

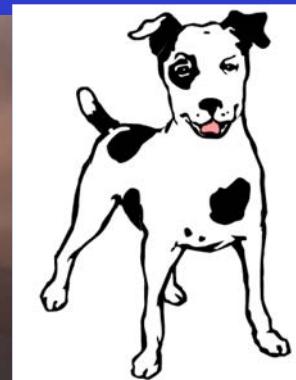
Whole-genome phylogenetics, gene regulation and the origin of evolutionary novelty

Feather photos: J. Trimble, MCZ

“Beast Legends”: A six part adventure in science and myth



Griffin



yap films
off the
leash tv



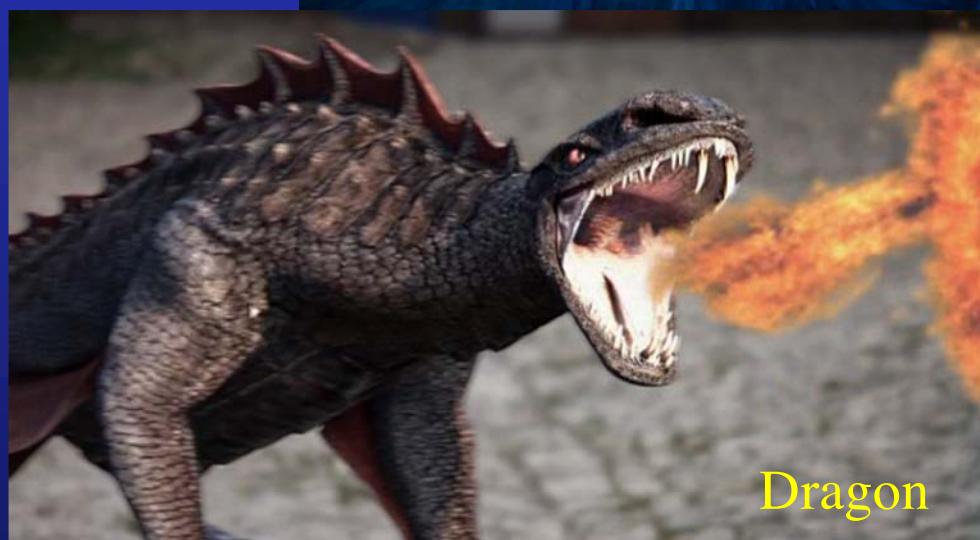
Fijian shark god



Wild man



Kraken



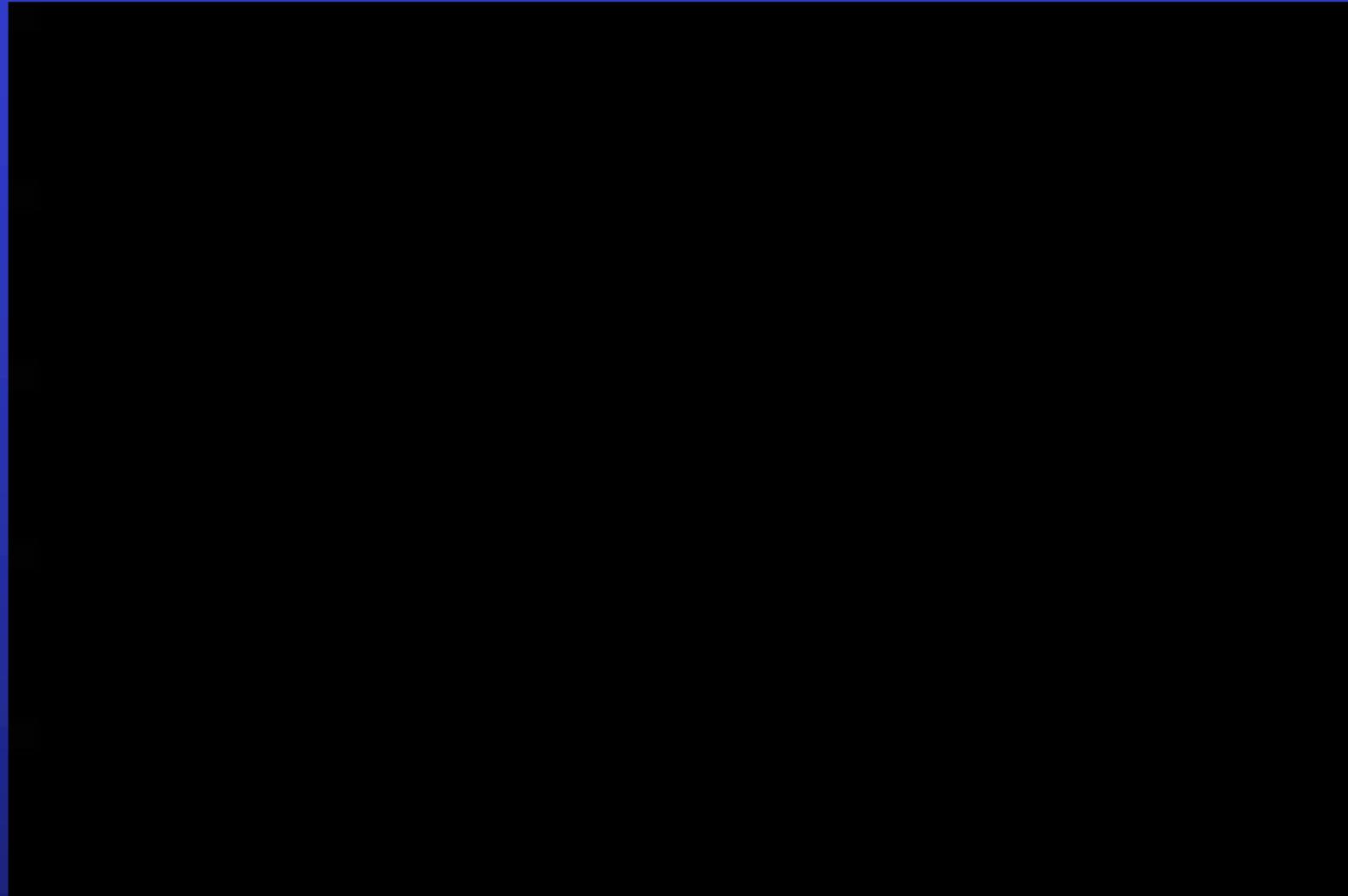
Dragon



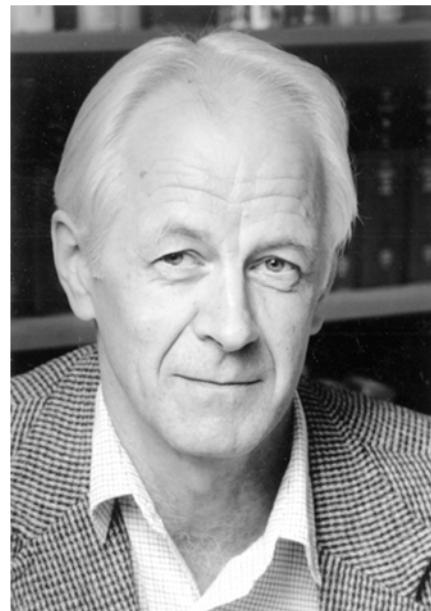
Terror bird

Graphics by Invisible Pictures, Inc.

Beast Legends – Griffin episode



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.

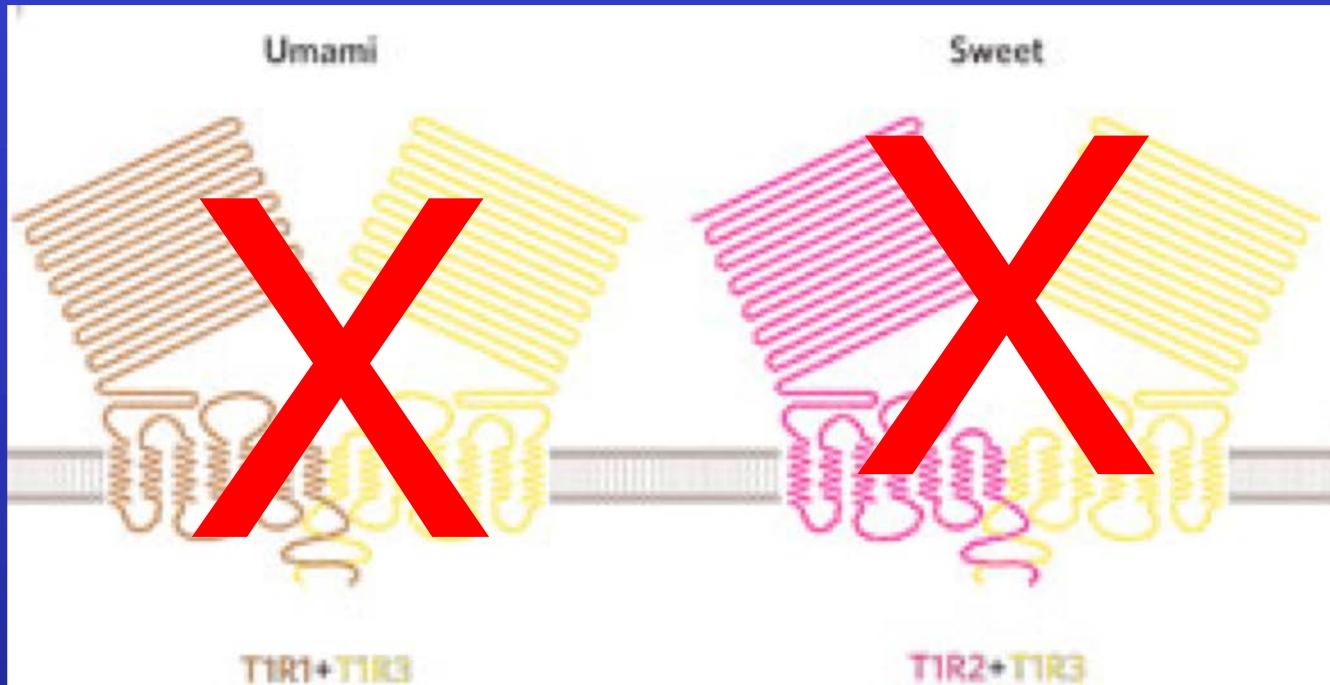


Mary-Claire King and A. C. Wilson

SCIENCE

11 April 1975, Volume 188, Number 4184

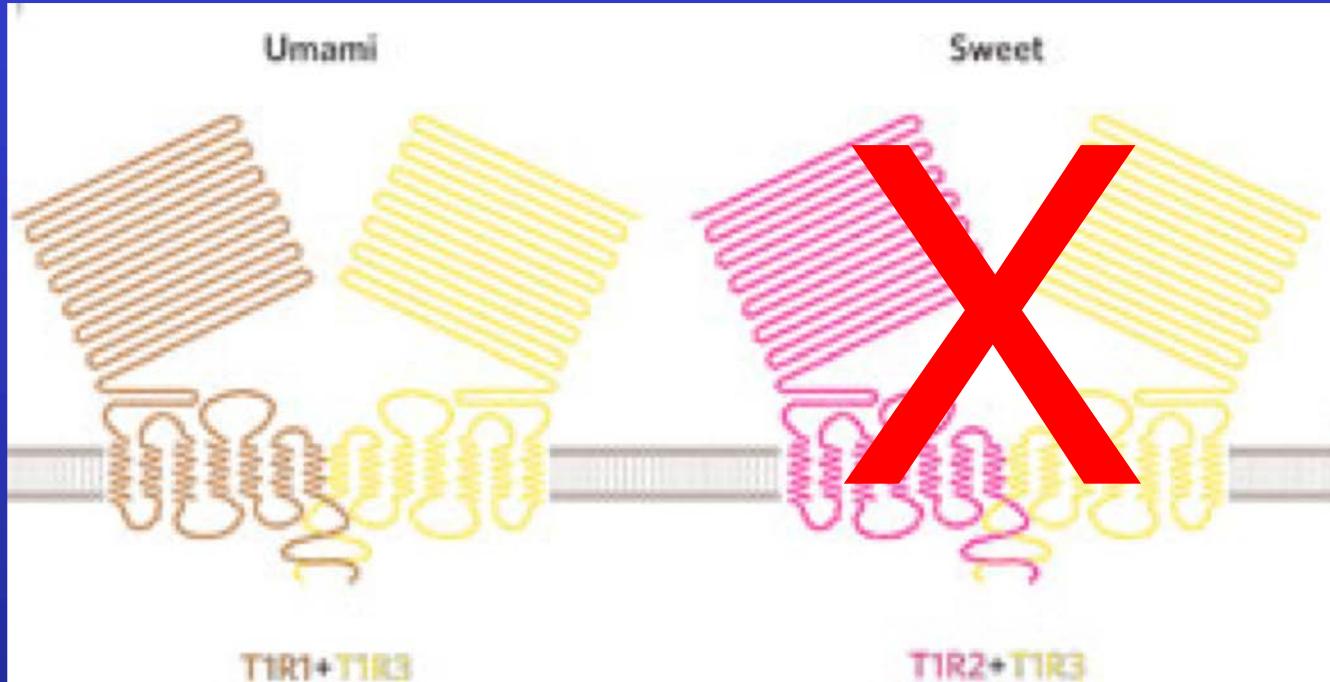
Taste receptors in mammals



Taste receptors on
the tongue

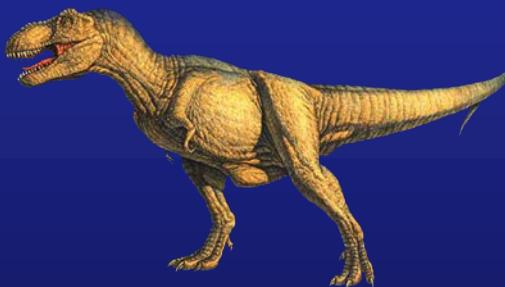


Birds inherited only the umami (meat) receptor from their dinosaur ancestors



el receptor del gusto dulce se perdió

incluso en colibríes !



Hummingbirds can taste **sugar** due to **changes** in the gene other birds use to taste meat (or insects)



Can taste **sugar**?

