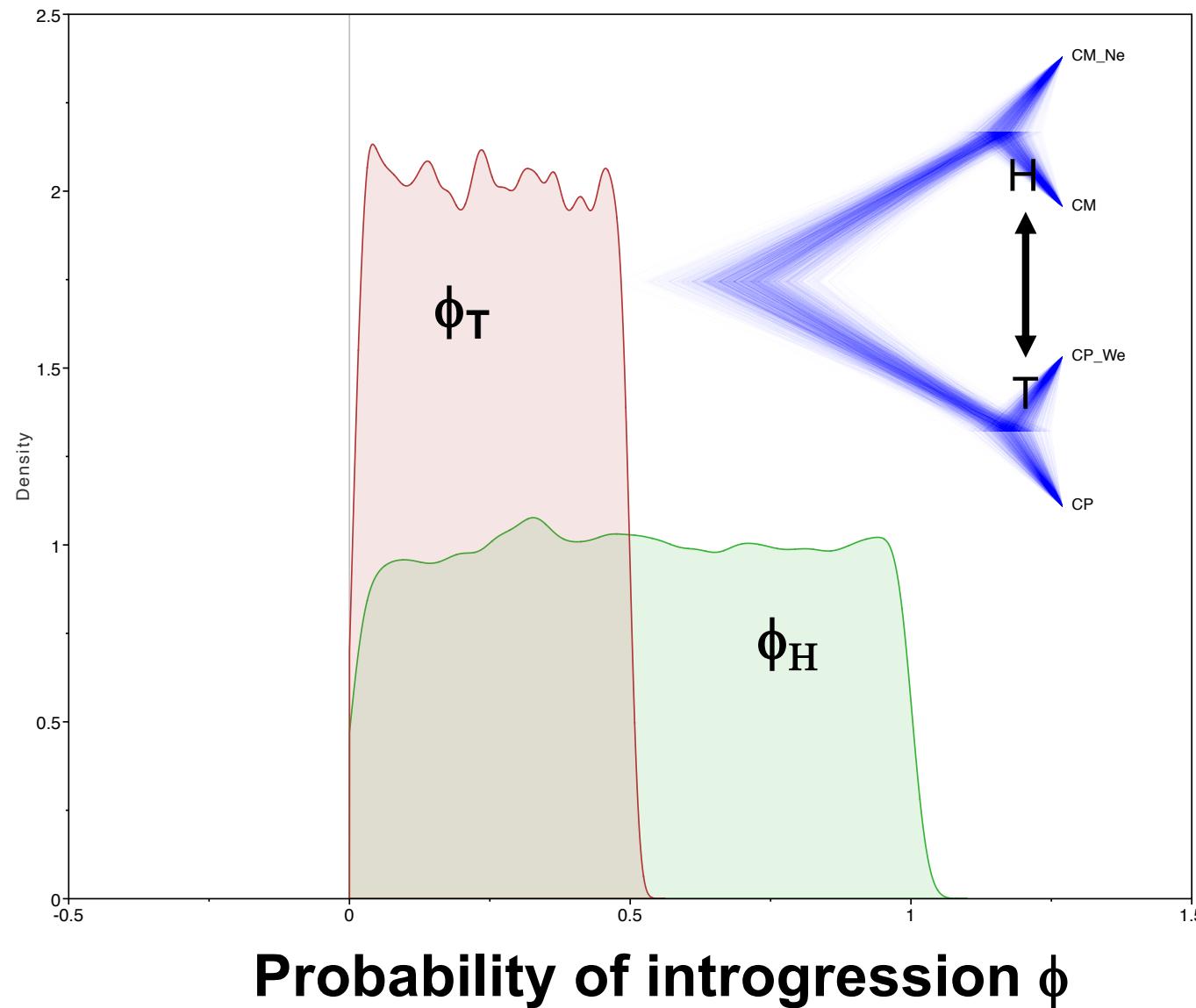
Phylogeny and gene flow with bpp

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

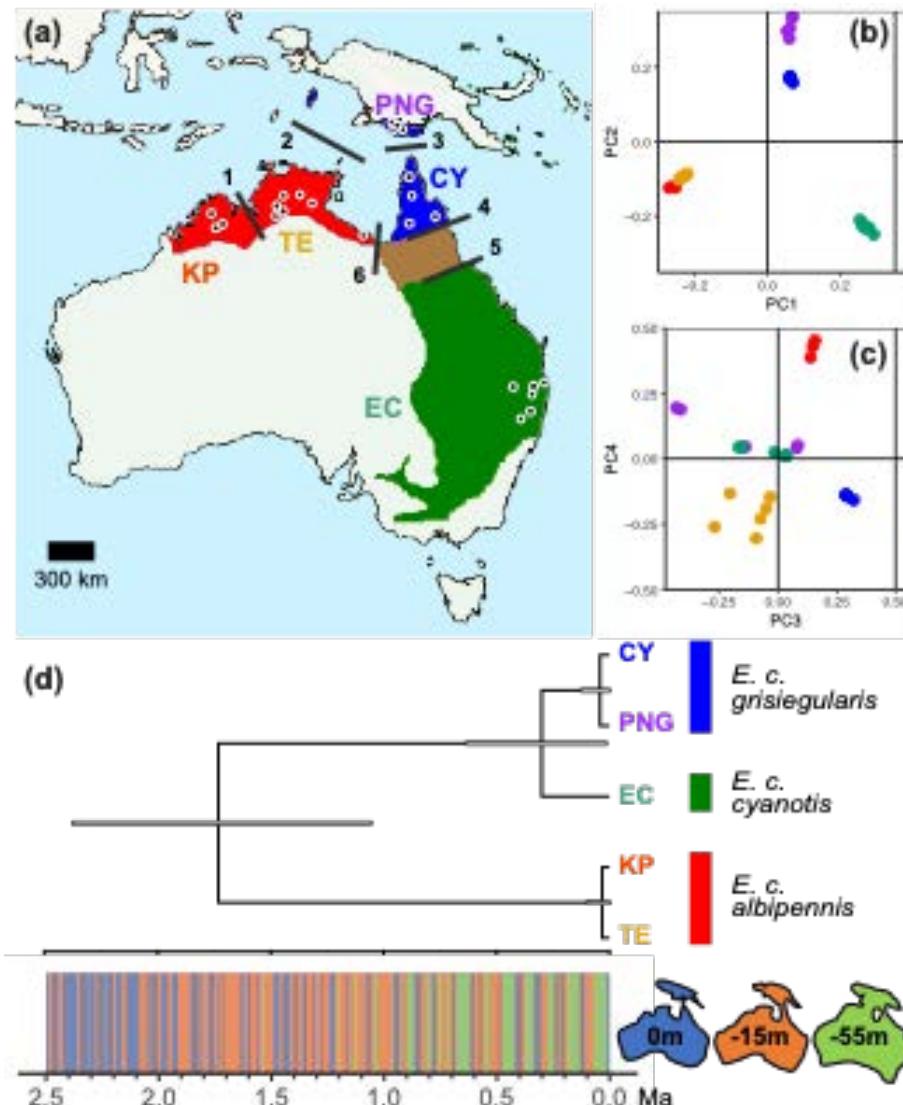


Divergence time (years), gen = 1 yr., $\mu = 2.2\text{e-}09$

BPP: flat posterior distributions of gene flow probability



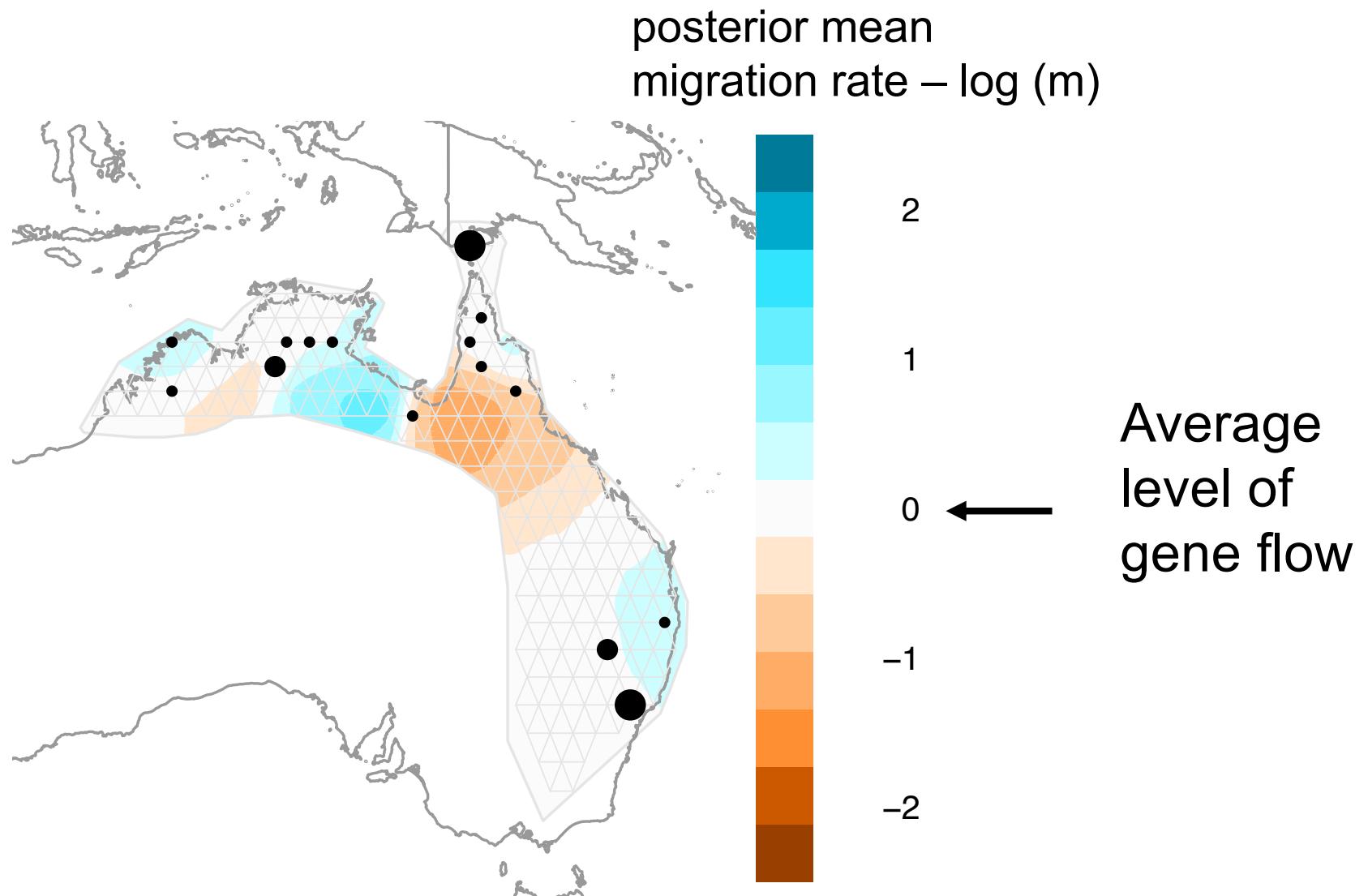
Whole-genome phylogeography of a widespread Australian honeyeater



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*)

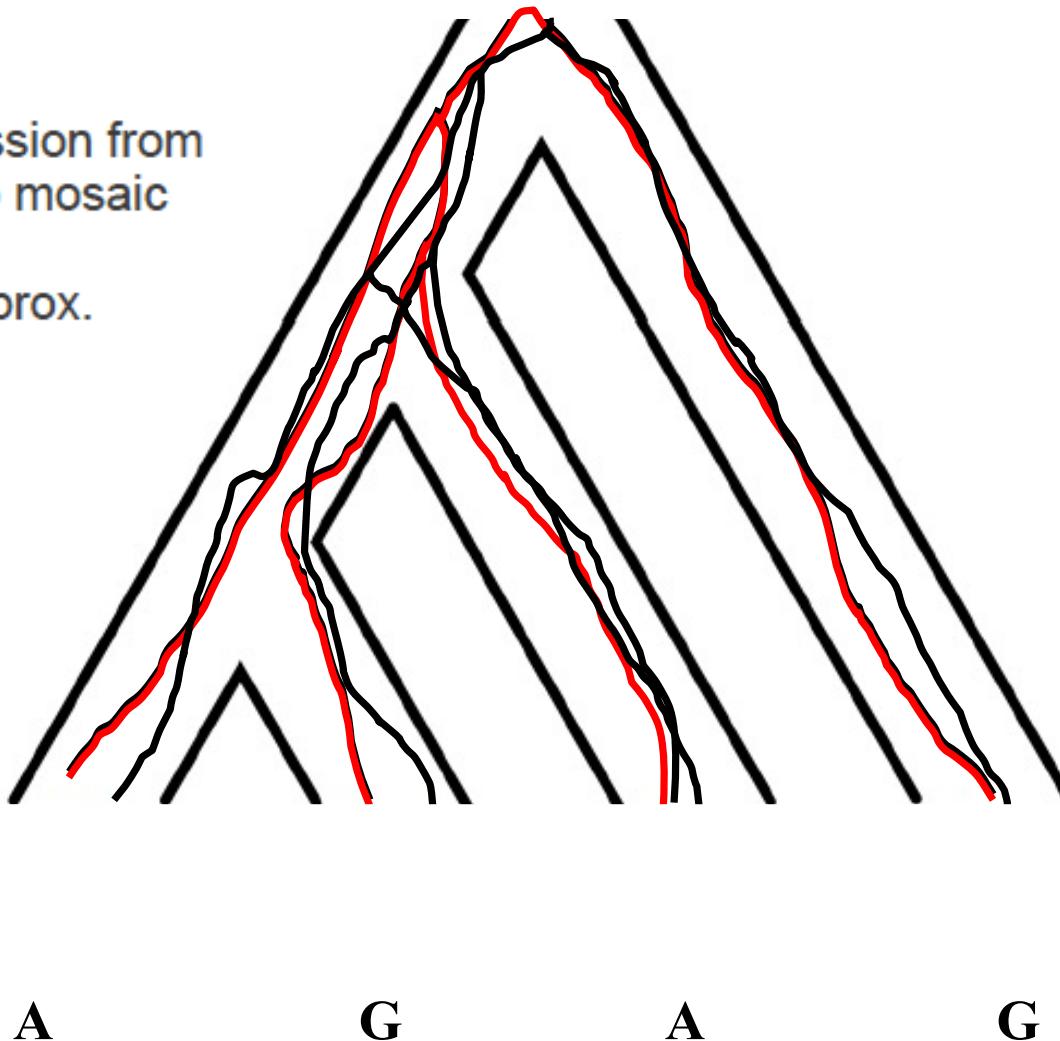
Effective migration surface



ABBA-BABA: Genome-wide test of introgression

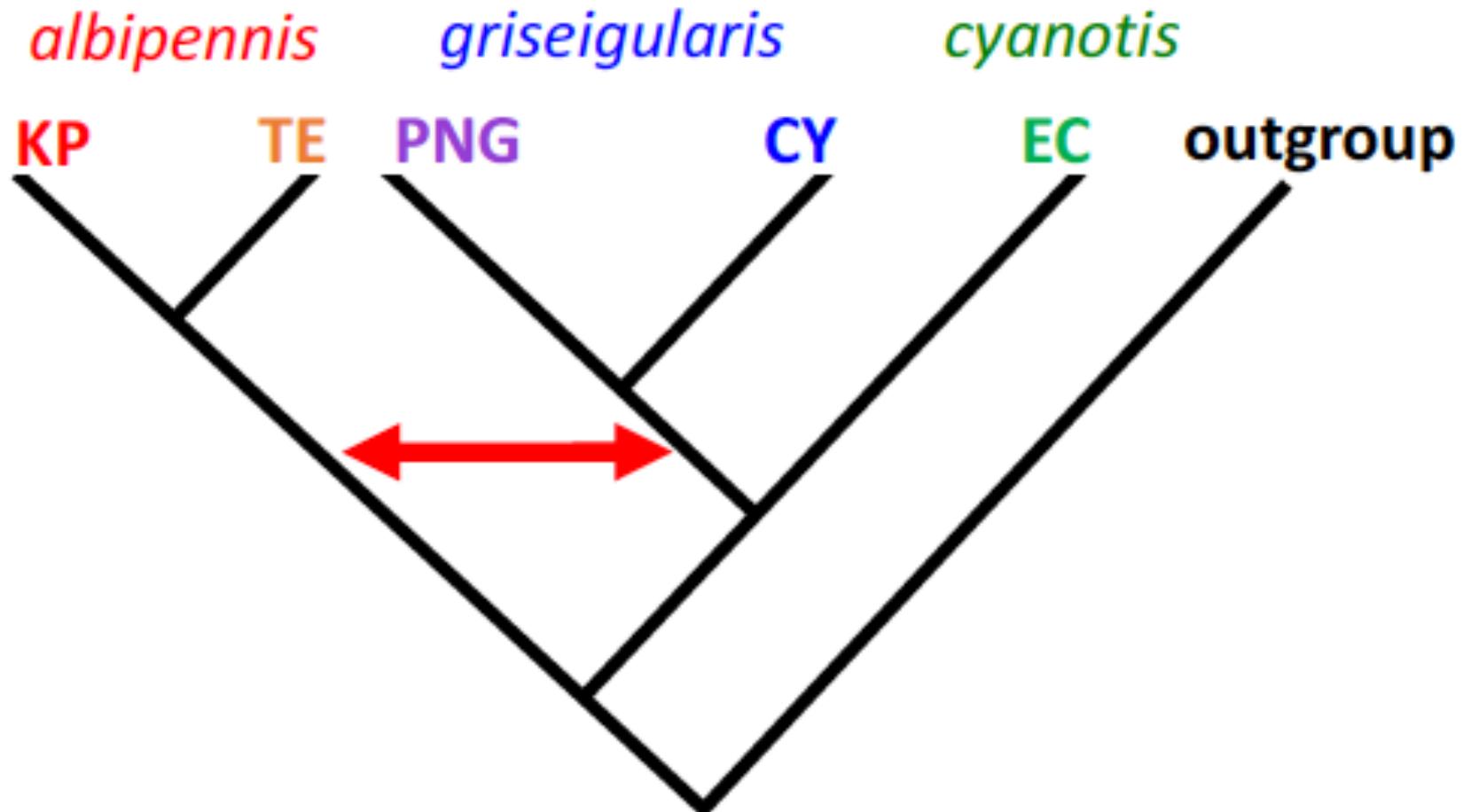
$$D = \frac{\text{Num. ABBA} - \text{Num. BABBA}}{\text{Num SNPs}}$$

