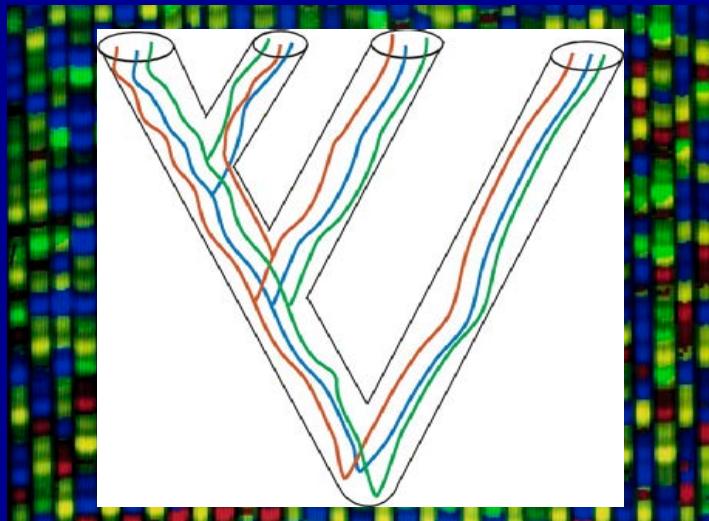
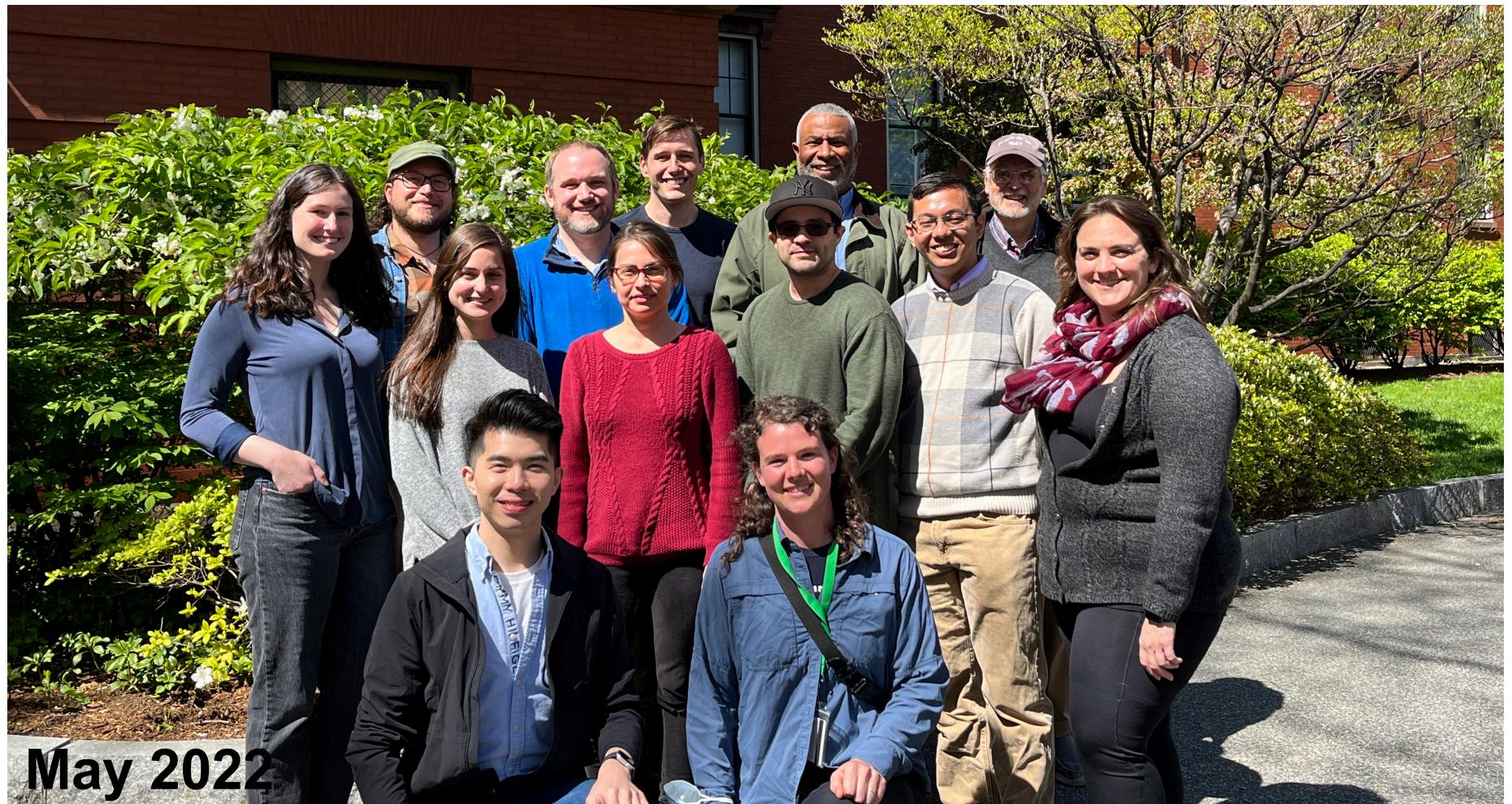
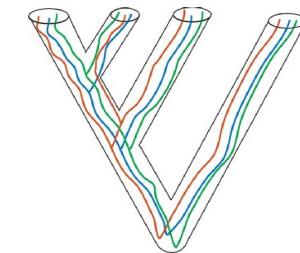
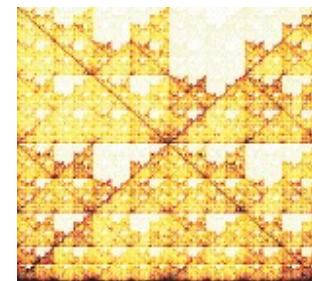
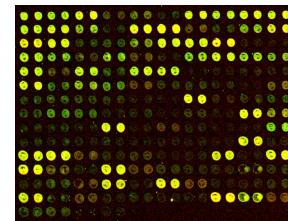
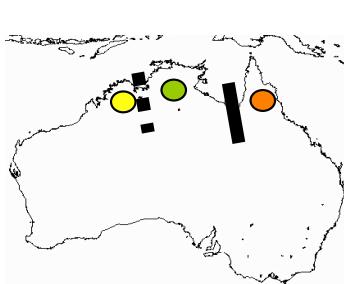


The continuum between phylogeography and phylogenetics



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!



76 days
3848 miles



@ScottVEdwards1



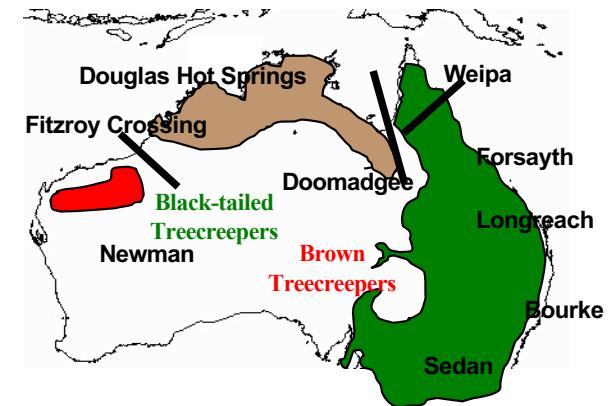
Adventure
Cycling
Association

A 1980s conundrum in evolutionary biology

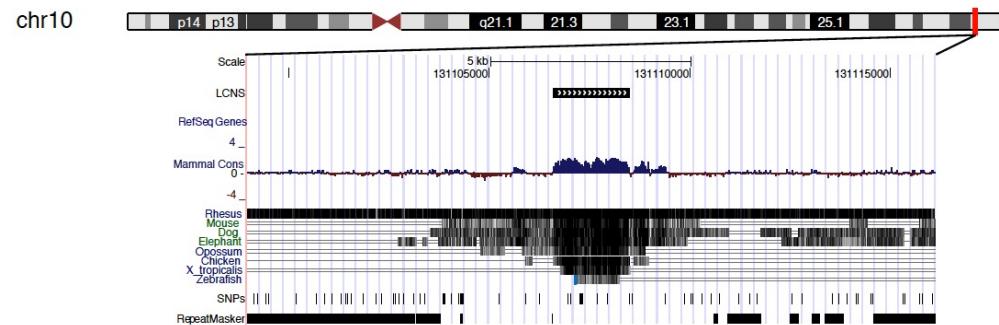


Talk overview

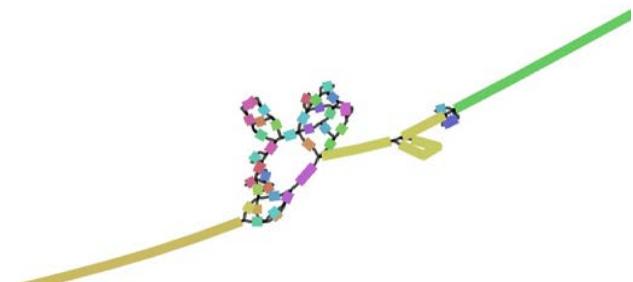
Part I: Reticulation and the emerging continuum between phylogeography and phylogenetics



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



Part III: Pangenomes: the future of evolutionary genomics



Case studies



Babblers



Honeyeaters



Treecreepers

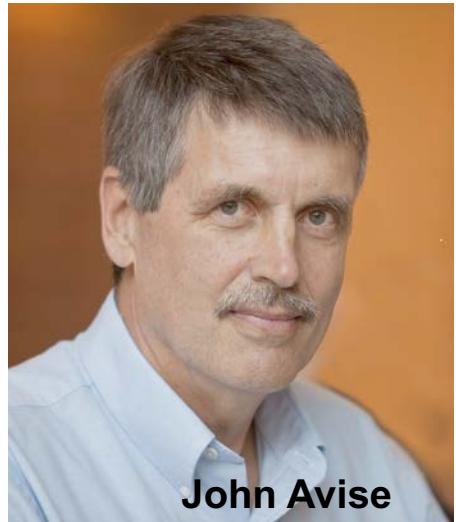


Paleognaths

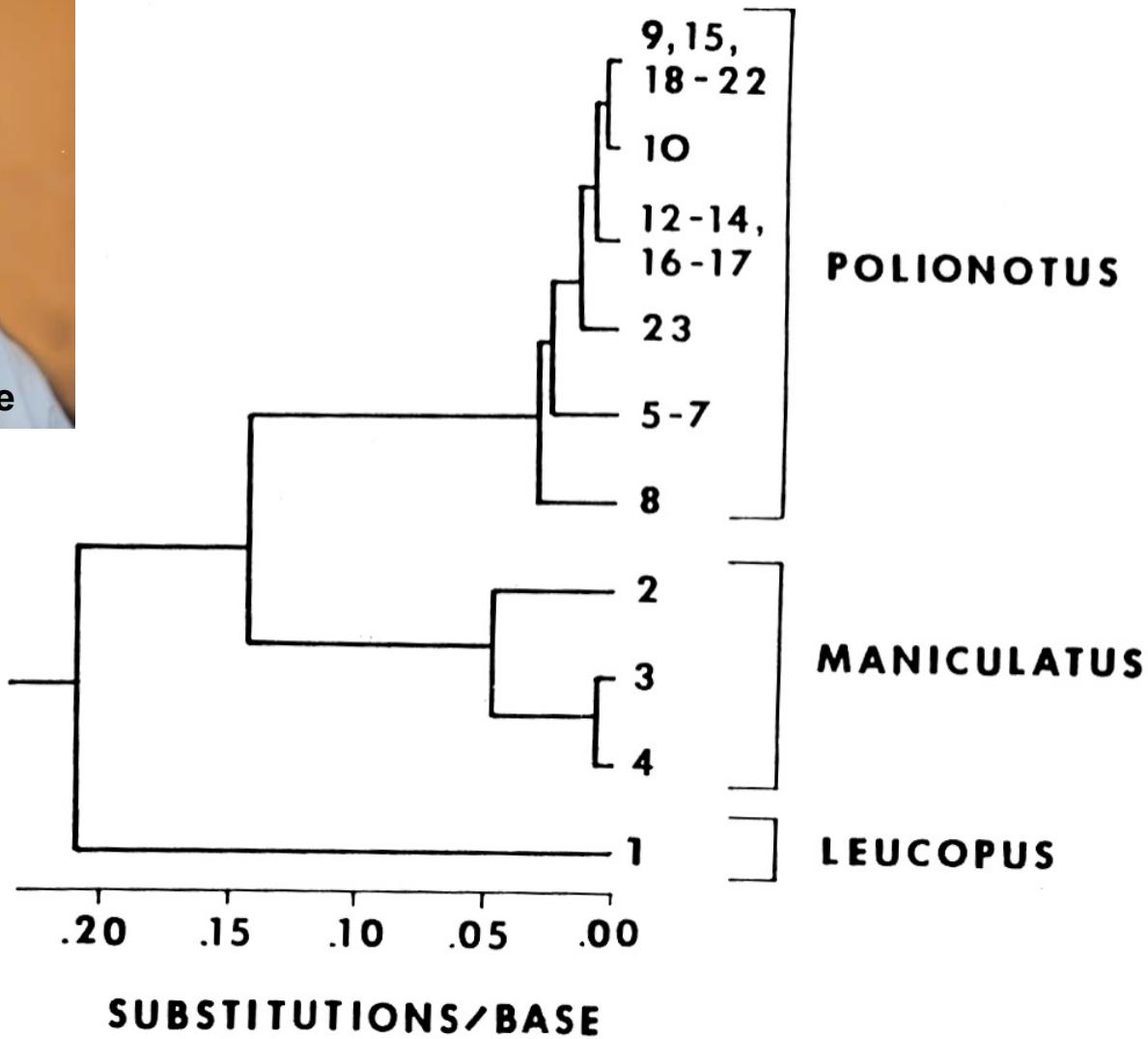


Scrub Jays

The first ‘gene tree’, 1979



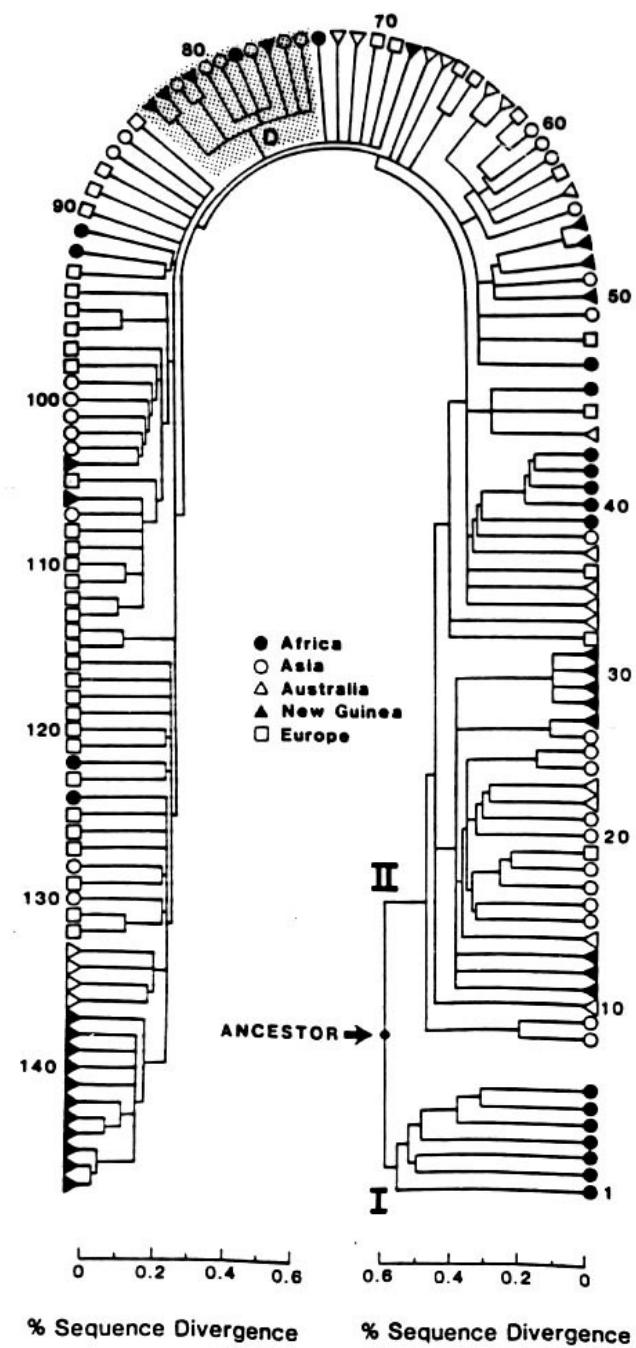
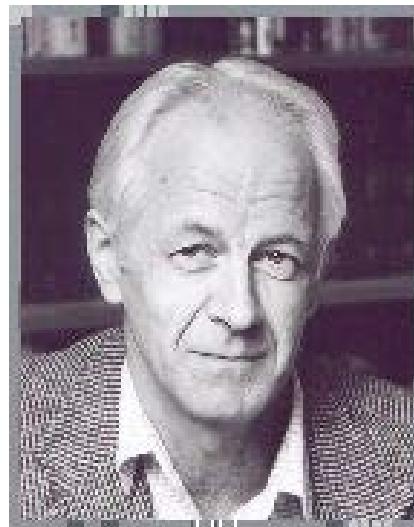
J. C. AVISE, R. A. LANSMAN AND R. O. SHADE



Mitochondrial DNA and human evolution

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

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



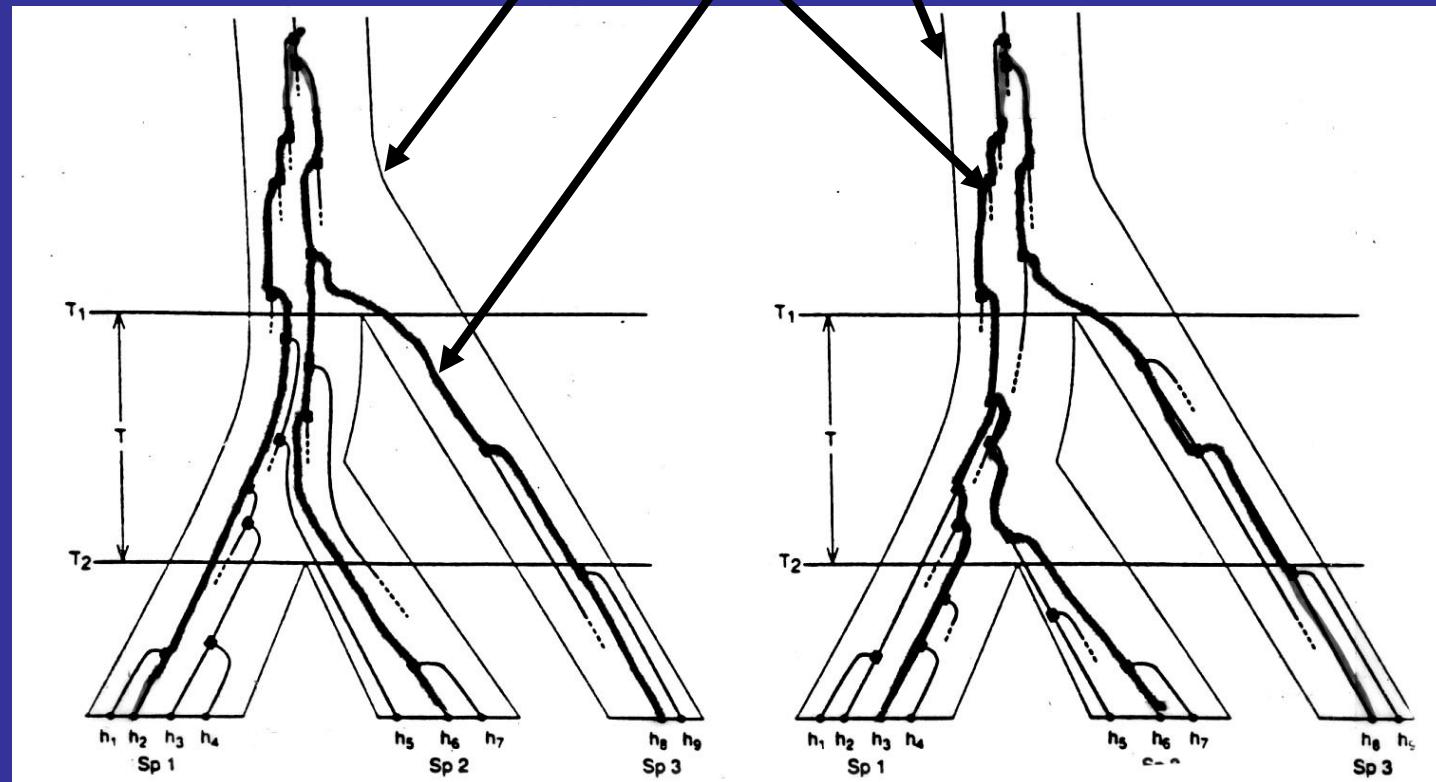
Stochastic gene tree variation during rapid radiations

“incomplete lineage
sorting”

Species tree

Gene tree

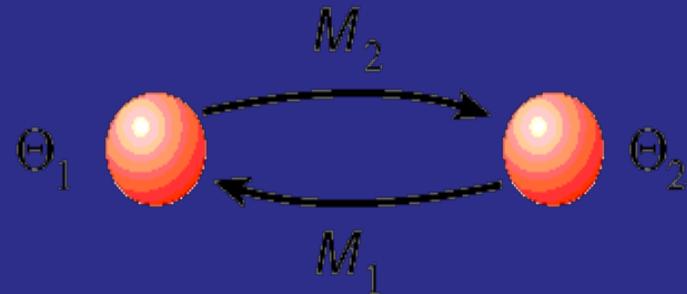
“deep coalescence”