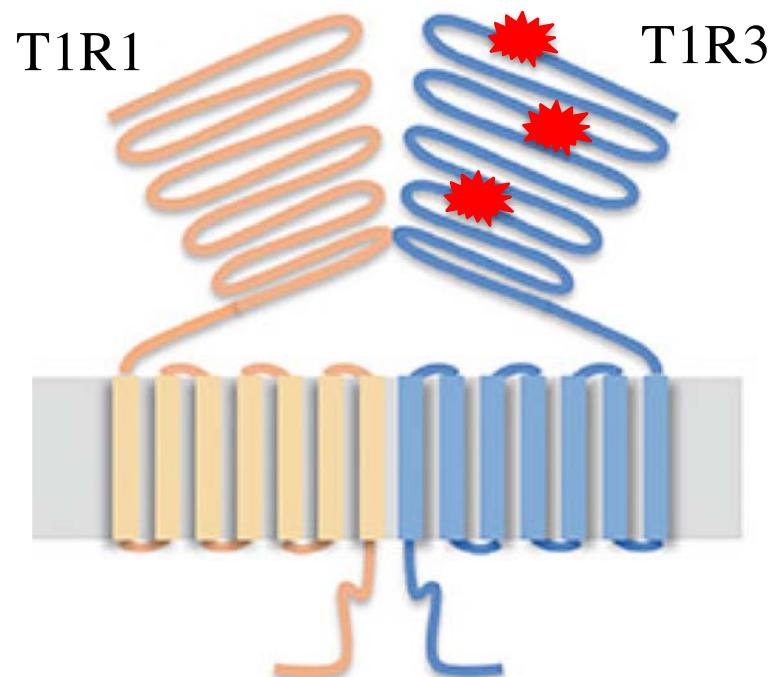


X



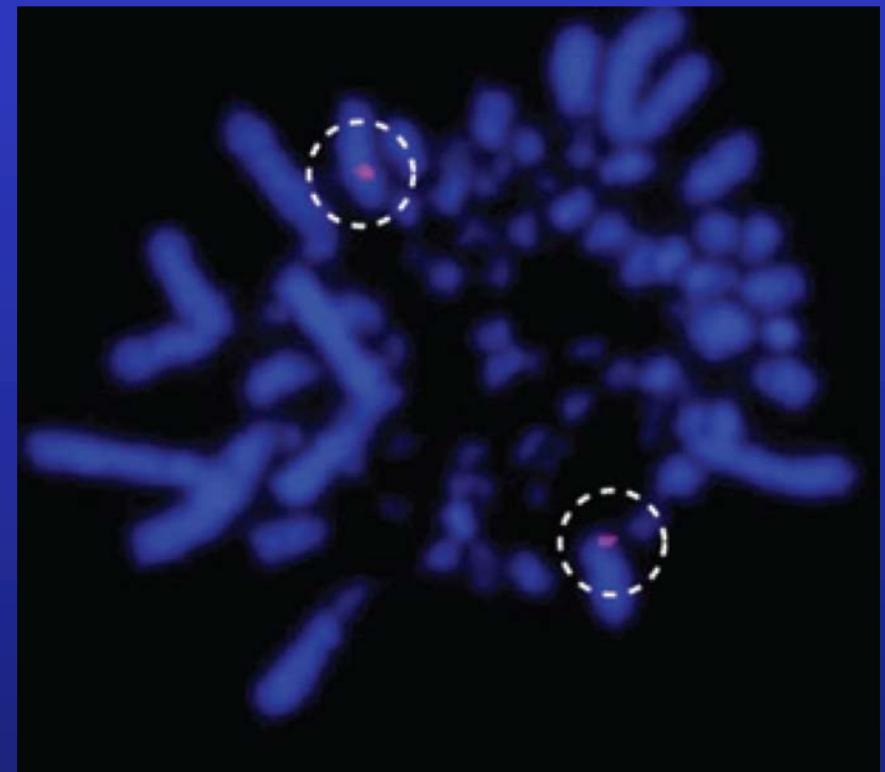
X

Taste receptor



Baldwin et al. 2014. *Science* 345: 929-933

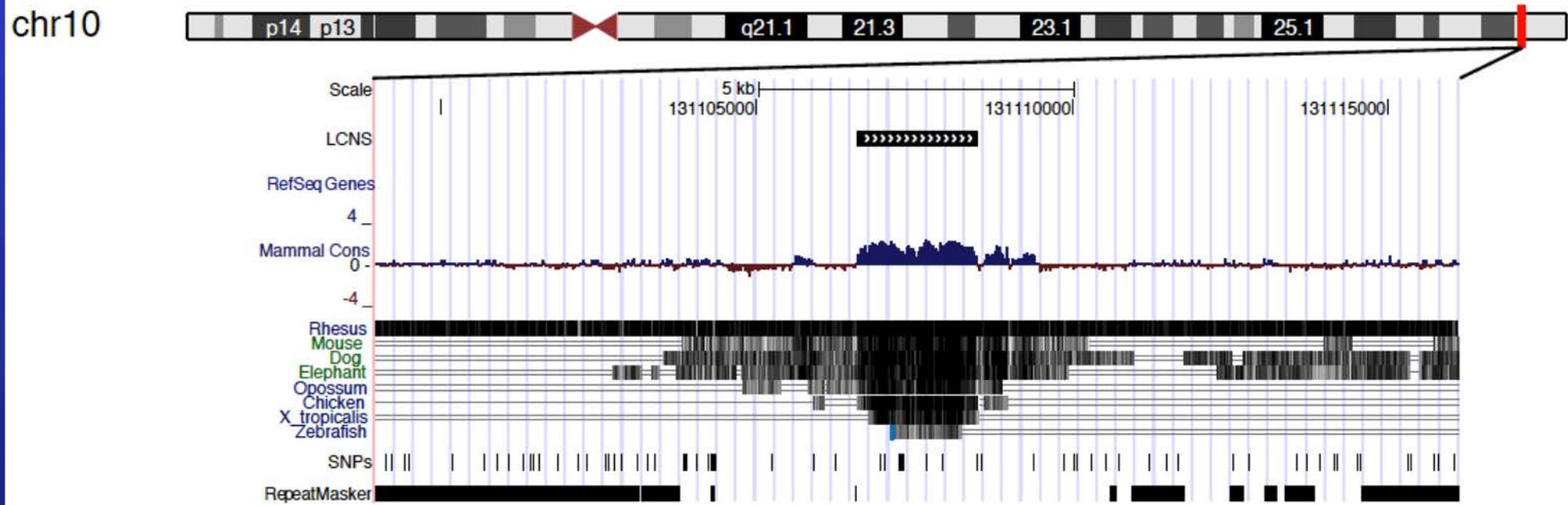
Non-coding ‘Dark matter’ of the genome: a regulatory network?



Karyotype of an Emu

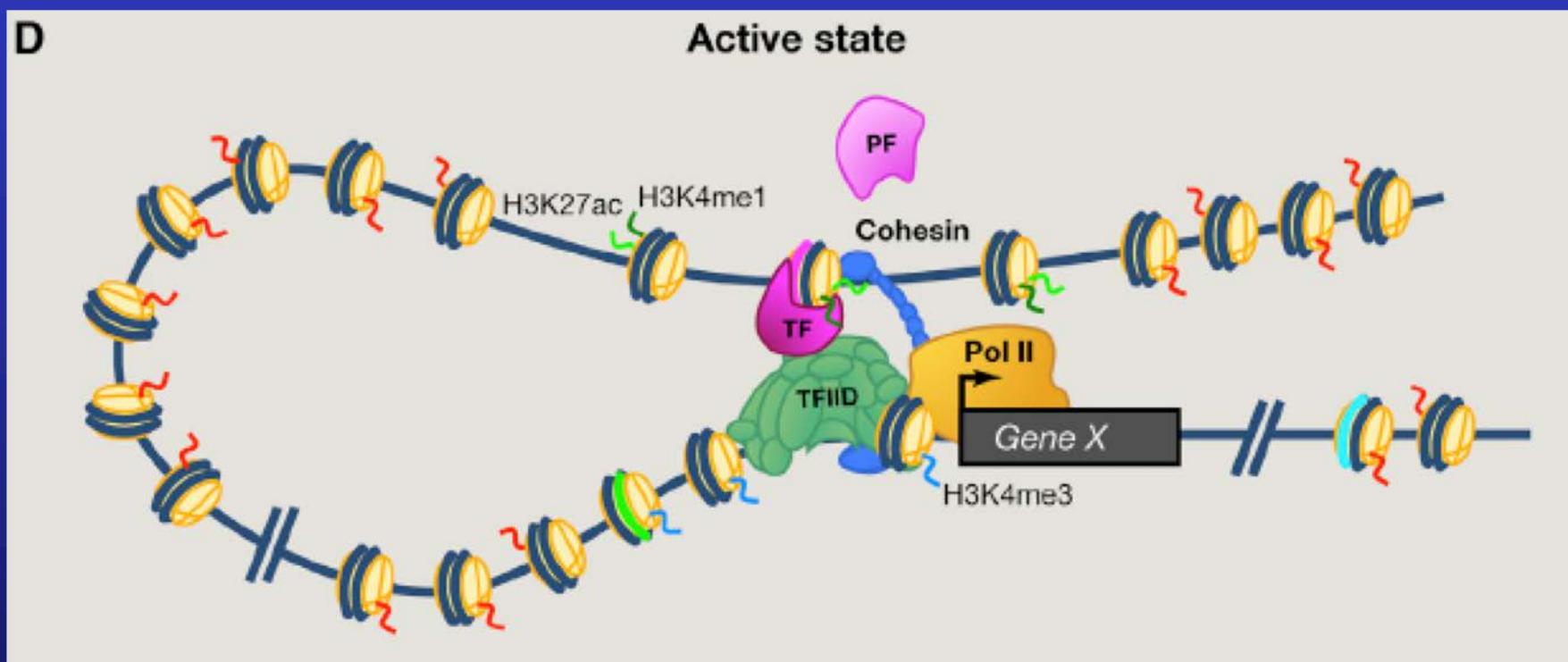
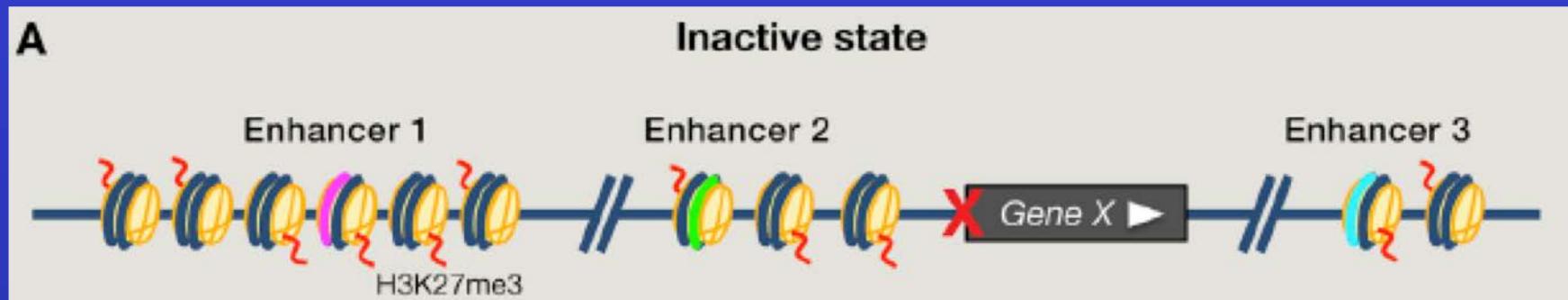
CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements