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



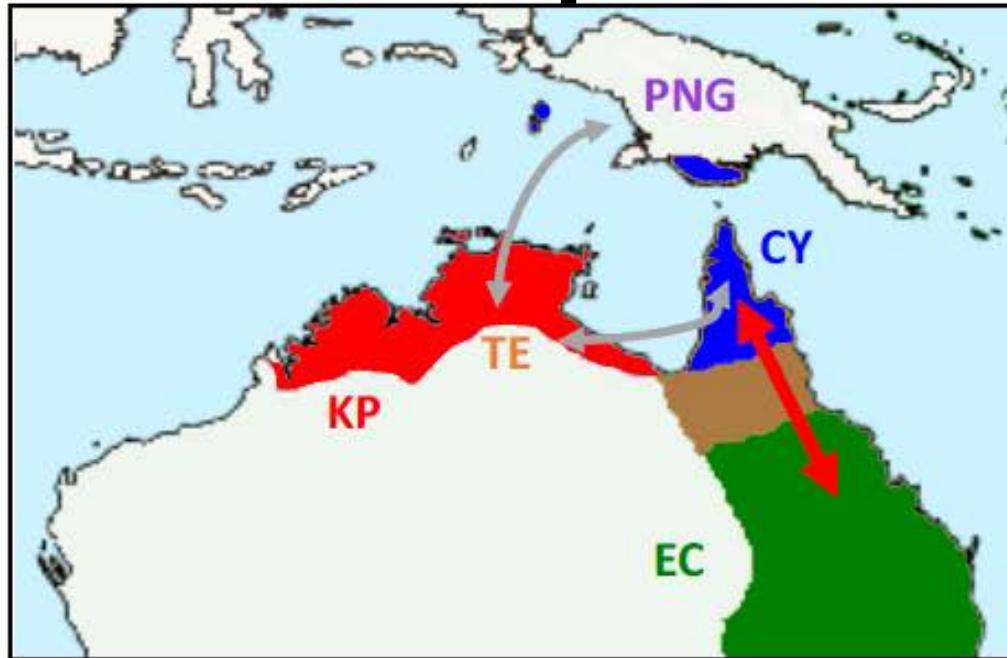
ABBA-BABA tests reveal signals of past introgression

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

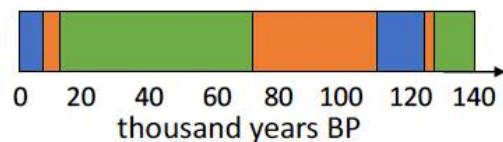


Burley et al. 2022. *Molecular Ecology* 32, 1248-1270

Phylogeographic shadows of Lake Carpentaria



Pleistocene
Lake Carpentaria



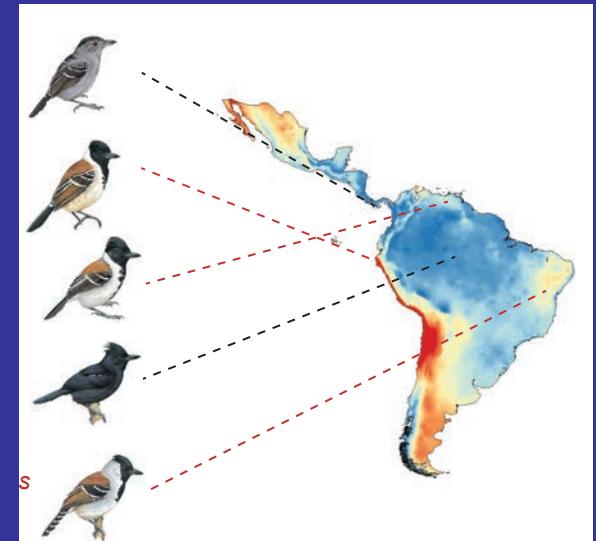
Opinion

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}

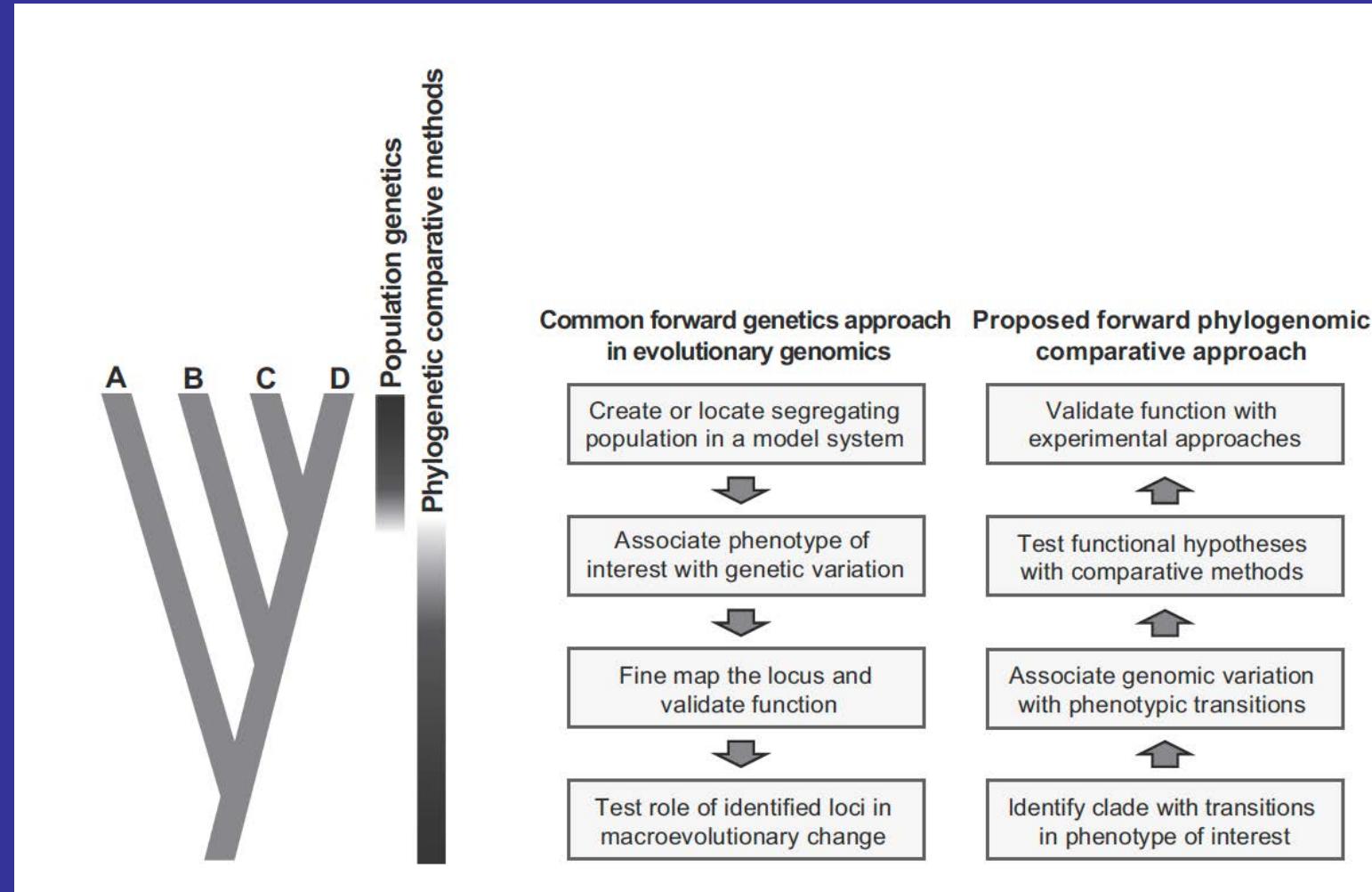
Smith, et al. 2020. *TREE* 35: P415-525

- 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

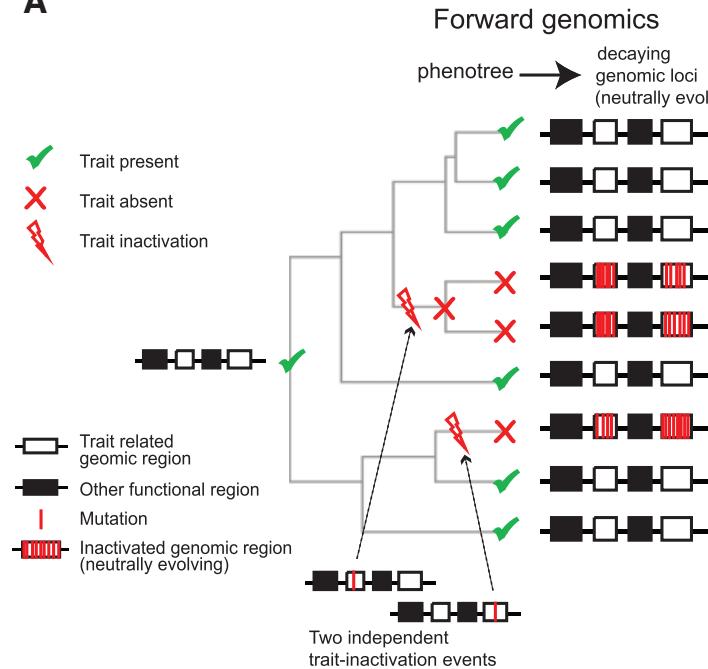
Using phylogenies to connect genotype to phenotype



Genomic signatures of trait associations

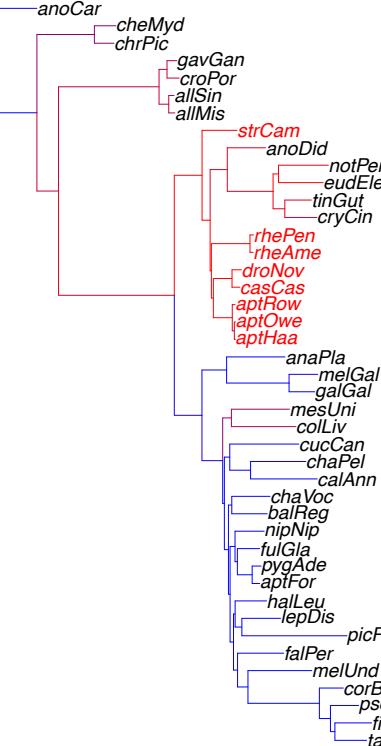
Deletion or inactivation of conserved noncoding elements

A

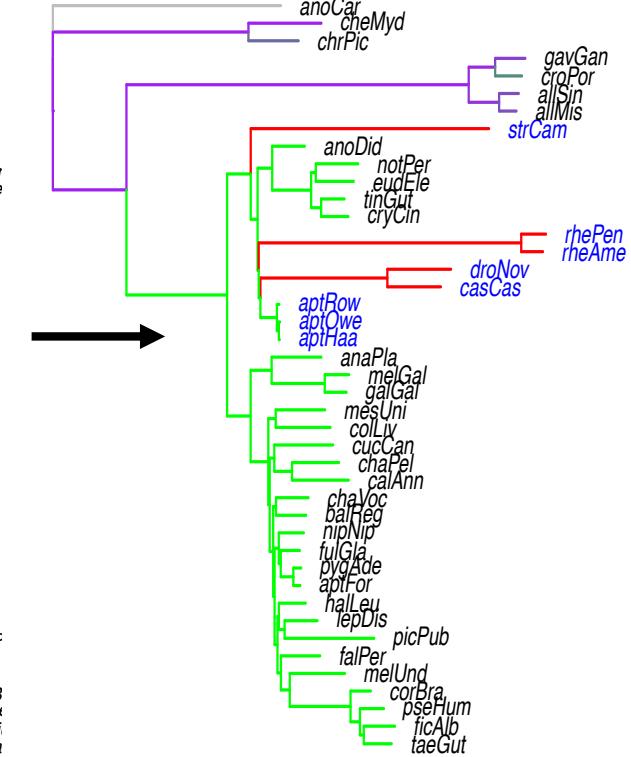


Acceleration of clade-wide conserved noncoding elements

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



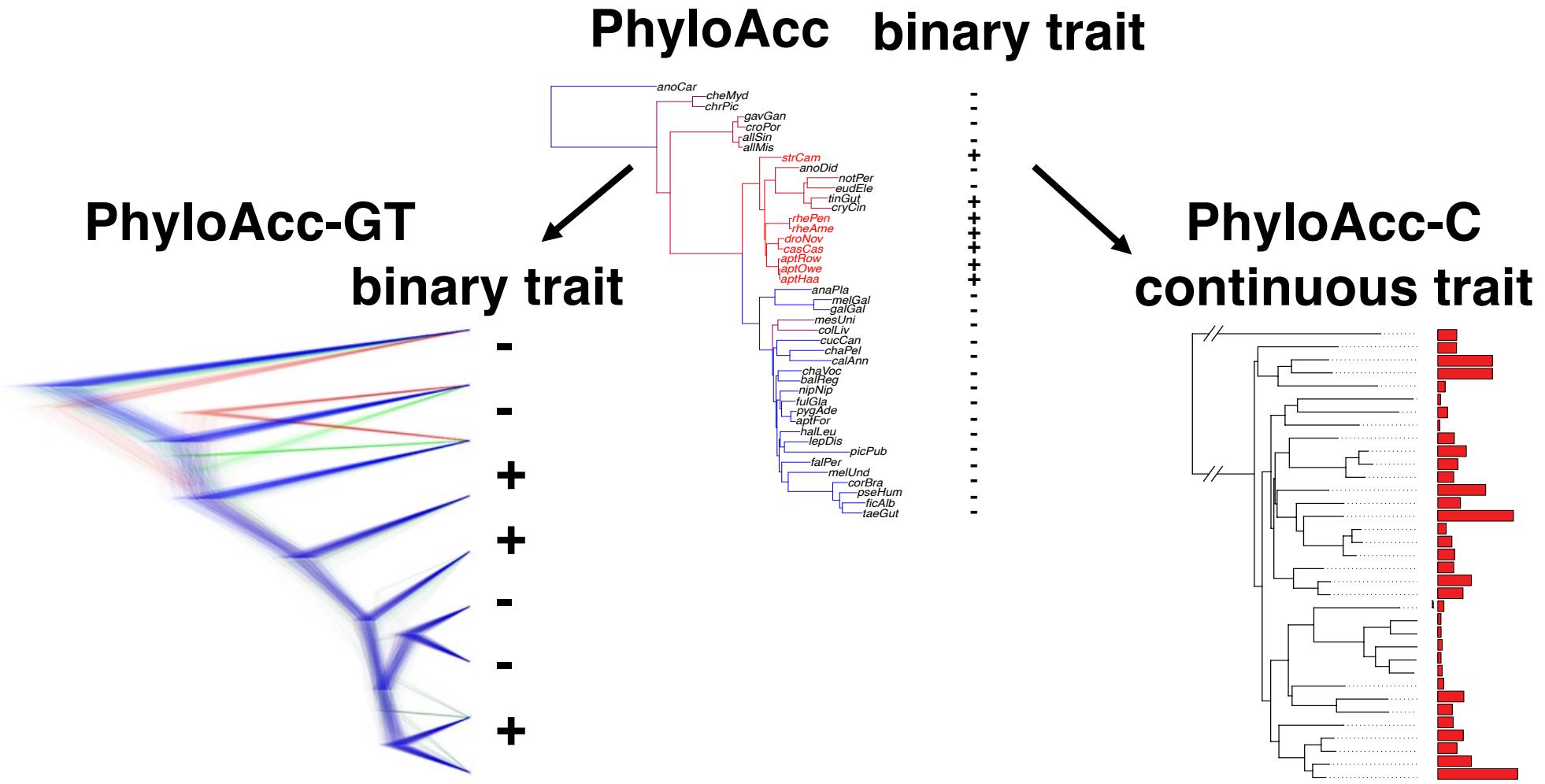
MOTOO KIMURA

Photo by William D. Pevsner

**Motoo Kimura
(1924-1994)**

Functionally less important plurals or gentives will have a high substitution rate than functionally important ones

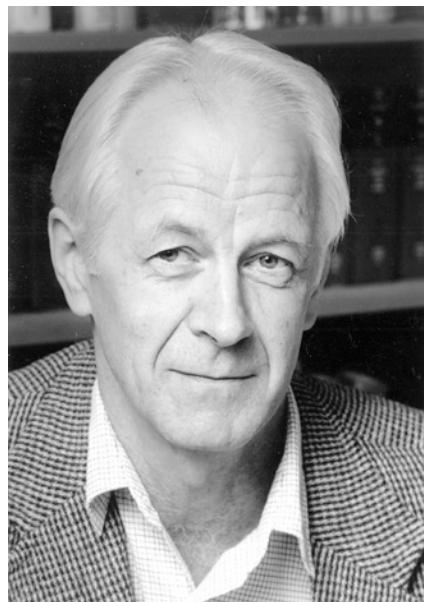
The growing PhyloAcc software family connects genomic and trait variation via phylogenies



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

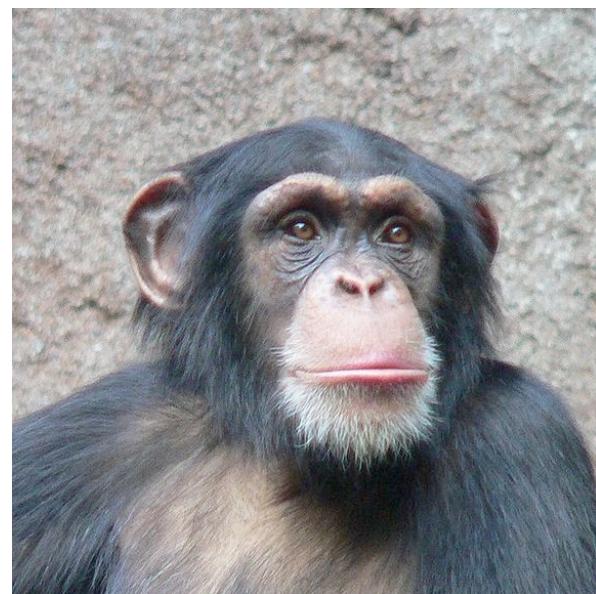
Yan et al. 2023. bioRxiv. <https://doi.org/10.1101/2022.12.23.521765>

Evolutionary change: genes or gene regulation?