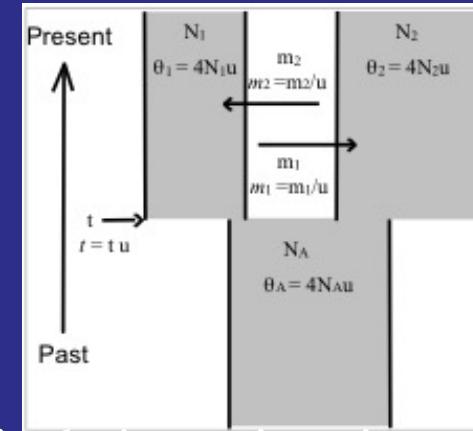
↑
 $2N_e$
↓
 $T/2N_e$
↓

Multilocus models in phylogeography

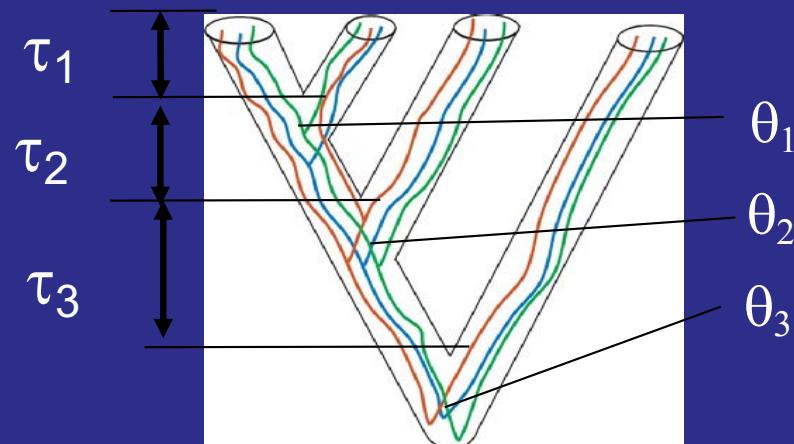
Population size: $\theta = 4N\mu$ Divergence time: $\tau = \mu t$ Gene flow: $M = m/\mu$



MIGRATE: Beerli 2006 *Bioinformatics*



IM: Hey and Nielsen 2004 *Genetics*



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

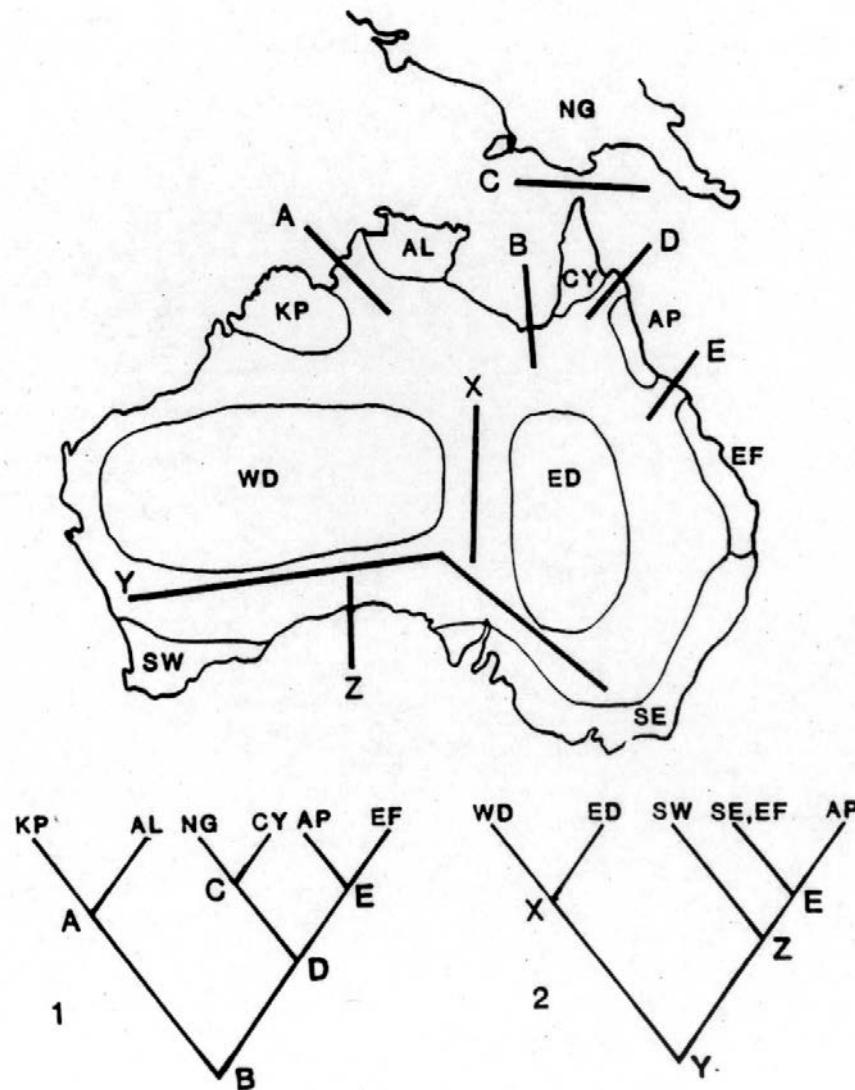
BEAST*: Heled and Drummond. 2010. *Mol. Biol. Evol.*

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



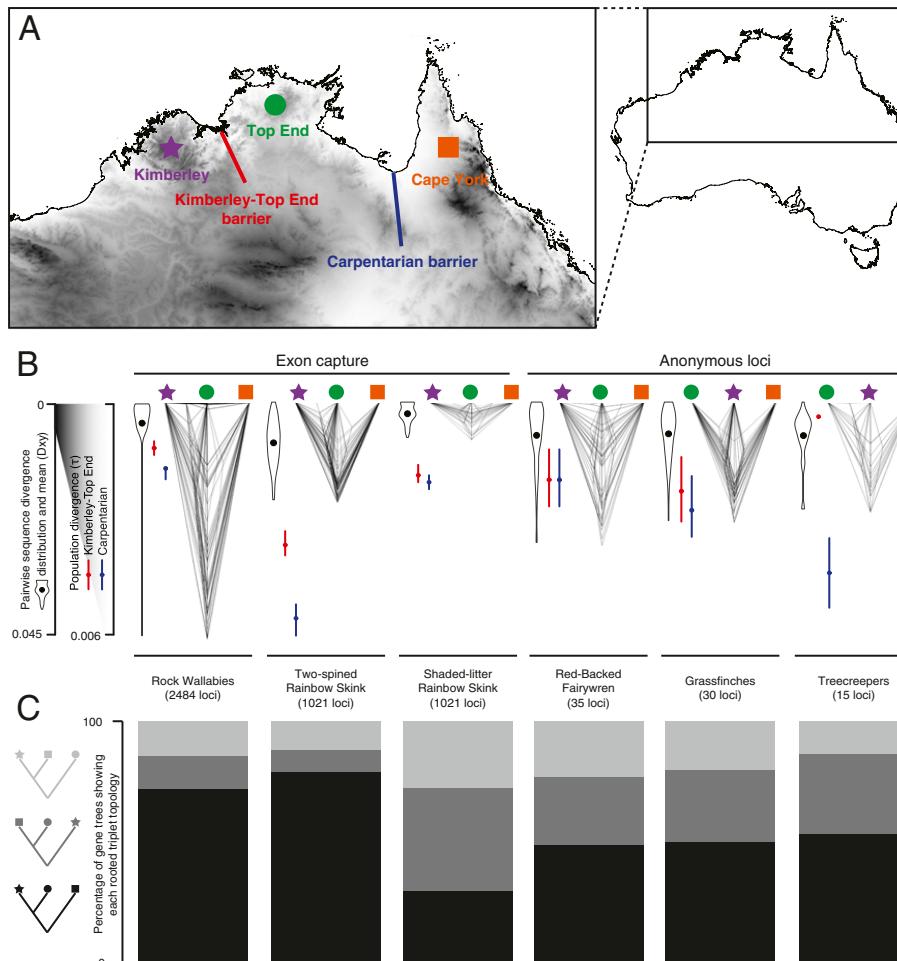
Stinking hot

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



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}

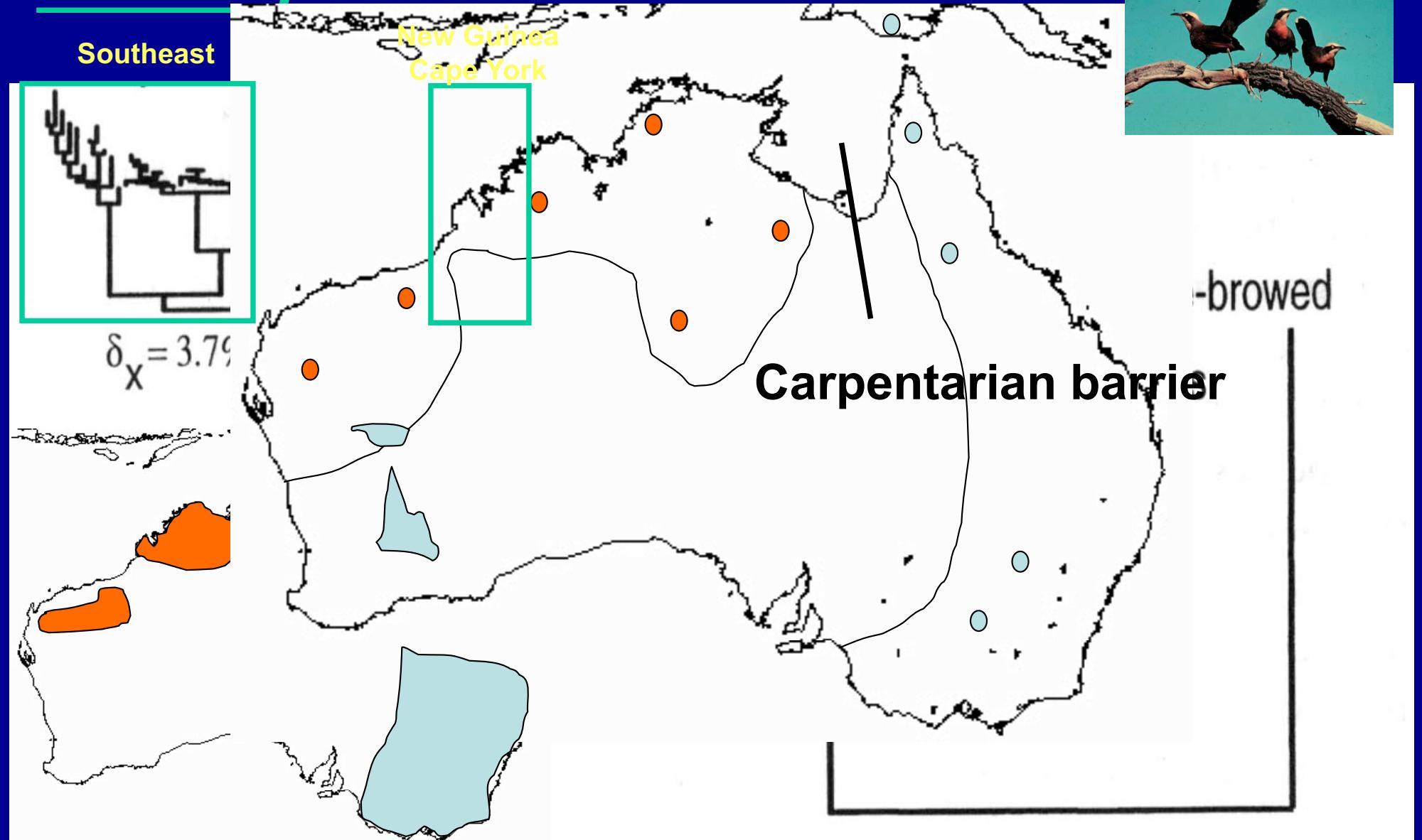


Edwards et al. 2016. PNAS 113: 8025-8032

Mitochondrial genetic differences between red-bellied

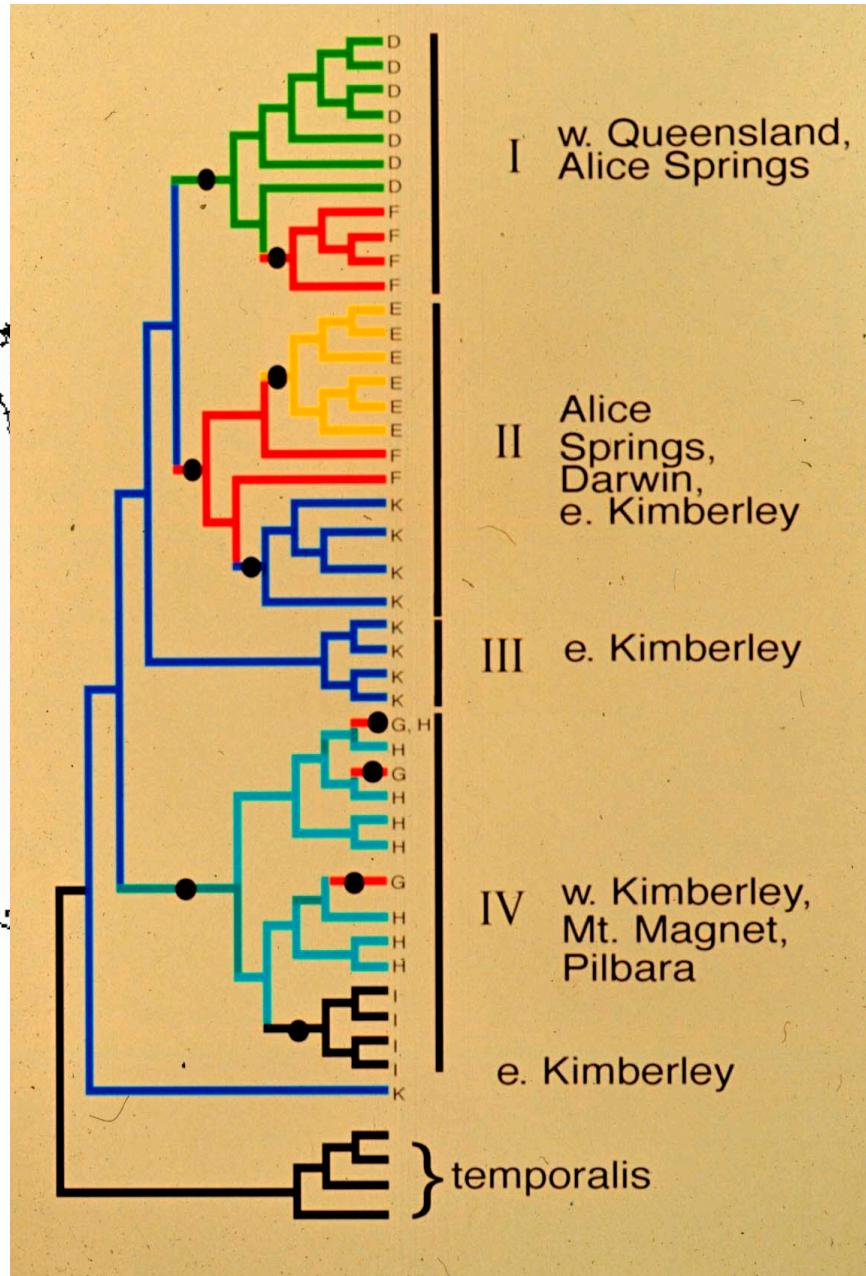
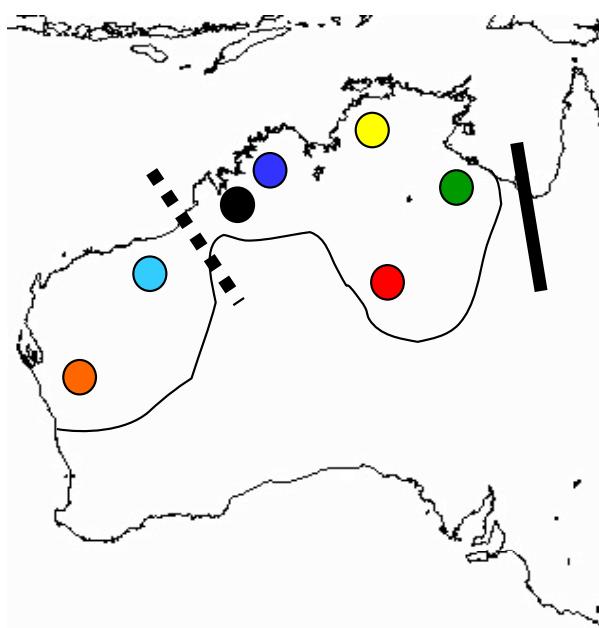
temporalis

rubeculus



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

Gene flow erodes population monophyly



I w. Queensland,
Alice Springs

II Alice
Springs,
Darwin,
e. Kimberley

III e. Kimberley

IV w. Kimberley,
Mt. Magnet,
Pilbara

e. Kimberley

} temporalis



- 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

Treecreepers (*Climacteris*)

Black-tailed treecreeper

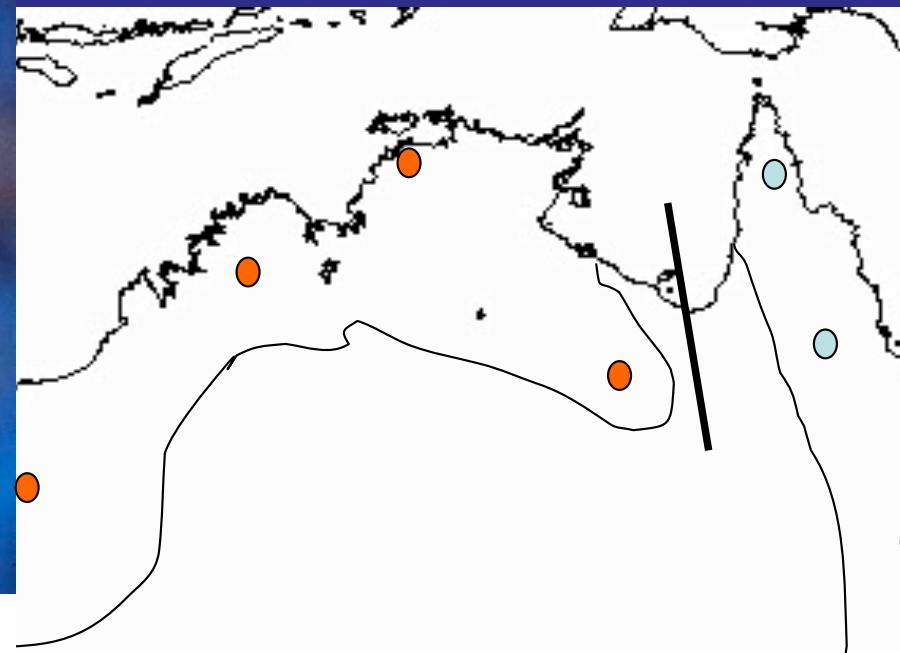


C. melanura

Brown treecreeper



C. picumnus



12 gene trees in Australian treecreepers

Brown Treecreeper
(eastern) lineages

Black-tailed (western)
Treecreeper lineages

AL3

AL5

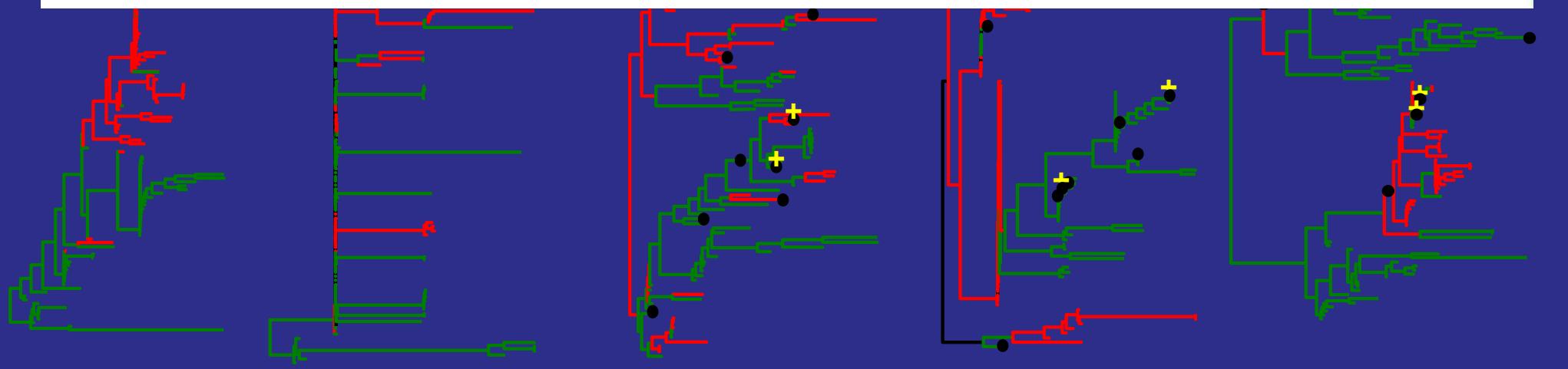
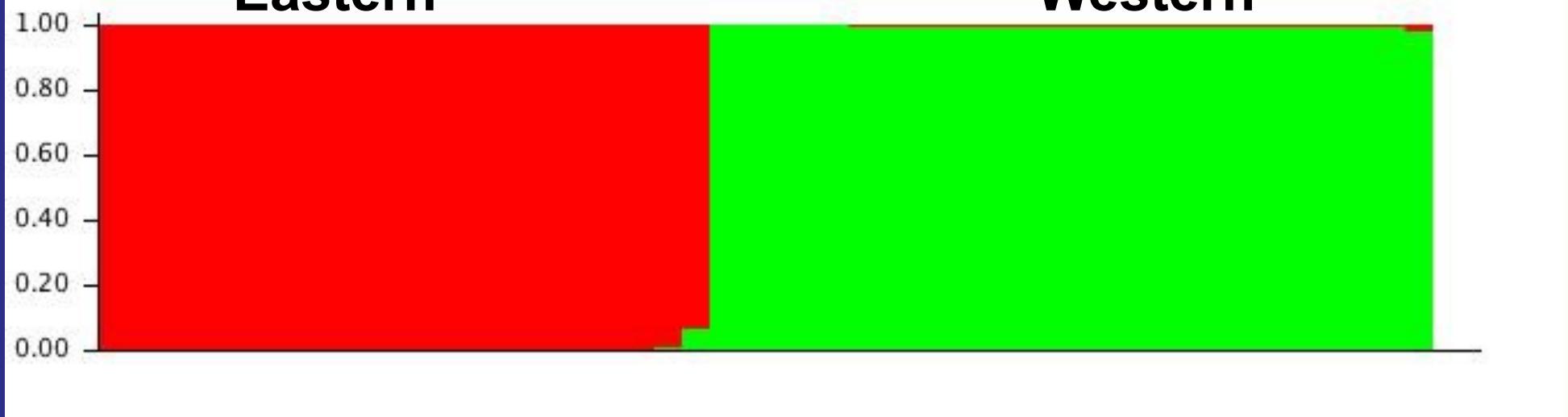
AL7