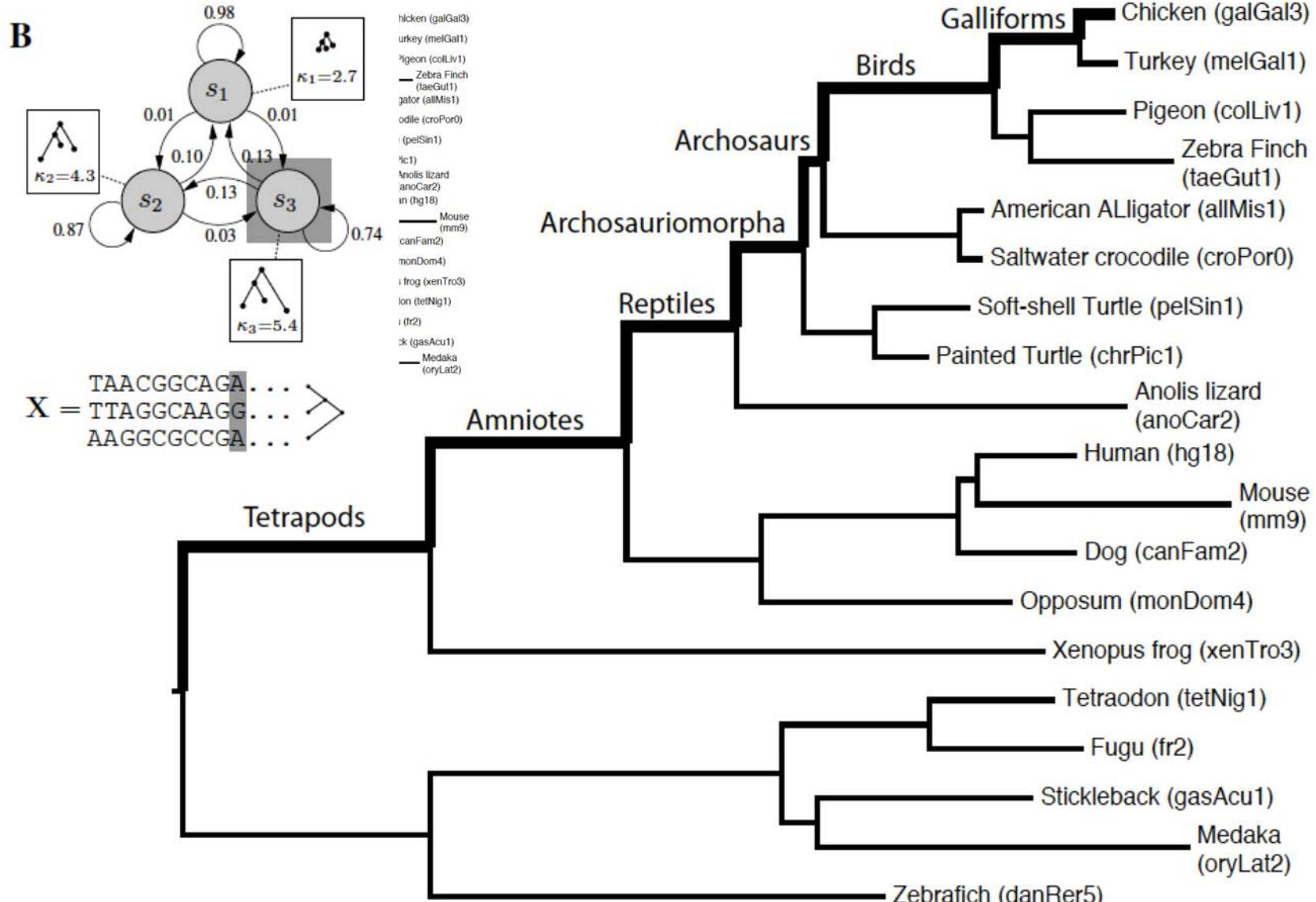


View of a segment of human chromosome 10 using UCSC Genome Browser

Noncoding enhancers: long-range control of gene expression



Phylogenetic hidden Markov model detects CNEEs using Phastcons*



*Siepel et al. 2005. *Genome Res.* 15:1034-1050

A role for gene regulation in the origin of feathers



Sinosauropeltaopteryx



Archaeopteryx

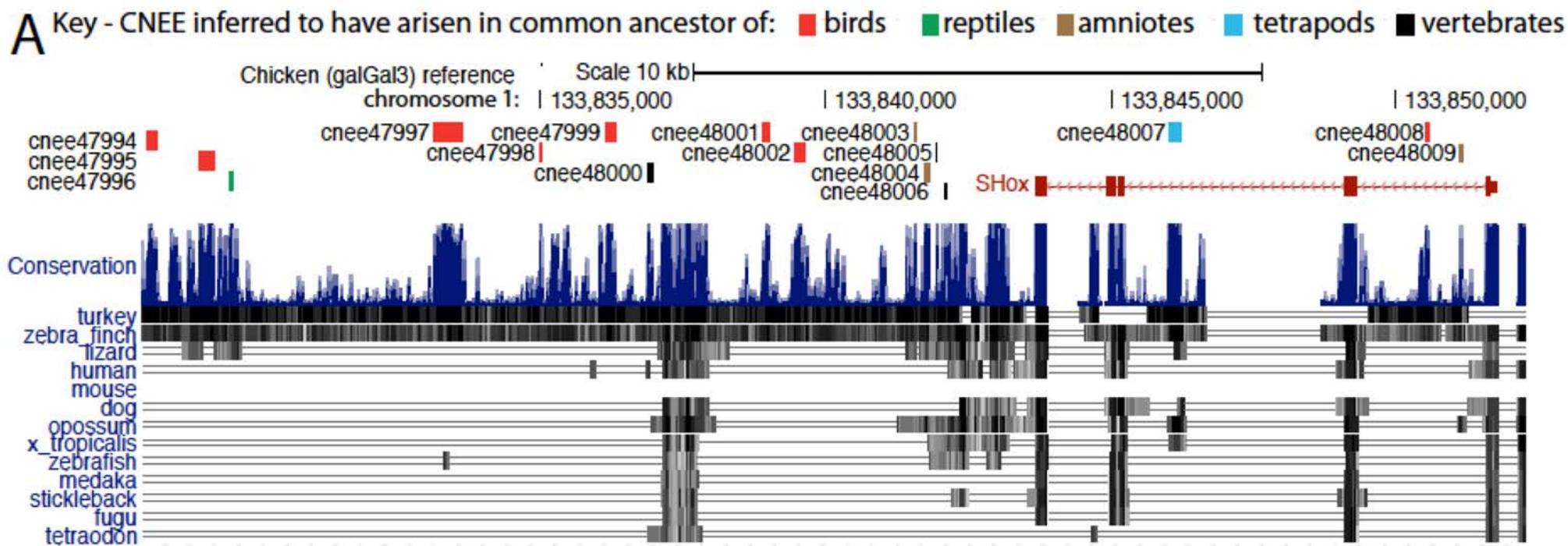


Quanguo et al. (2010) *Science*
Anchiornis

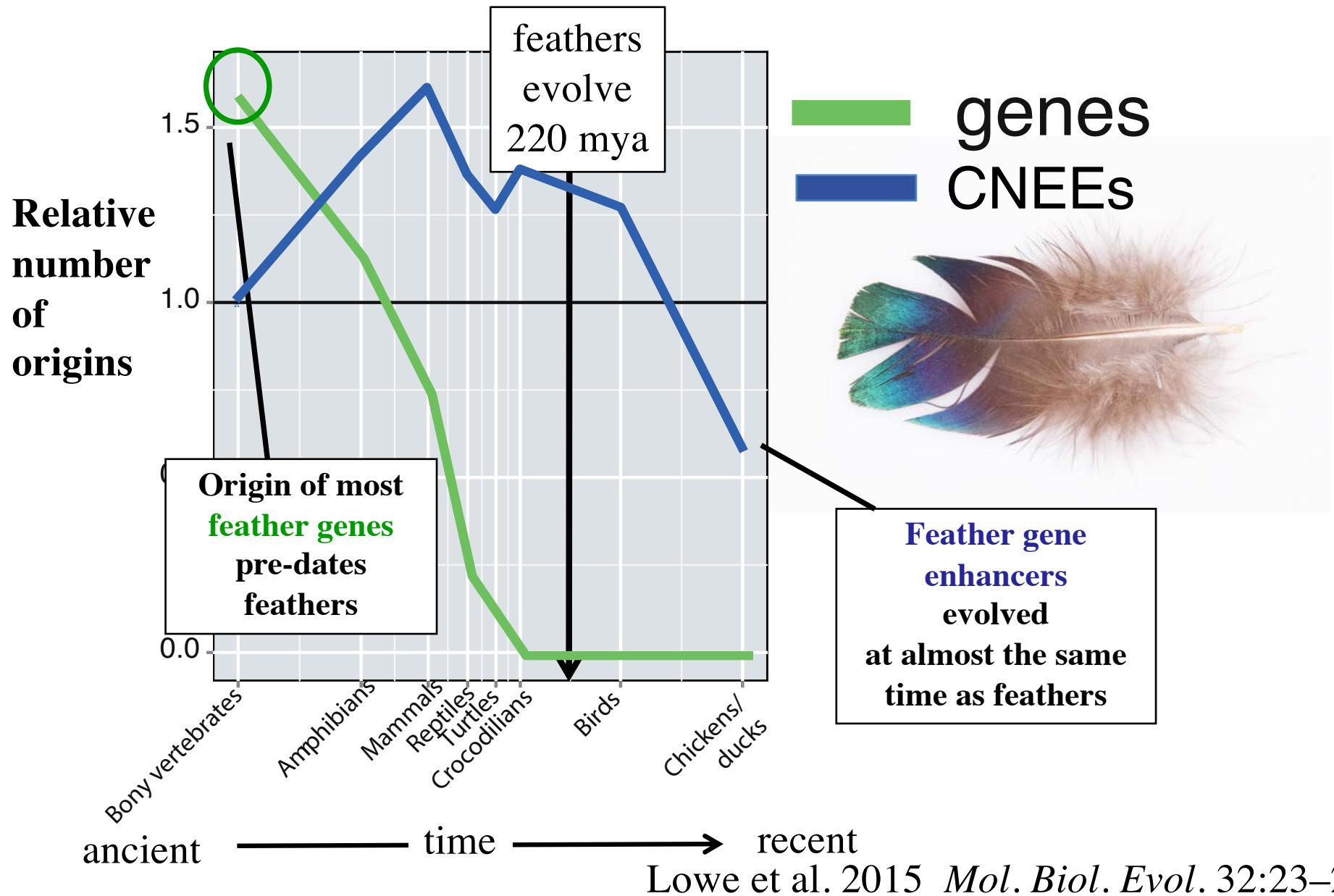
Feather photos: J. Trimble, MCZ

Conserved non-exonic elements (CNEEs) act as enhancers for feather genes

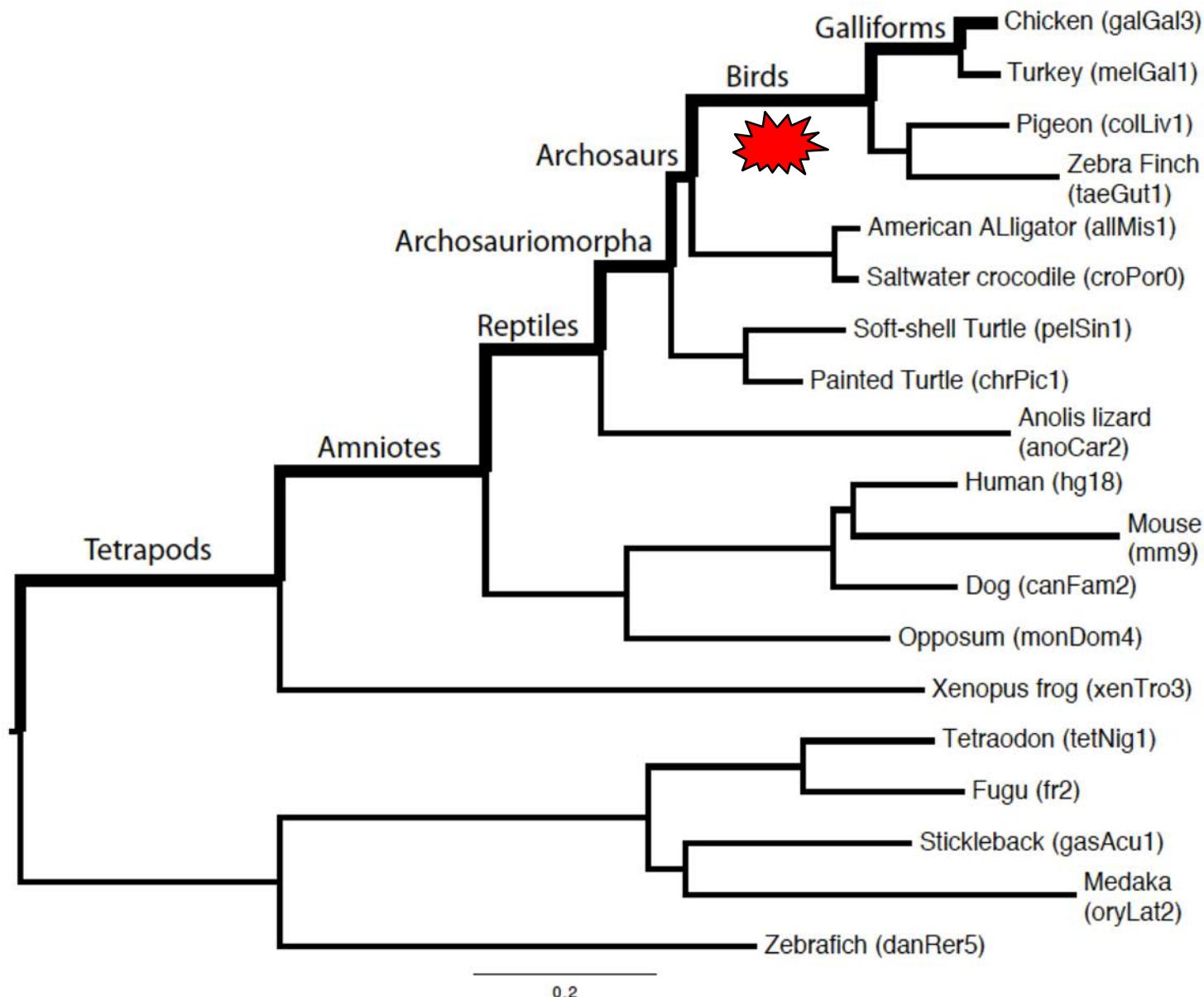
Bird, amniote- and tetrapod-specific CNEEs near SHox



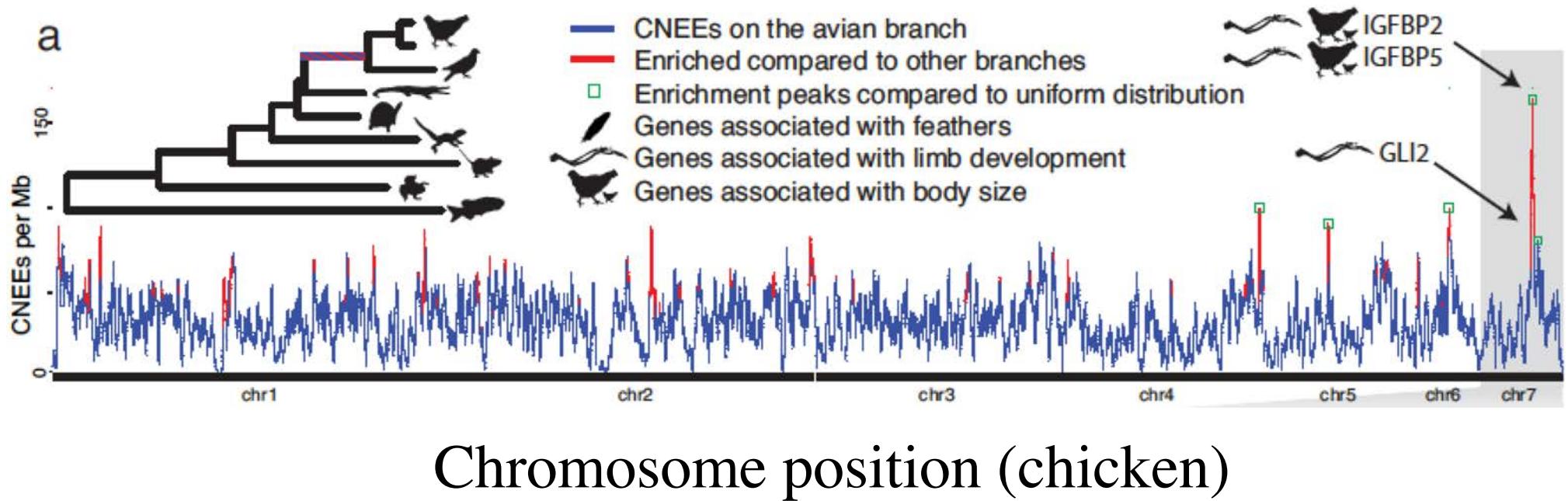
High origination rates of feather CNEEs, but not feather genes, when feathers evolved