Evolution at Two Levels in Humans and Chimpanzees

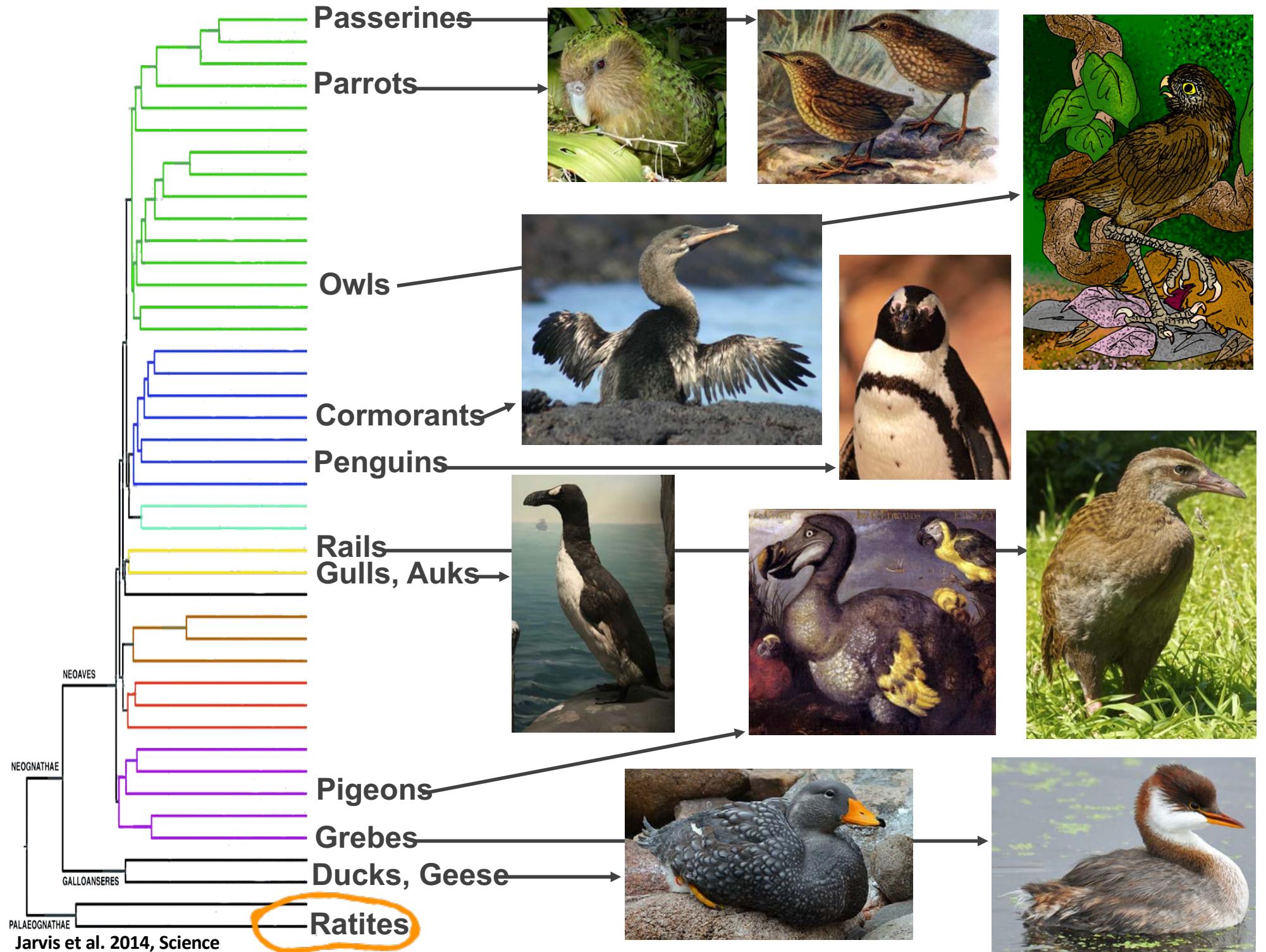
Their macromolecules are so alike that regulatory mutations may account for their biological differences.



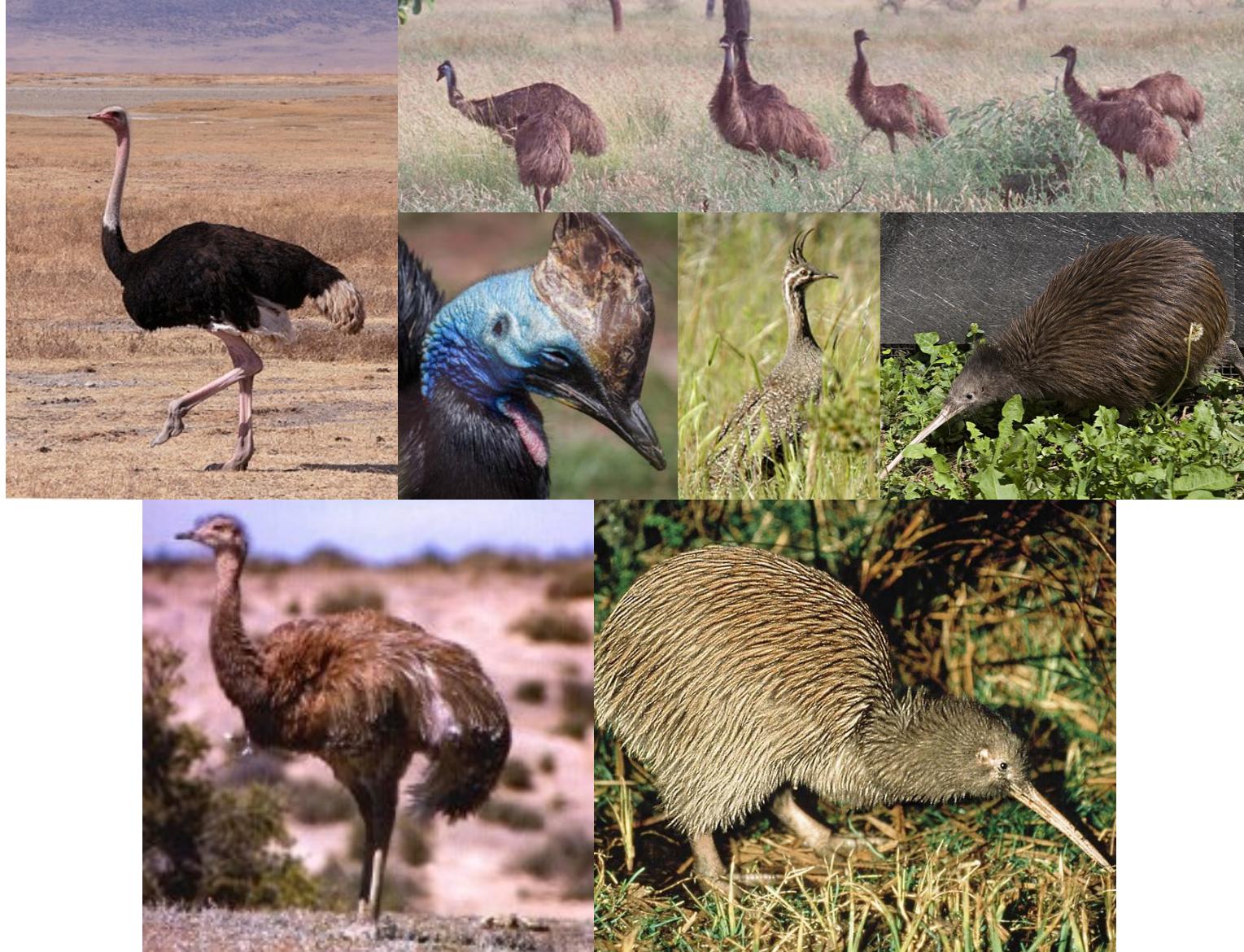
Mary-Claire King and A. C. Wilson

SCIENCE

11 April 1975, Volume 188, Number 4184



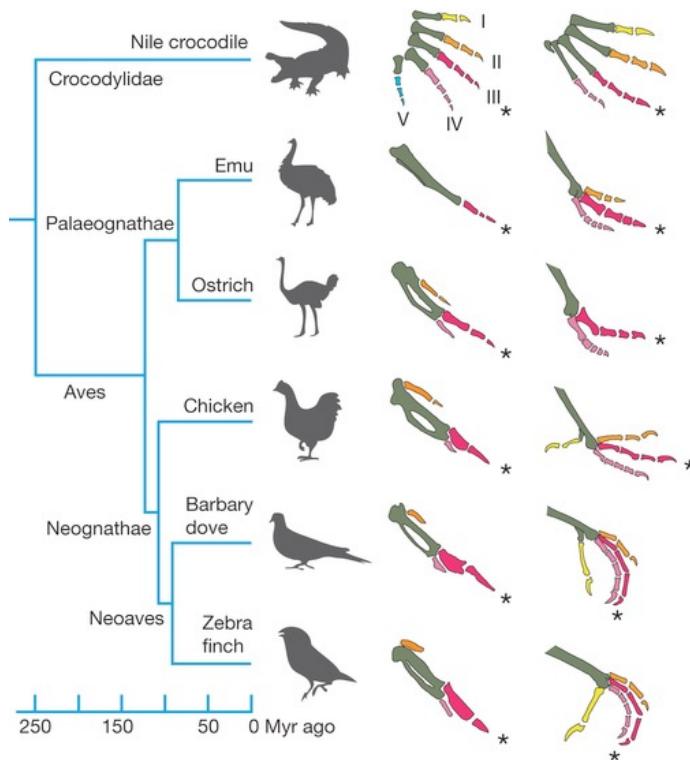
CNEEs and the convergent evolution of flightlessness in Palaeognathae



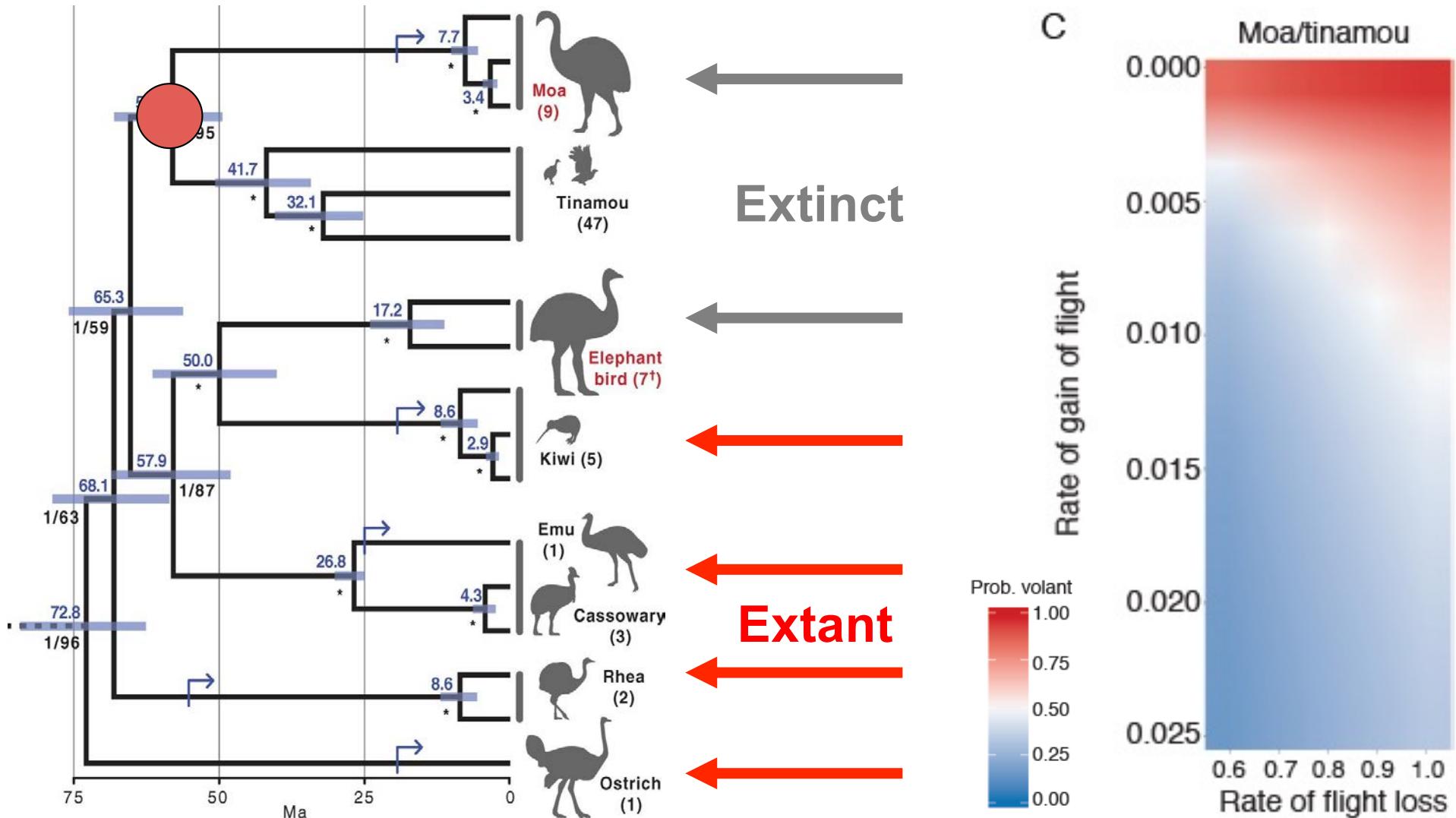
Skeletal modifications for flightlessness