AL14

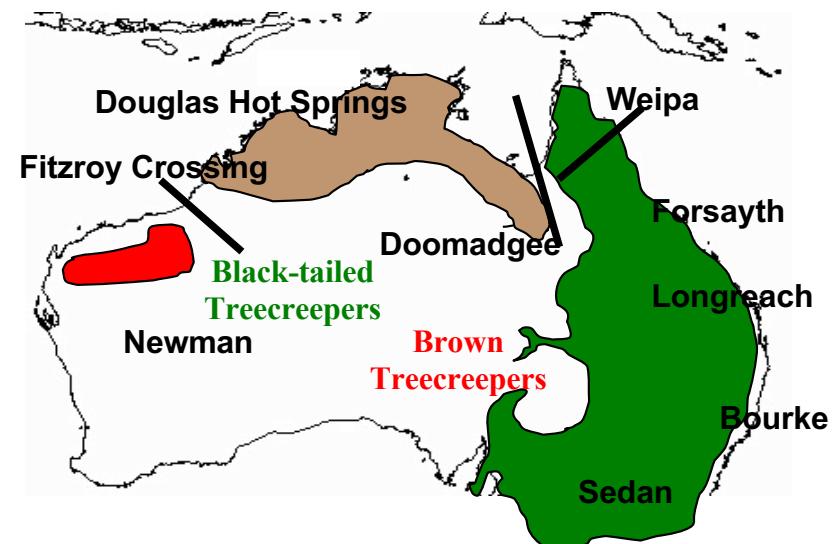
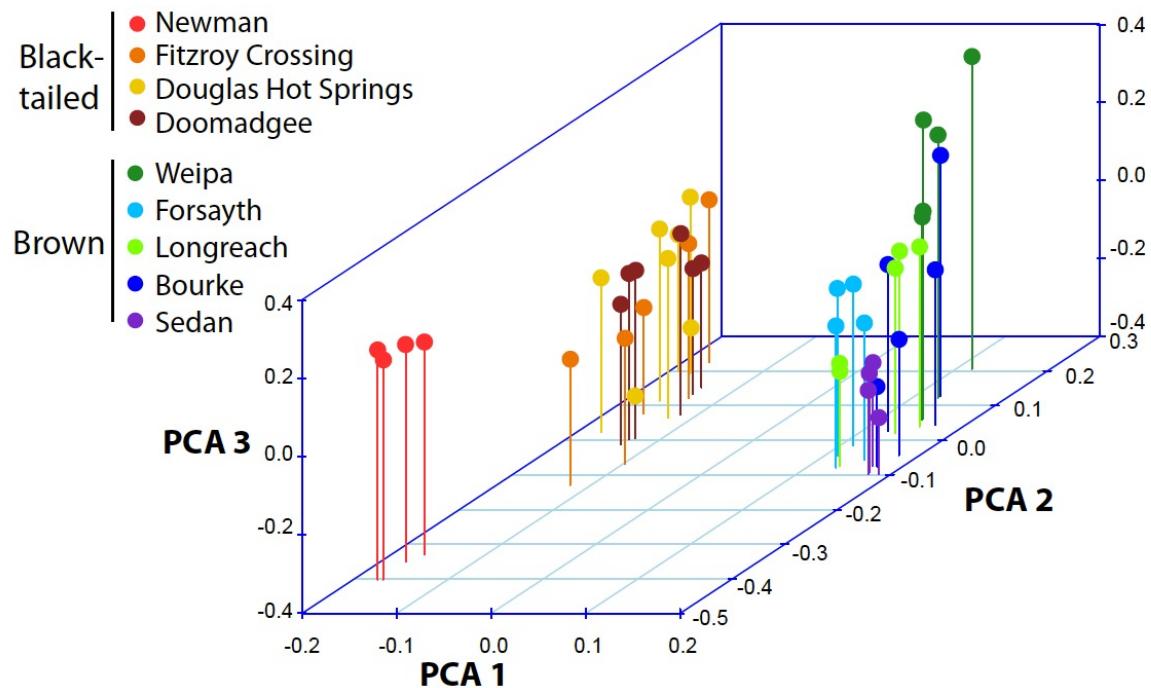
AL16

Eastern

Western

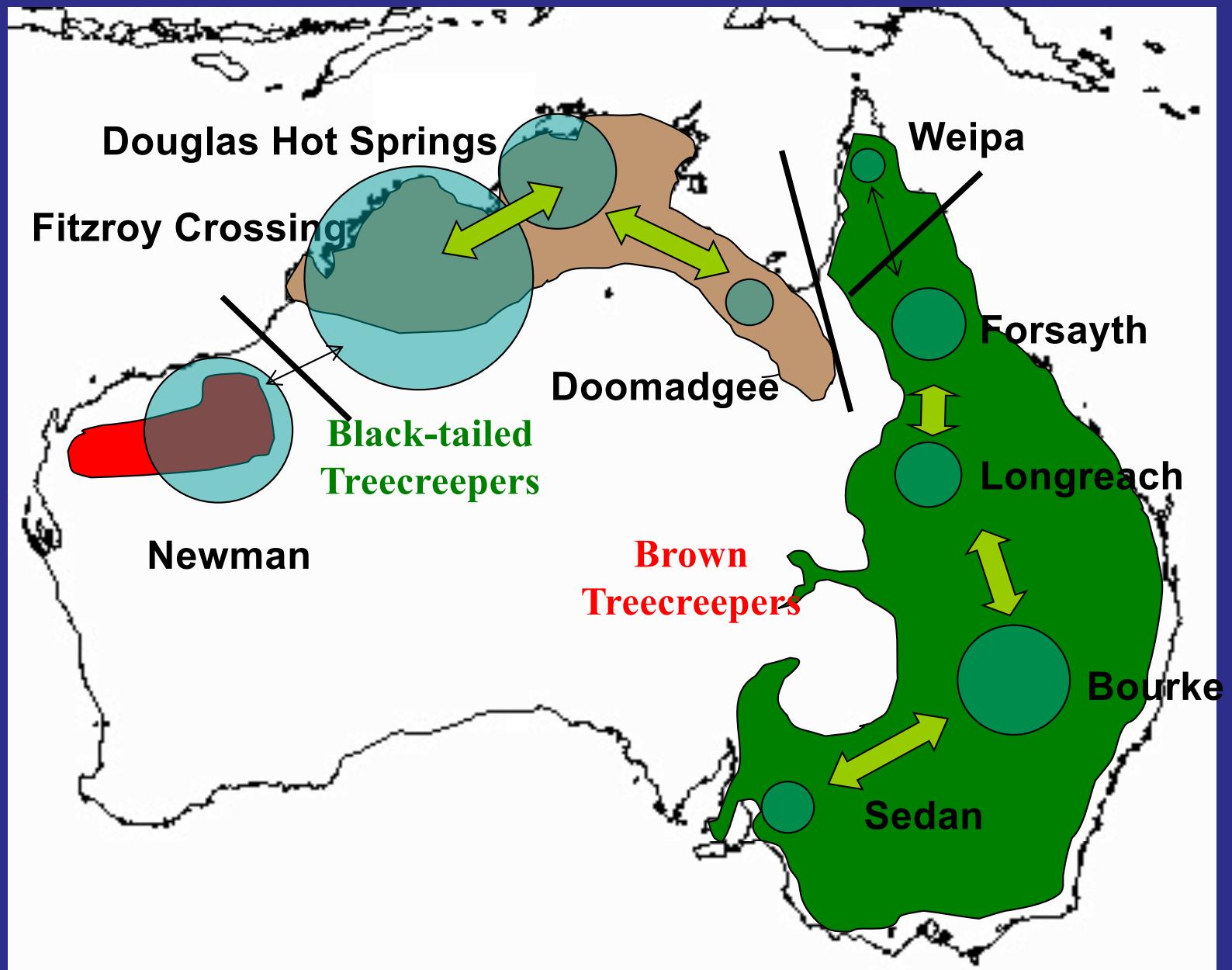


Correspondence between PCA and geography



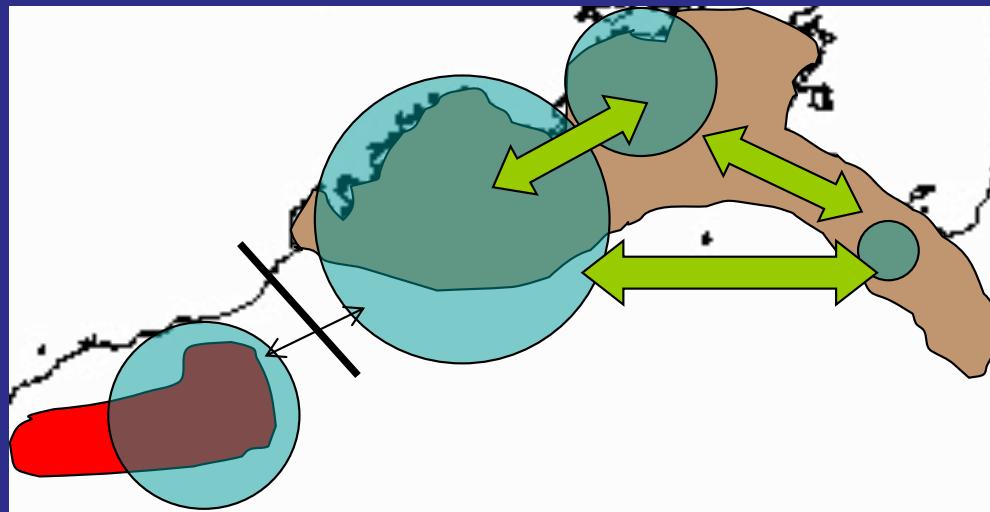
Treecreeper populations are connected but variable in size (MIGRATE)

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

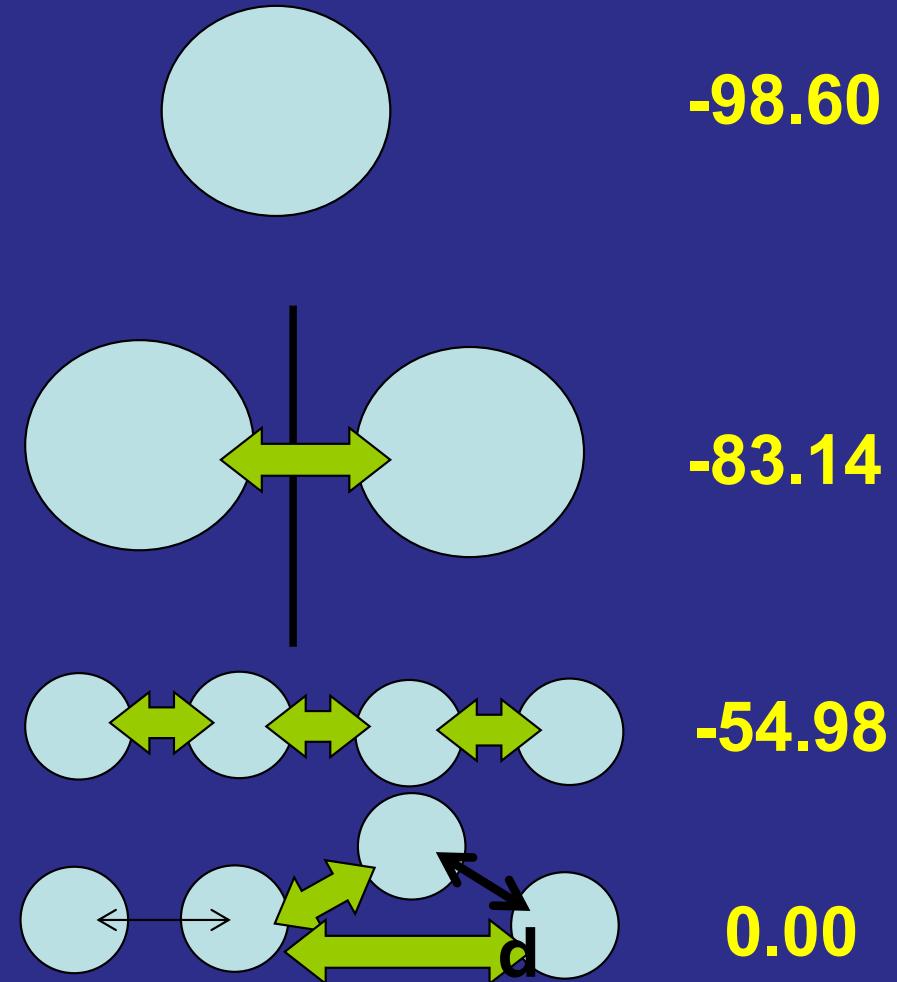


Using Bayes Factors to evaluate phylogeographic models

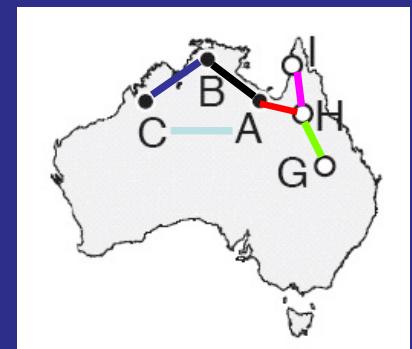
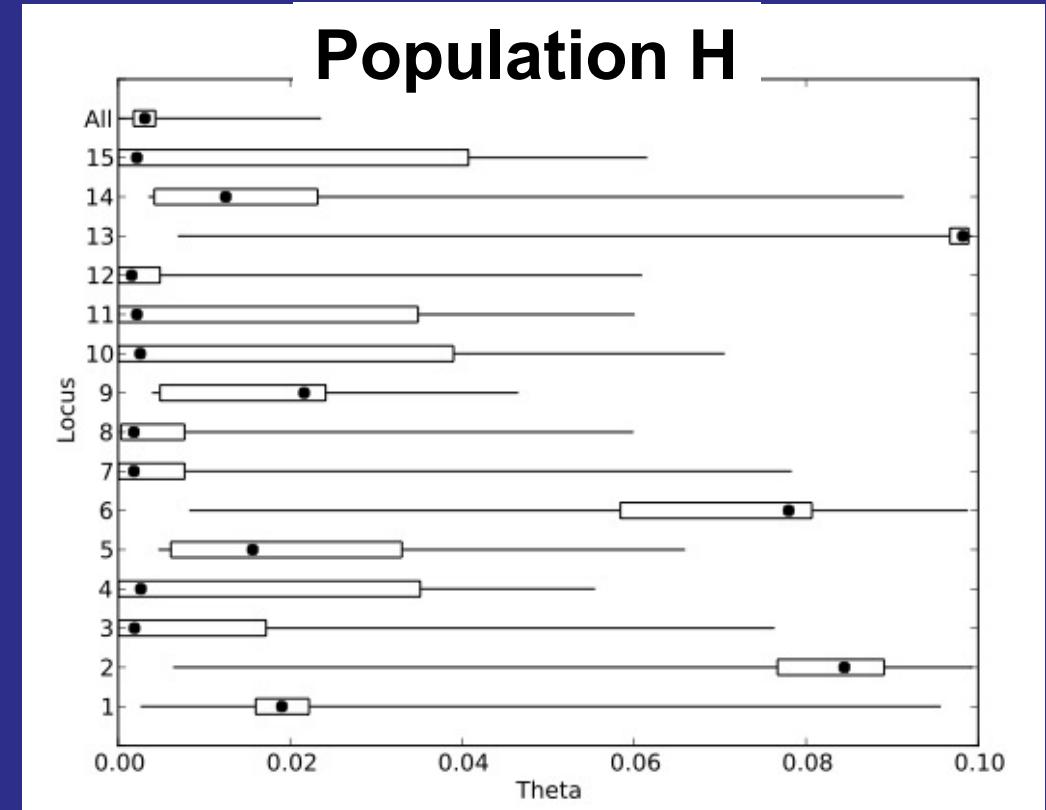
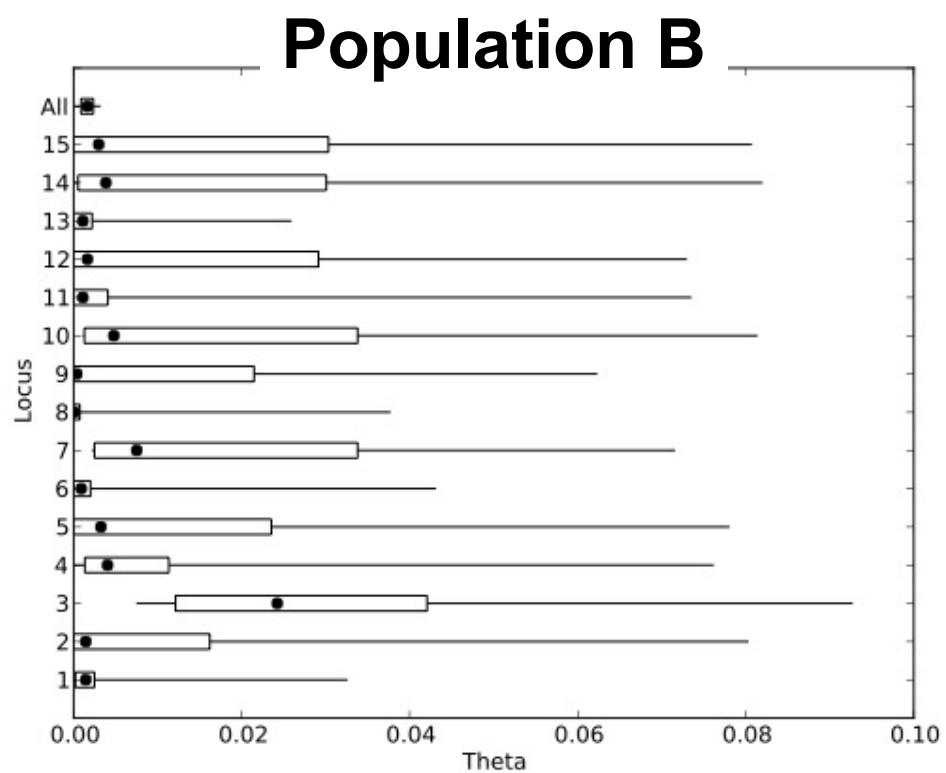
$$2(mL_1 - mL_2)$$



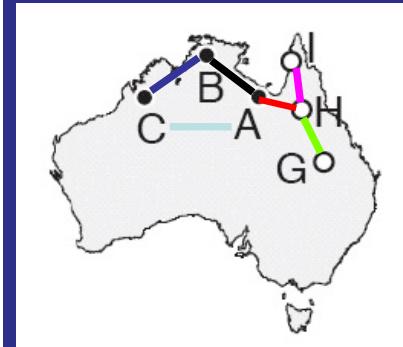
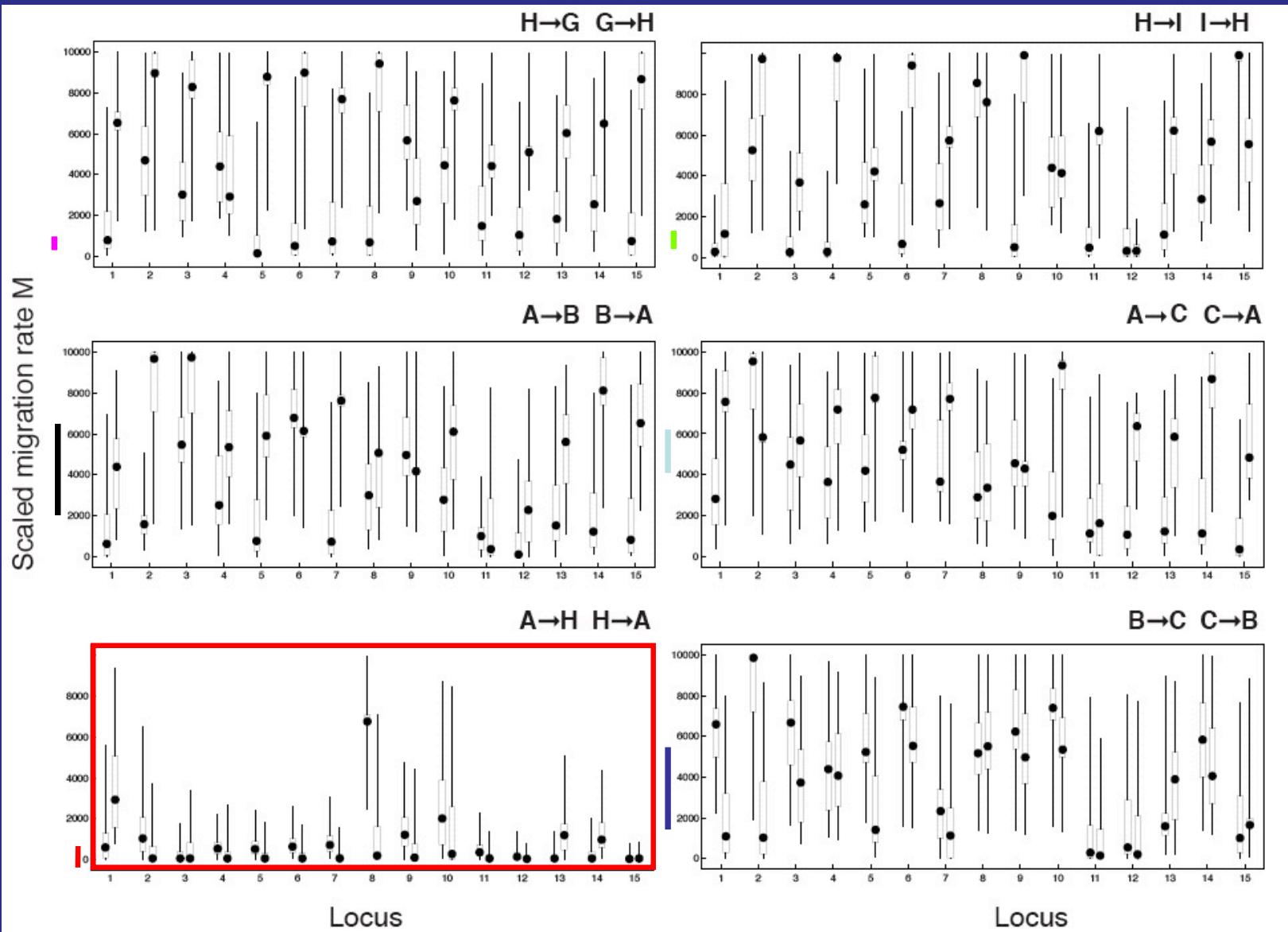
mL = marginal likelihood



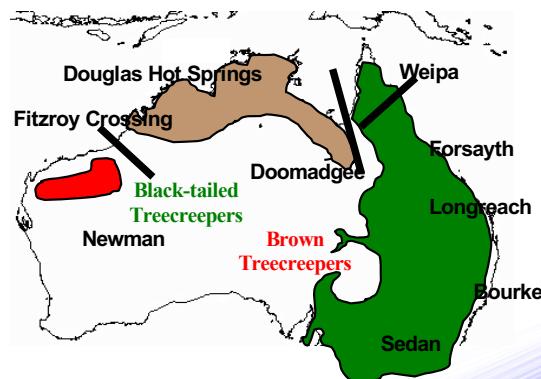
Higher precision estimates of demographic parameters (θ) with more loci