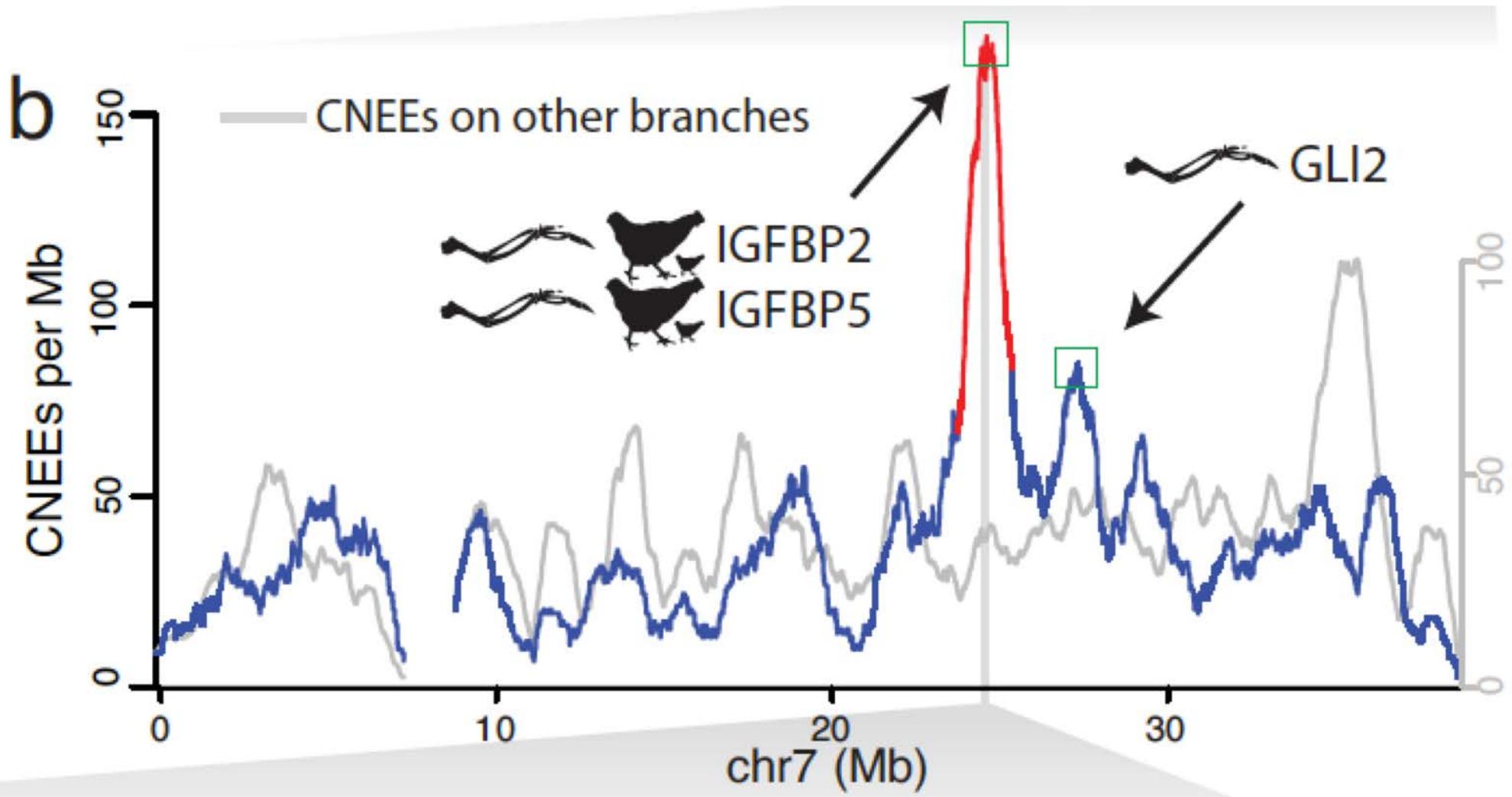
Bird-specific regulatory evolution: what makes a bird a bird?



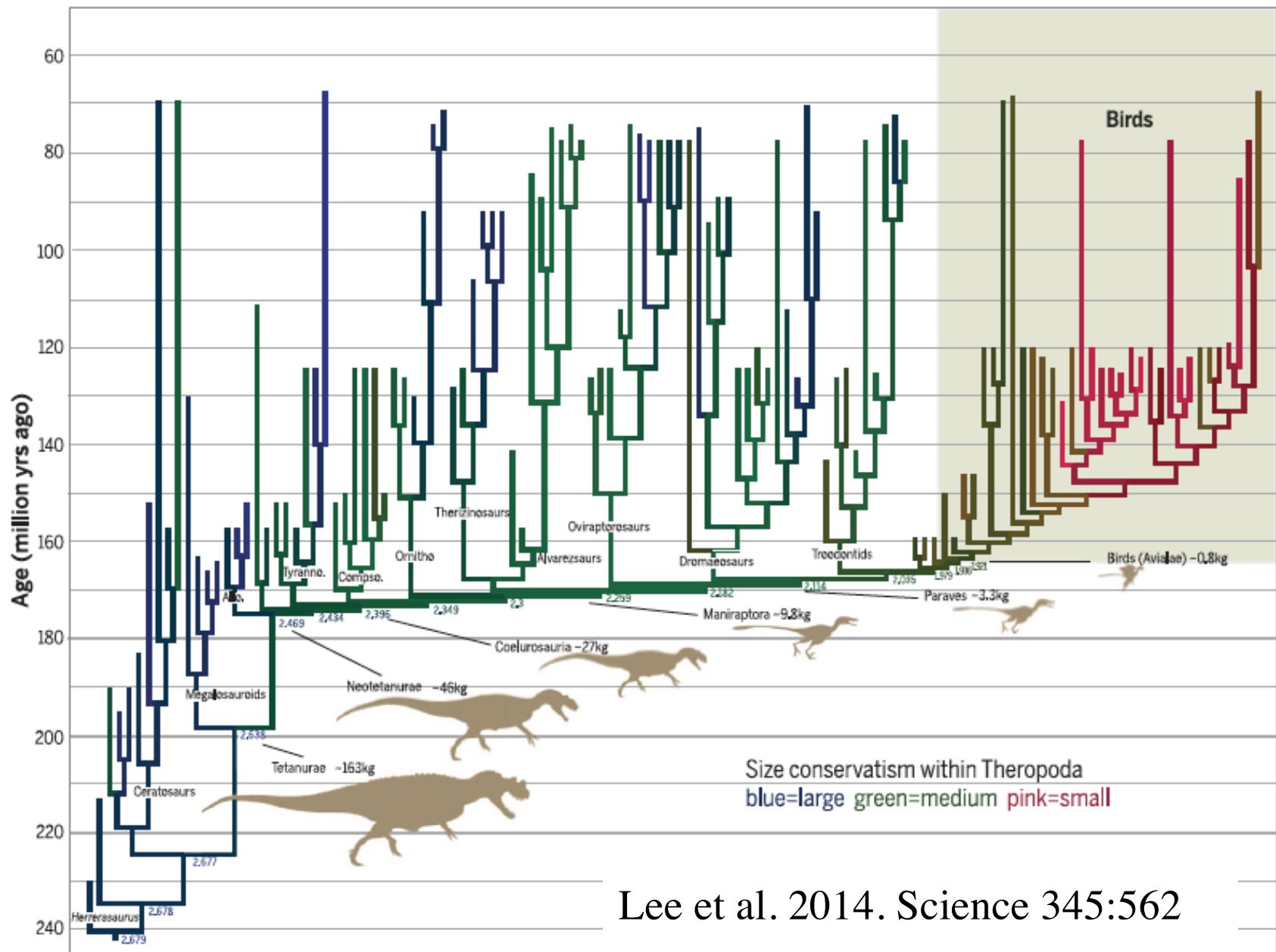
Bird-specific CNEEs associated with genes for limb and body size evolution



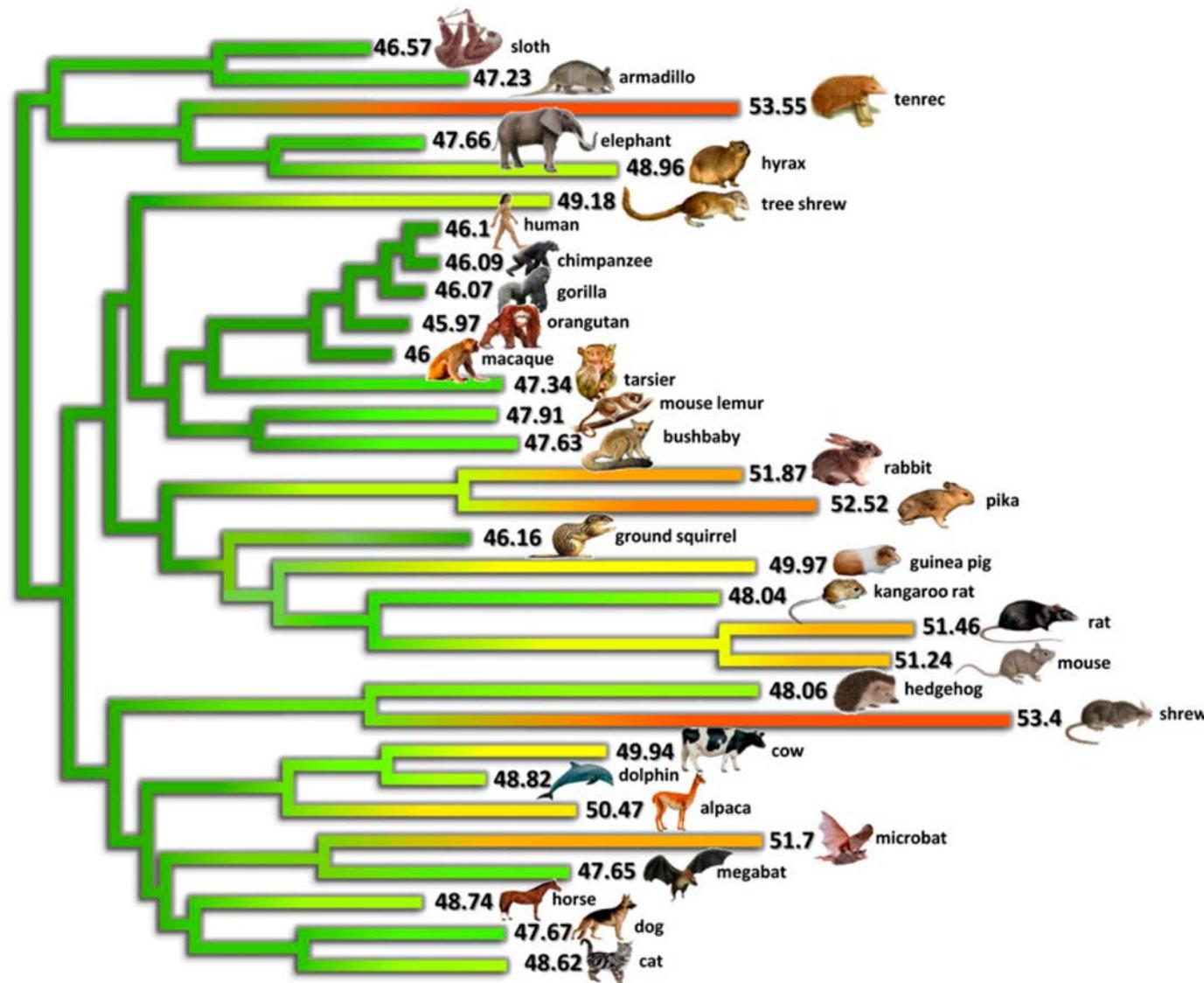
Bird-specific CNEEs associated with genes for limb and body size evolution



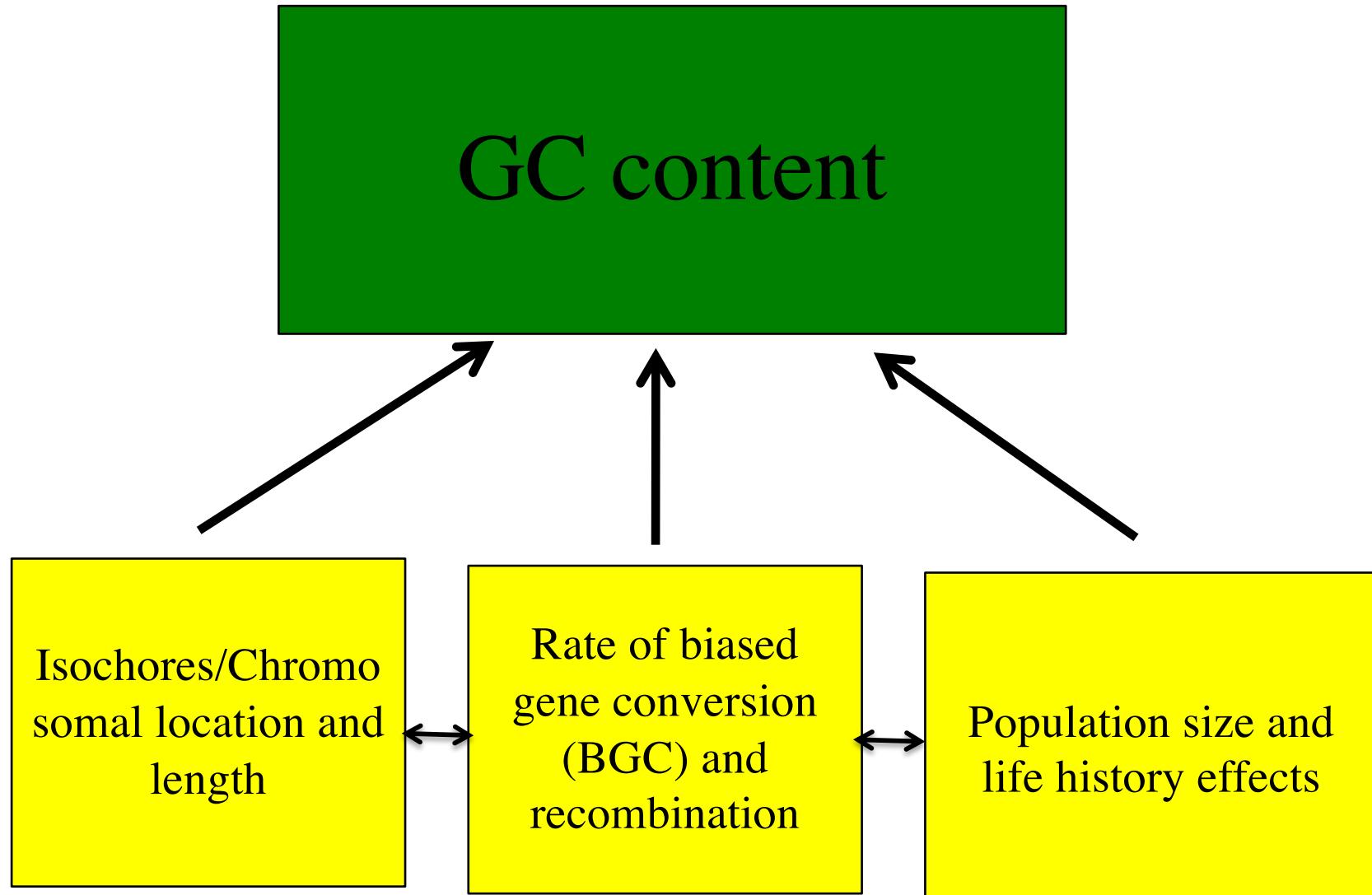
Bird-specific CNEEs associated with genes for limb and body size evolution