Forelimb Hindlimb



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



11 new paleognath genomes



Little Spotted Kiwi



Little bush moa



Great-spotted Kiwi



Lesser Rhea



ern
ary

Emu



Elegant-crested
Tinamou



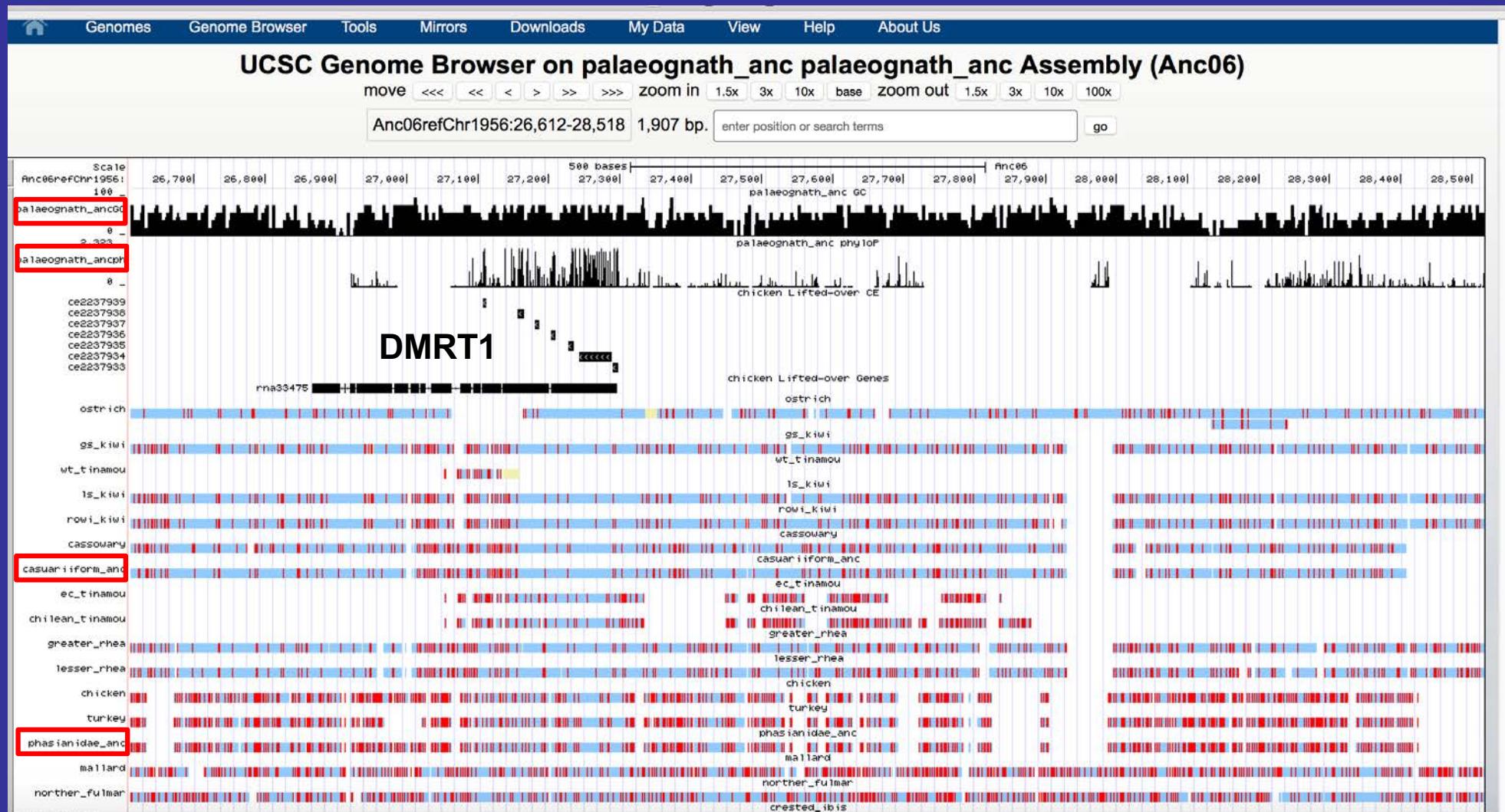
Thicket tinamou



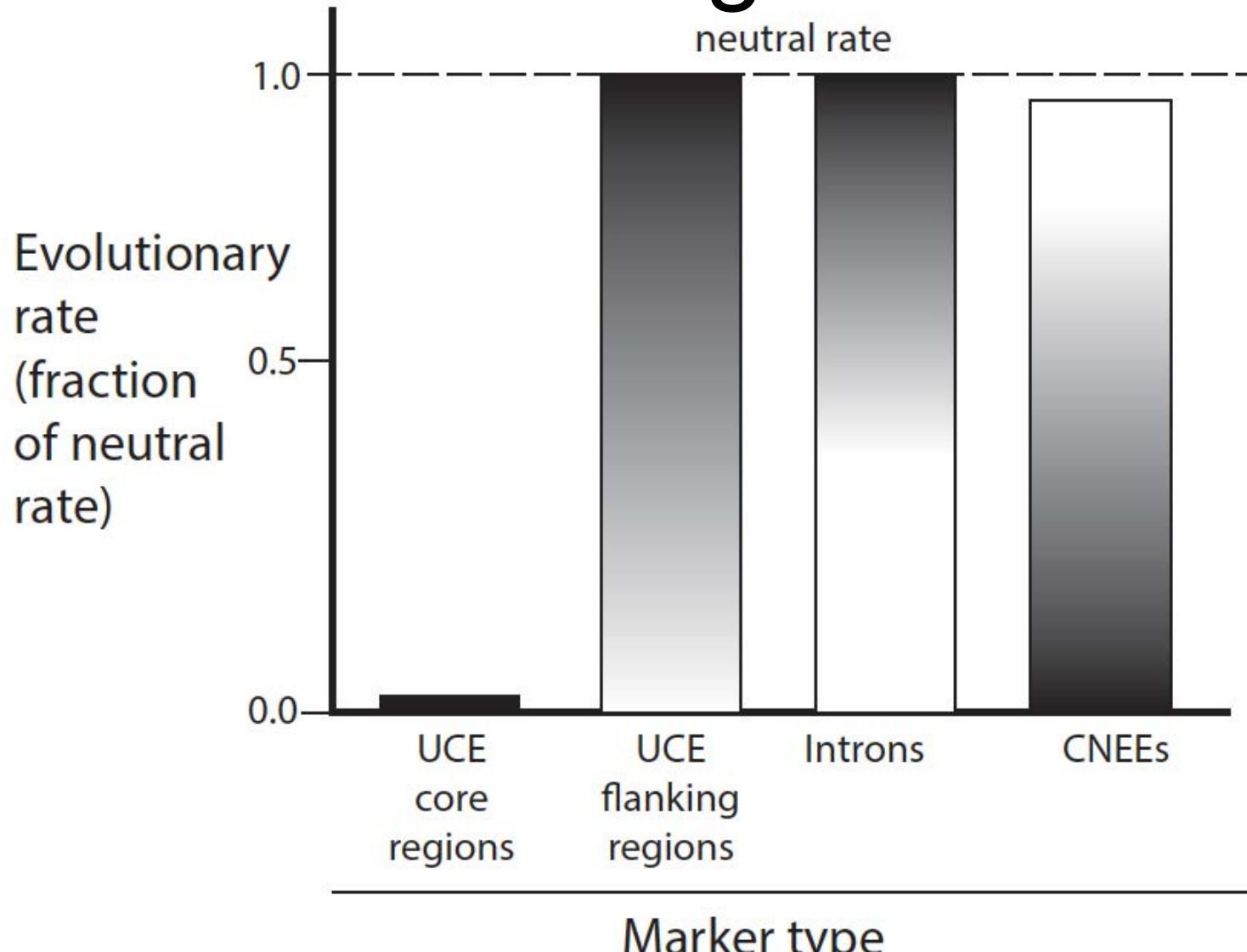
Chilean tinamou

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

42-species whole genome alignment for birds using ProgressiveCactus

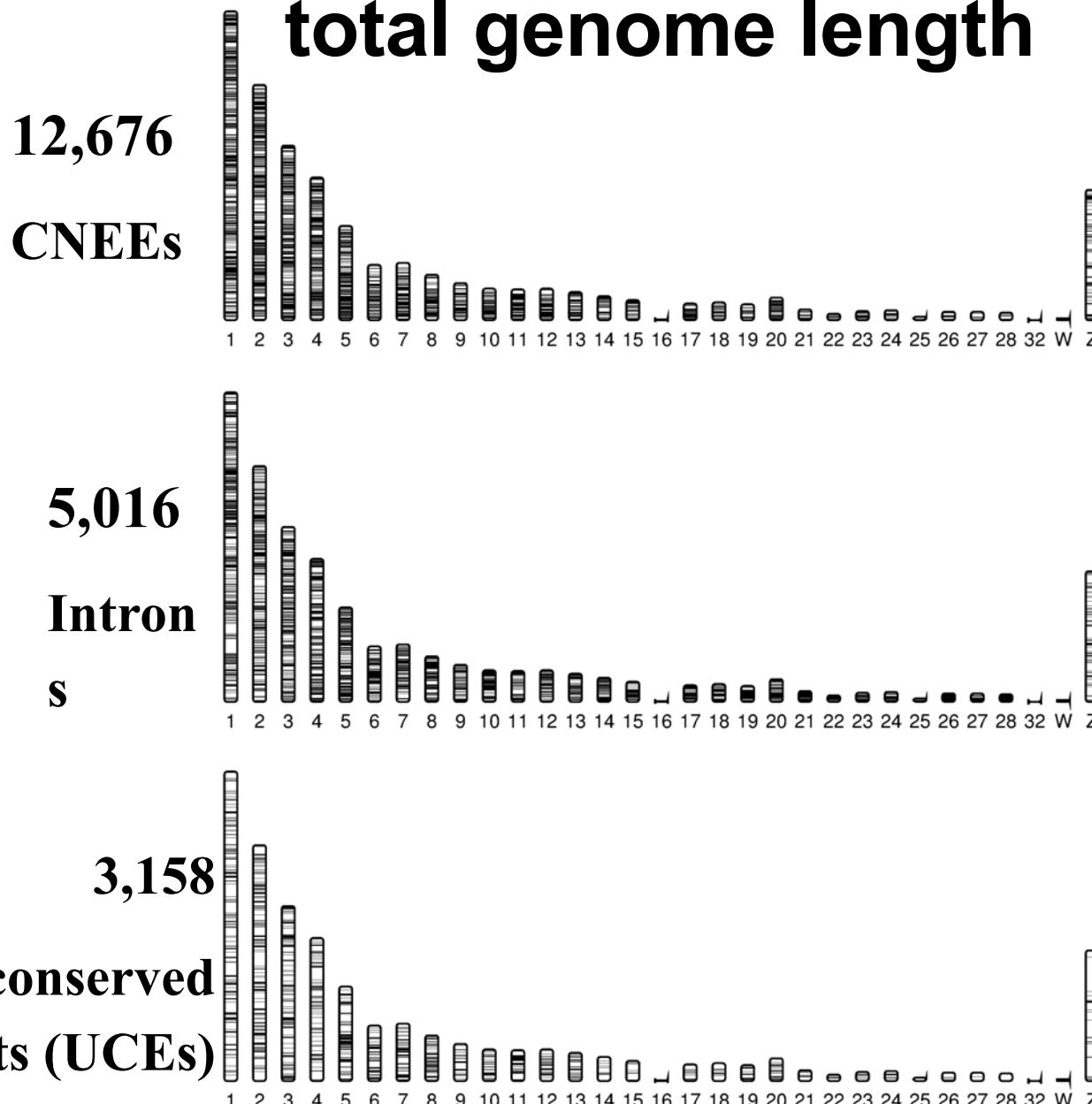


Relative rates of different noncoding markers



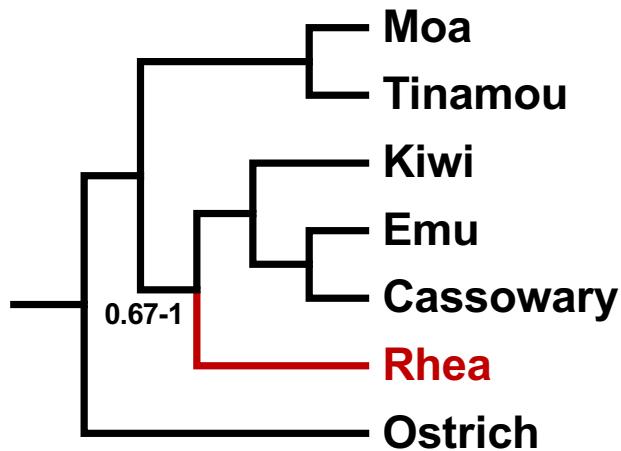
Edwards et al. 2017. *Syst. Biol.* 66: 1028

Phylogenomic markers cover c. 3% of total genome length

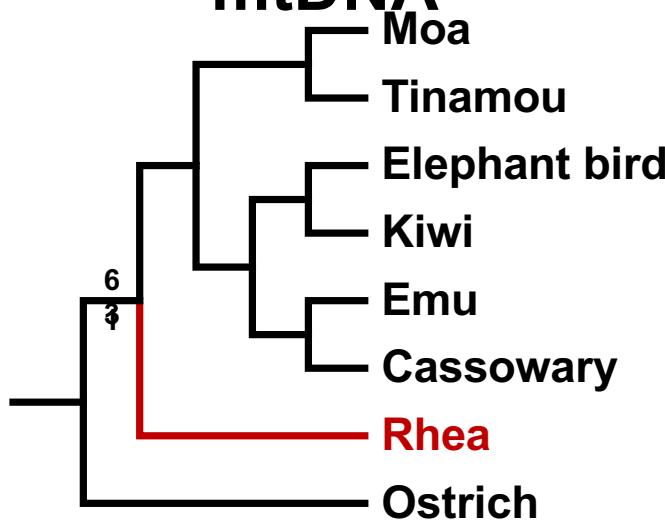


Relationships of rheas unclear

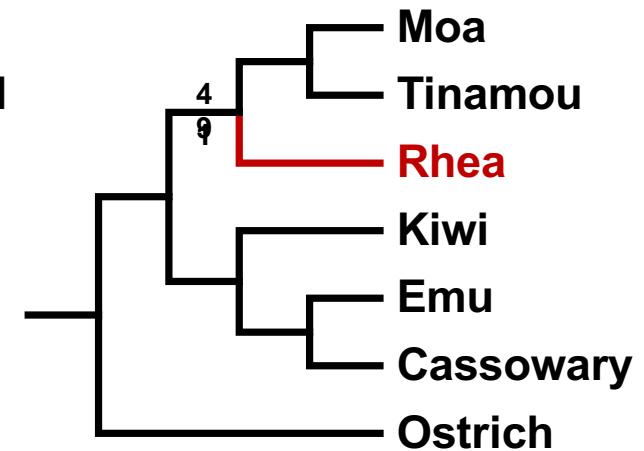
Haddrath & Baker
(2012)
27 nuclear loci



Mitchell et al.
(2014)
mtDNA



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

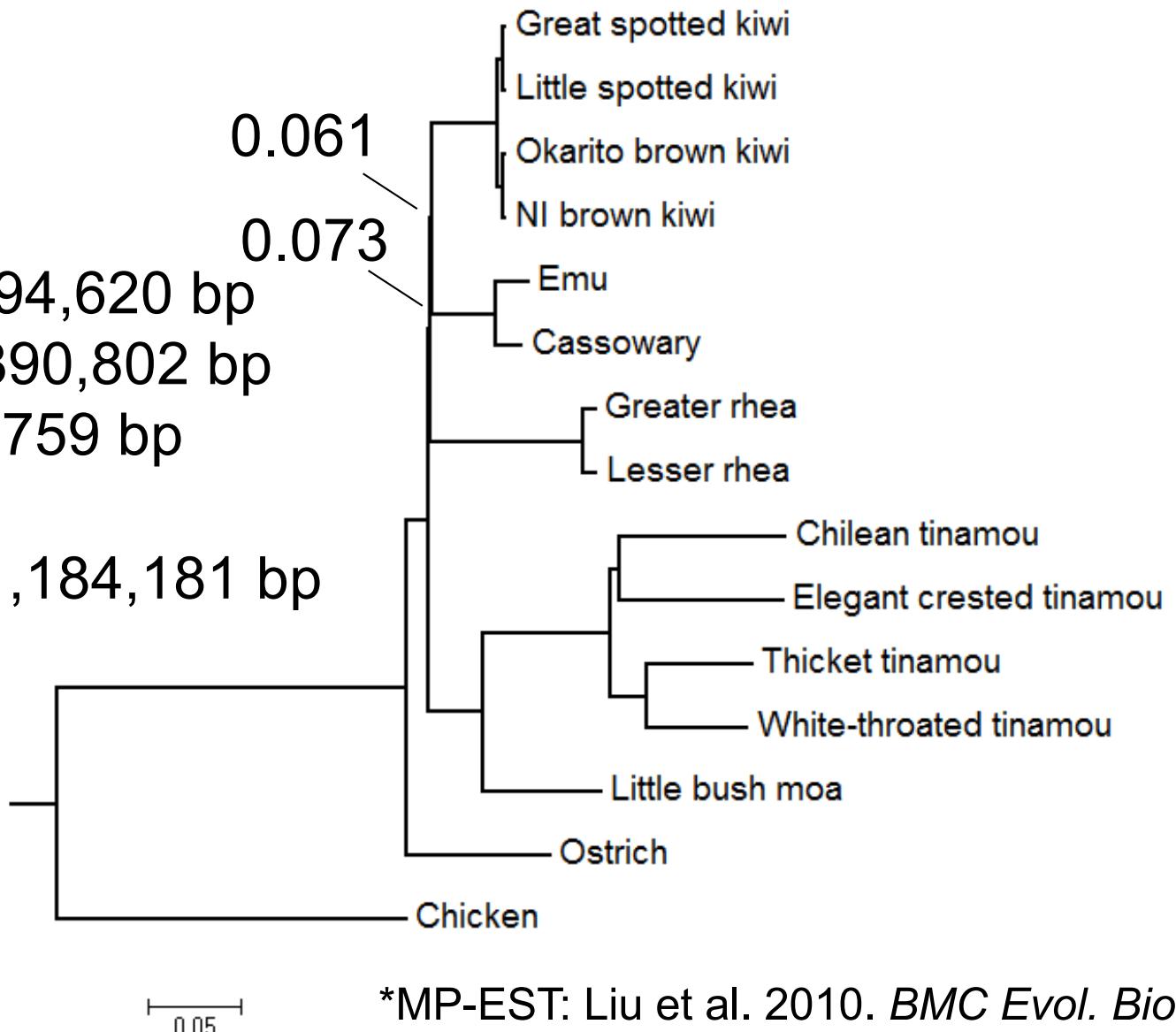


Coalescent* analyses resolve the position of rheas and reveal an ancient rapid radiation

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

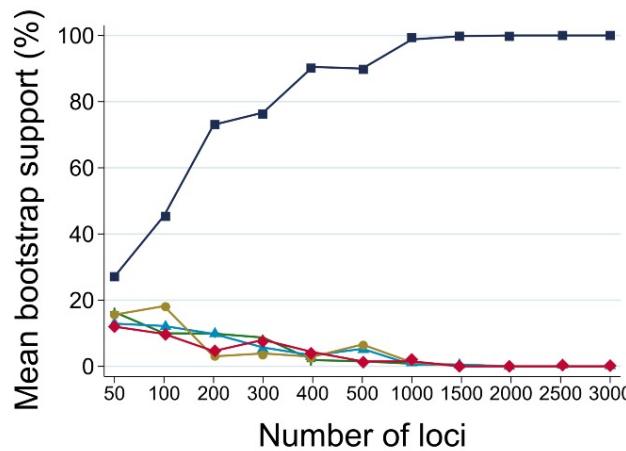
Total: 20,850 loci; 41,184,181 bp

Branch lengths in coalescent units

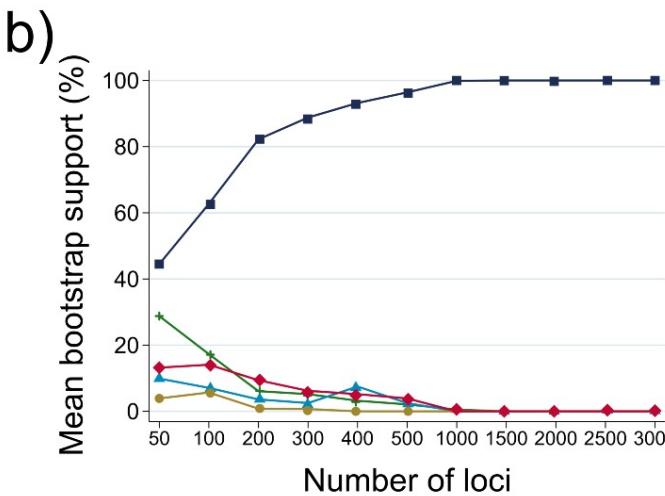


Consistent accumulation of phylogenetic signal using MP-EST

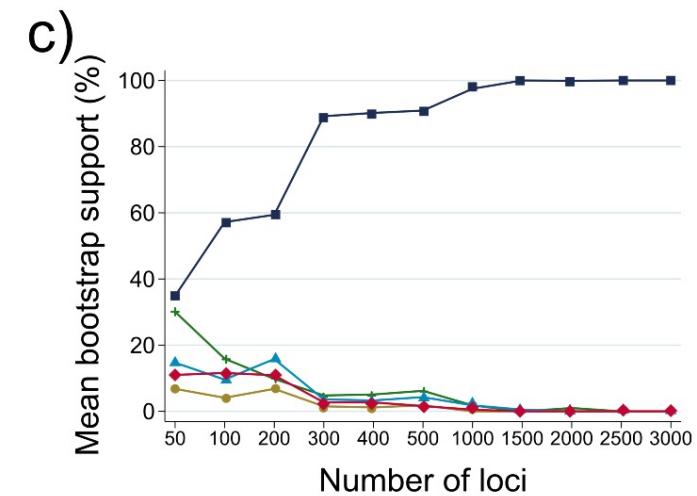
CNEEs



Introns



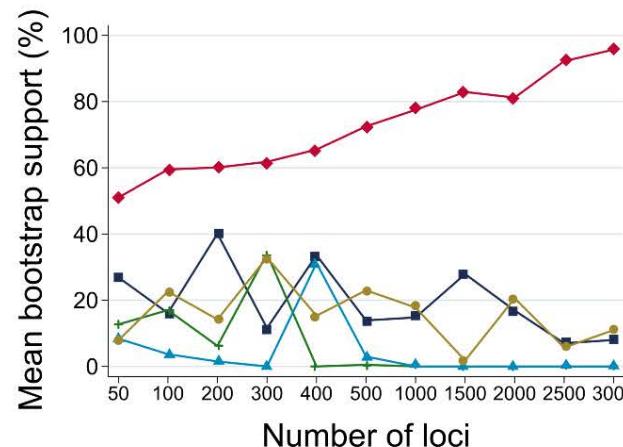
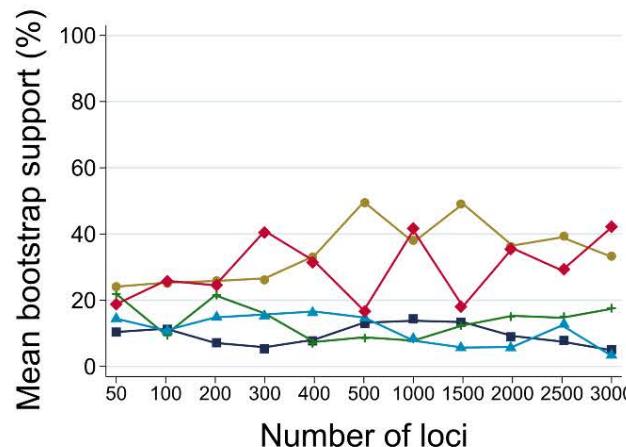
UCEs