Multilocus estimates of migration rate

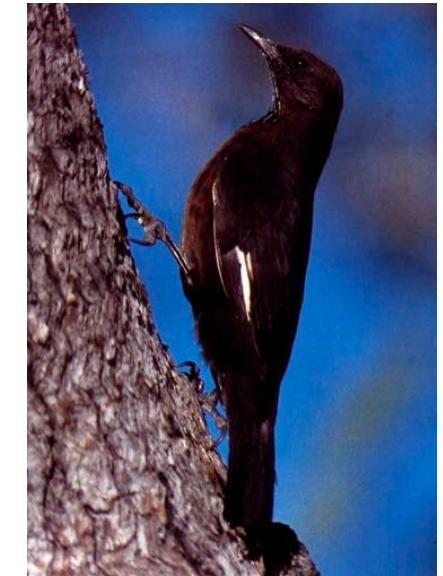


Phylogeny and divergence times



CM_Ne

CM



2000000 1500000 1000000 500000 0

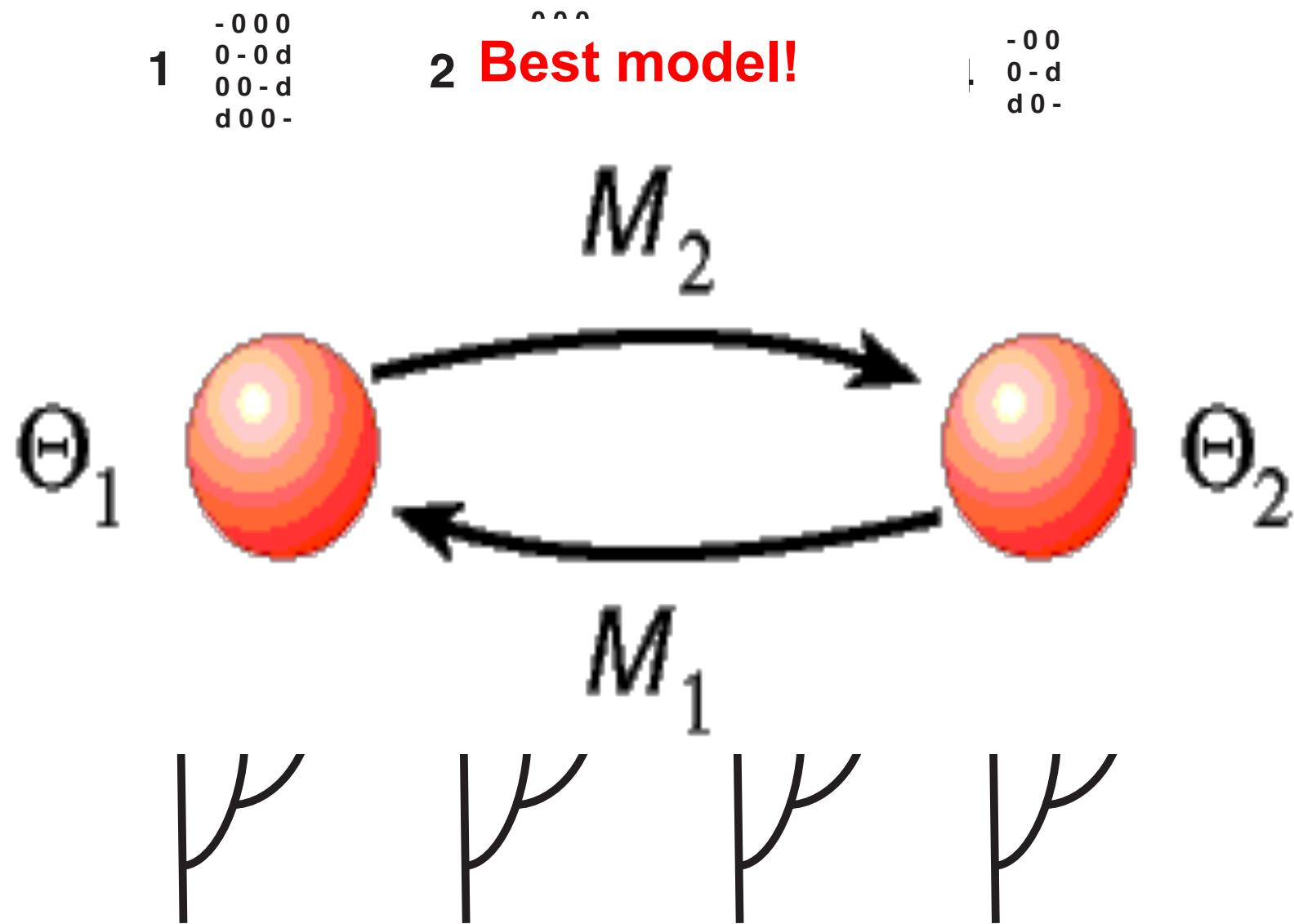
CP_We

CP



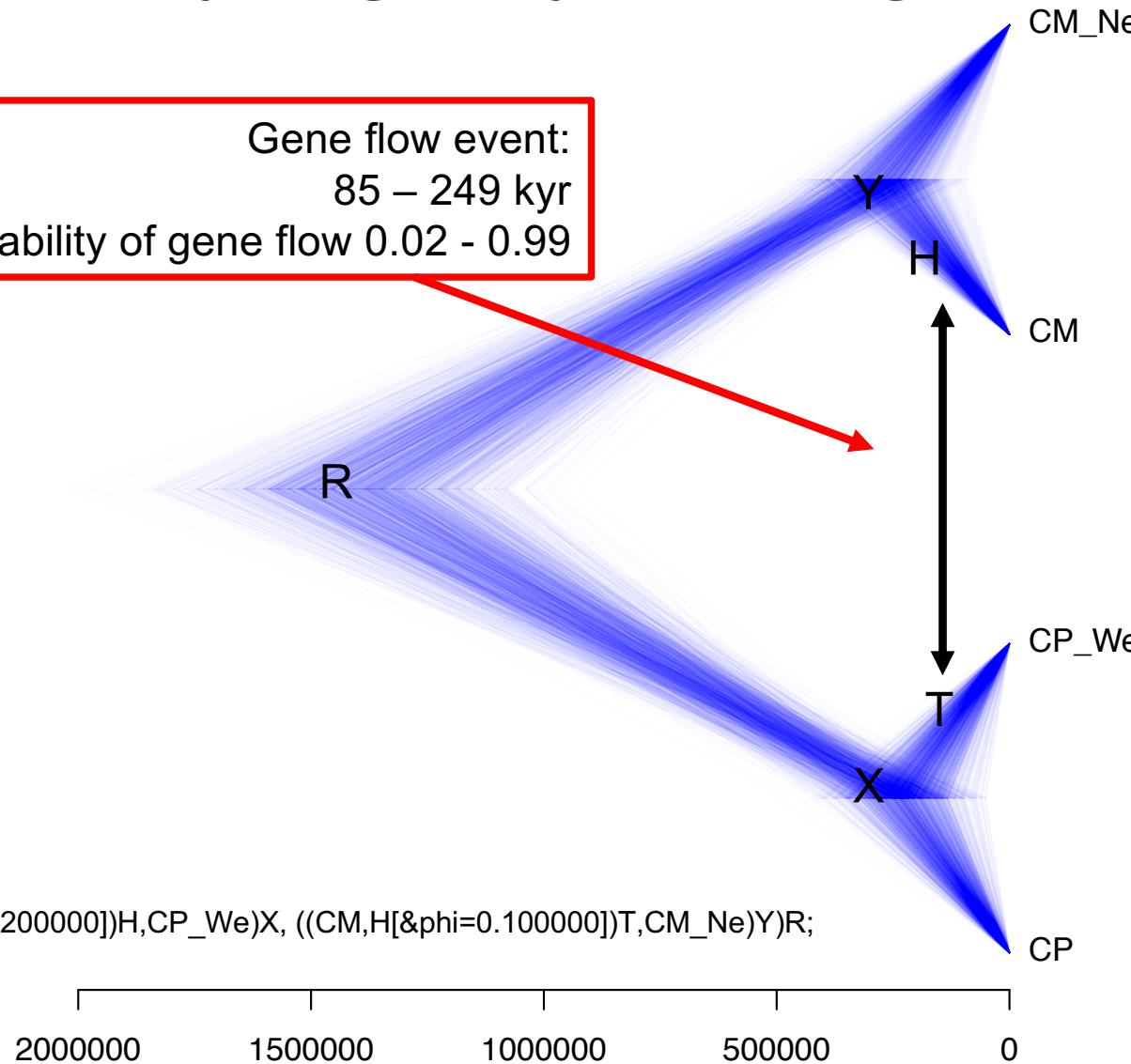
Divergence time (years), gen = 1 yr., $\mu = 2.2\text{e-}09$

Model selection using Migrate



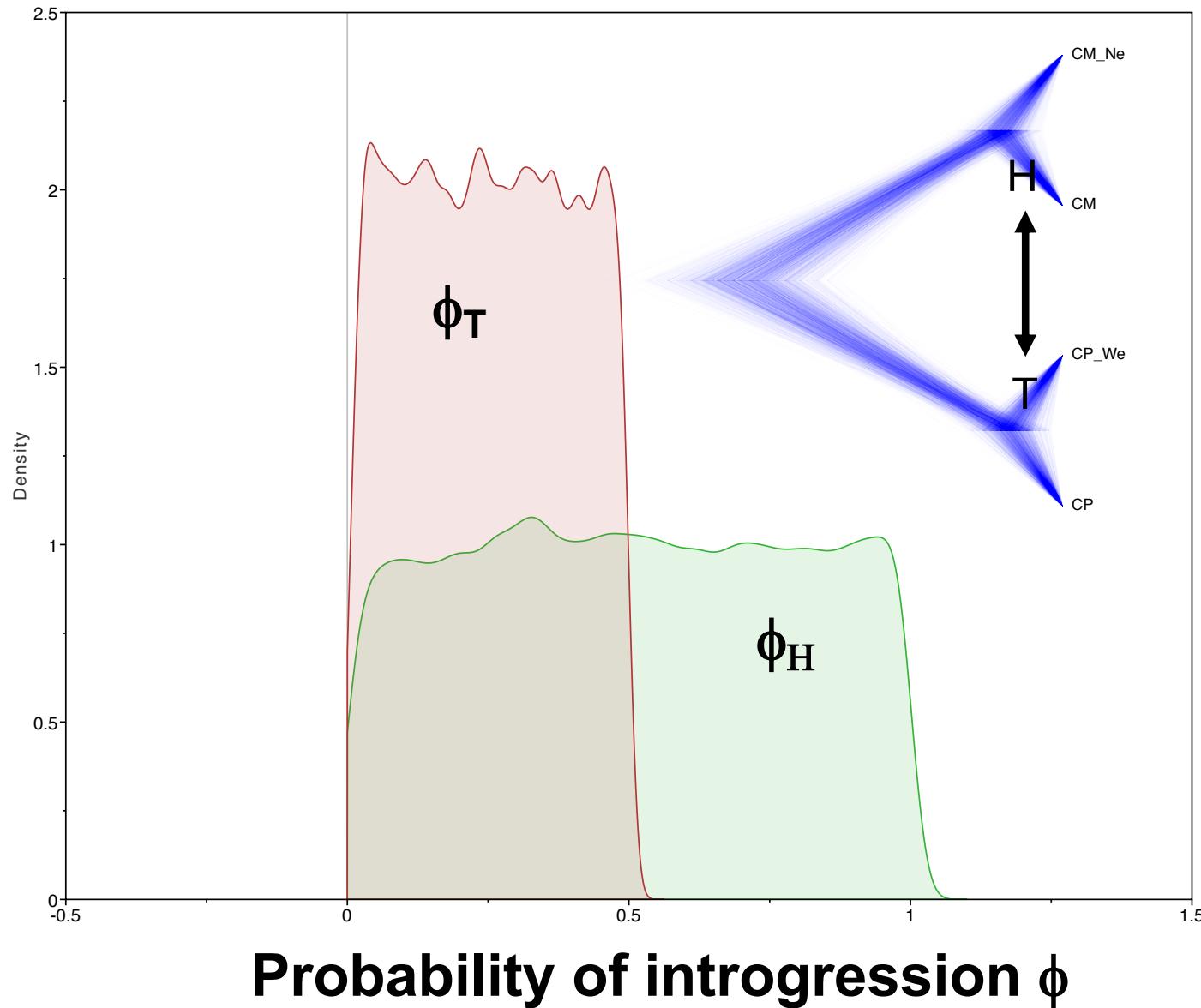
Phylogeny and gene flow

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

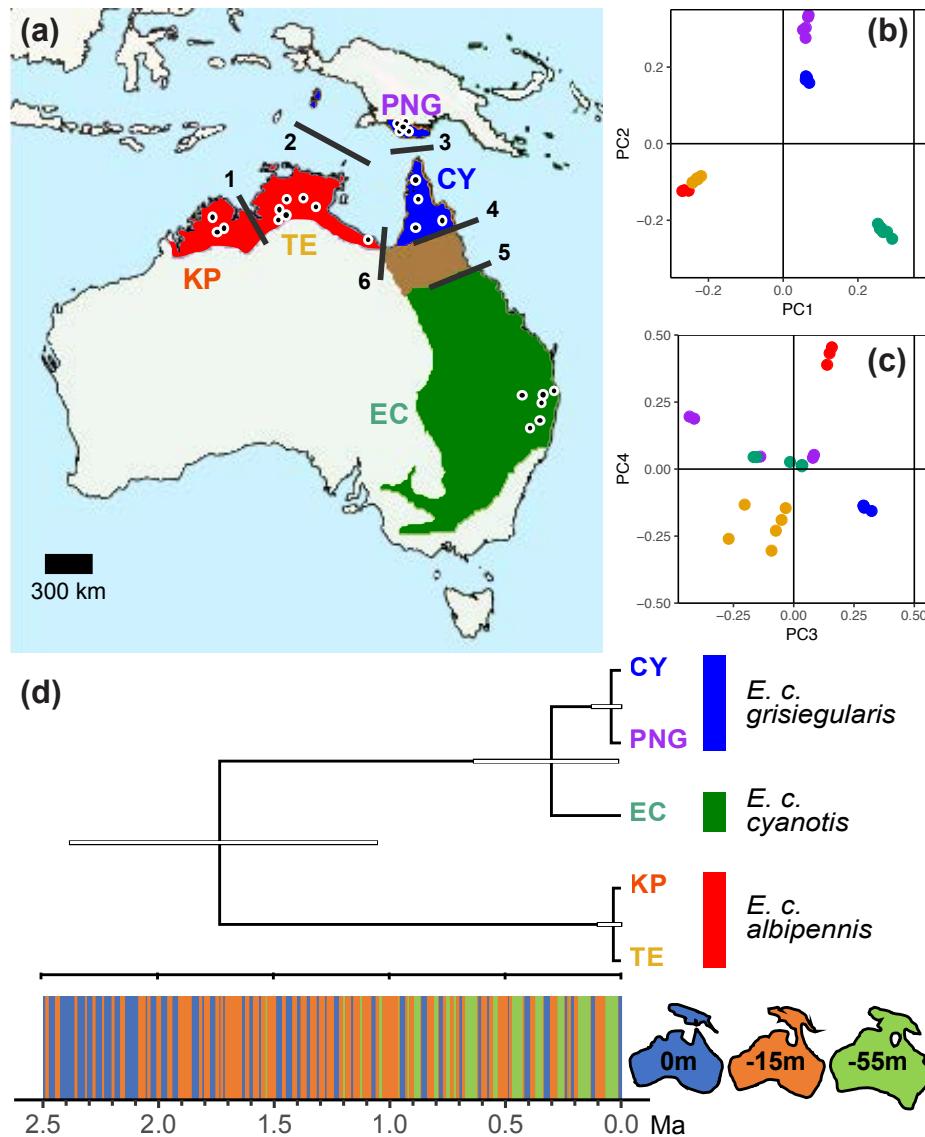


Divergence time (years), gen = 1 yr., $\mu = 2.2\text{e-}09$

BPP: flat posterior distributions of gene flow probability



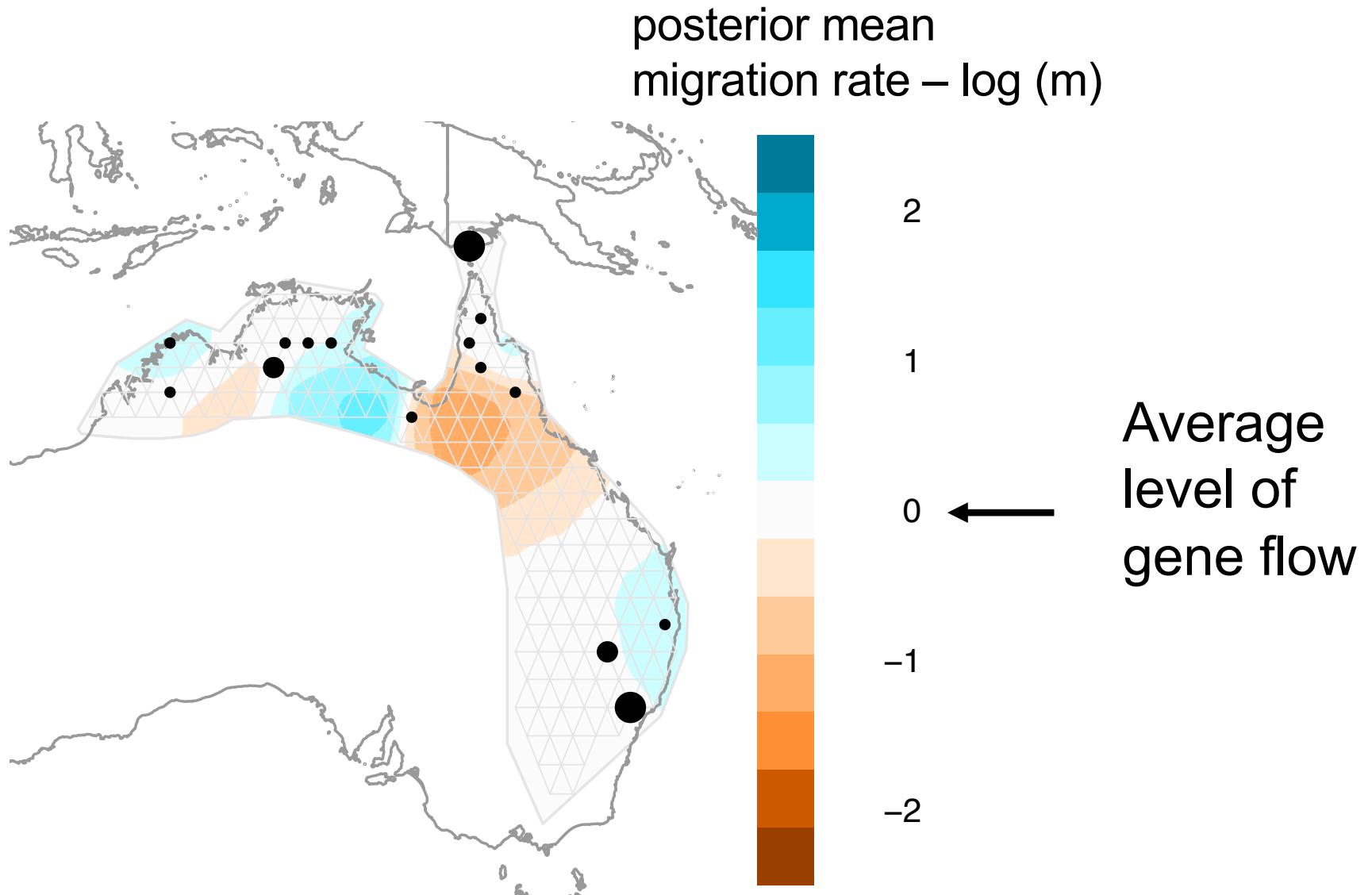
Whole-genome phylogeography of a widespread Australian honeyeater



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*)