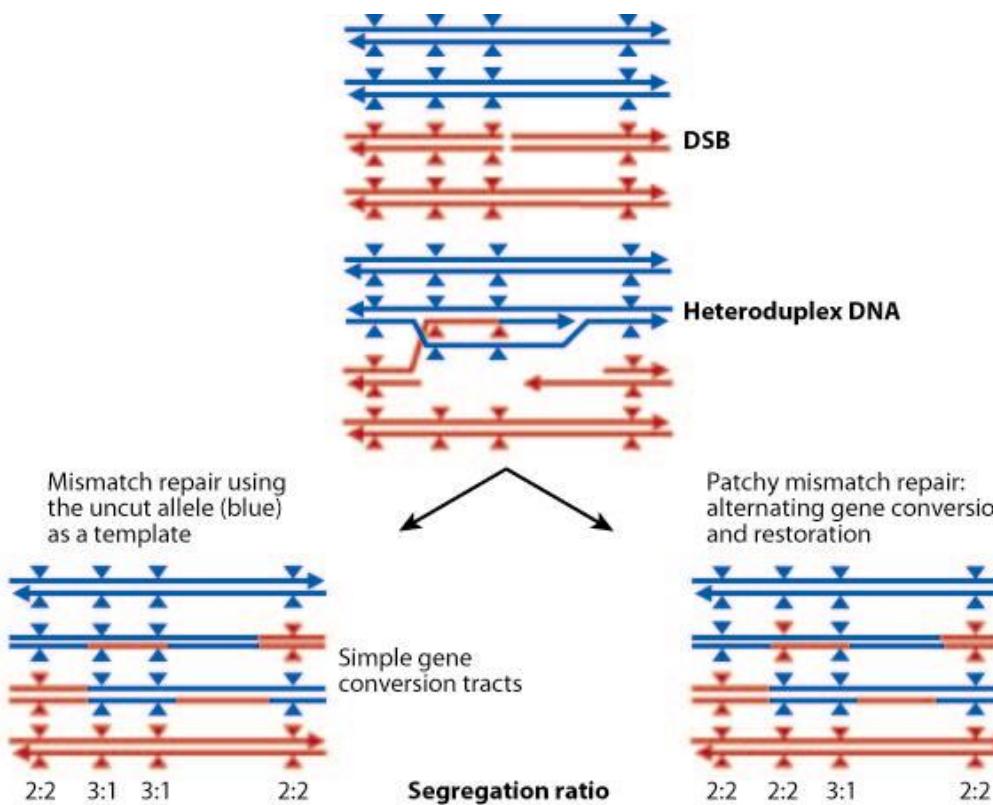
Correlates of GC content in ~1000 mammalian coding regions



Mechanistic hypotheses for high GC content in a lineage



Consequences of repair of double strand breaks induced by BGC or recombination



Frequency of GC alleles among gametes from a heterozygous individual:

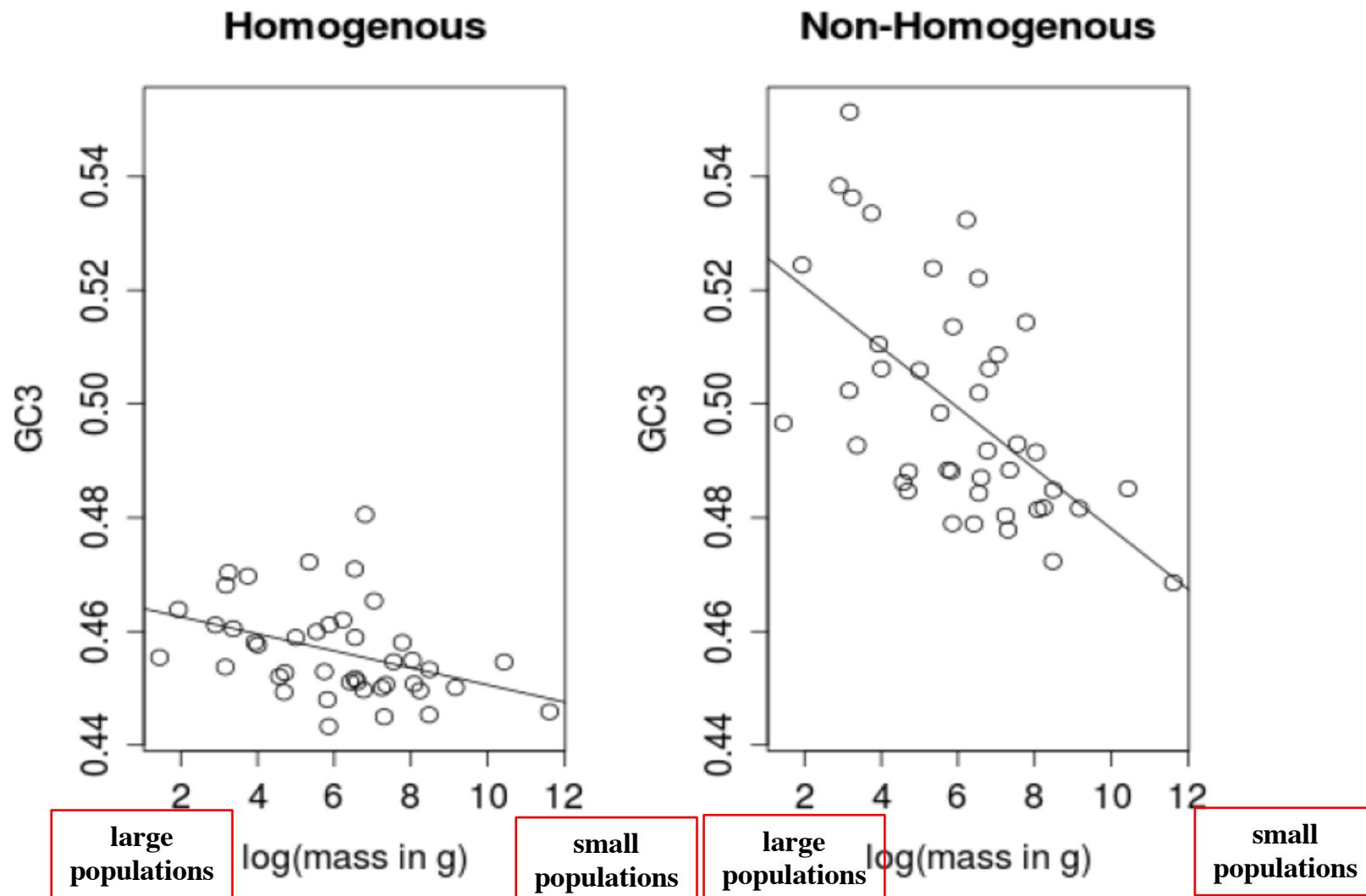
$$x_{GC} = \frac{1}{2}(1 + b)$$

$$P(AT \rightarrow GC) = \frac{1 - e^{-2b}}{1 - e^{-4Neb}}$$

(b =BGC bias;
 Ne = effective population size)

A Duret L, Galtier N. 2009.
R Annu. Rev. Genomics Hum. Genet. 10:285–311

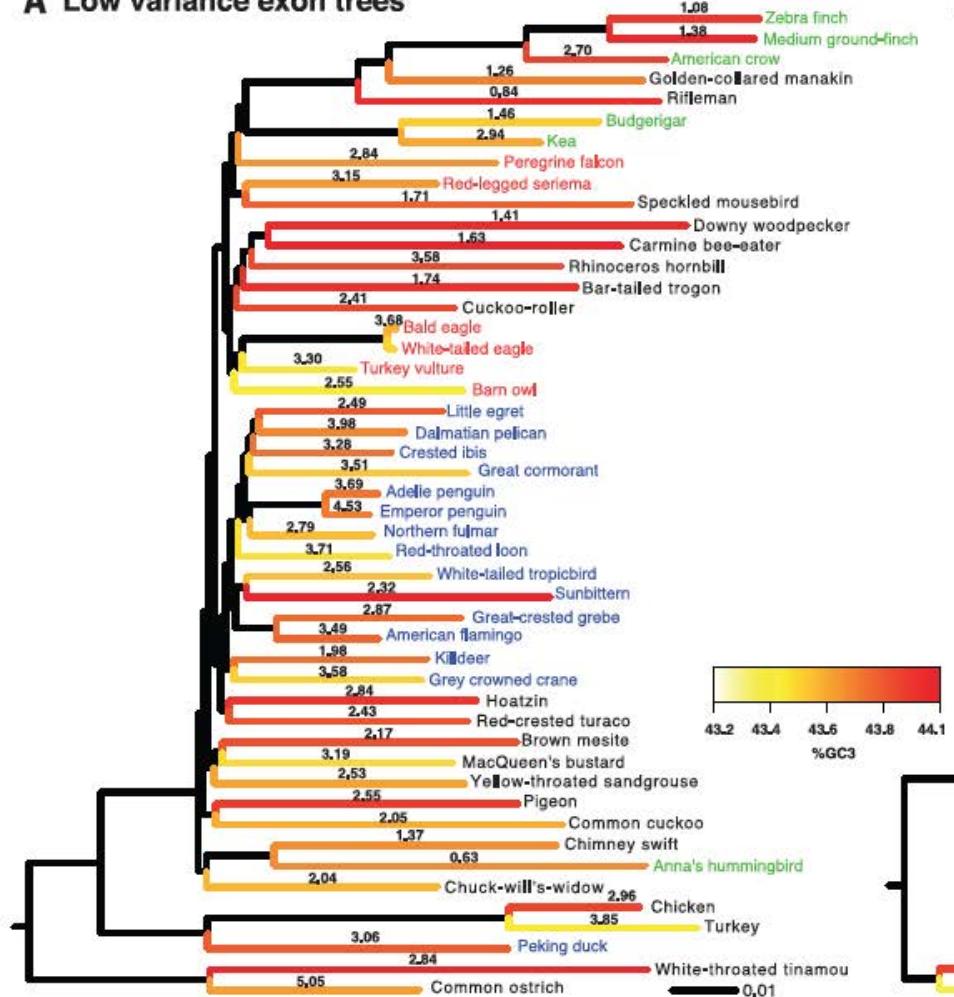
Covariation of GC content and body mass in birds



Effects of GC variation on phylogenomic analysis

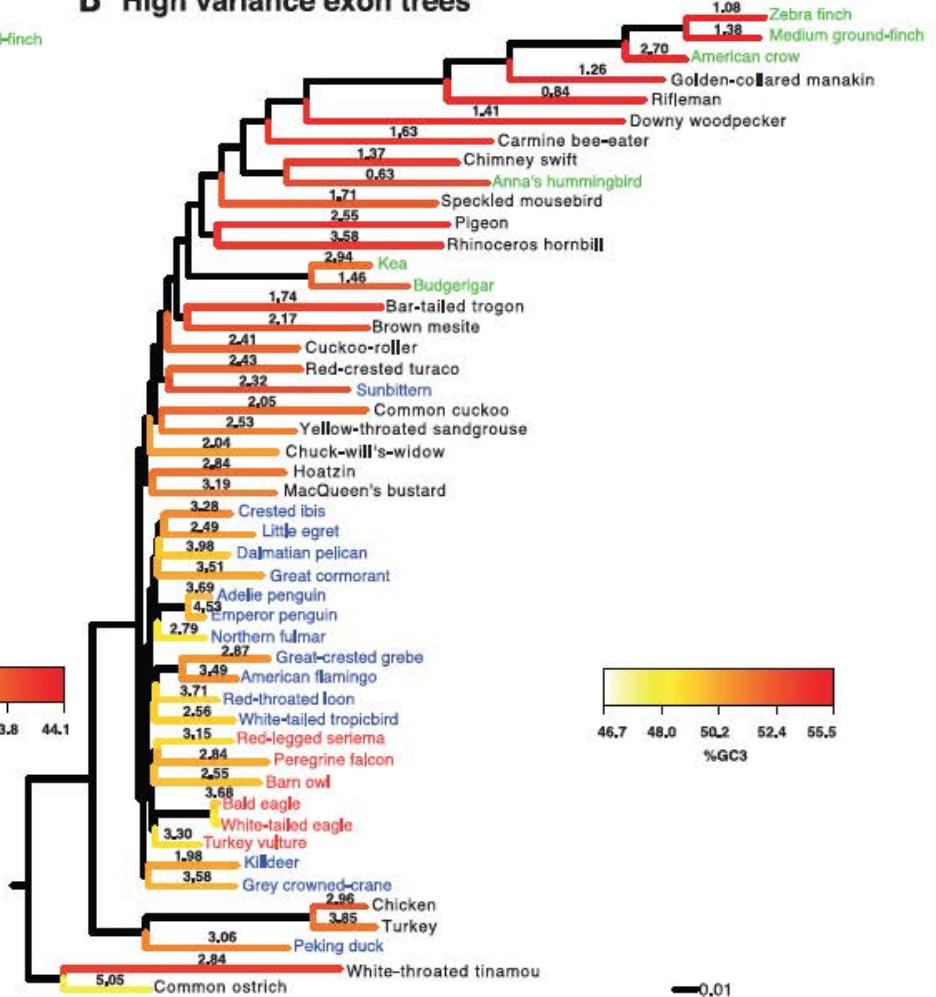
Including only genes with low-GC variation among lineages