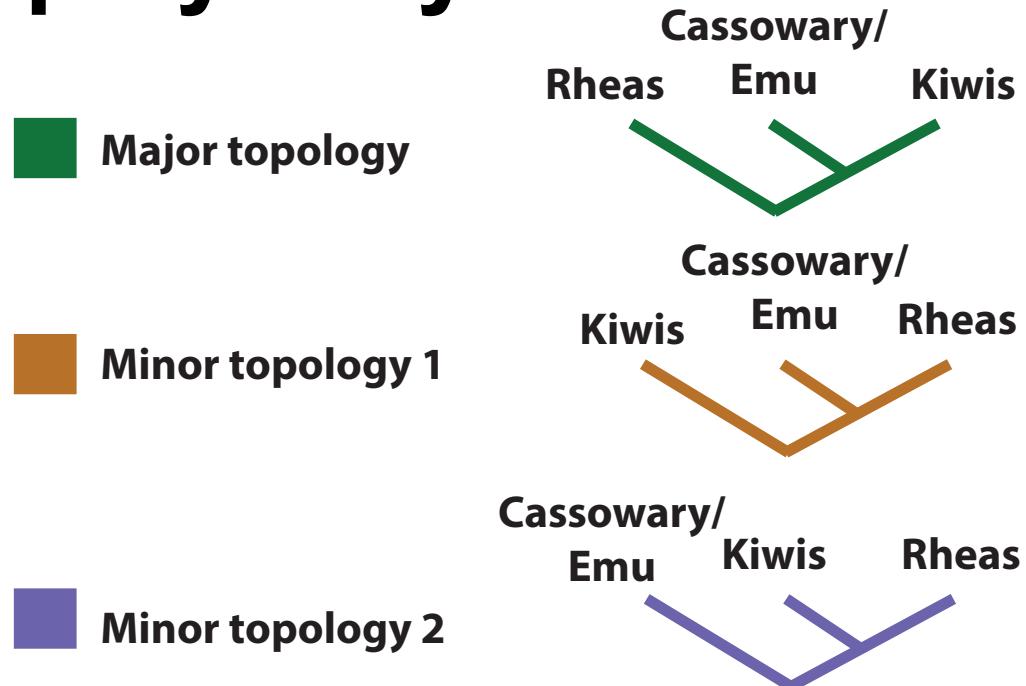
Rheas sister to:

- Emu + Cassowary + Kiwi
- Kiwi
- Emu + Cassowary

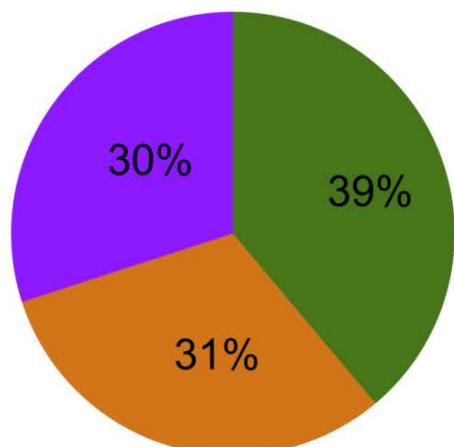
- Moa + Tinamous
- Other non-ostrich palaeognaths



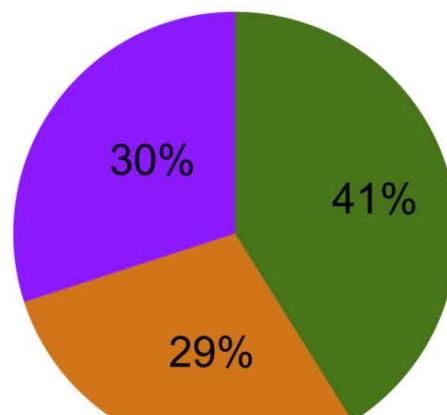
Gene tree distribution suggests a near polytomy at base of ratites



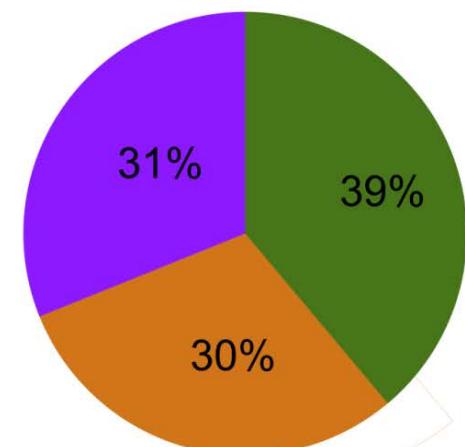
CNEEs



Introns

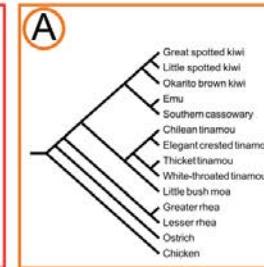
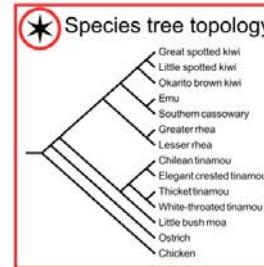
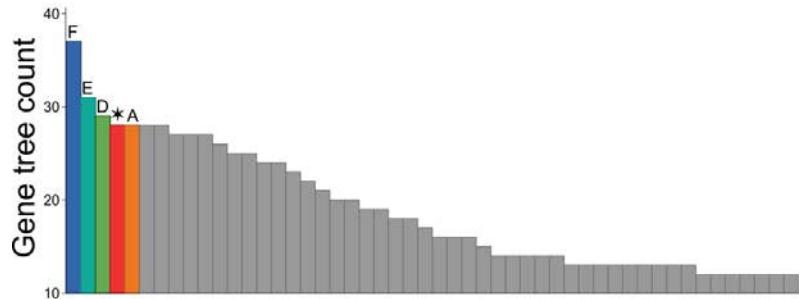


UCEs

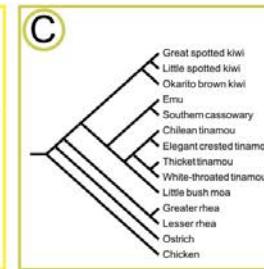
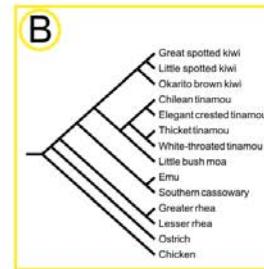
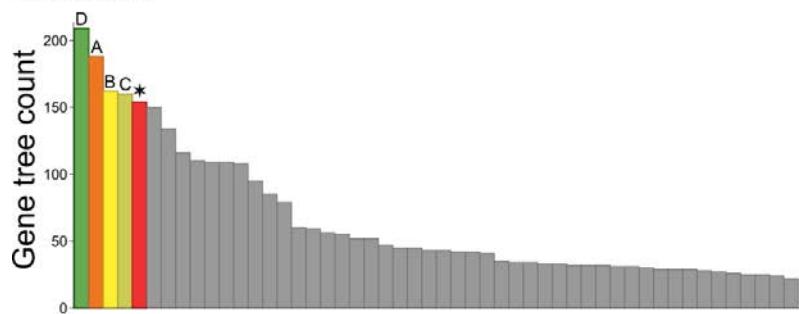


Anomaly zone: most common gene tree does not match the species tree

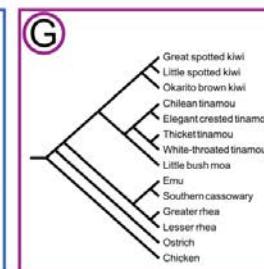
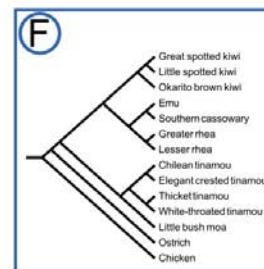
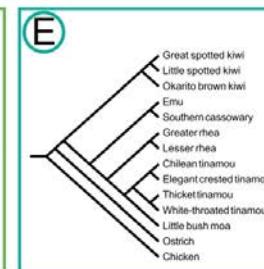
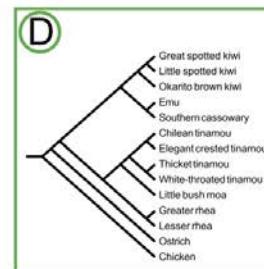
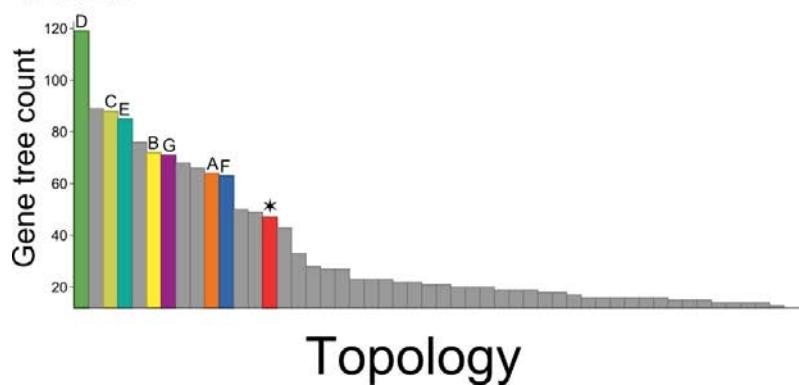
CNEEs



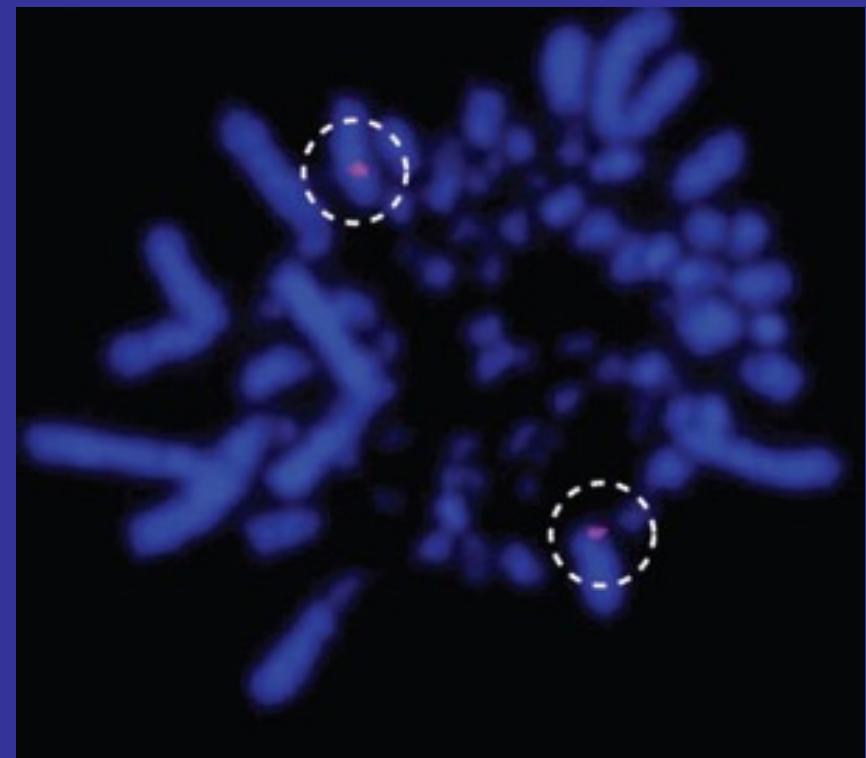
Introns



UCEs



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

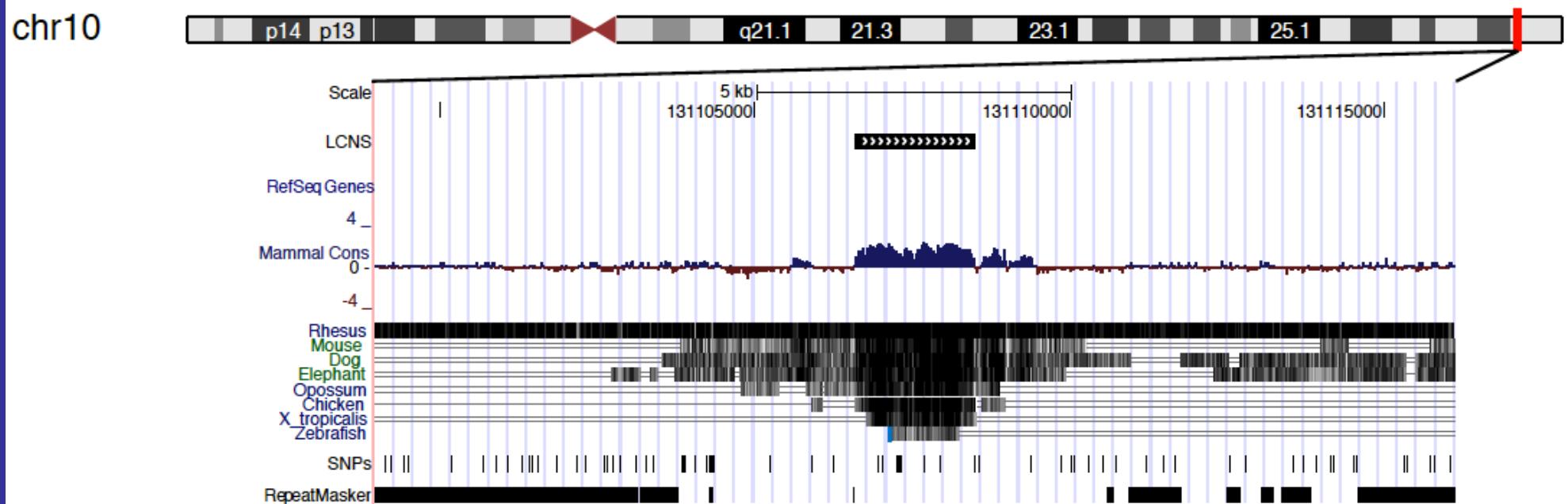


Karyotype of an Emu

CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements

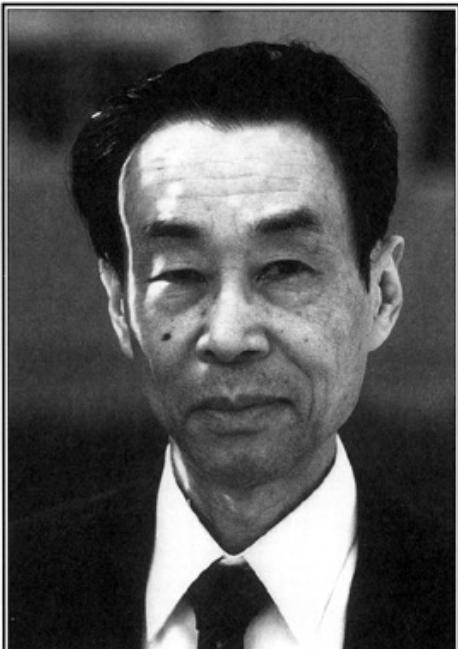
284,001 long (* > 50 bp) CNEEs in data set