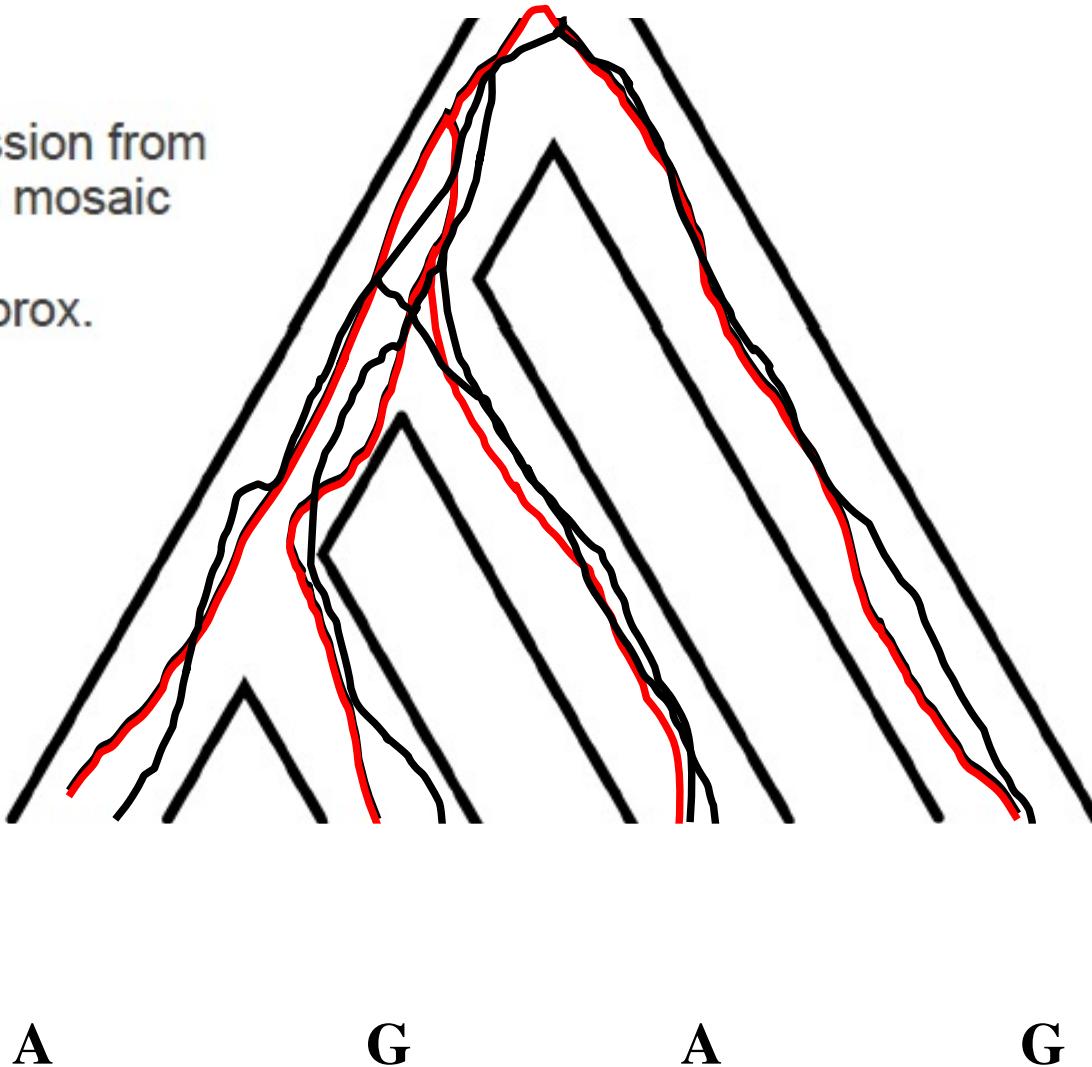
Effective migration surface



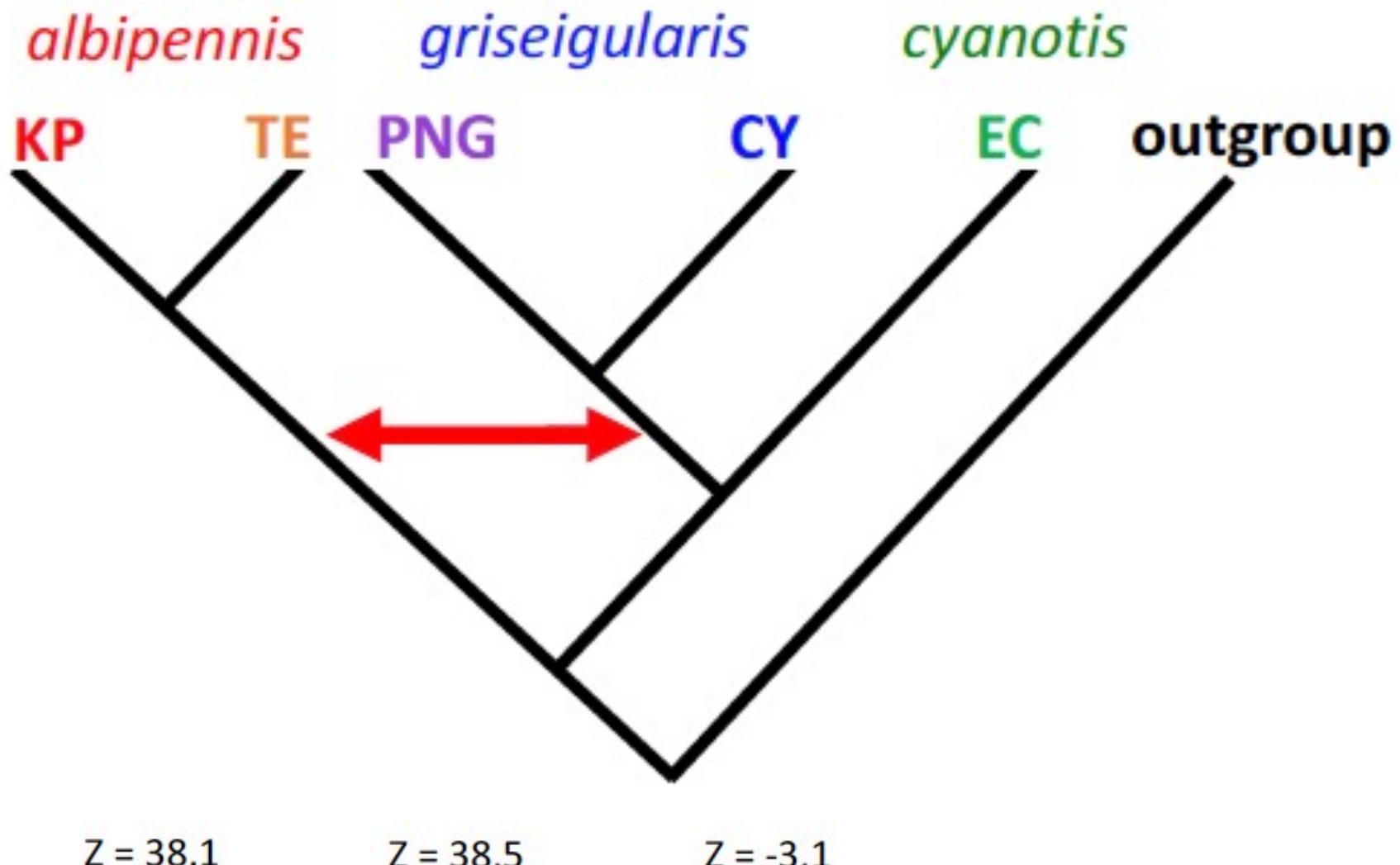
ABBA-BABA: Genome-wide test of introgression

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

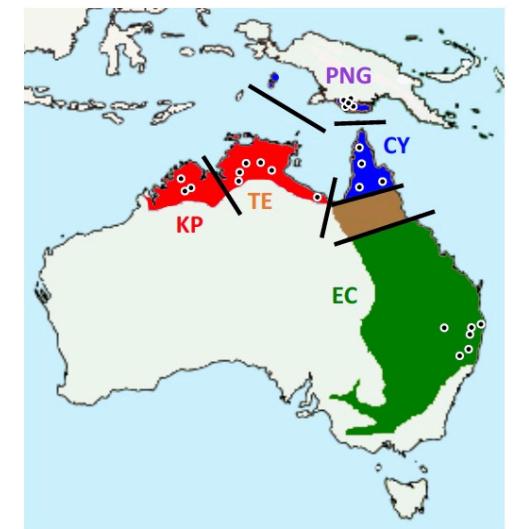
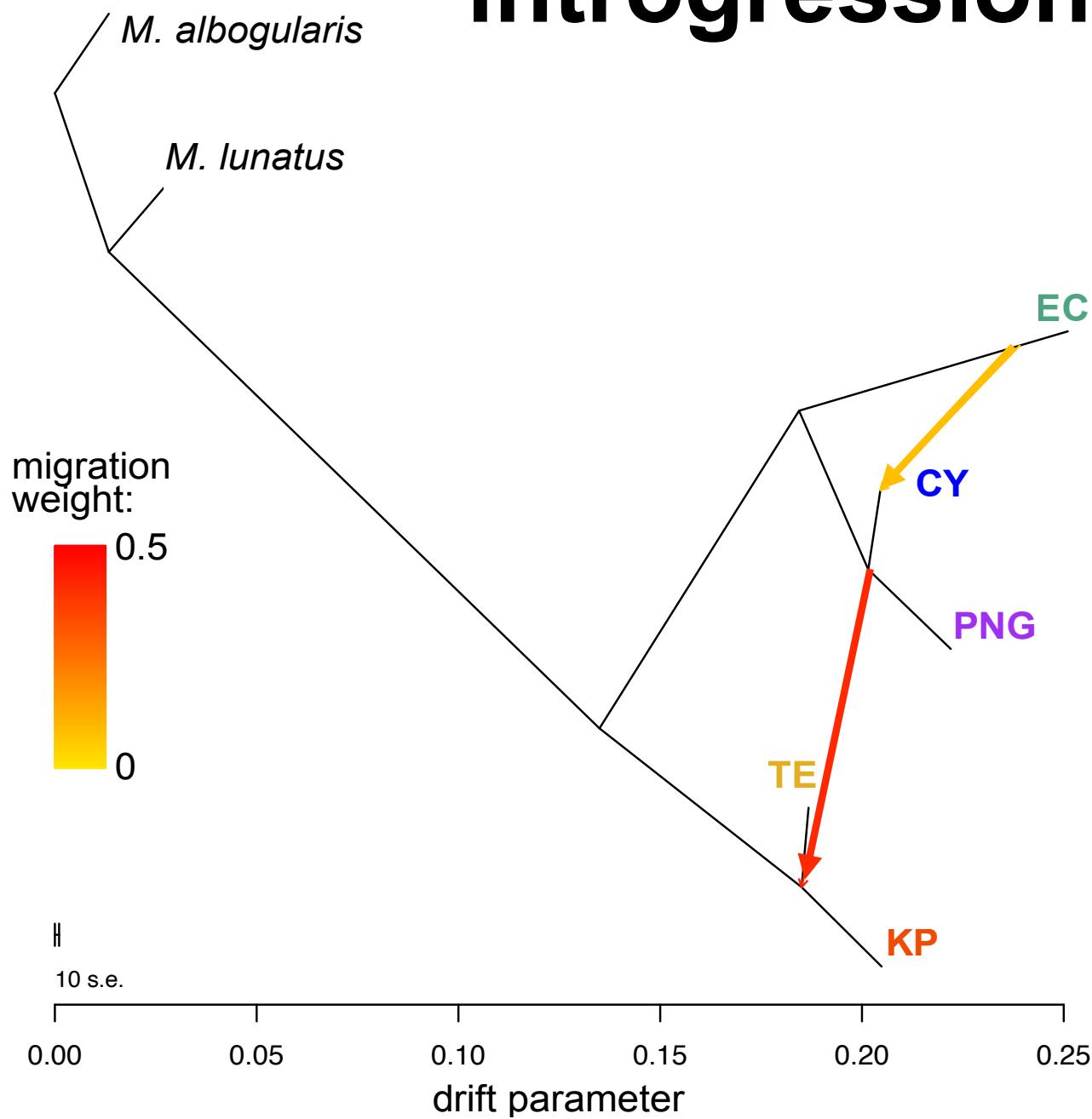
- $D > 0$: introgression from *Z. chrysops* into mosaic population
- (D assumed approx. Normal)



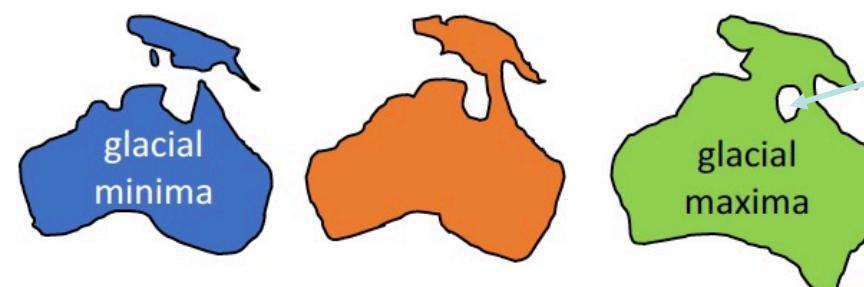
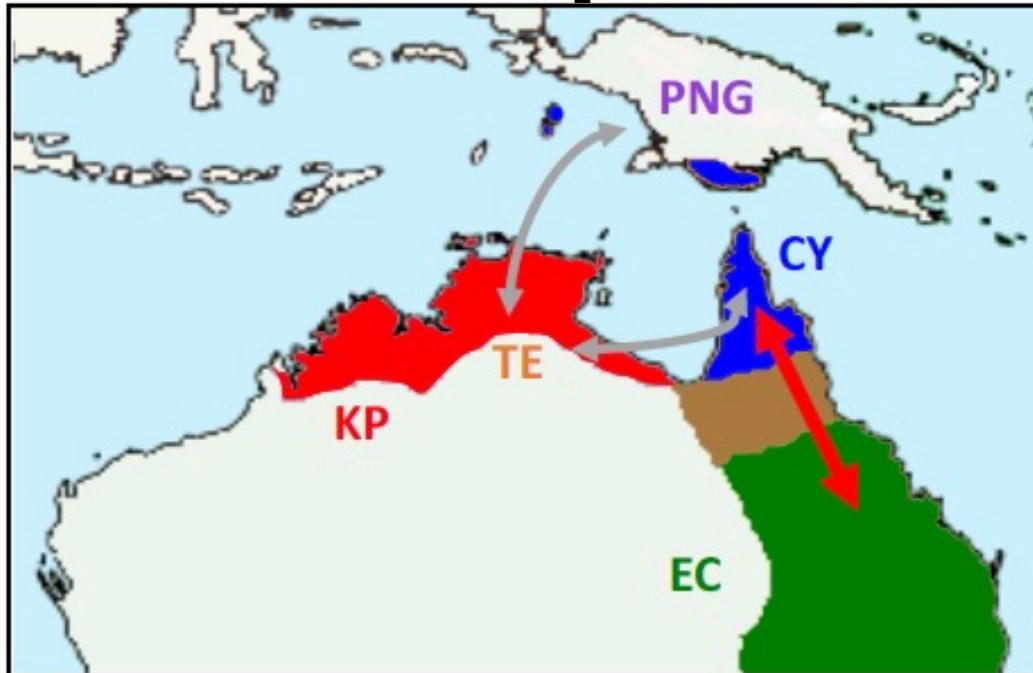
ABBA-BABA tests reveal signals of past introgression



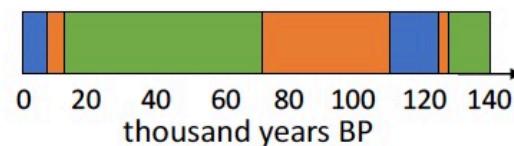
TreeMix analysis also suggests introgression



Phylogeographic shadows of Lake Carpentaria



Pleistocene
Lake Carpentaria



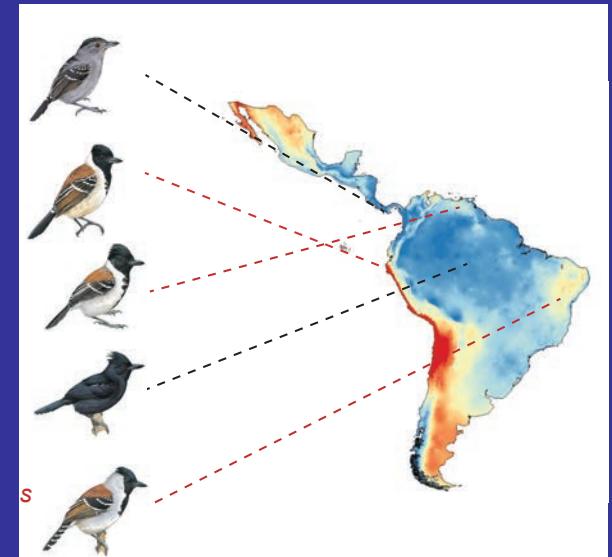
Opinion

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}

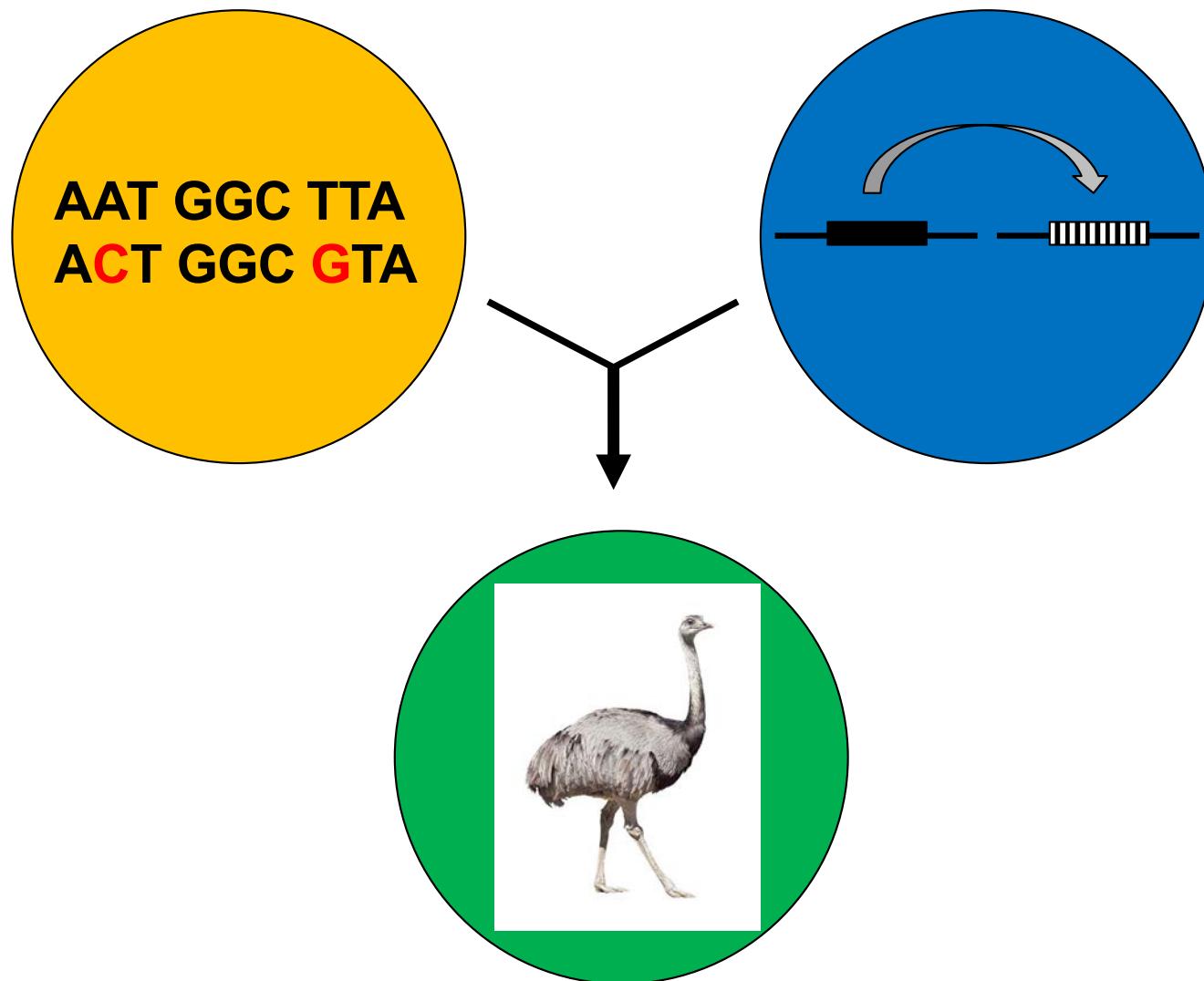
Smith, et al. 2020. *TREE* 35: P415-525

- 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.



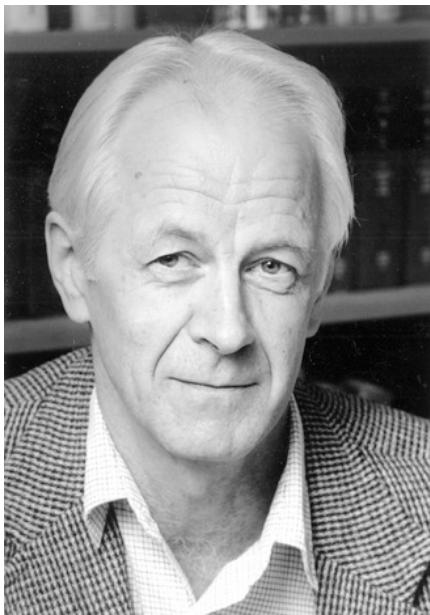
Gustavo Bravo, in prep.
Antbirds, Thamnophilidae

What are the relative roles of gene sequence change versus gene regulatory change in phenotypic evolution?

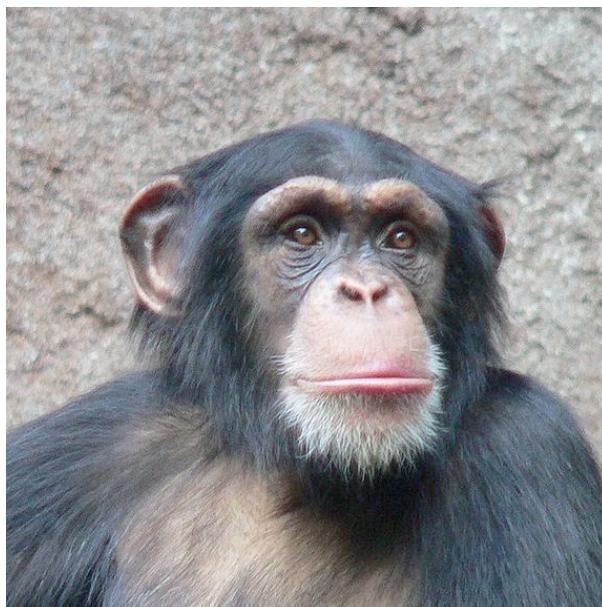


Evolutionary change: genes or gene regulation?

Evolution at Two Levels in Humans and Chimpanzees



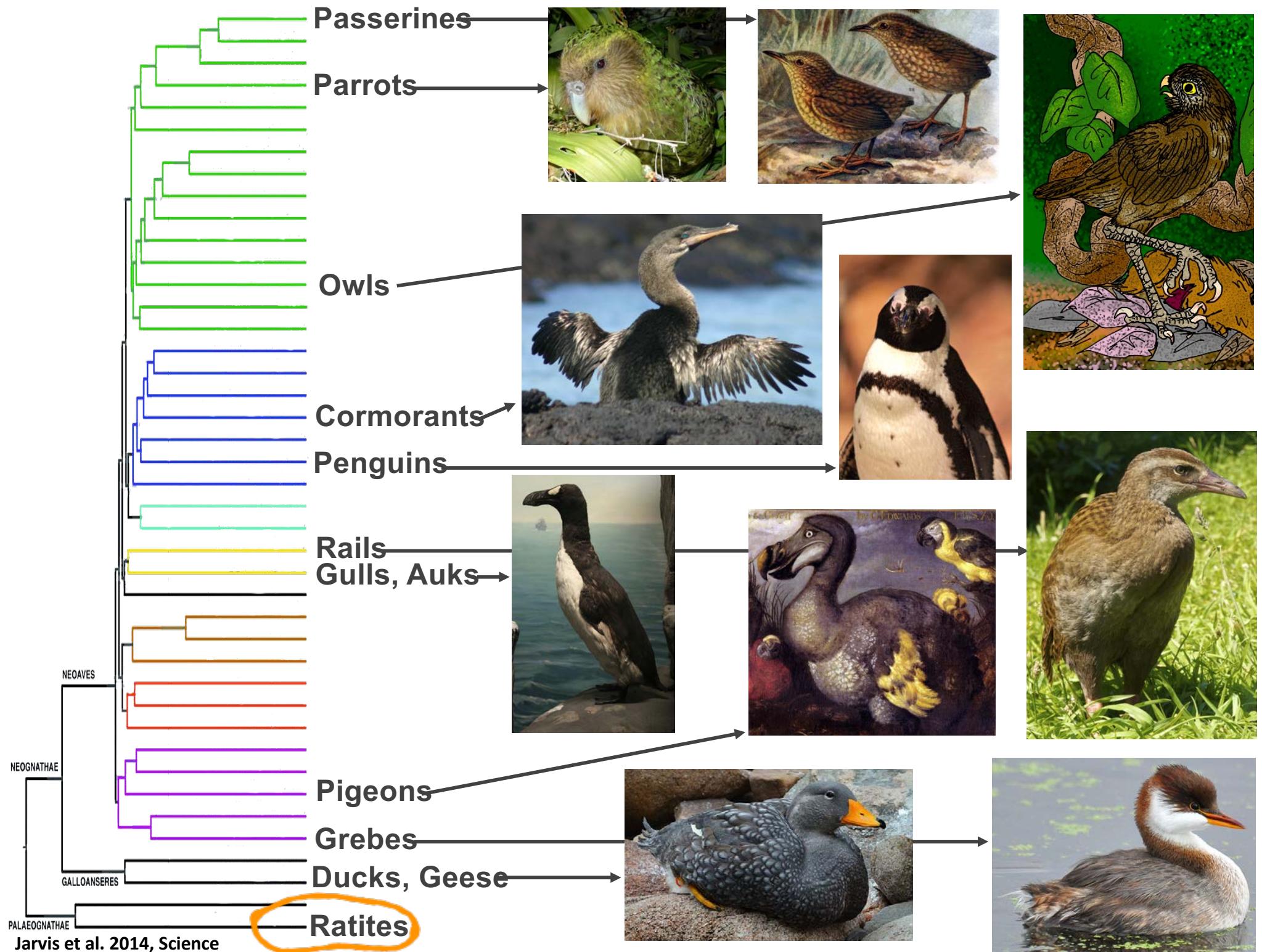
Their macromolecules are so alike that regulatory mutations may account for their biological differences.