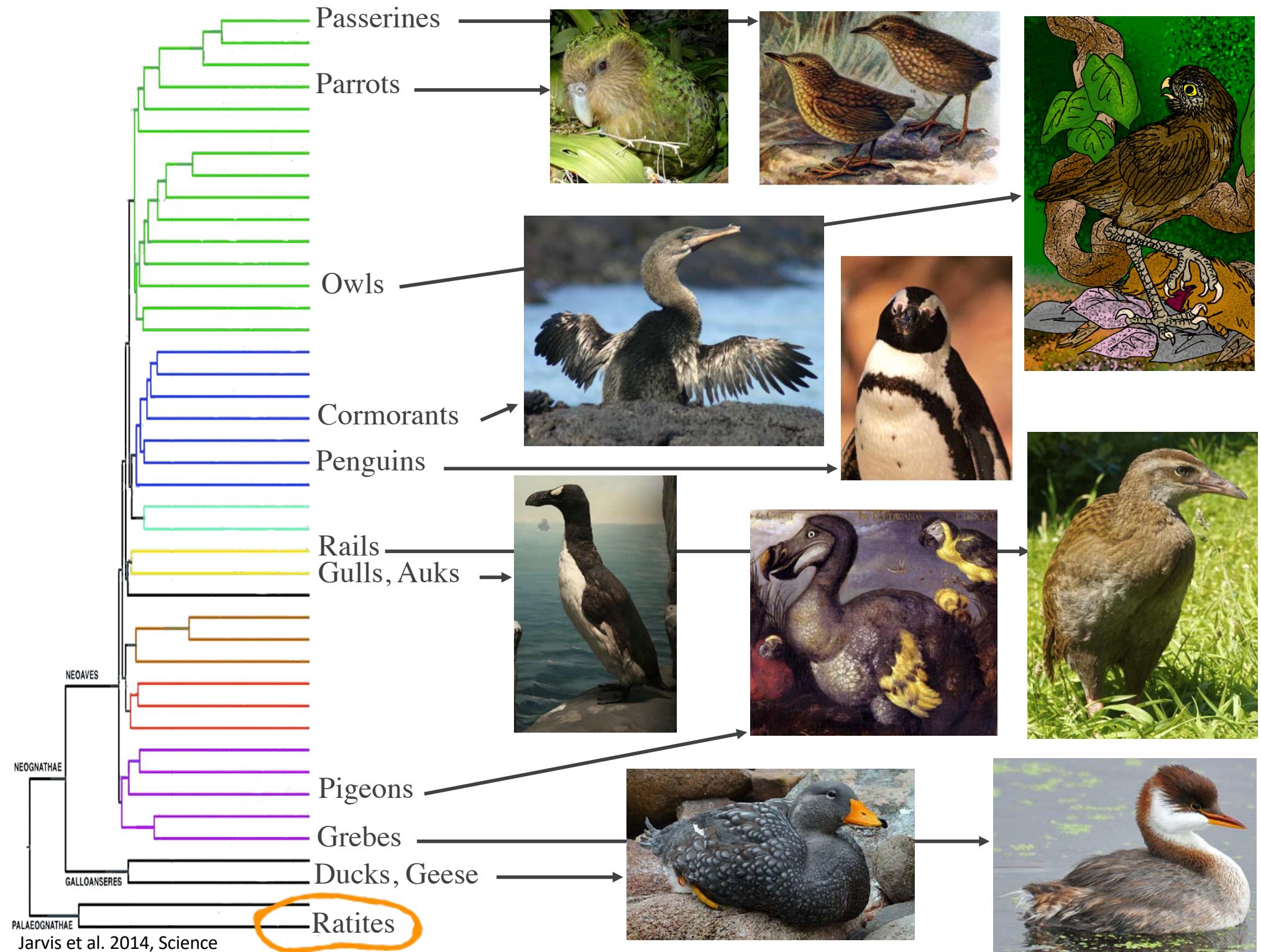
A Low variance exon trees



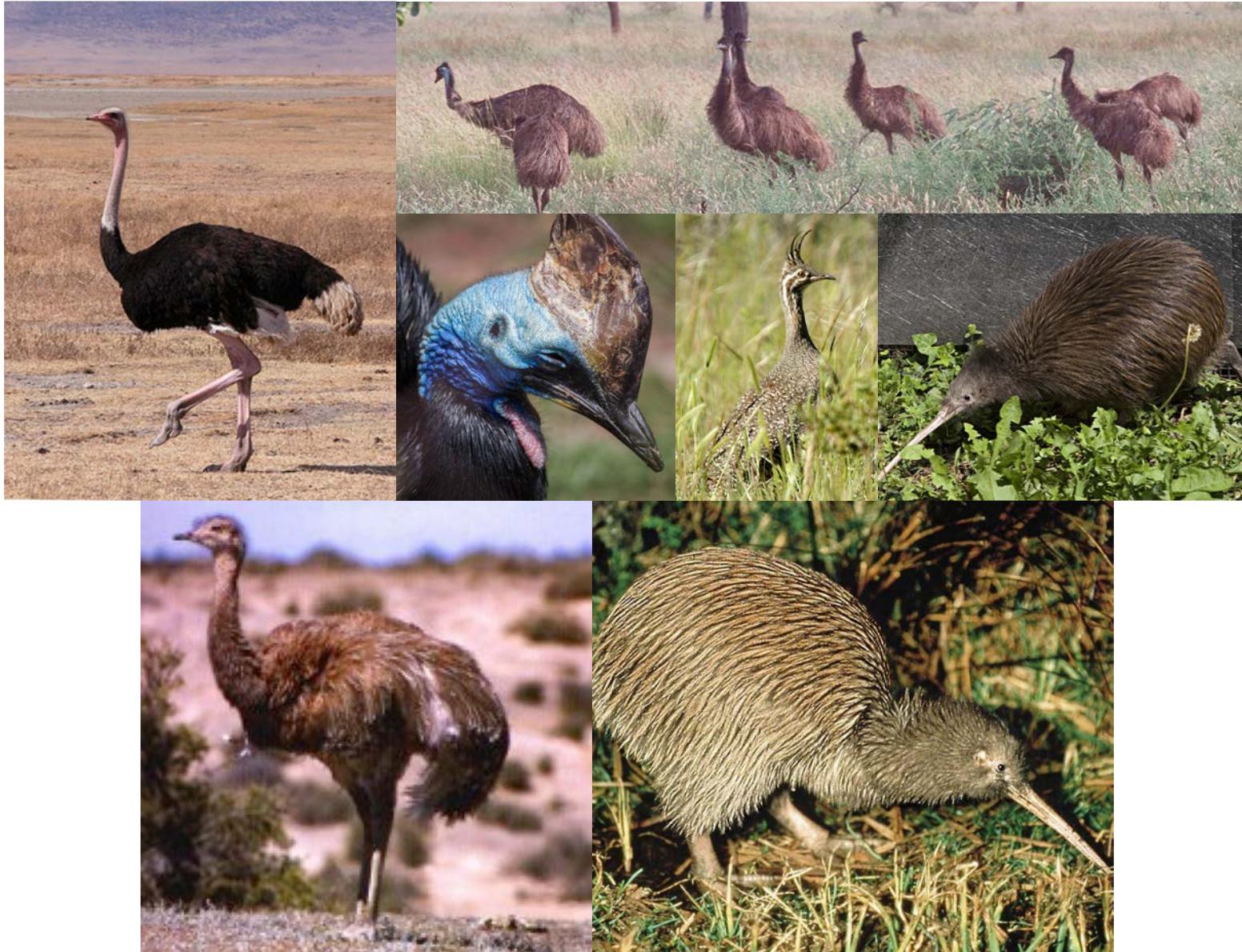
Including only genes with high-GC variation among lineages

B High variance exon trees

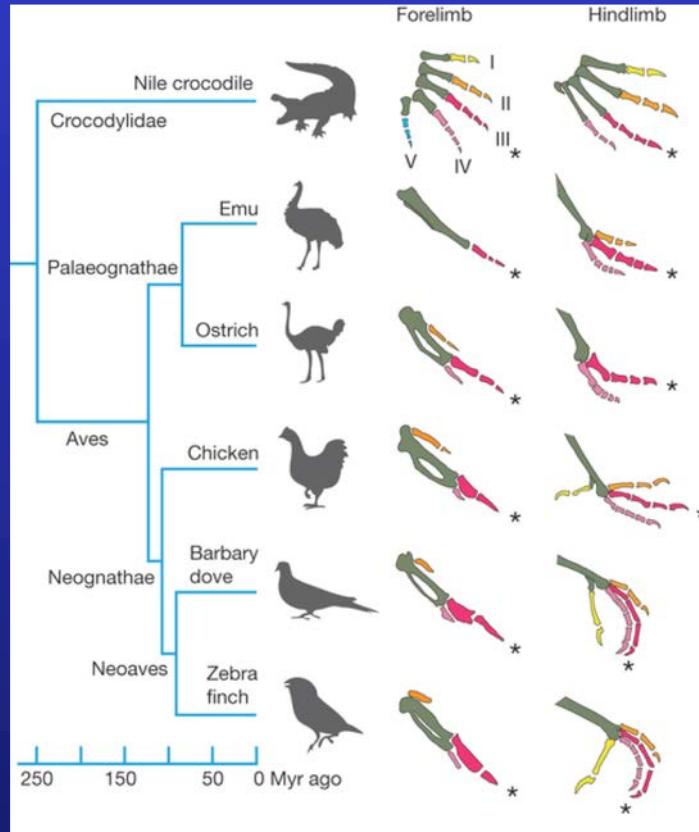




CNEEs and the convergent evolution of flightlessness in Palaeognathae



Skeletal modifications for flightlessness



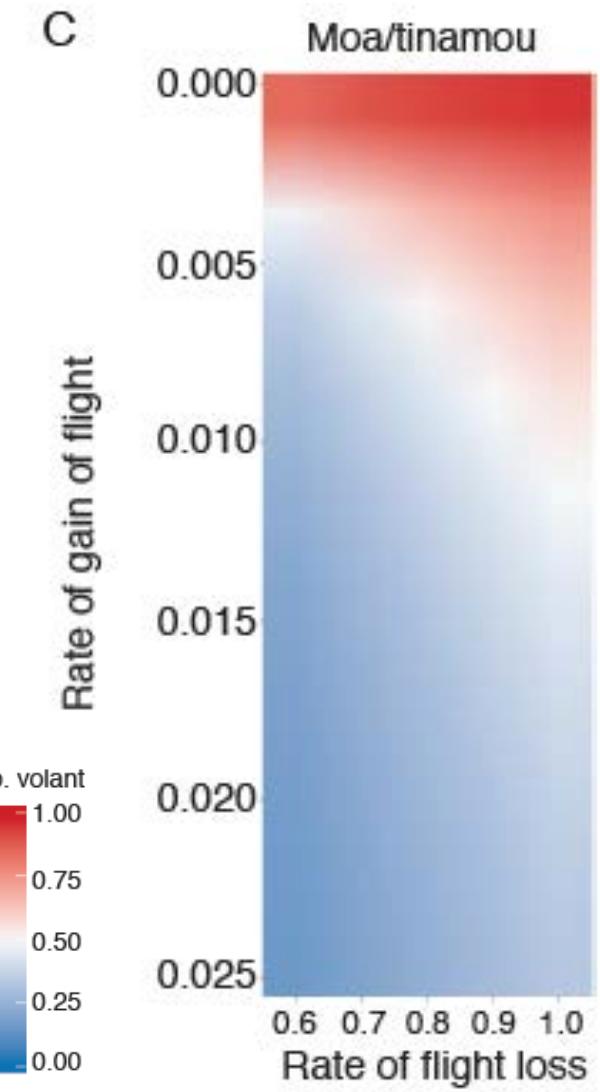
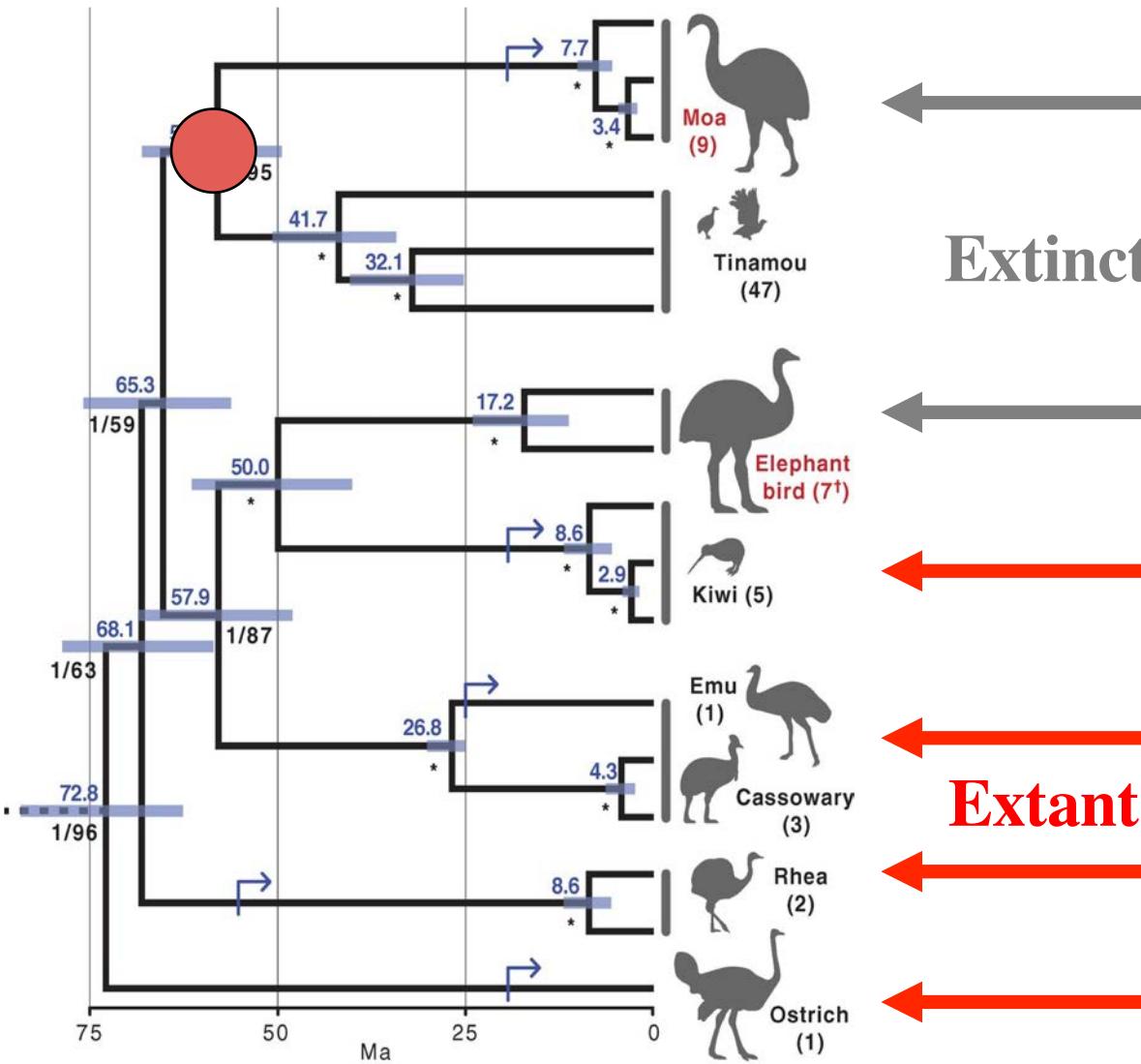
Little-spotted kiwi sternum