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

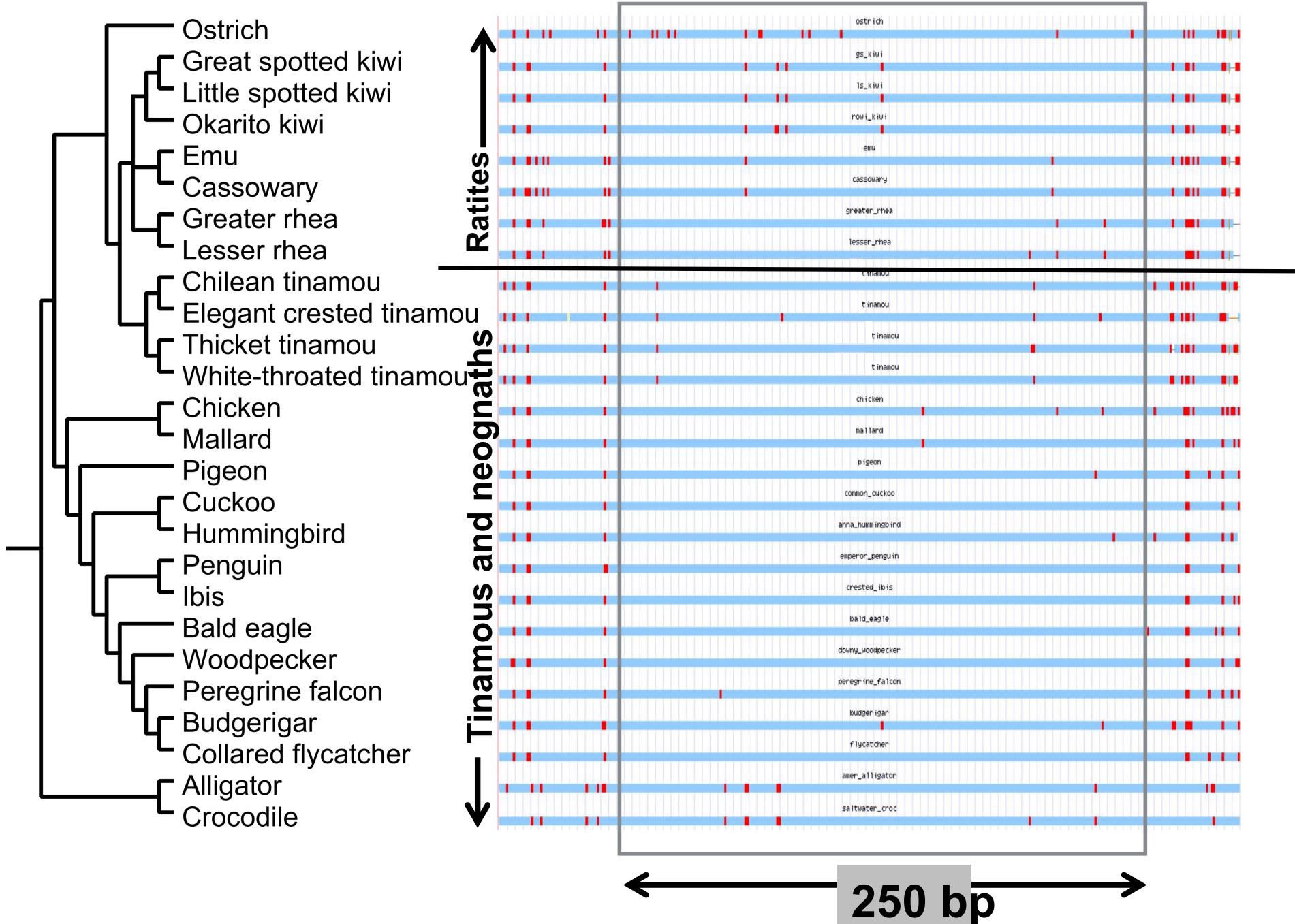


MOTOO KIMURA

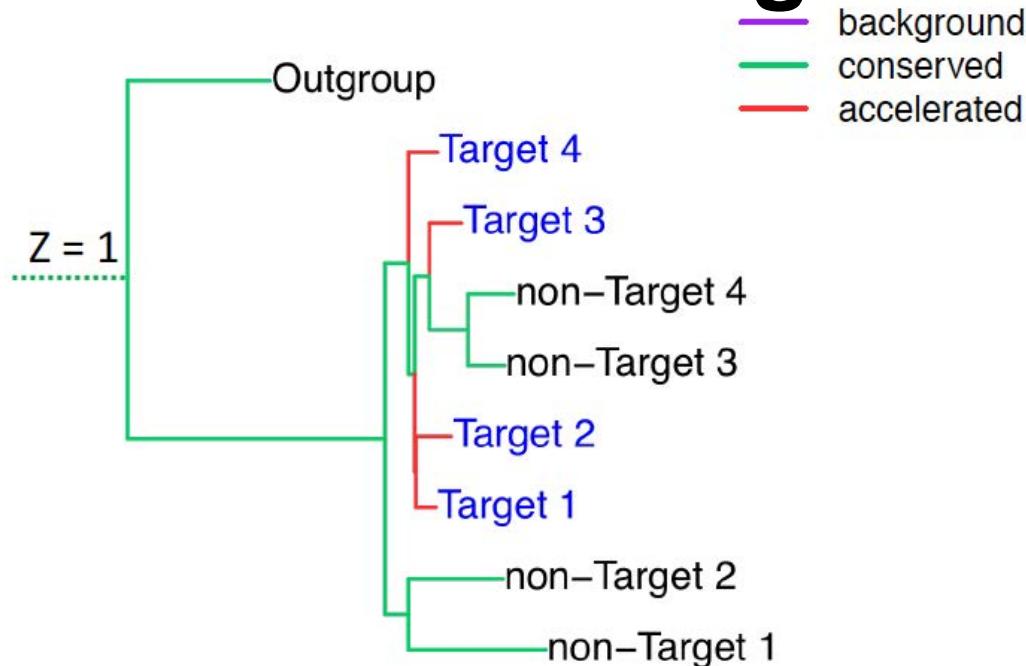
Functionally less important neutral sites or genes will have a high substitution rate than functionally important ones

Motoo Kimura
(1924-1994)

Convergent loss of function of CNEEs in ratite lineages



Branch-specific Bayesian model of noncoding rate accelerations



for noncoding element i

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

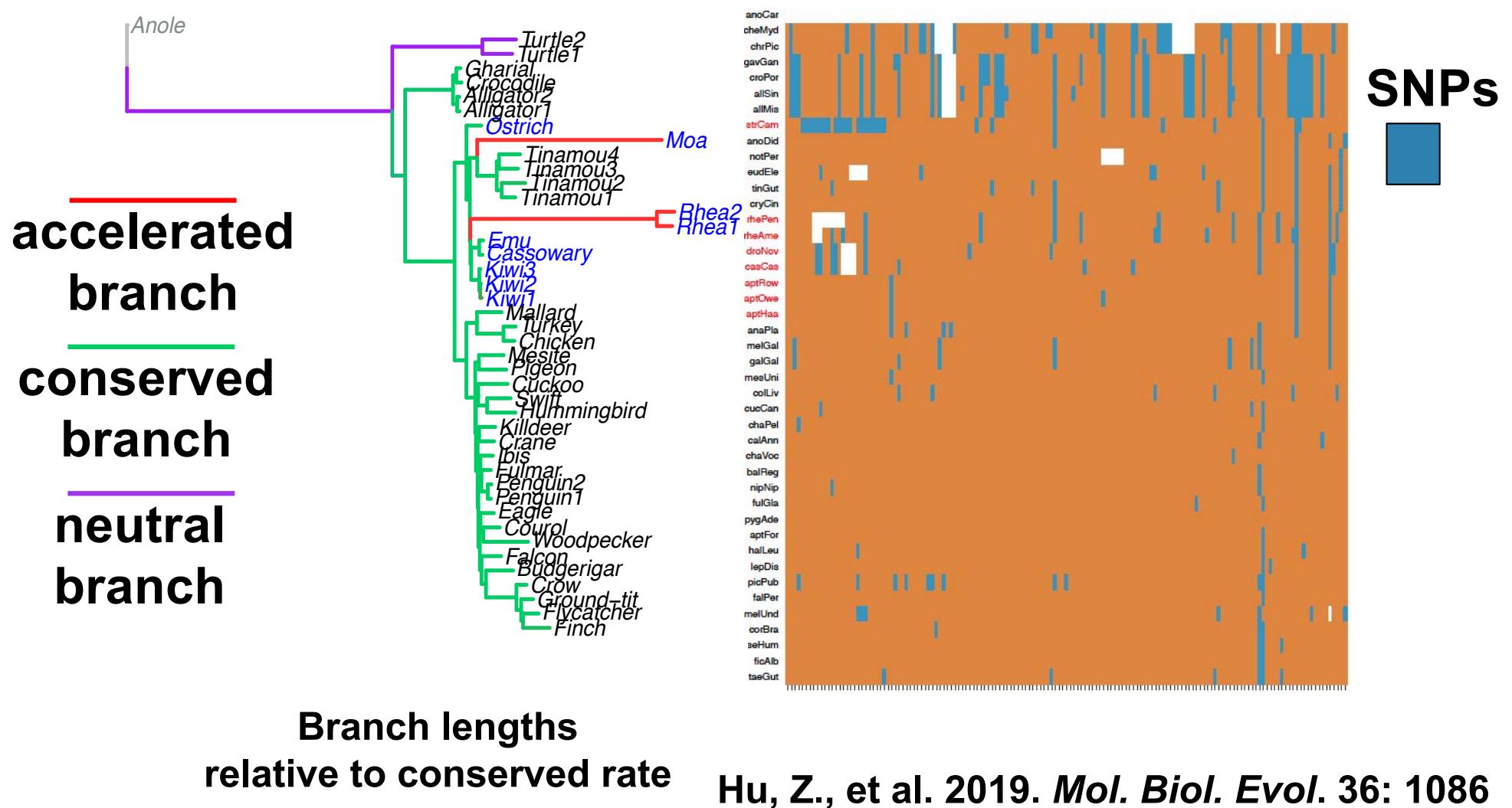
α = probability of gain of conserved state

β = probability of loss of conserved state

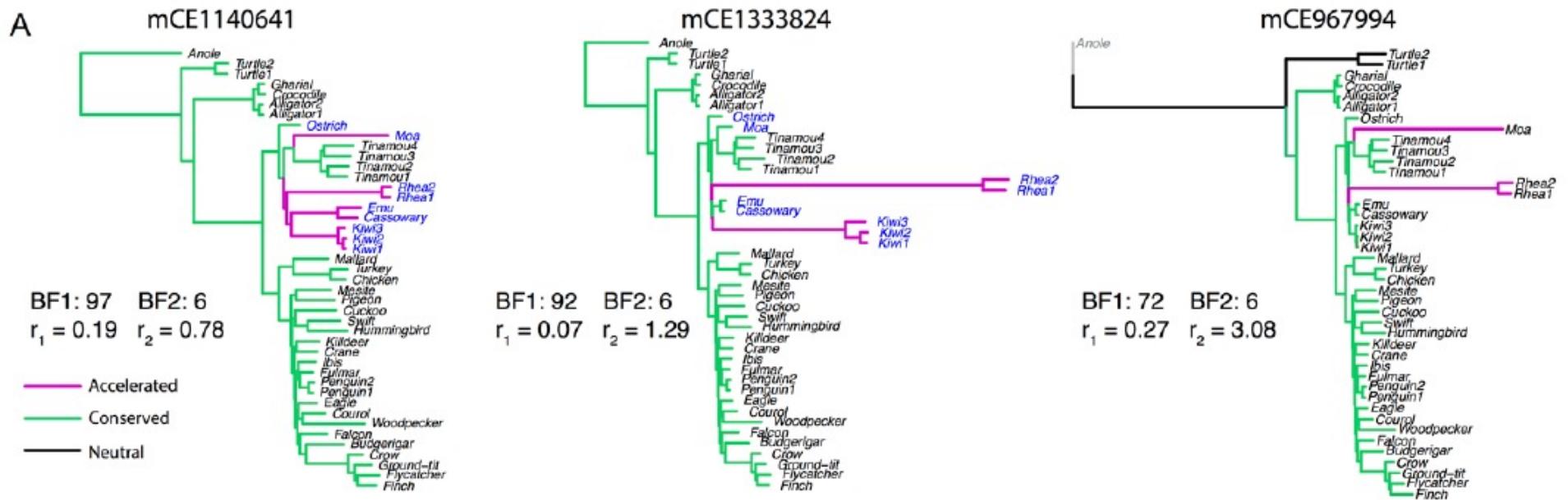
For branch s ,
$$\begin{cases} r_s = r_1, & \text{if } Z_s = 1, \text{ conserved} \\ r_s = r_0 = 1, & \text{if } Z_s = 0, \text{ background} \\ r_s = r_2, & \text{if } Z_s = 2, \text{ accelerated} \end{cases}$$

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

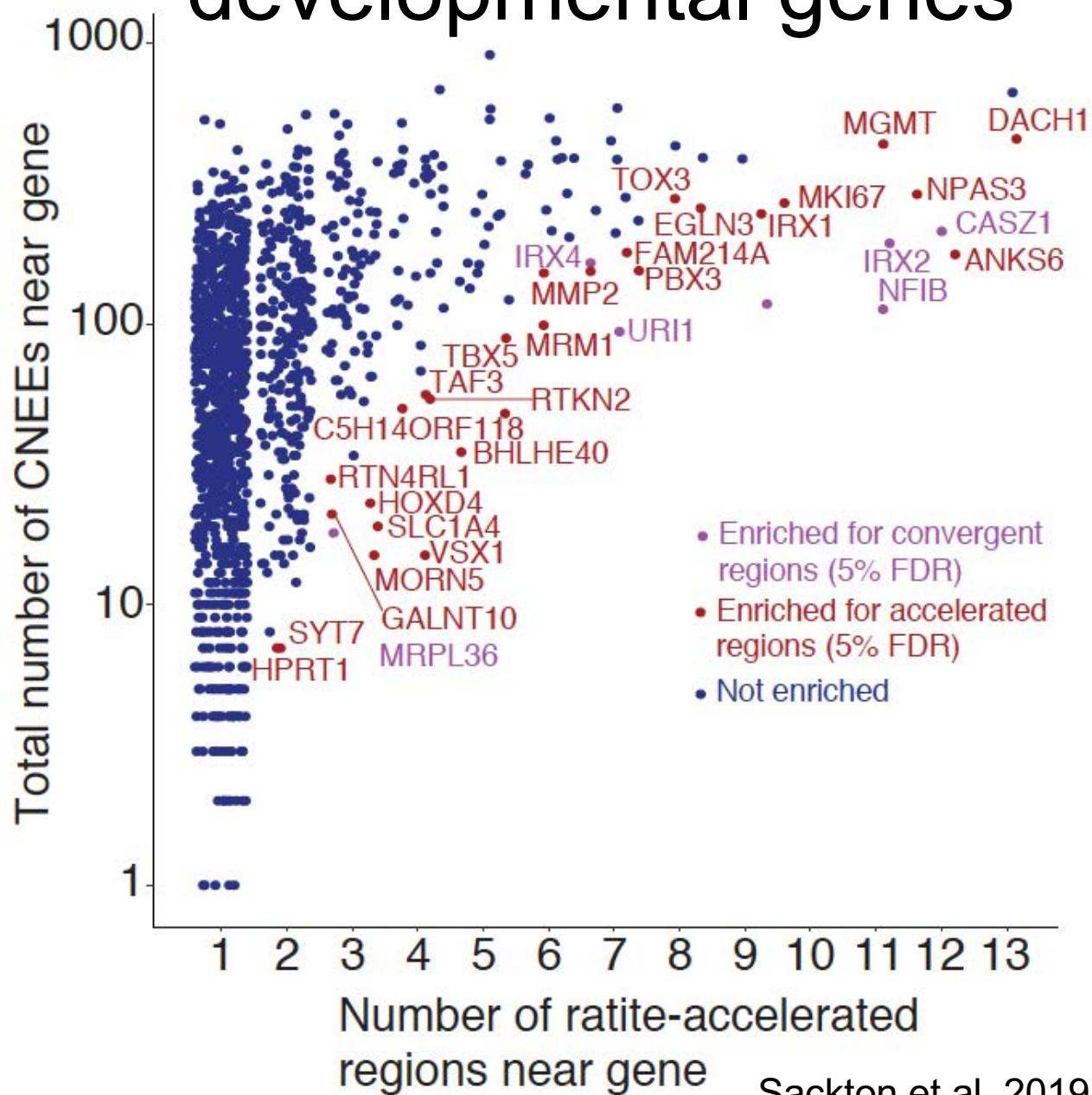
A convergently accelerated CNEE detected with a novel Bayesian method



Additional examples of convergently accelerated CNEEs



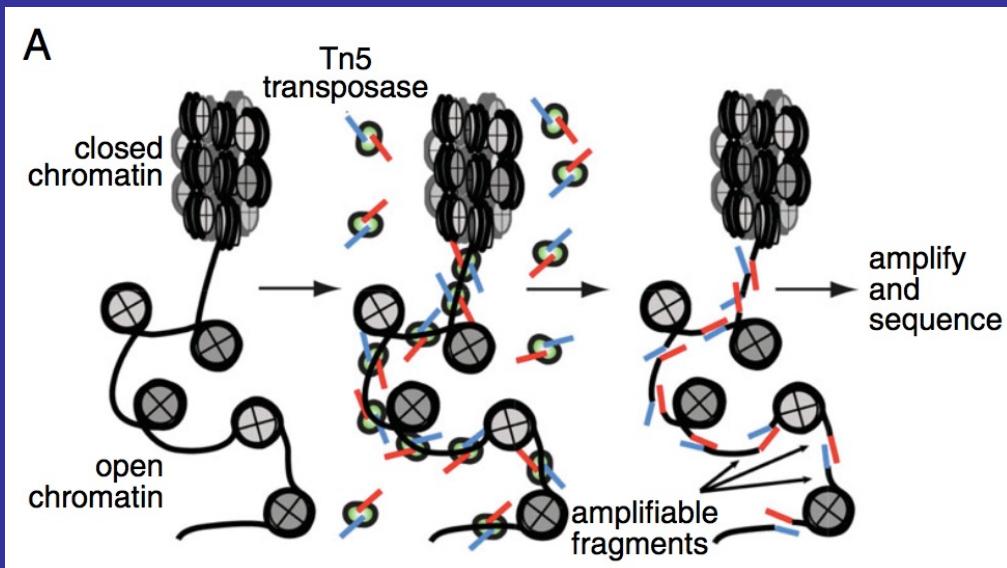
Rapid regulatory evolution near developmental genes