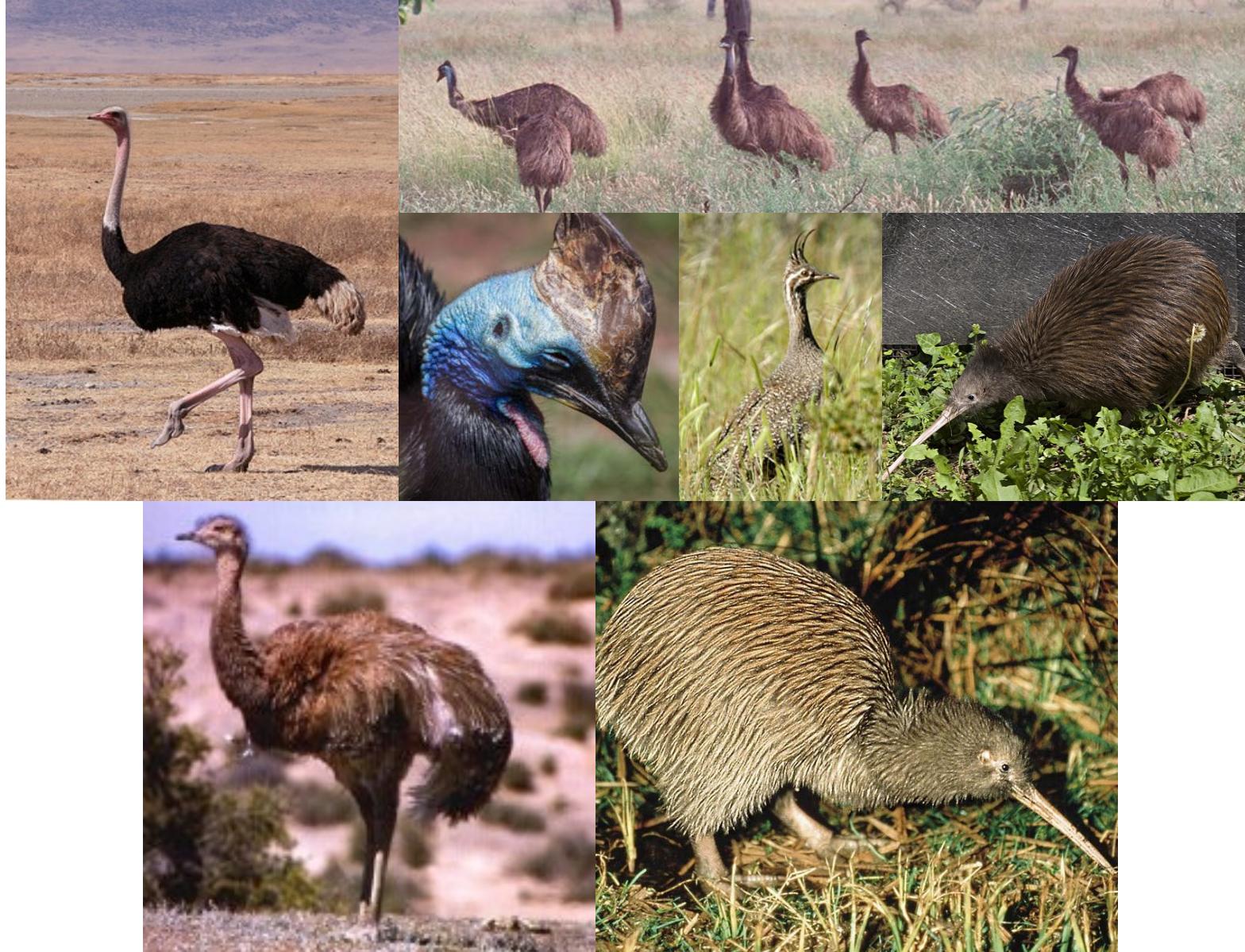
Mary-Claire King and A. C. Wilson

SCIENCE

11 April 1975, Volume 188, Number 4184



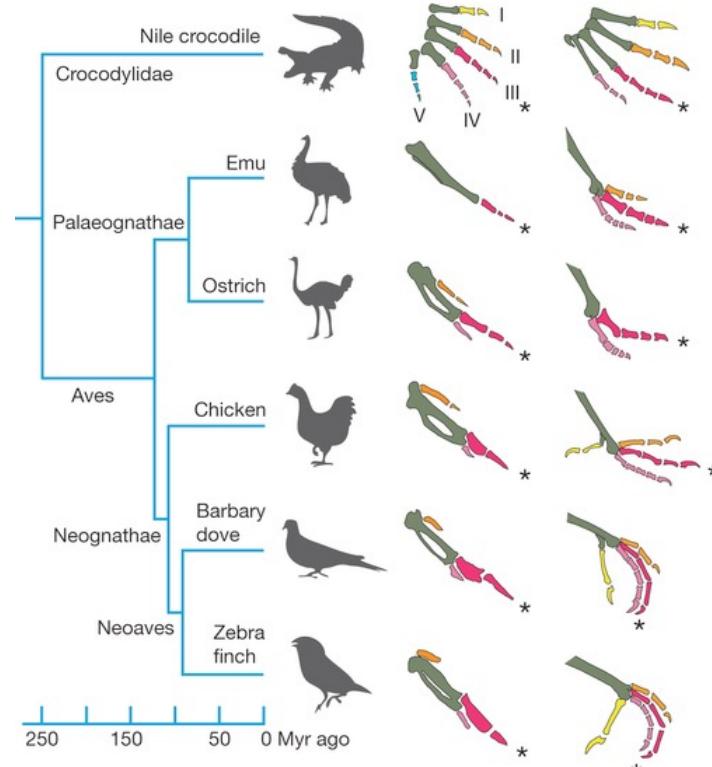
CNEEs and the convergent evolution of flightlessness in Palaeognathae



Skeletal modifications for flightlessness



Forelimb Hindlimb

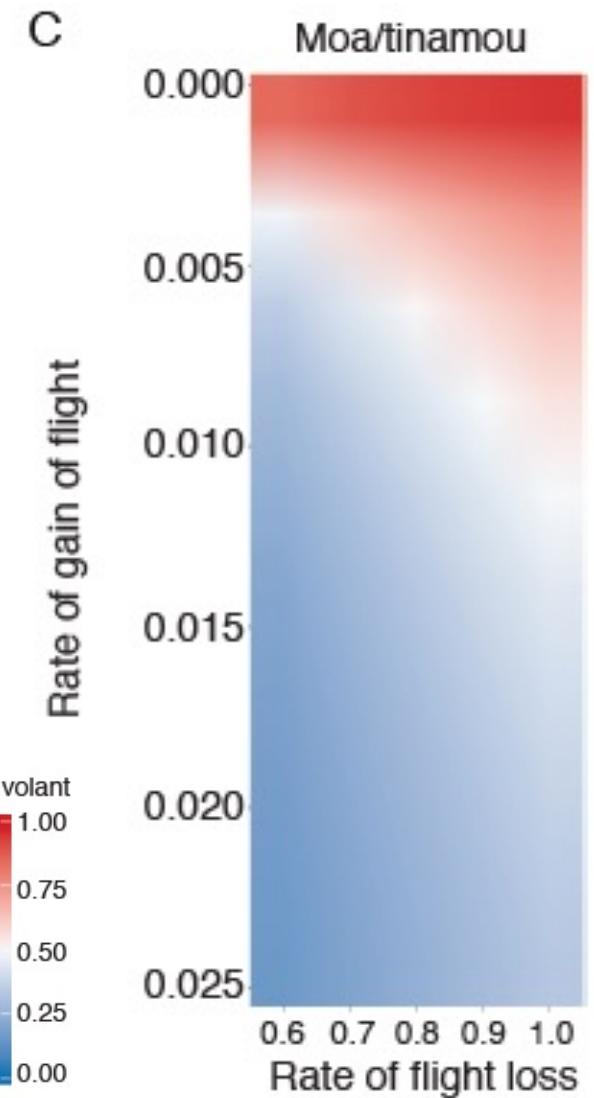
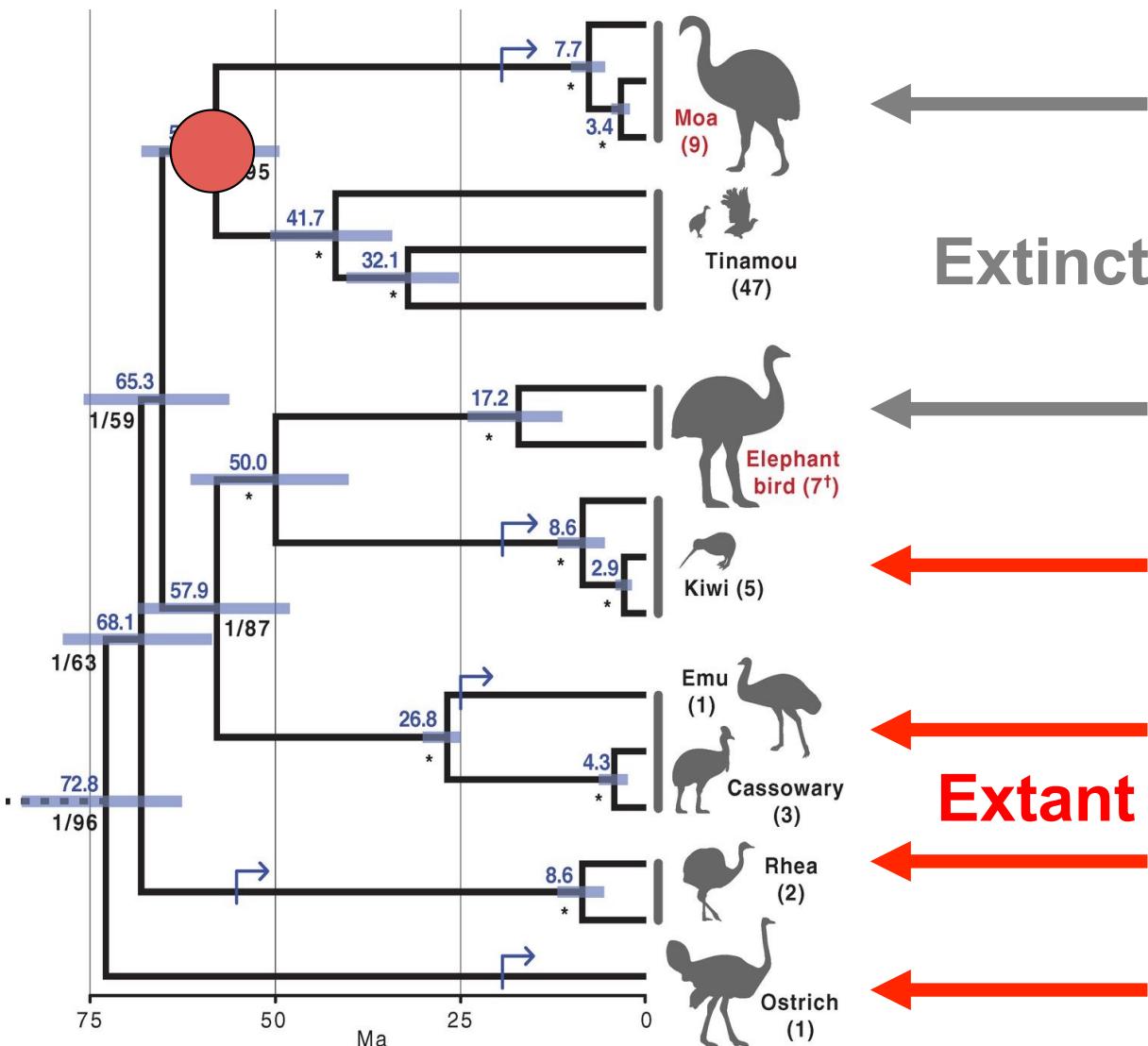


Little-spotted kiwi sternum



Emu and ostrich keelless sterna

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



11 new paleognath genomes



Little Spotted Kiwi



Little bush moa



Great-spotted Kiwi



Lesser Rhea



Emu



Elegant-crested
Tinamou



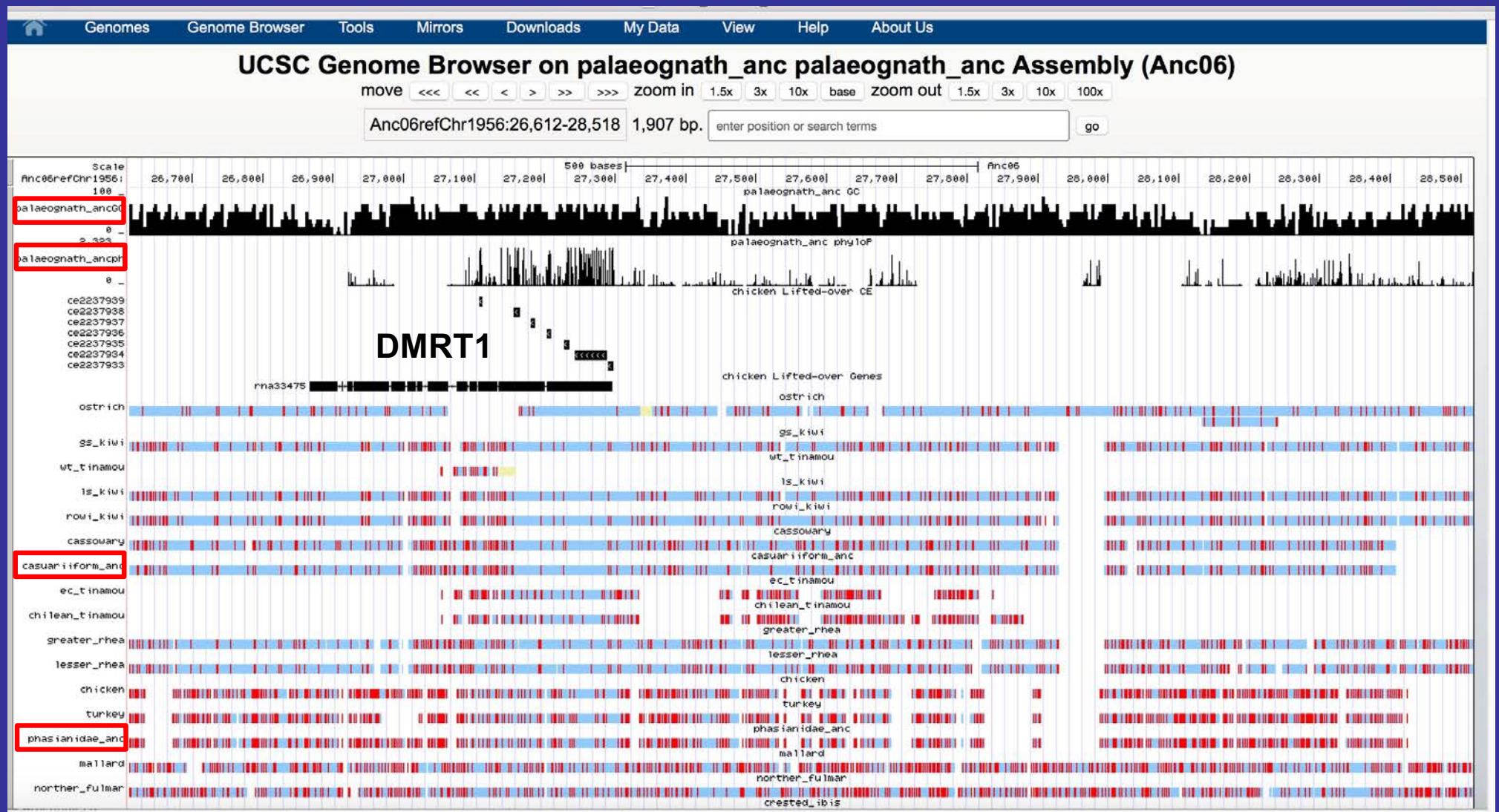
Thicket tinamou



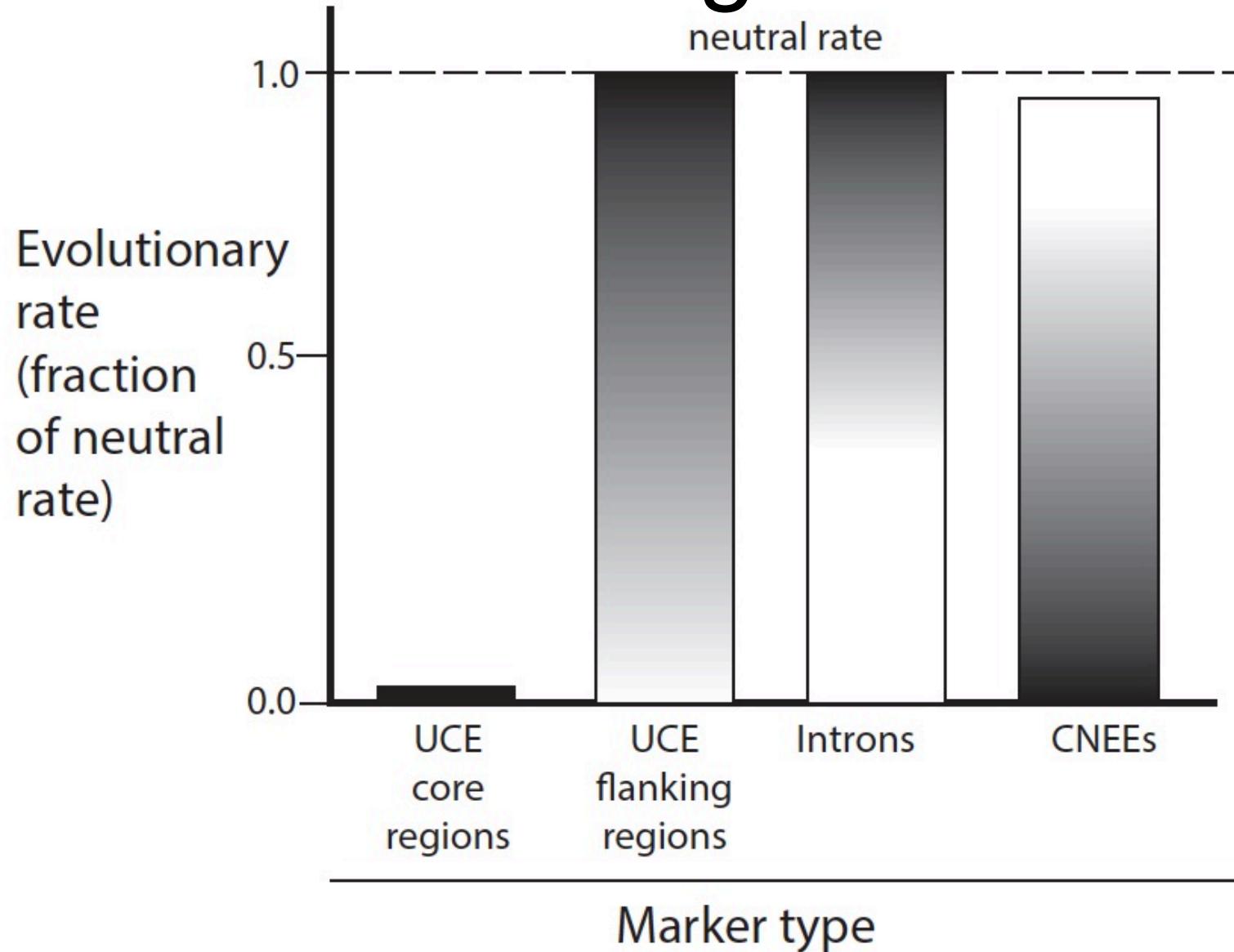
Chilean tinamou

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

42-species whole genome alignment for birds using ProgressiveCactus



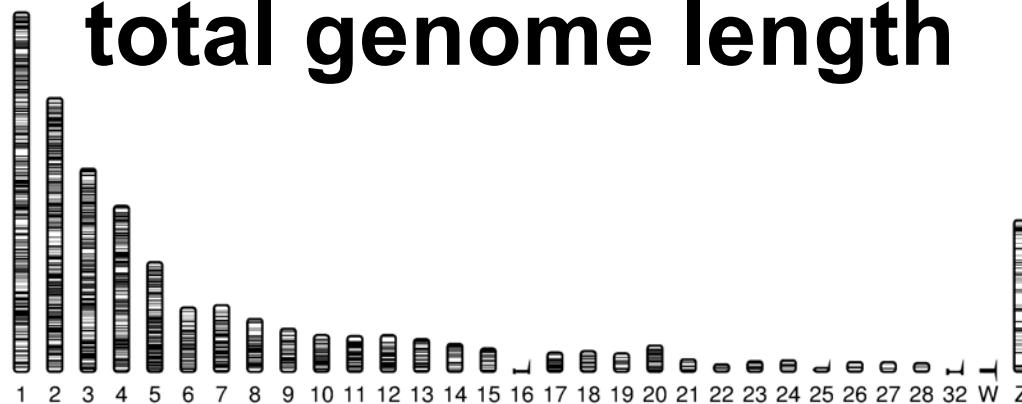
Relative rates of different noncoding markers