Emu and ostrich keelless sterna

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

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



11 new palaeognath genomes



Little Spotted Kiwi



Little bush moa



Great-spotted Kiwi



Lesser Rhea



Southern Cassowary



Emu



Elegant-crested Tinamou



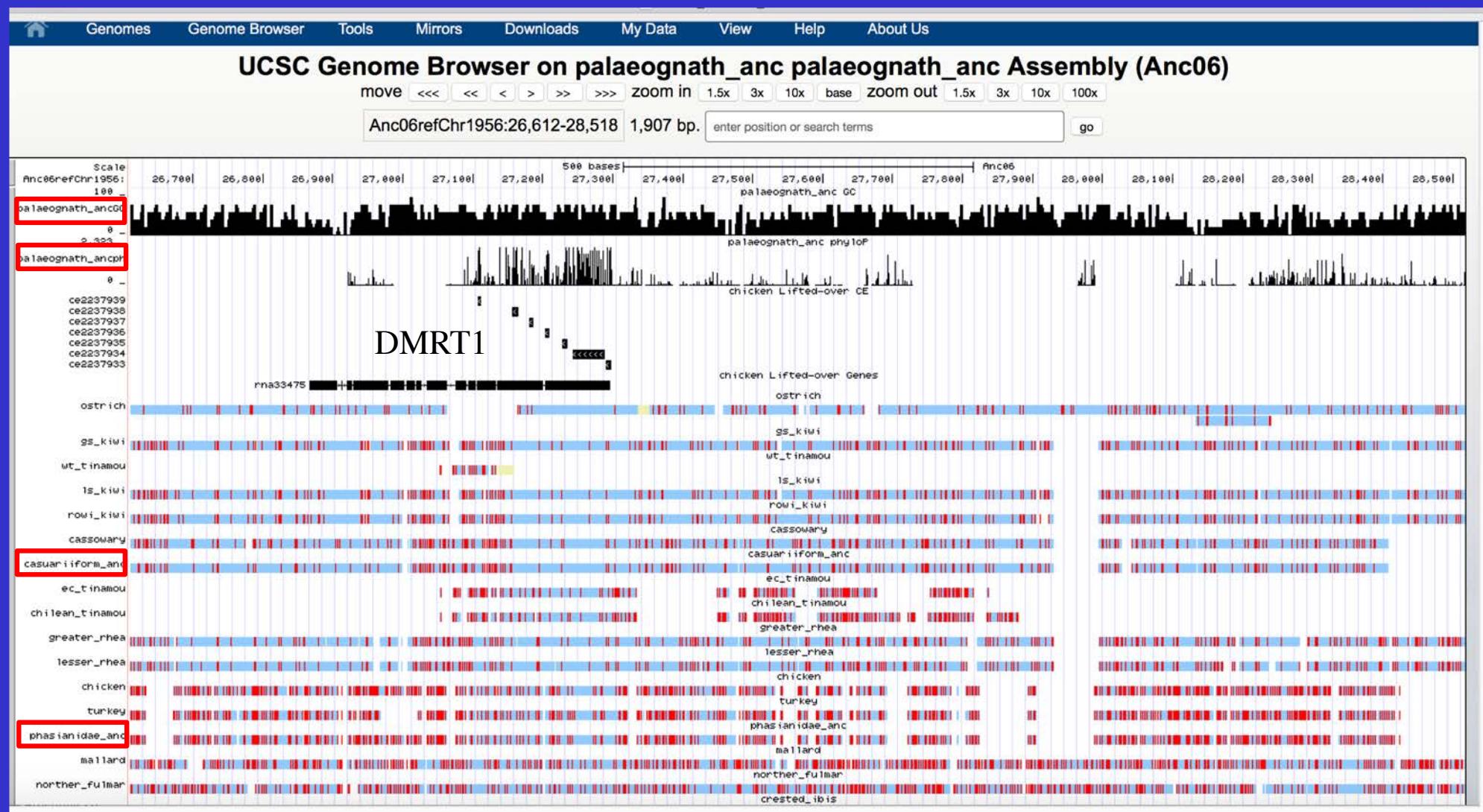
Thicket tinamou



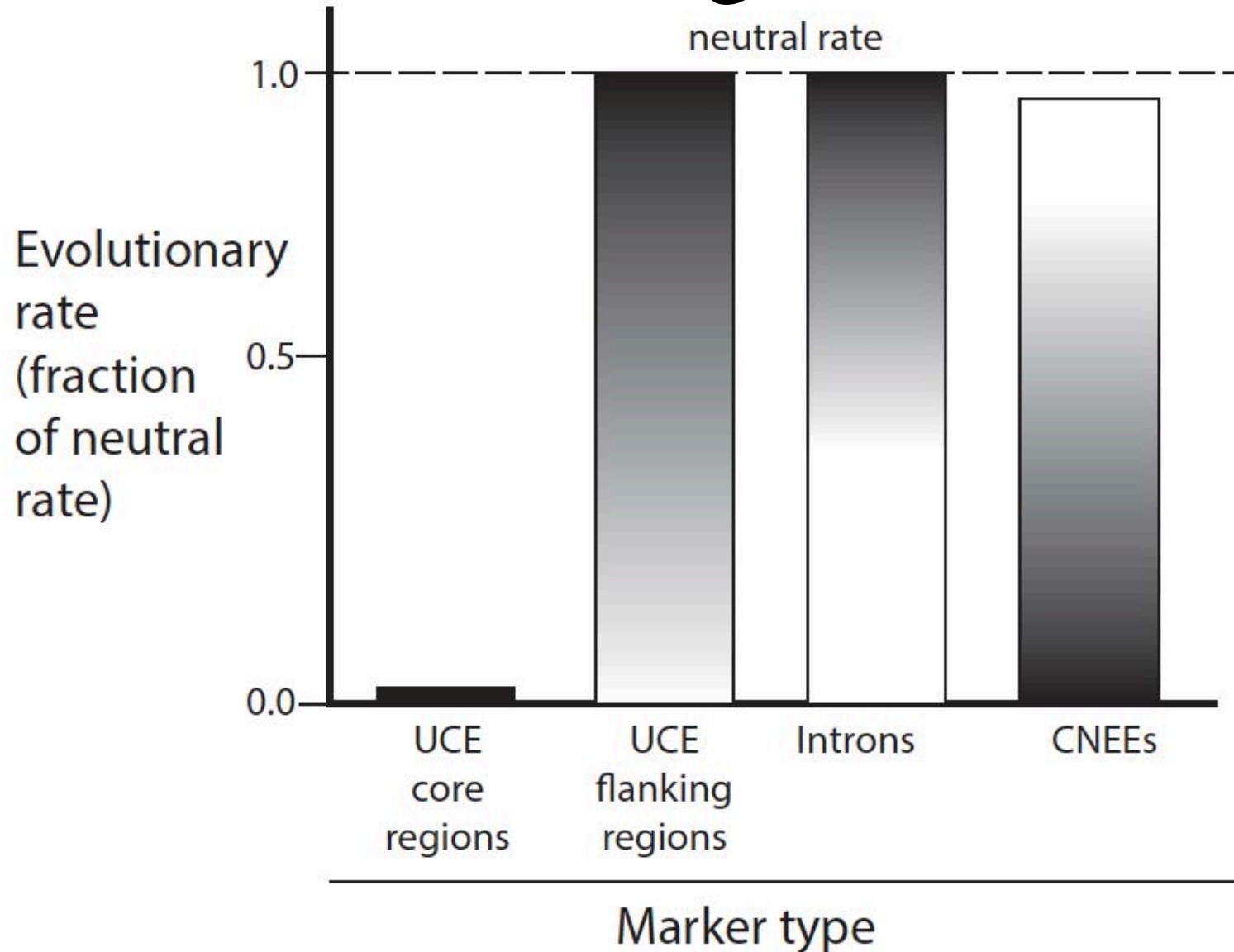
Chilean tinamou

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

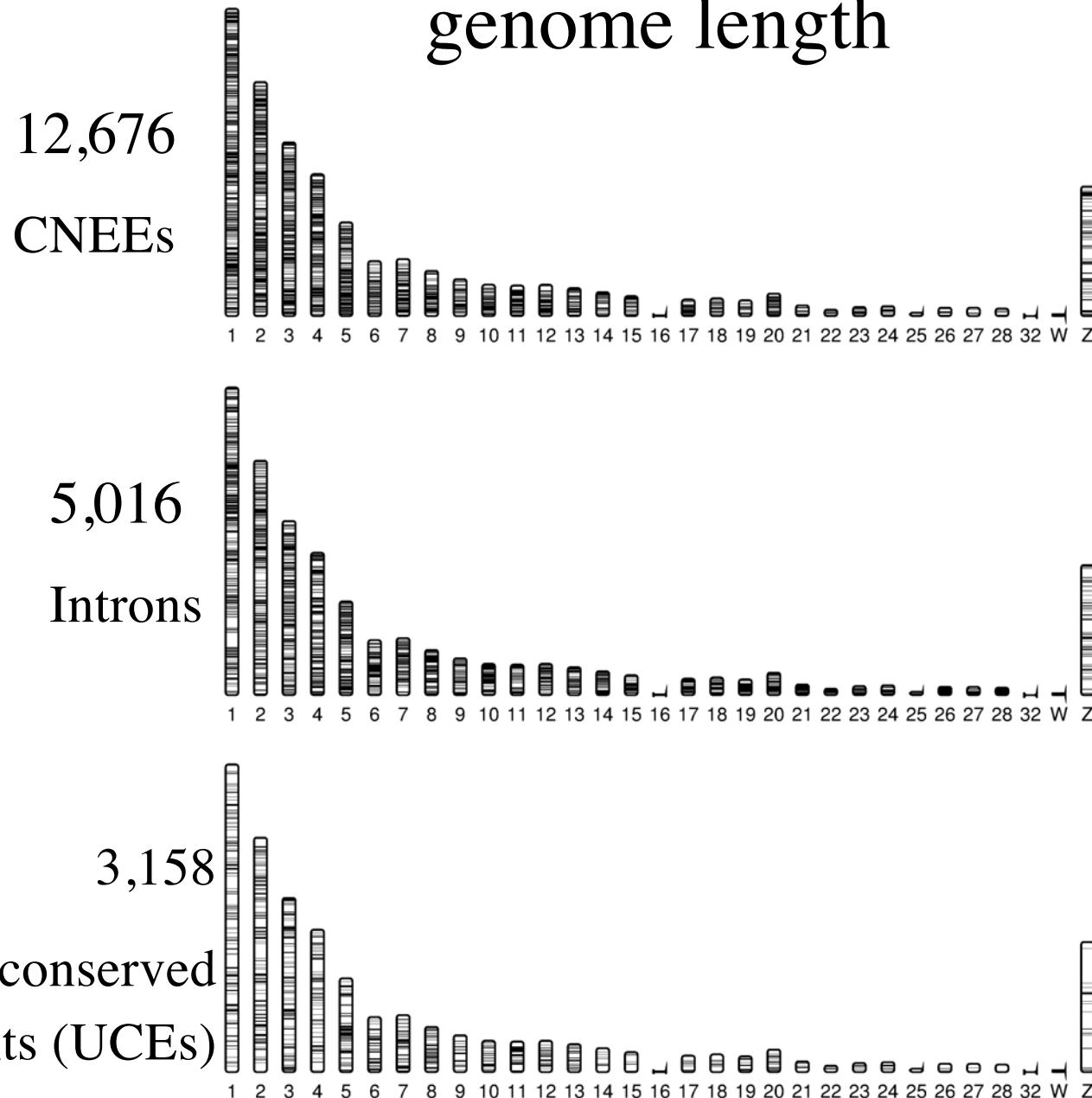
42-species whole genome alignment for birds using ProgressiveCactus