Assay for Transposase-Accessible Chromatin

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

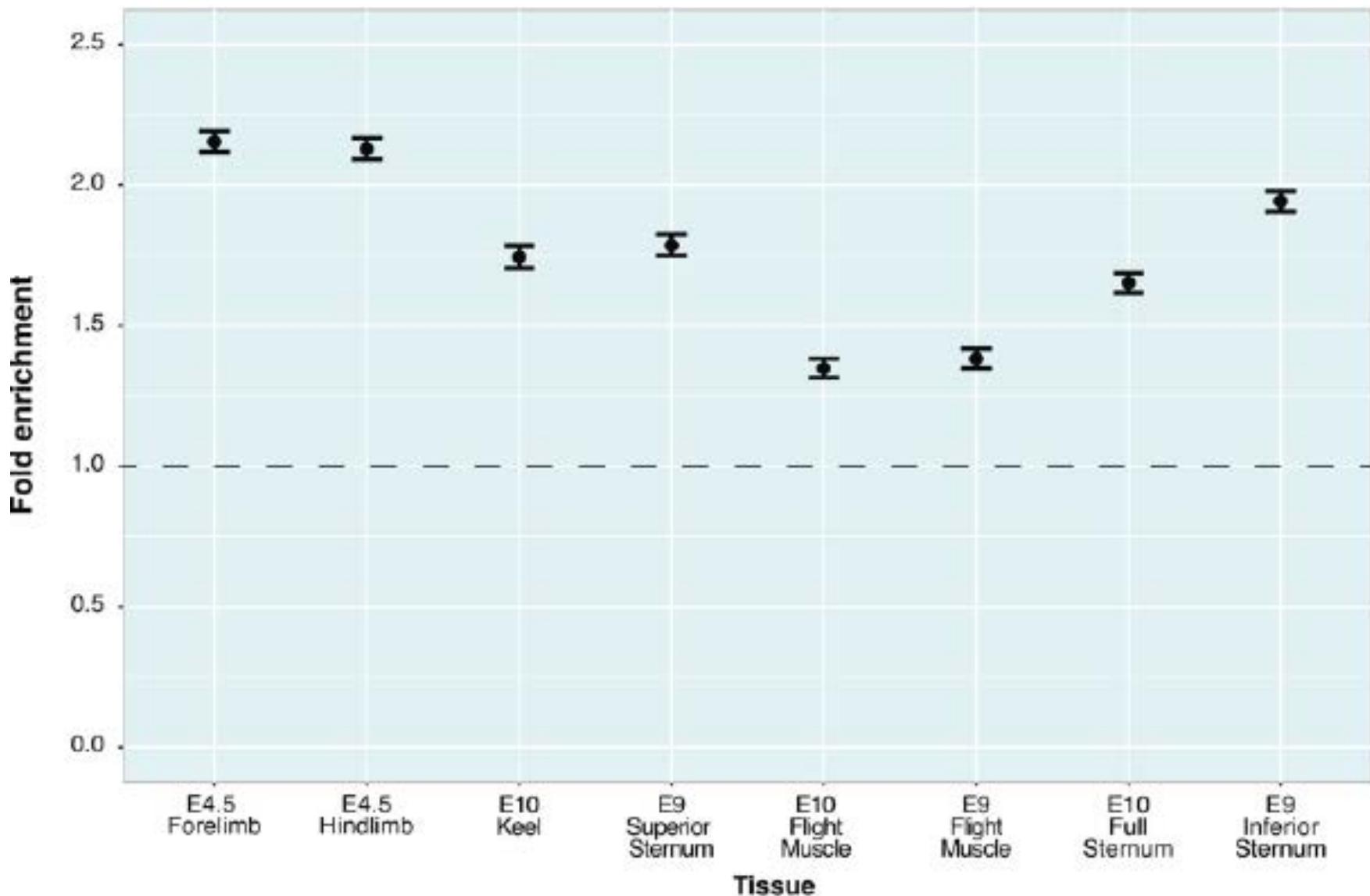
Stage HH24-25 chickens and rheas



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

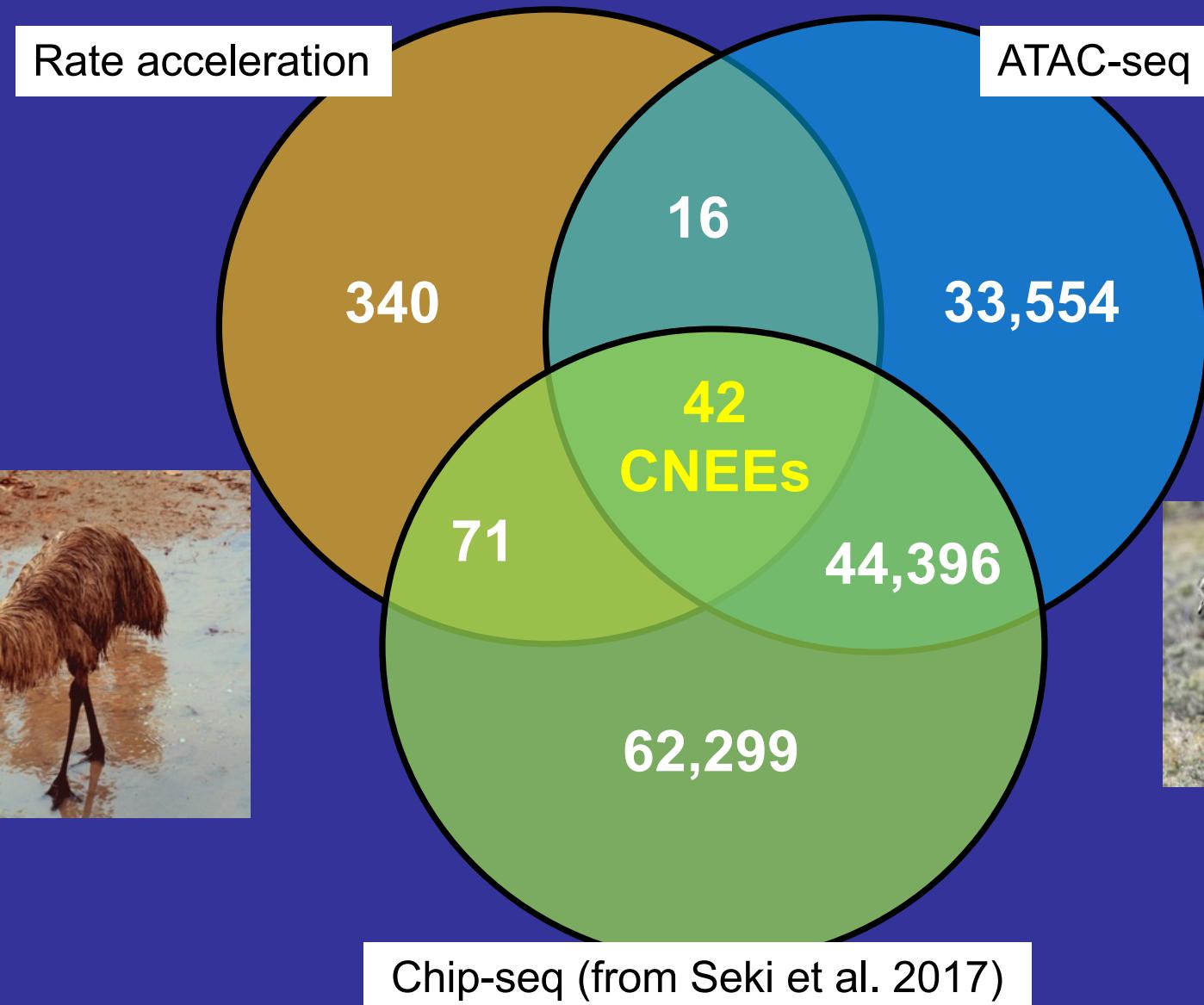


ATAC-seq: CNEEs in multiple flight-related tissues are active in early chick development

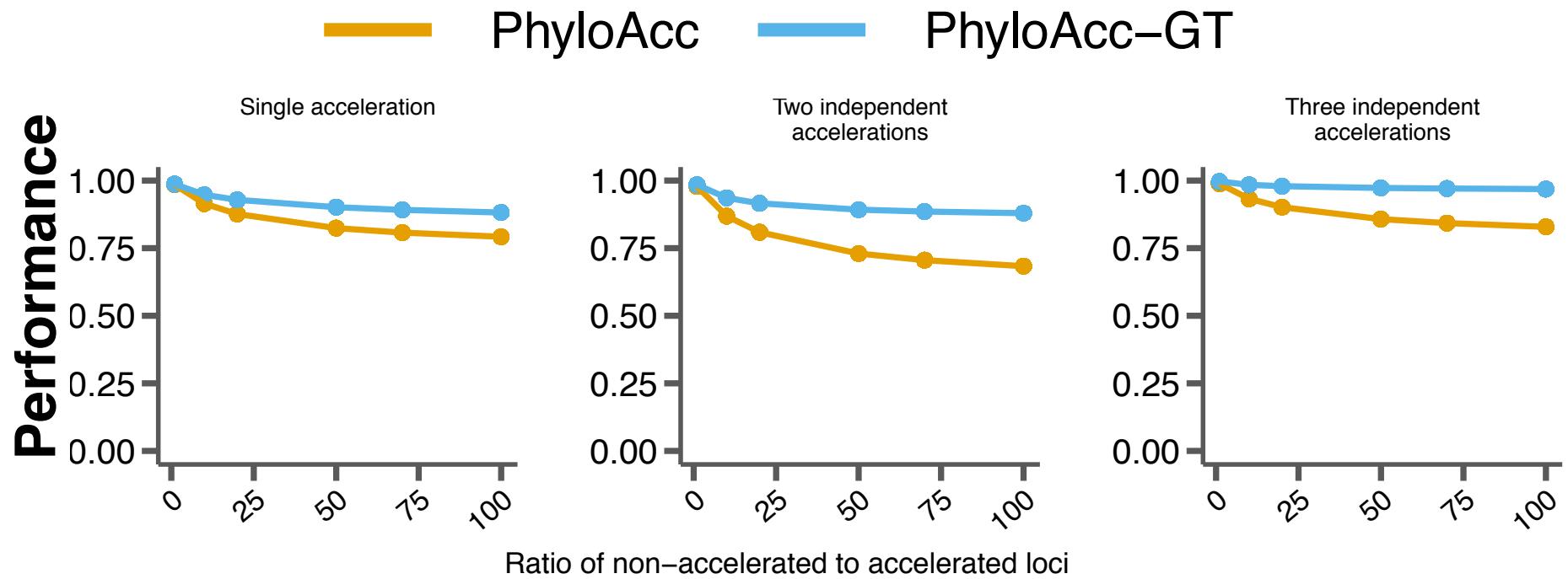


Sackton et al. 2019. *Science* 364: 74-78

Combined information from multiple sources suggests candidate enhancers for flightlessness phenotypes

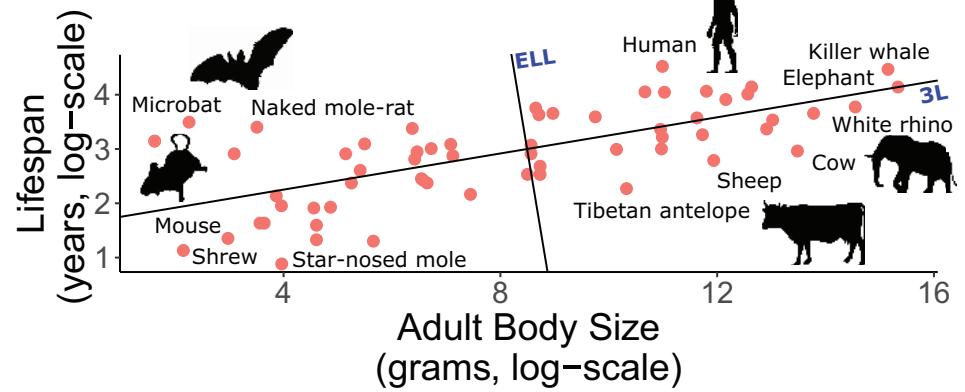
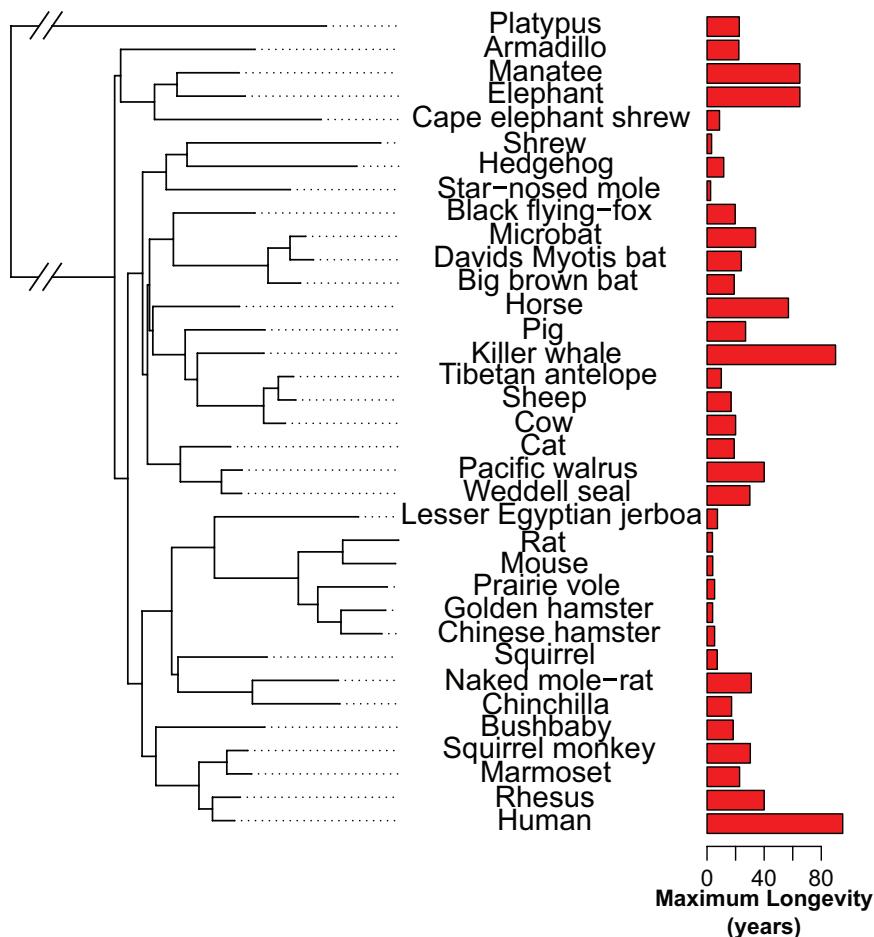


PhyloAcc-GT: Detecting accelerations with gene tree variation



PhyloAcc-C: Models connecting molecular rates and continuous traits

longevity in mammals



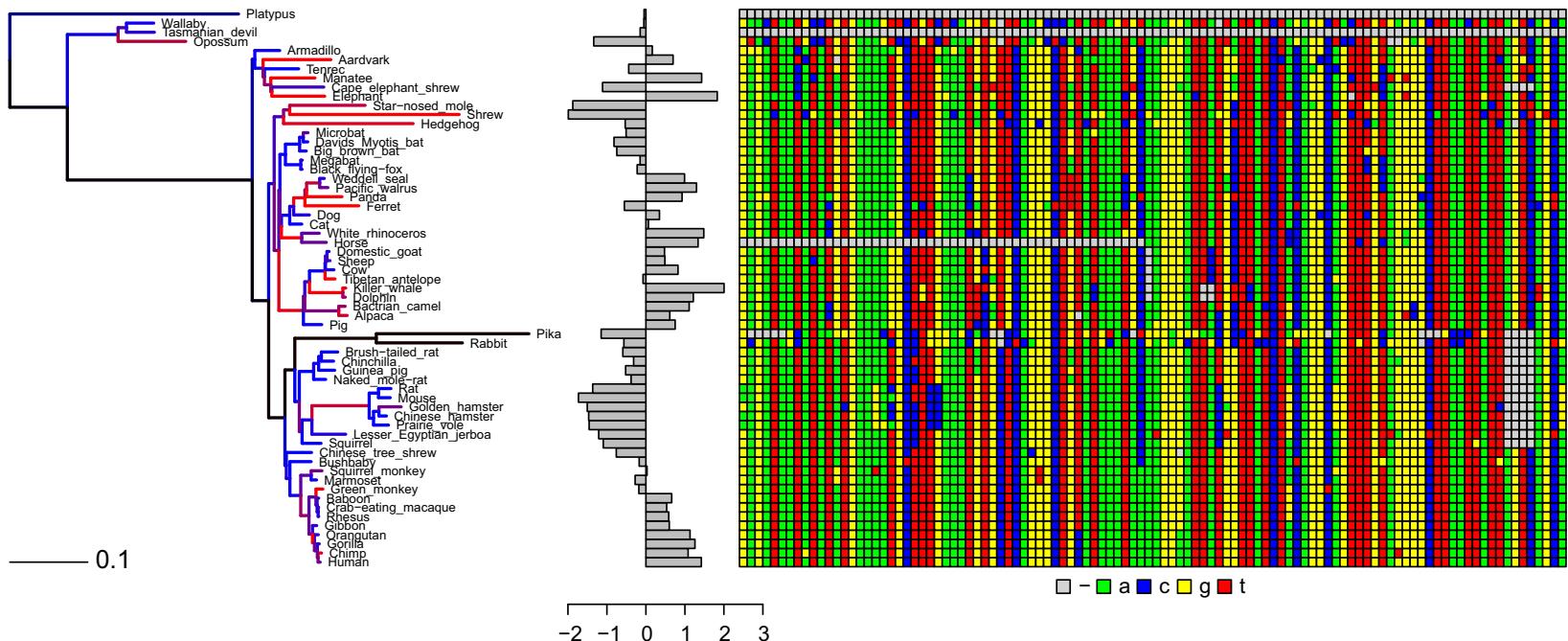
<https://github.com/phyloacc/PhyloAcc-C>

Kowalczyk et al. 2020. eLife.

Linking sequence evolution and trait data ...

A.