Phylogenomic markers cover c. 3% of total genome length

12,676

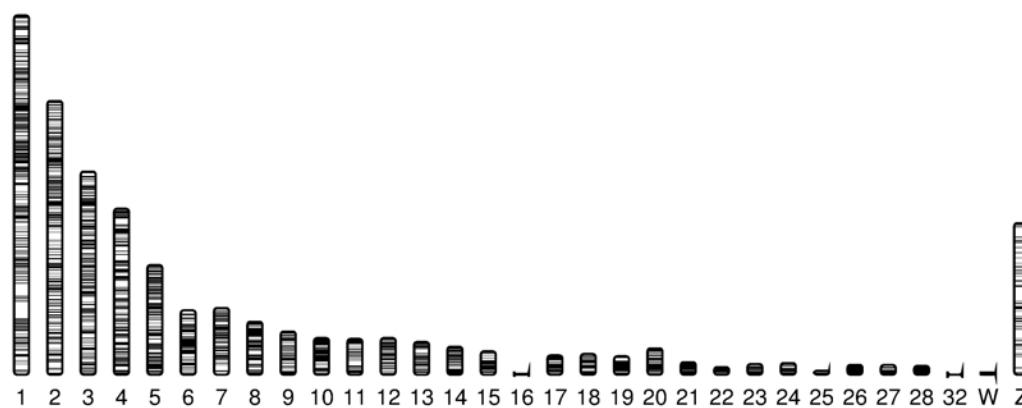
CNEEs



5,016

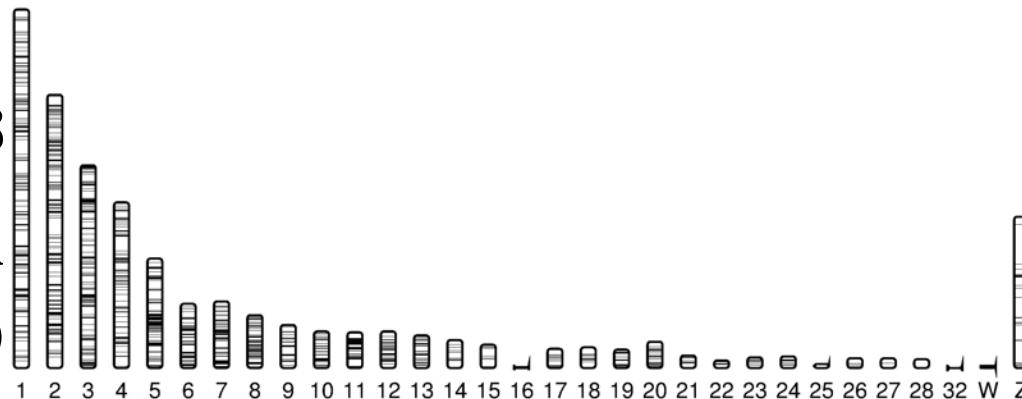
Intron

S



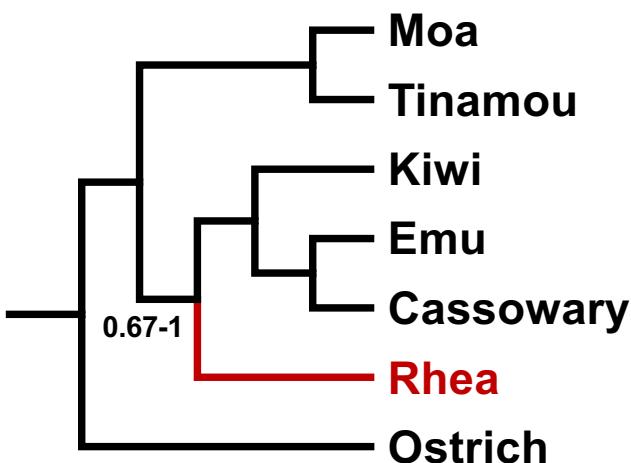
3,158

Ultraconserved
elements (UCEs)

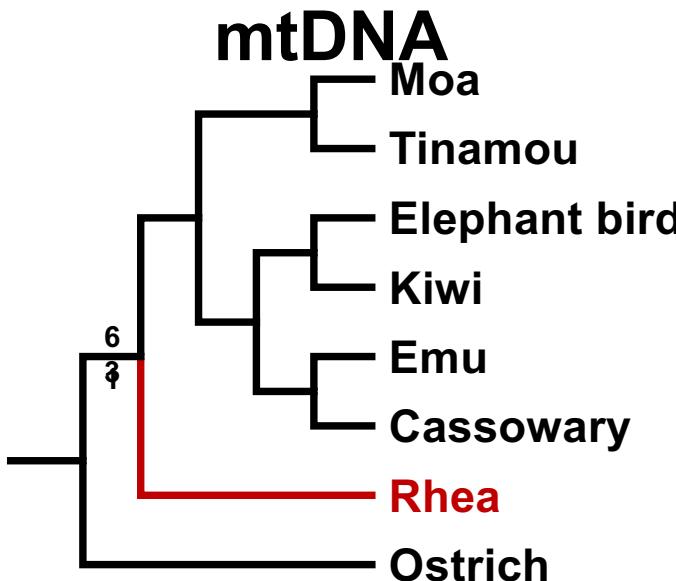


Relationships of rheas unclear

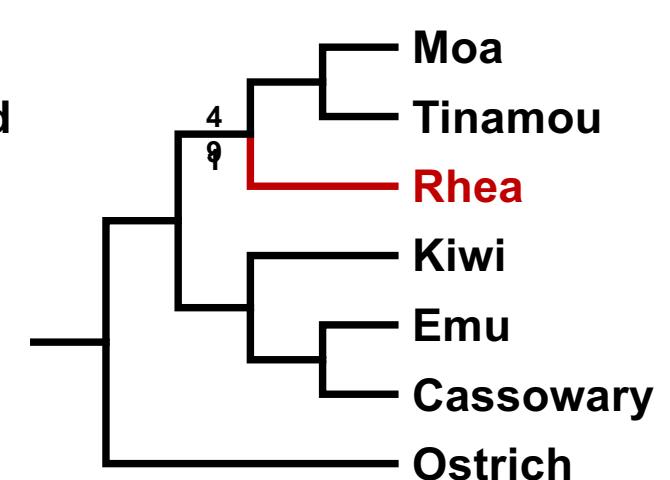
Haddrath & Baker
(2012)
27 nuclear loci



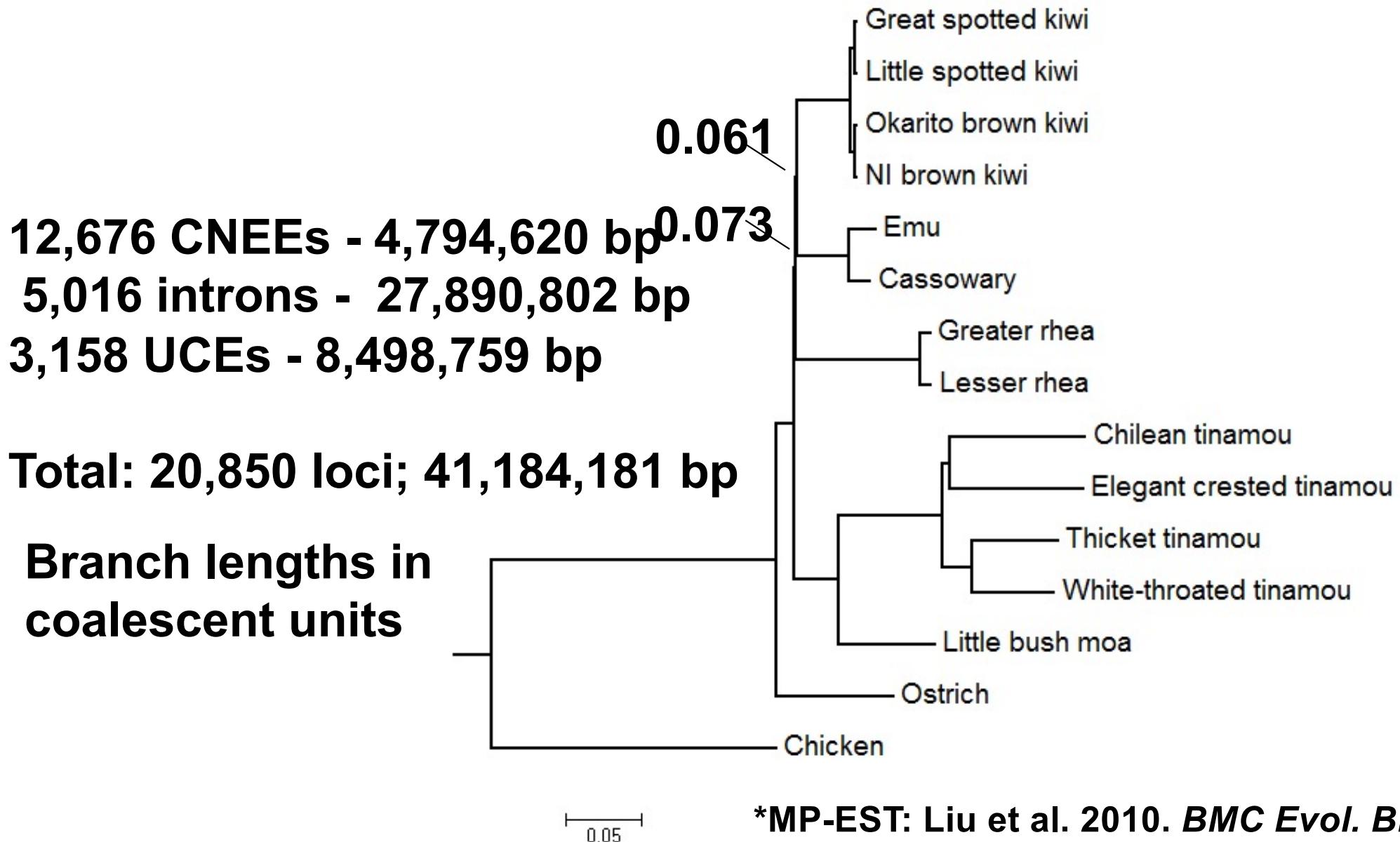
Mitchell et al.
(2014)



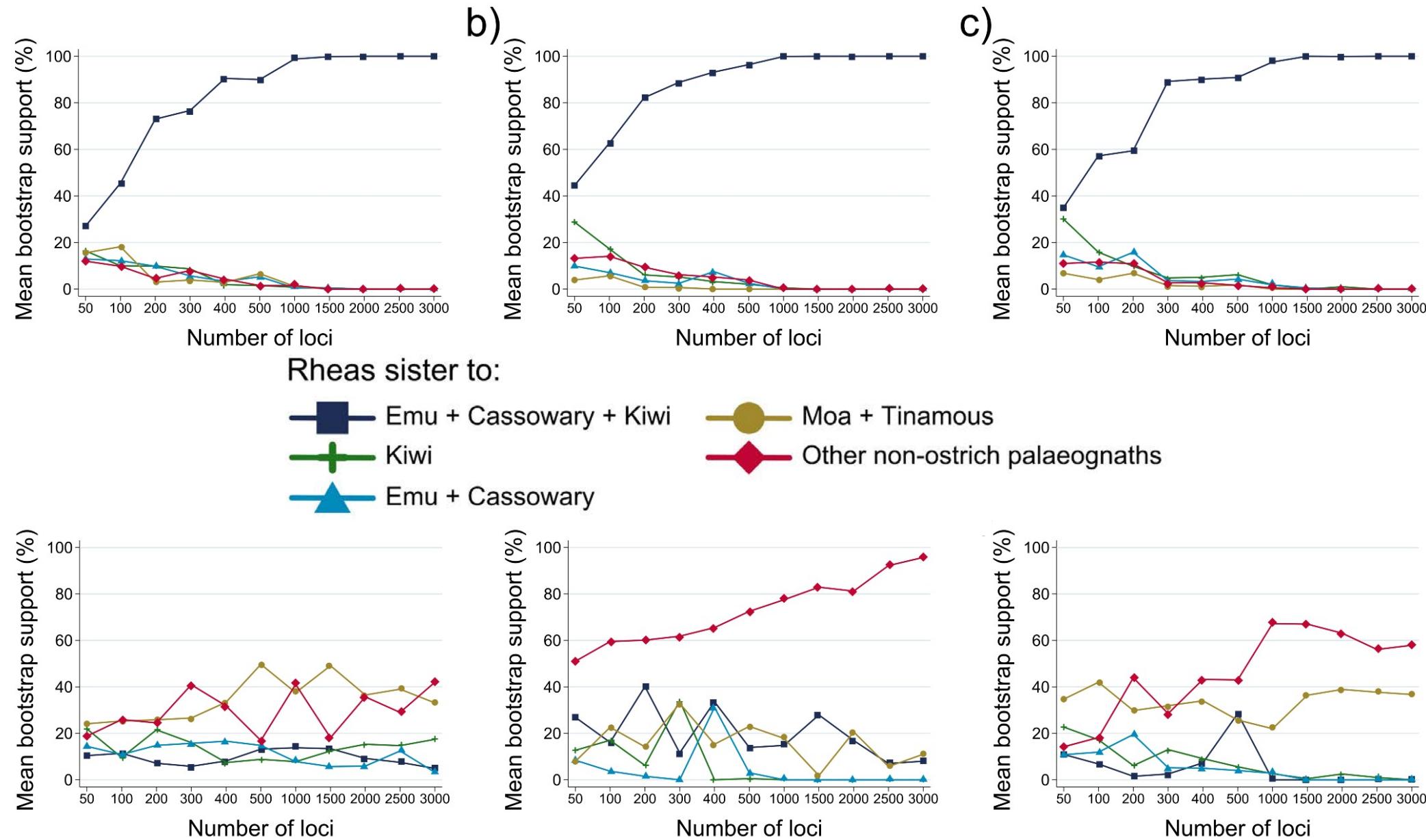
Smith et al. (2013)
60 nuclear loci +
mtDNA



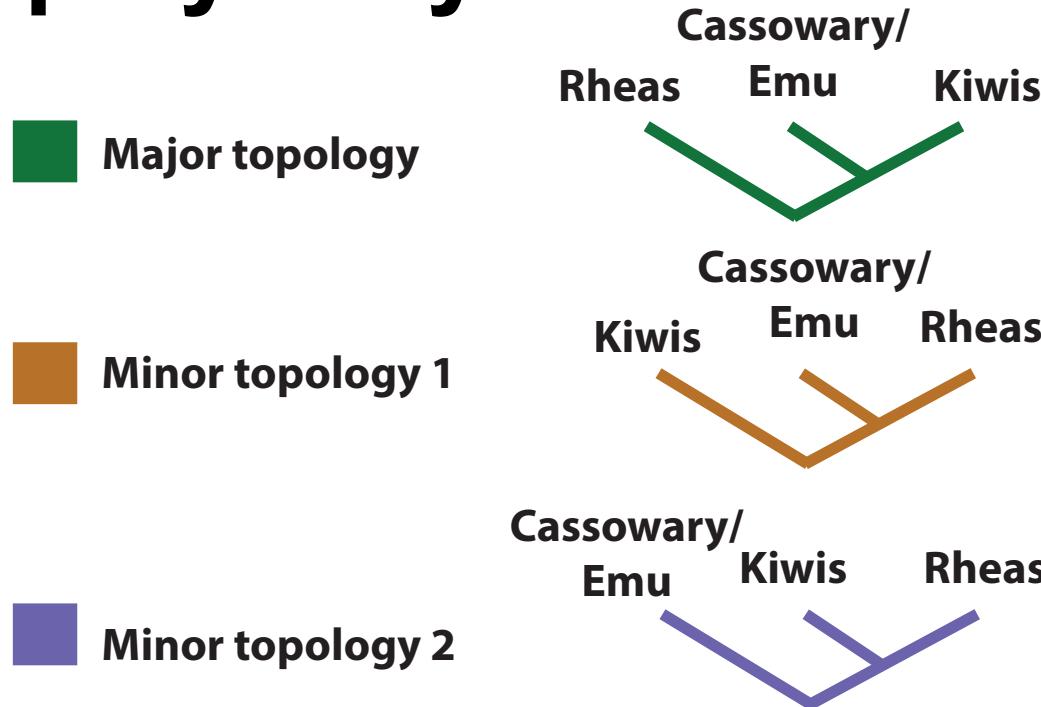
Coalescent* analyses resolve the position of rheas and reveal an ancient rapid radiation



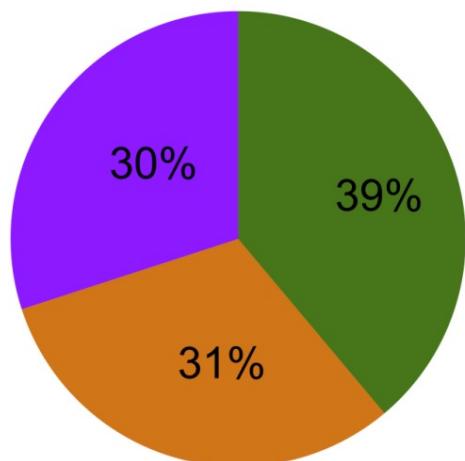
Consistent accumulation of phylogenetic signal using CNEEs, Introns, and UCEs



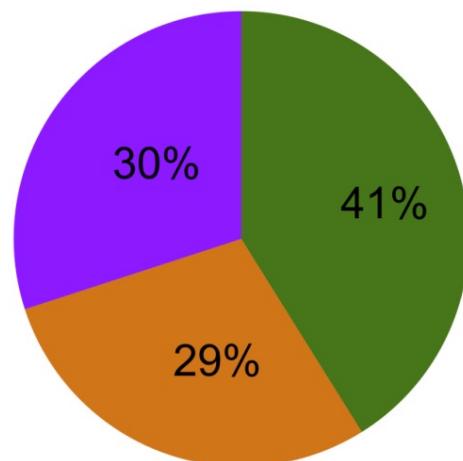
Gene tree distribution suggests a near polytomy at base of ratites



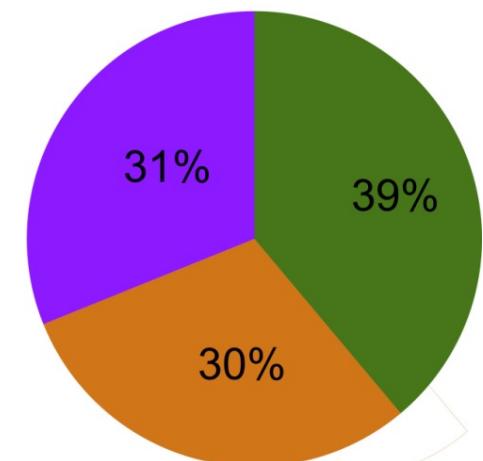
CNEEs



Introns

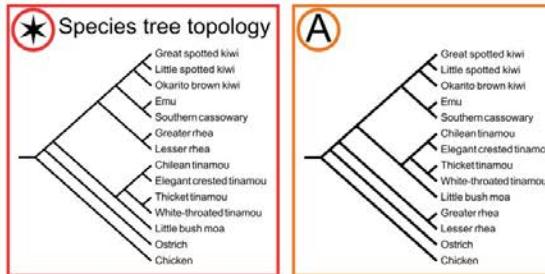
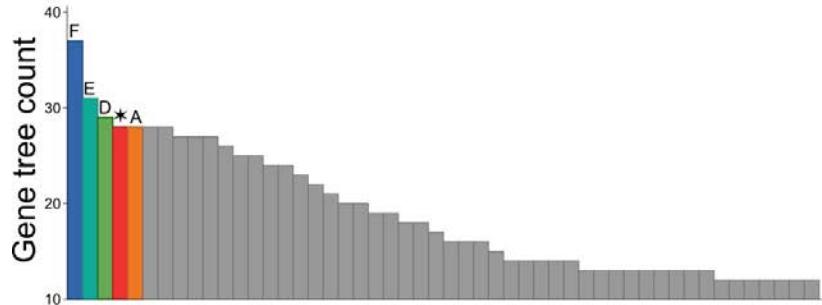


UCEs

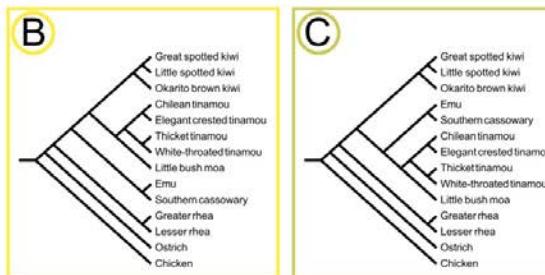
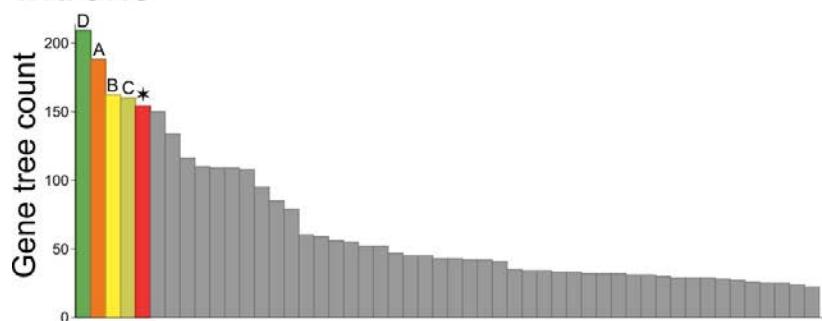


Anomaly zone: most common gene tree does not match the species tree

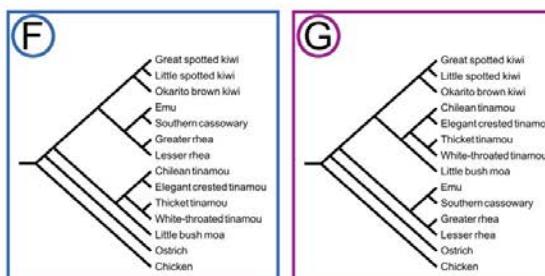
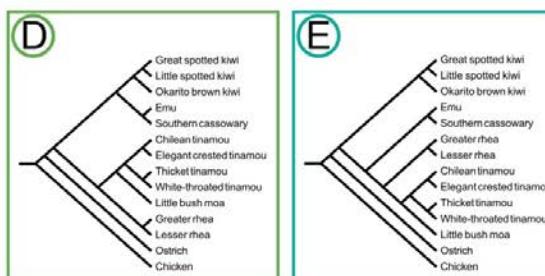
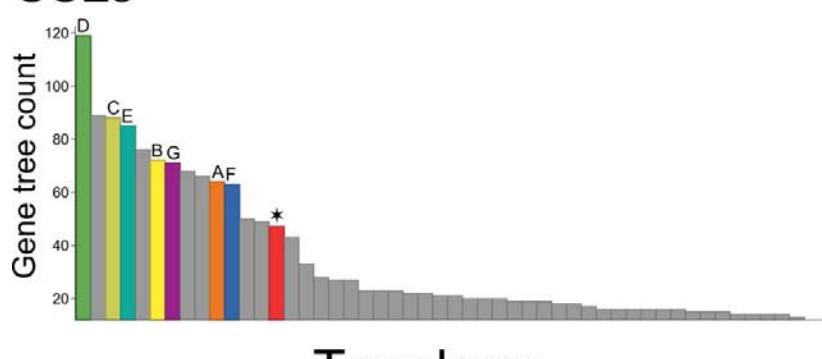
CNEEs



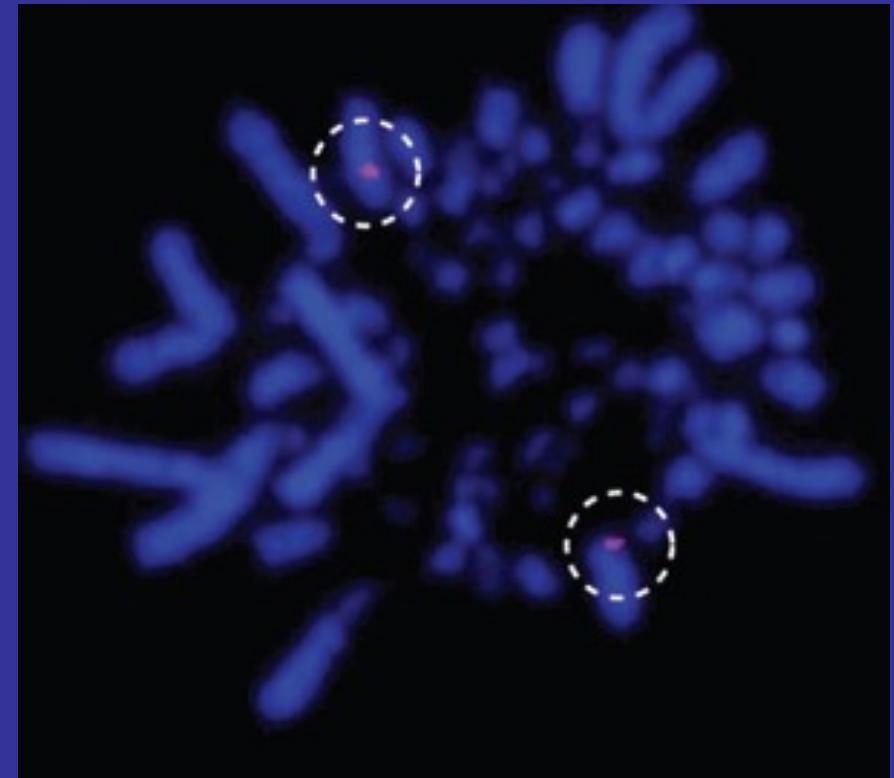
Introns



UCEs



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

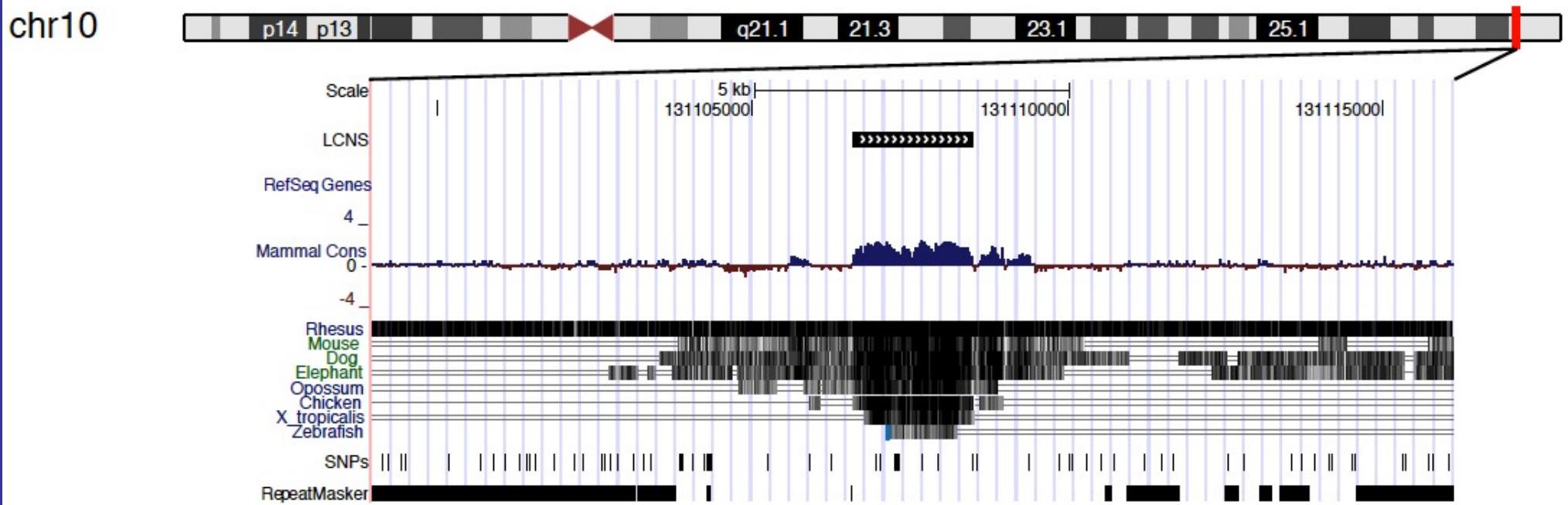


Karyotype of an Emu

CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements

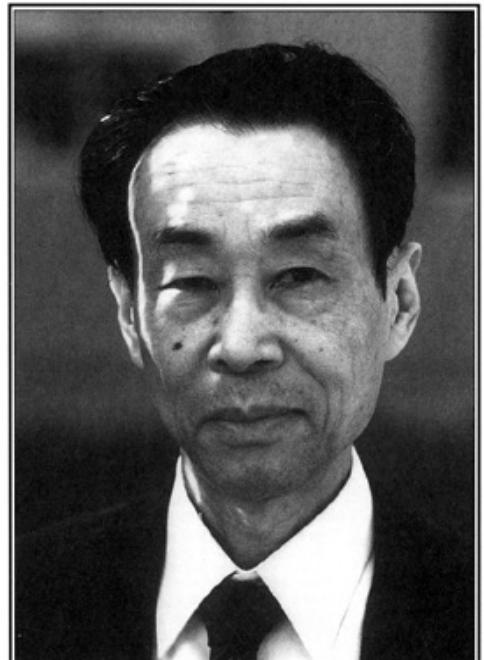
284,001 long (* > 50 bp) CNEEs in data set



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



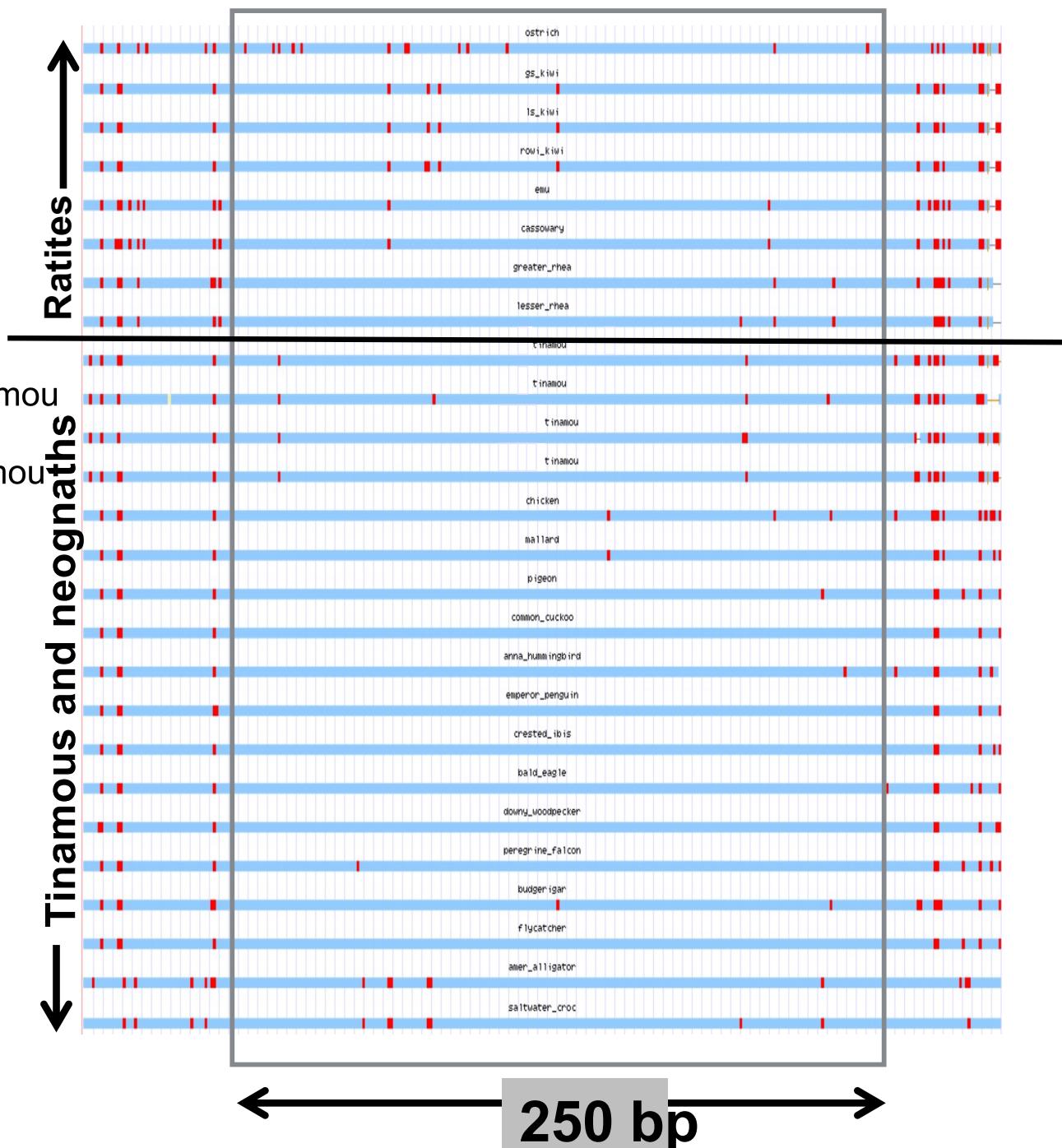
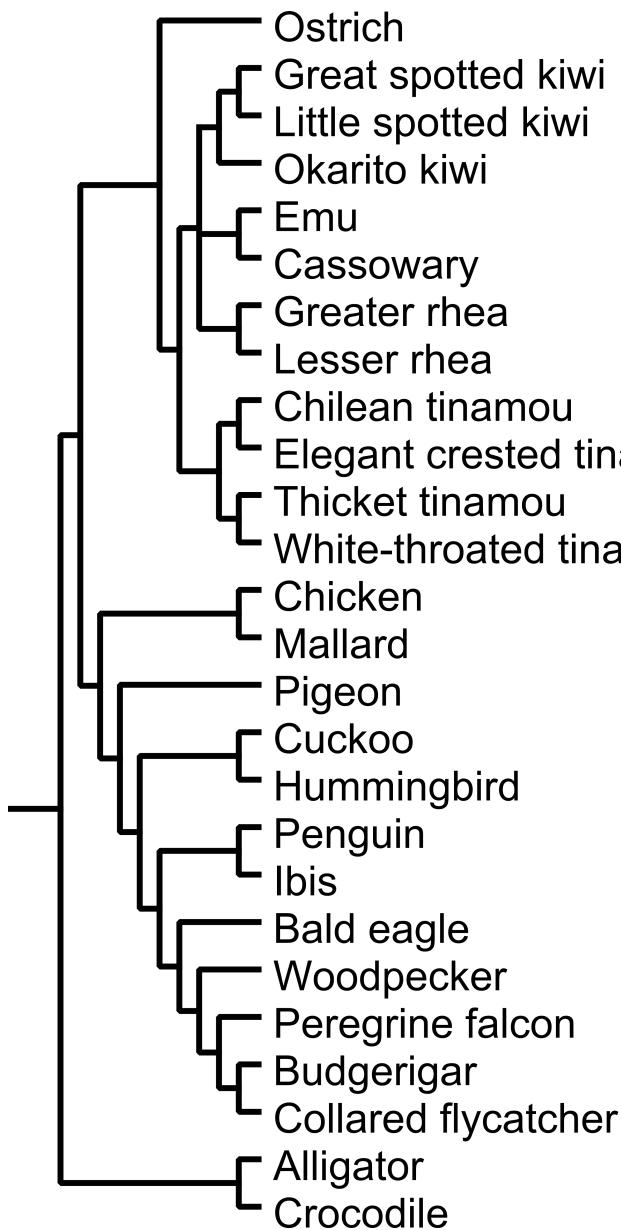
MOTOO KIMURA

Photo by William B. Proctor

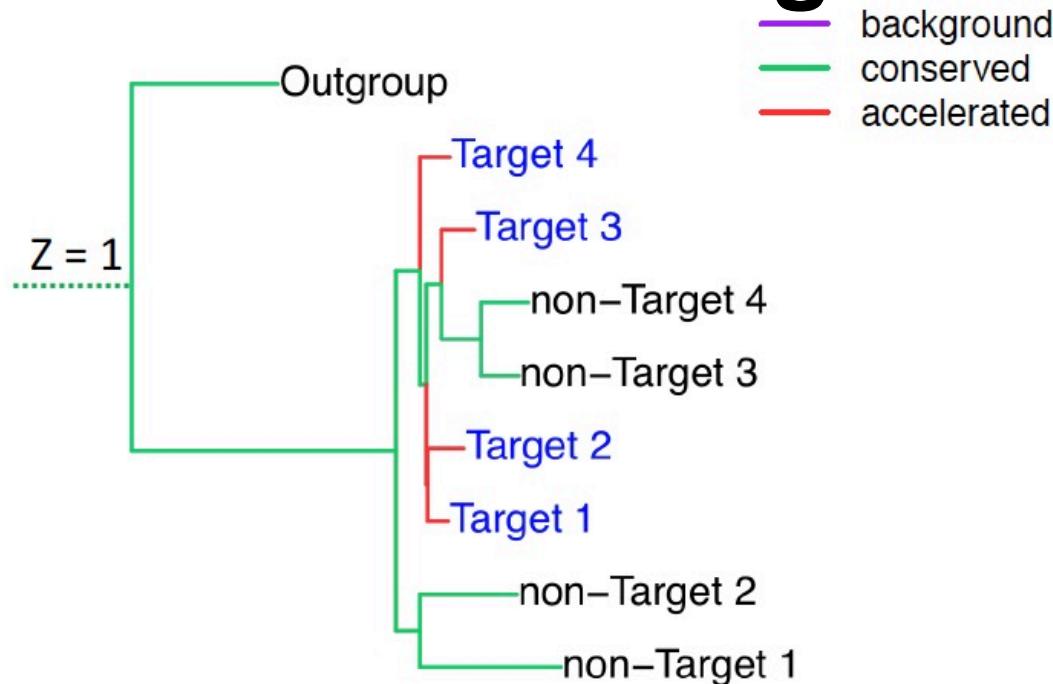
**Motoo Kimura
(1924-1994)**

Functionally less important neutral mutations or germline errors will have a high 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

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

α = probability of gain of conserved state

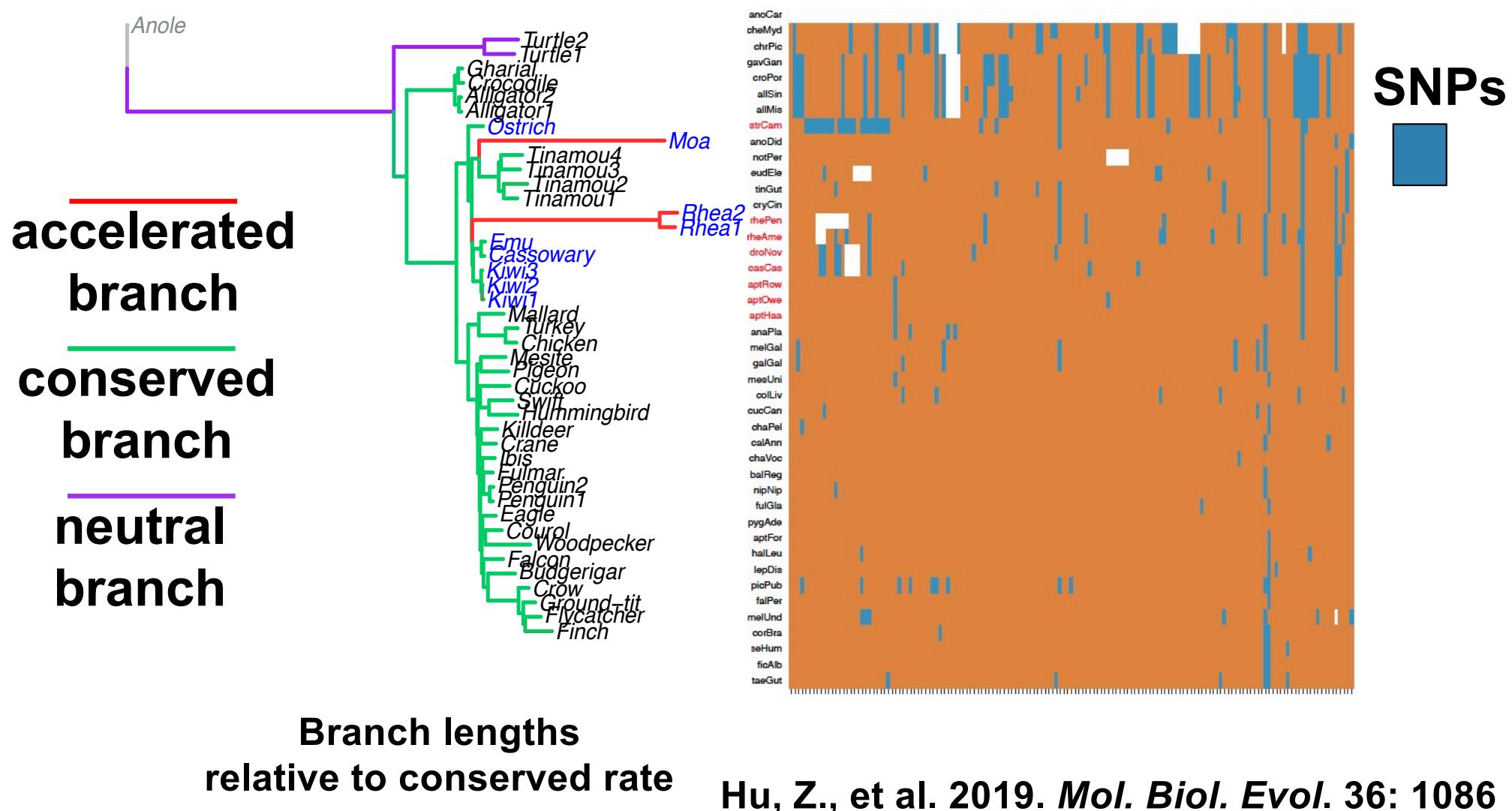
β = probability of loss of conserved state

For branch s ,

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

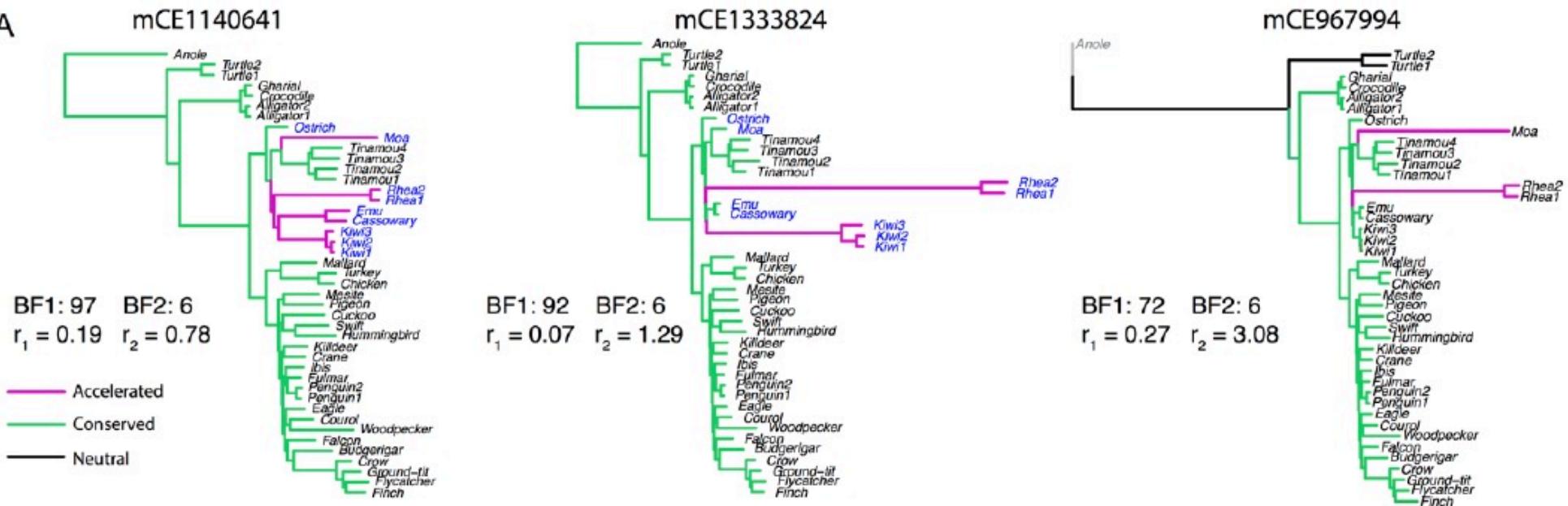
$$BF1 = \frac{P(Y|M_1)}{P(Y|M_0)} \text{ and } BF2 = \frac{P(Y|M_1)}{P(Y|M_2)}$$

A convergently accelerated CNEE detected with a novel Bayesian method



Additional examples of convergently accelerated CNEEs

A



Rapid regulatory evolution near developmental genes