Alignment of conserved noncoding element VCE277691



...using rate multipliers for CNEEs and traits

Molecular rates

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

$r_1 < 1 \text{ if } Z_s = 1 \text{ conserved}$

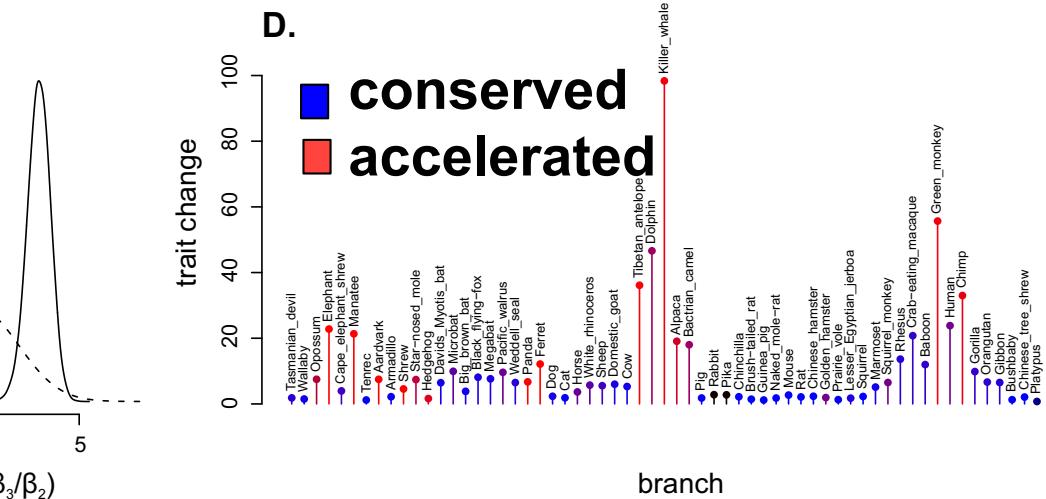
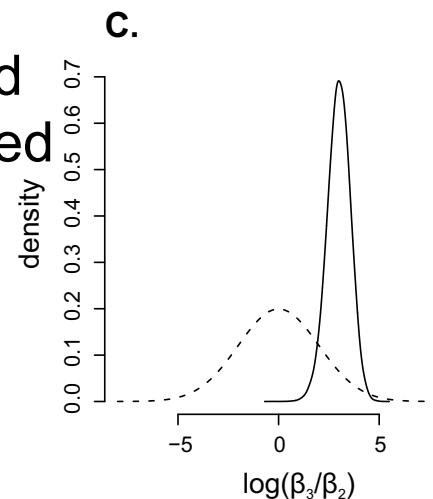
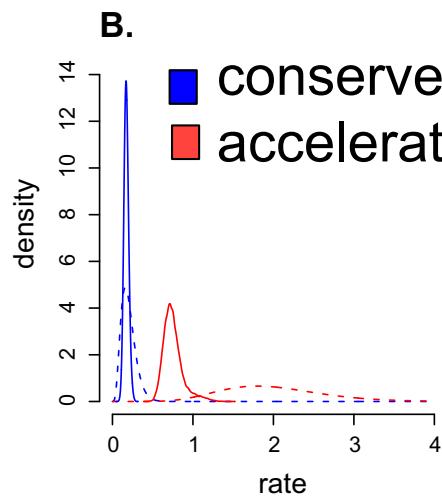
$r_2 > r_1 \text{ if } Z_s = 2 \text{ accelerated}$

Phenotypic rates

$$y_j | z_j \sim \text{Normal}(y_i, t_j v_{z_j})$$

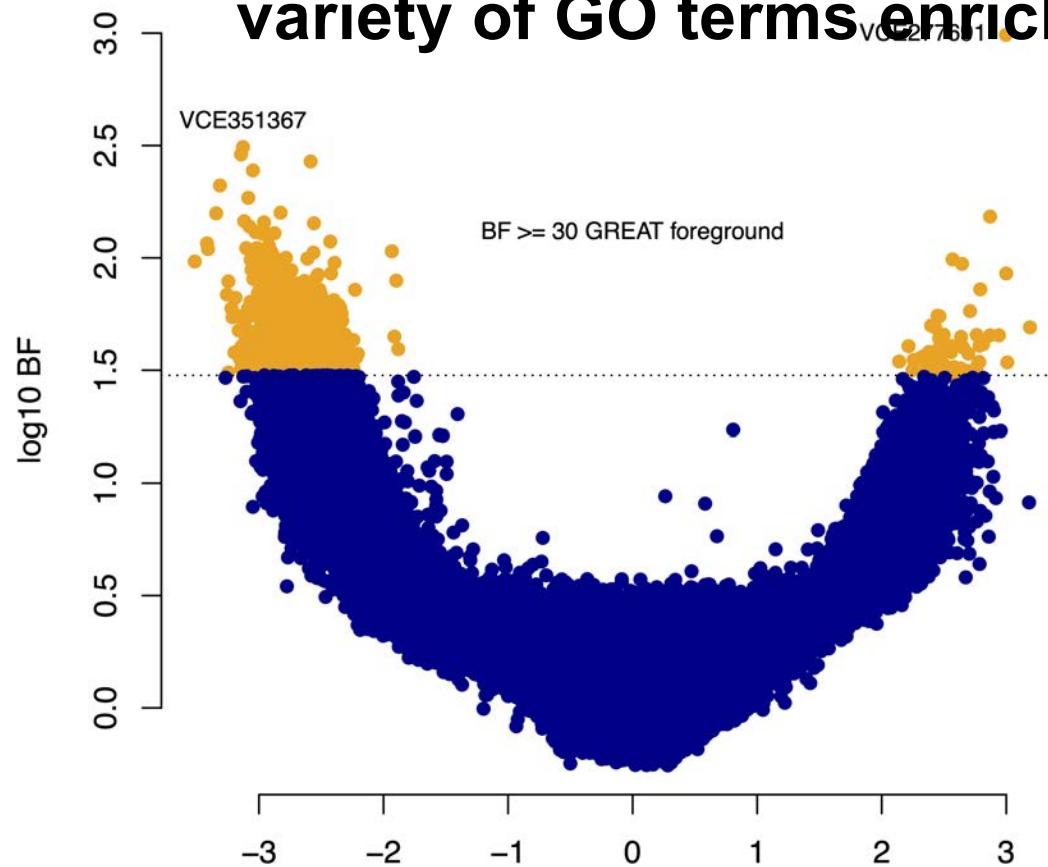
$$v_{z_j} = \begin{cases} \sigma^2 & \text{if } Z_s = 0 \text{ background} \\ \beta_2 \sigma^2 & \text{if } Z_s = 1 \text{ conserved} \\ \beta_3 \sigma^2 & \text{if } Z_s = 2 \text{ accelerated} \end{cases}$$

Results from alignment for VCE277691



CNEEs linked to longevity associated with genes with diverse functions

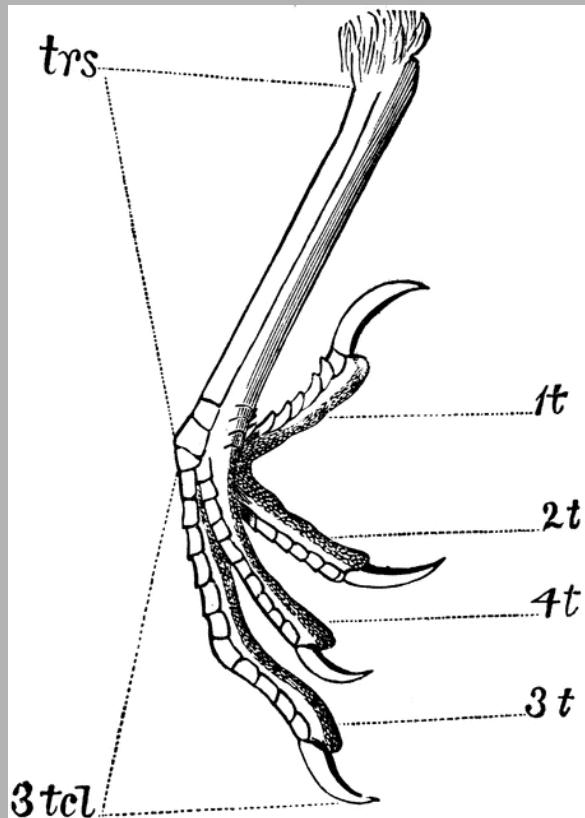
**136,859 CNEEs
1109 associated with longevity
variety of GO terms enriched**



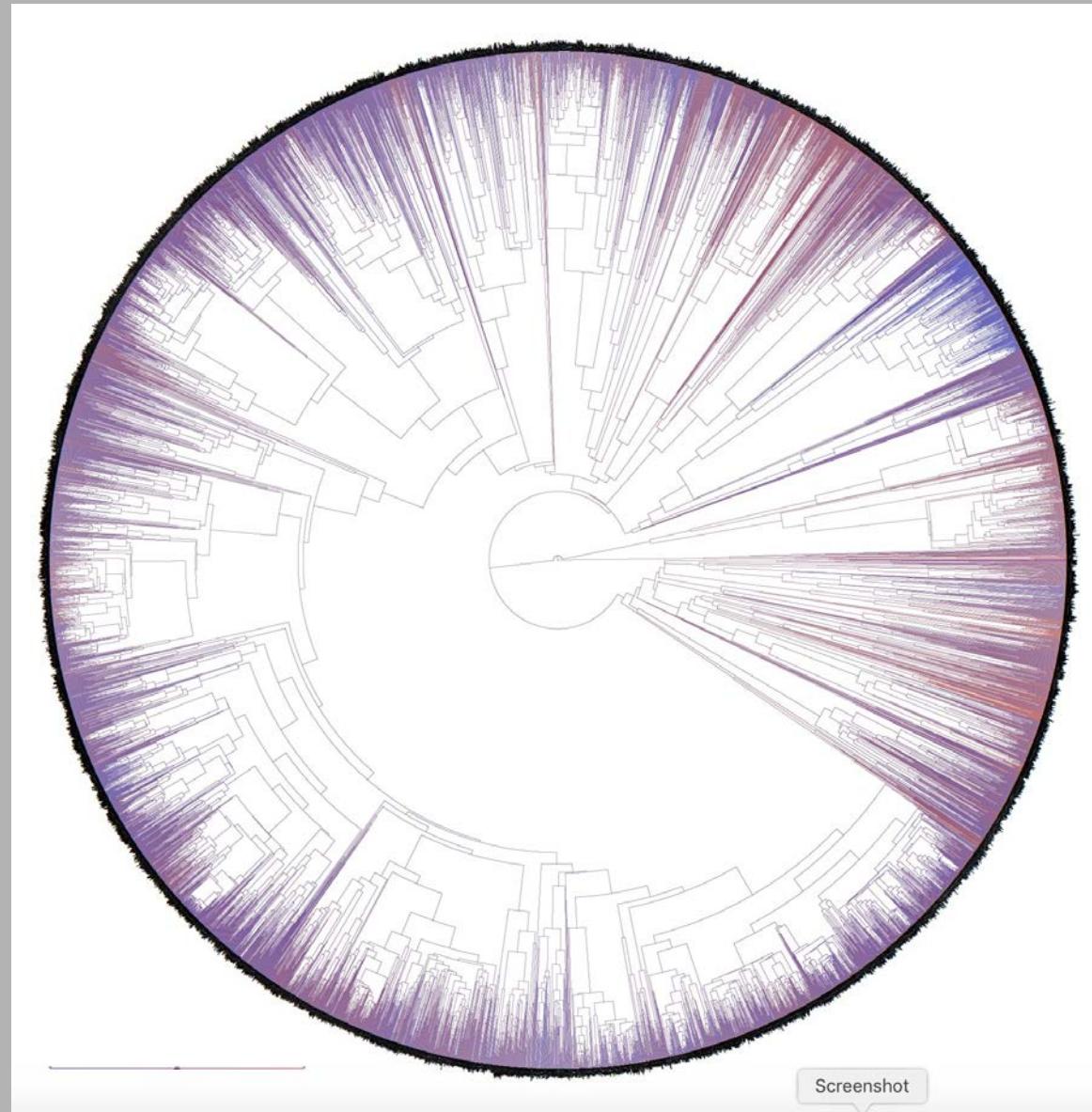
Tarsus length available for all 10,800 species of bird



Subir Shakya



short



Elliot Coues *Key to North American Birds* 1884

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

Four groups of birds exhibit shifts to shorter tarsus length

penguins



kingfishers

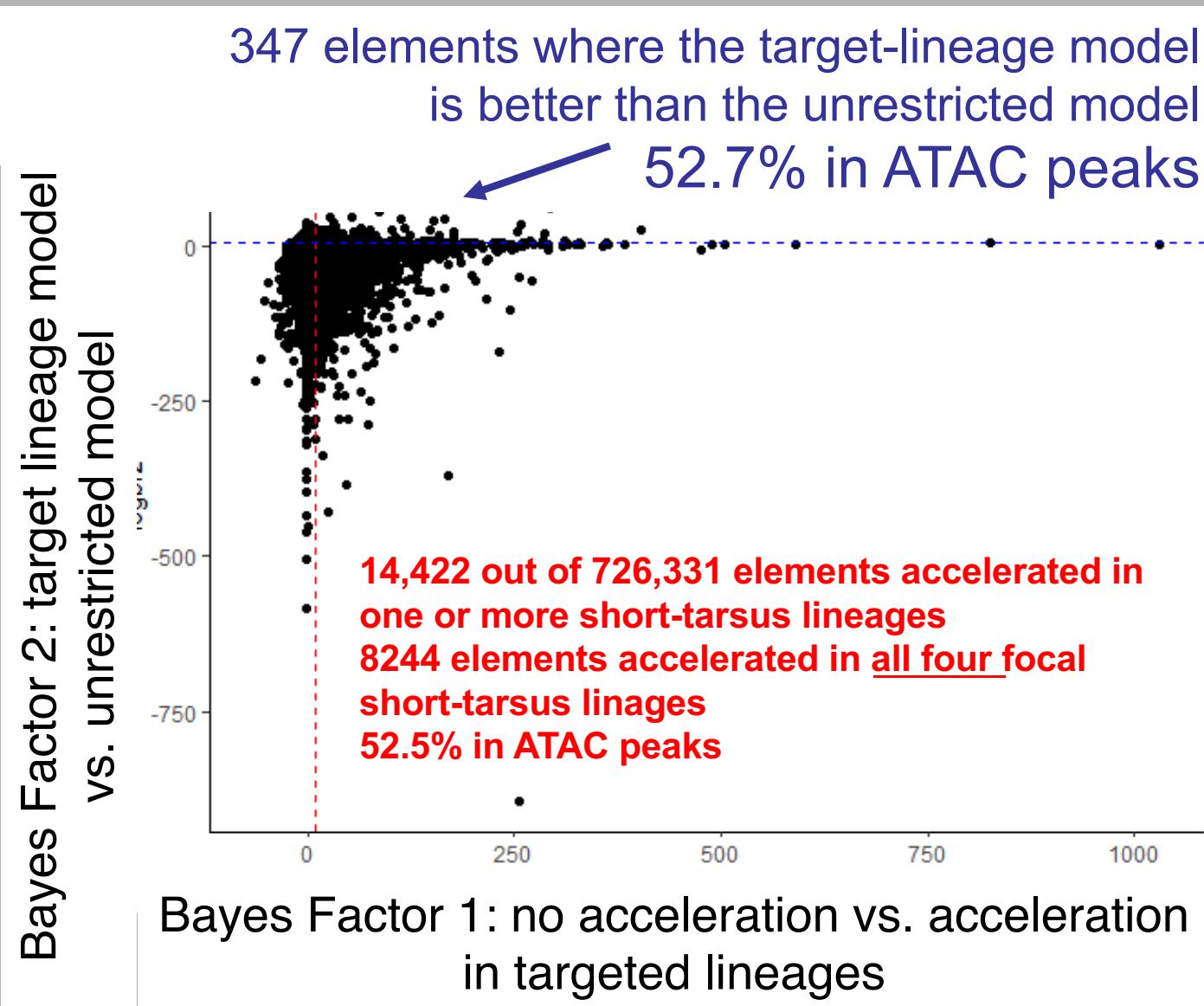


bulbuls

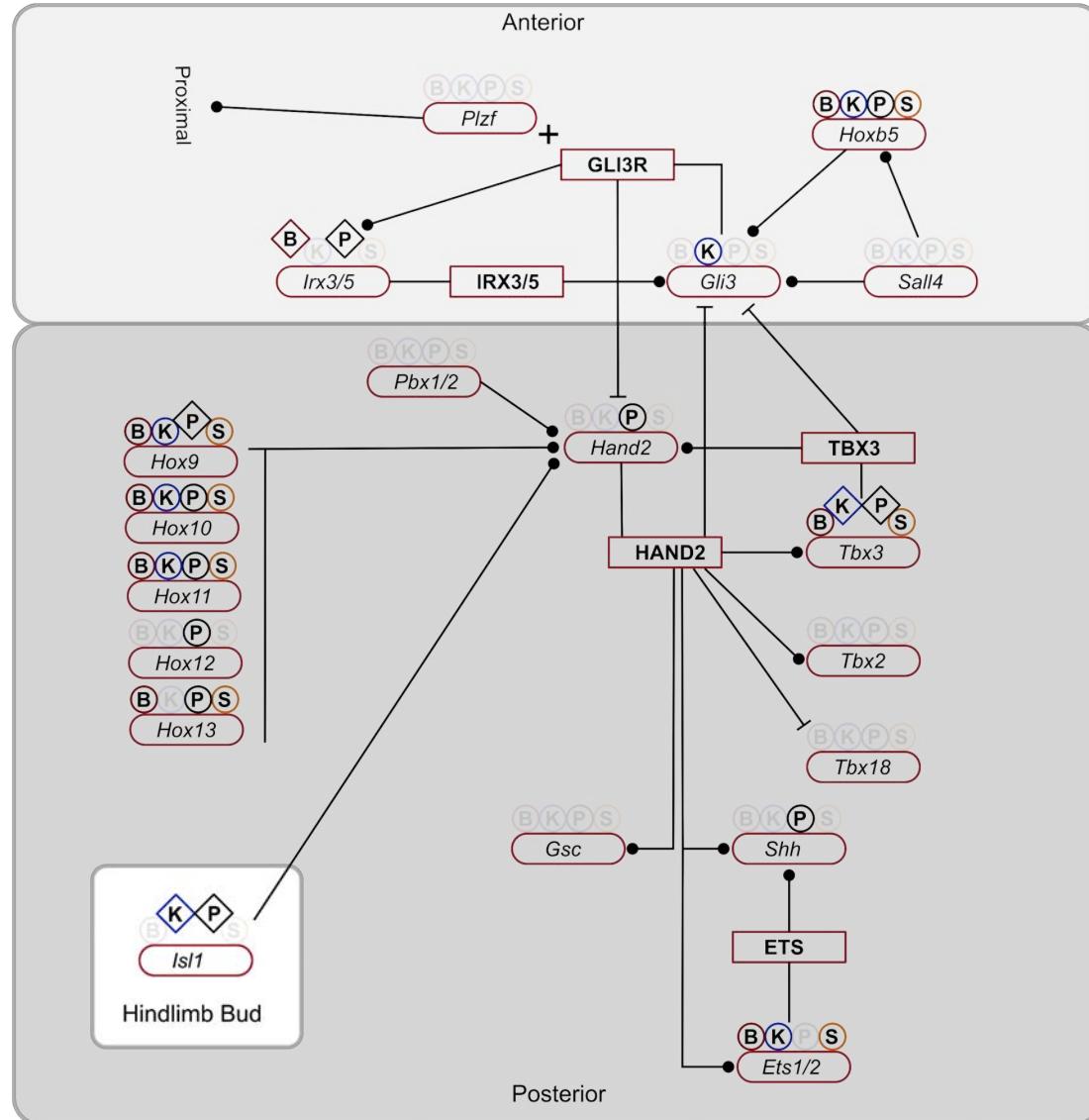


Avonet database: Tobias et al. 2022. *Ecol. Lett.*
Bayou: Uyeda and Harmon. 2014. *Syst. Biol.*

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



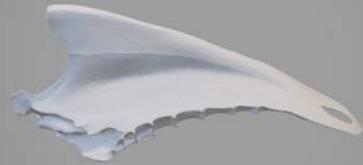
Accelerated conserved elements near genes involved in limb bud development, field positioning, AER formation and axis patterning



Limb Field Pre-patterning

- B** Bulbul
- K** Kingfisher
- P** Penguin
- S** Swallow
- Exclusively accelerated in target taxa**

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