Relative rates of different noncoding markers

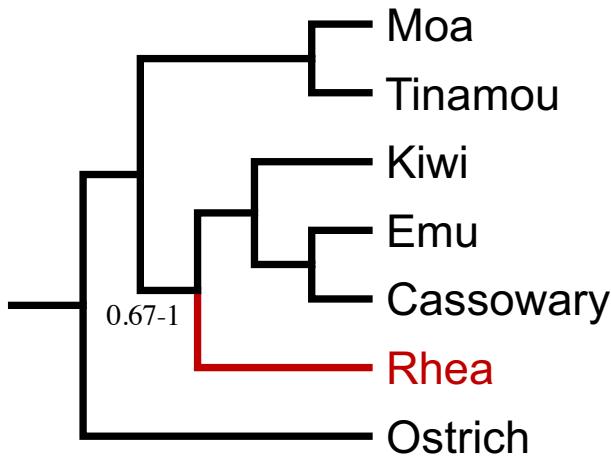


Phylogenomic markers cover c. 3% of total genome length

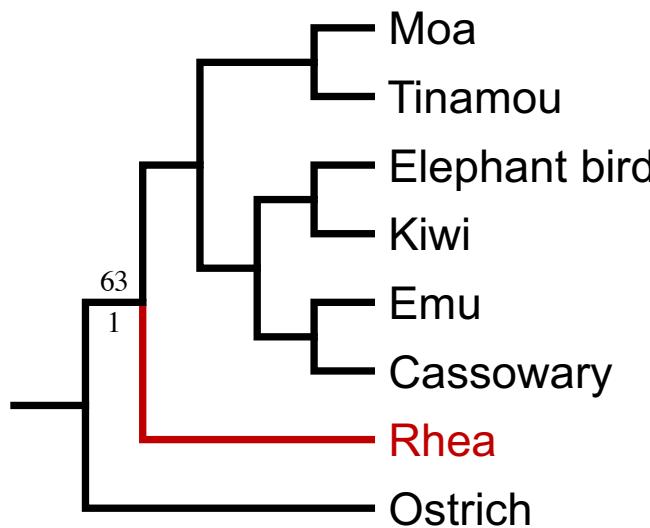


Relationships of rheas unclear

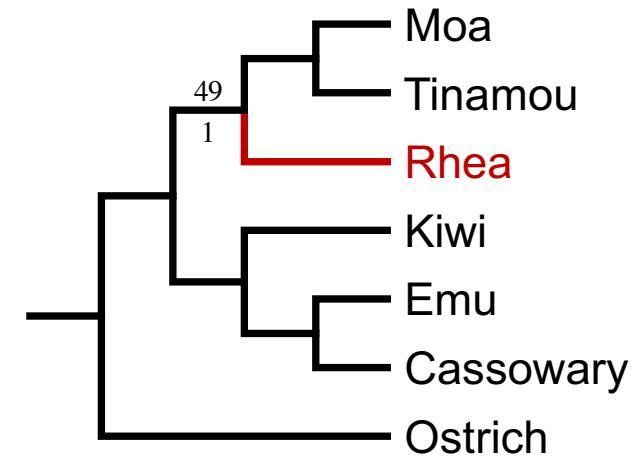
Haddrath & Baker (2012)
27 nuclear loci



Mitchell et al. (2014)
mtDNA



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

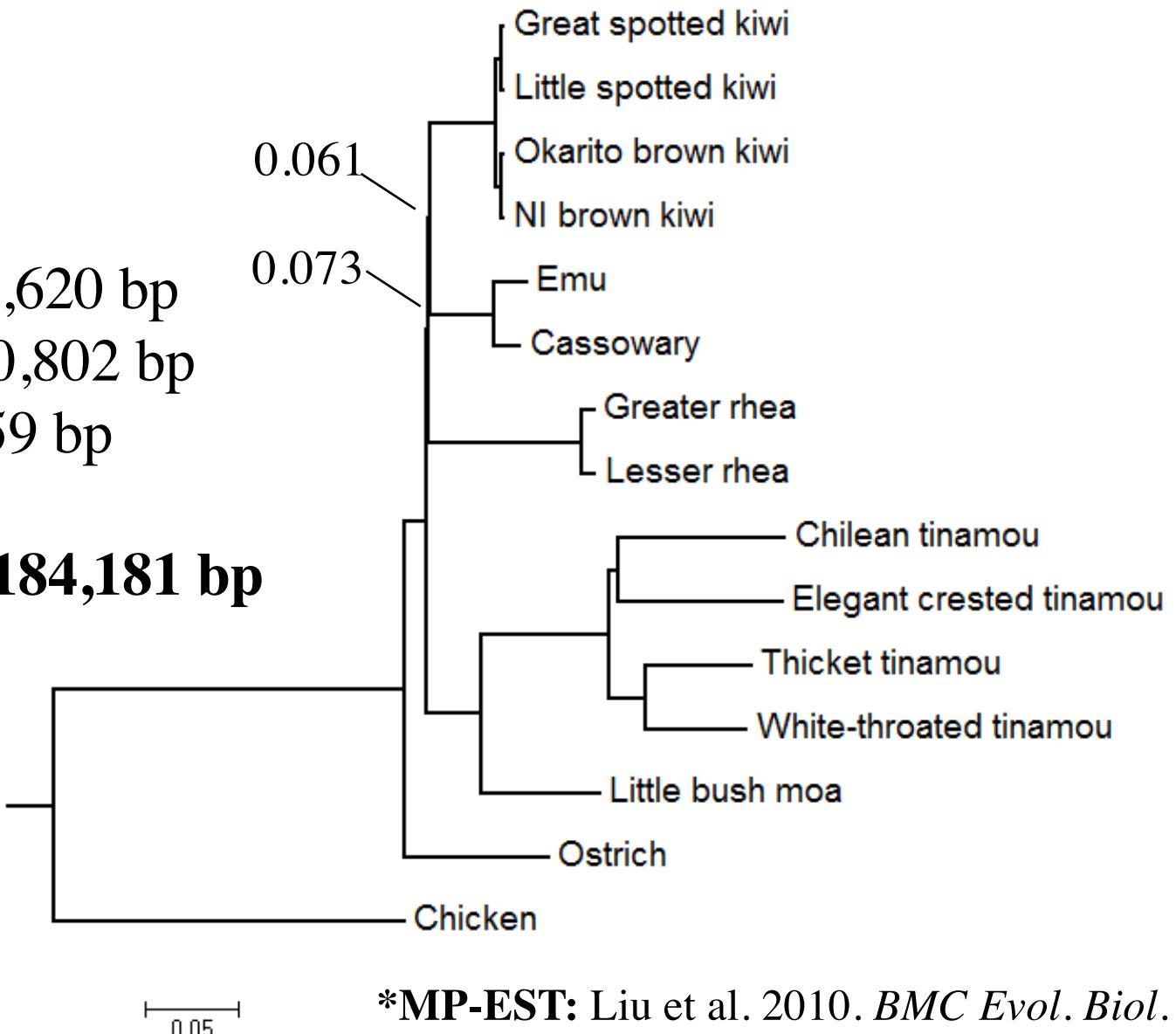


Coalescent* analyses resolve the position of rheas and reveal an ancient rapid radiation

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

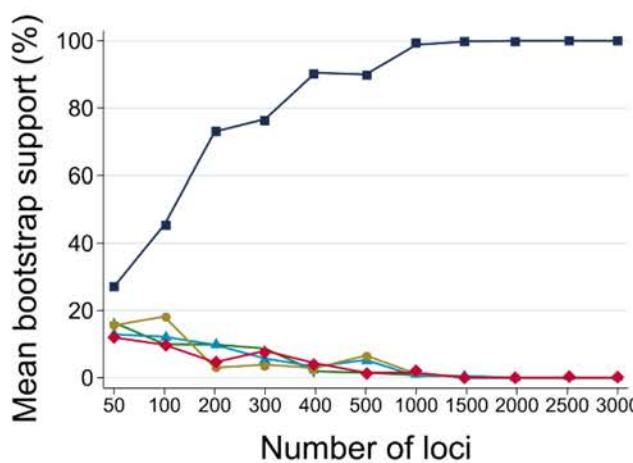
Total: **20,850 loci; 41,184,181 bp**

Branch lengths in coalescent units

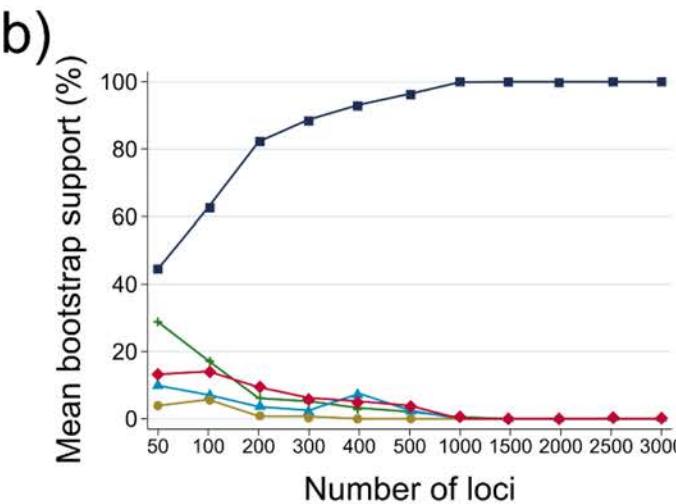


Consistent accumulation of phylogenetic signal using MP-EST

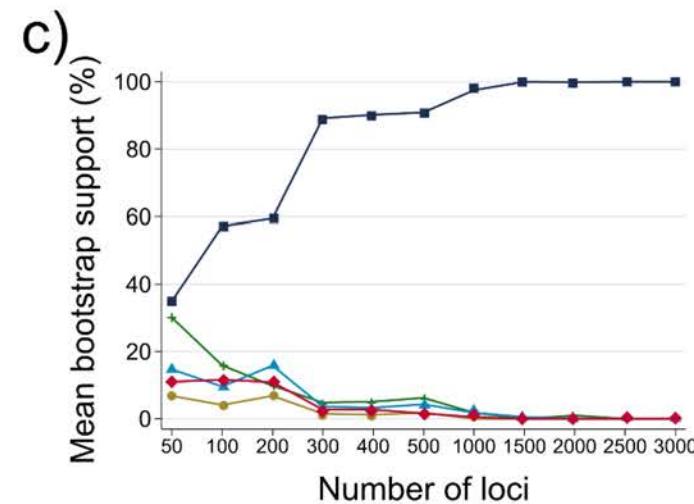
CNEEs



Introns



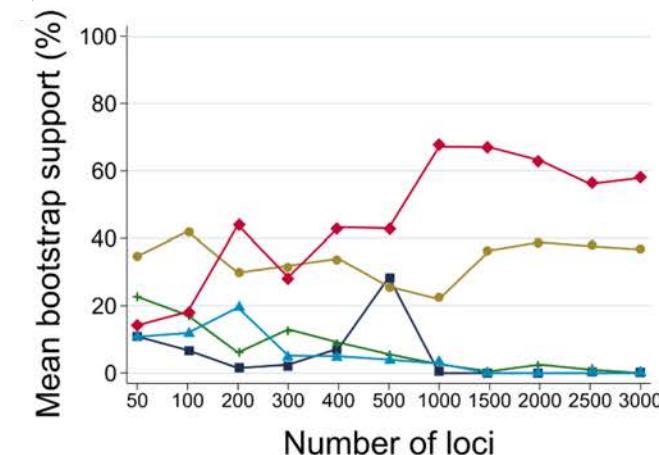
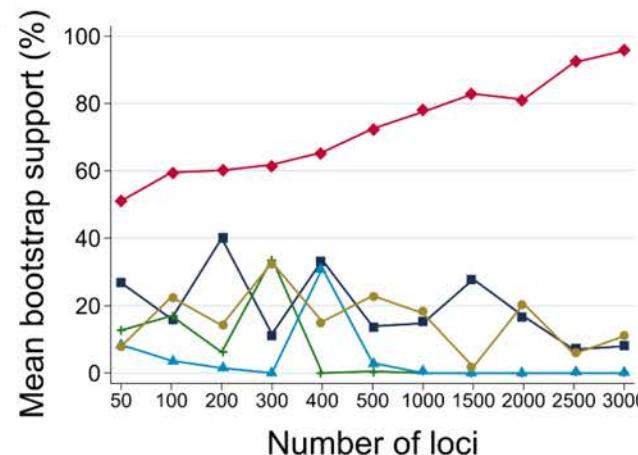
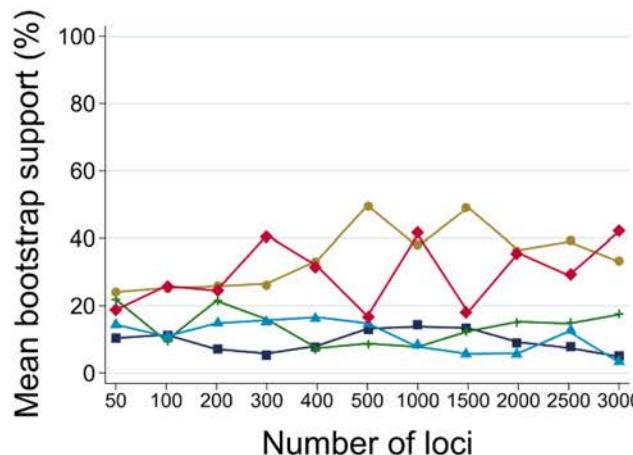
UCEs



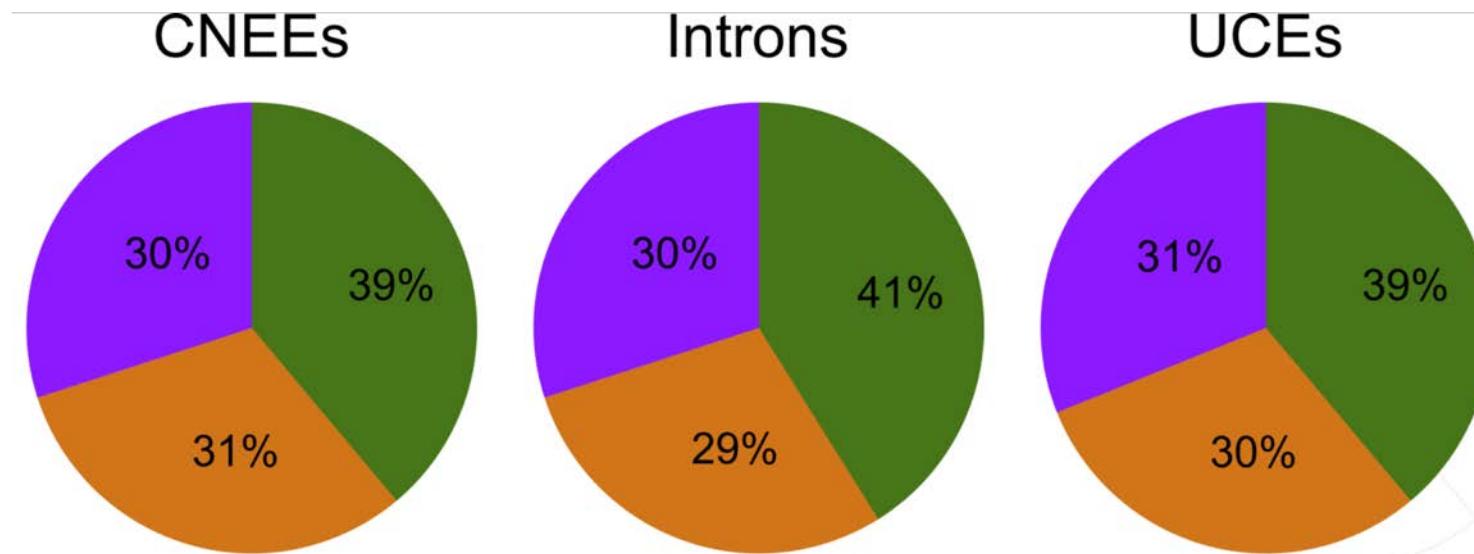
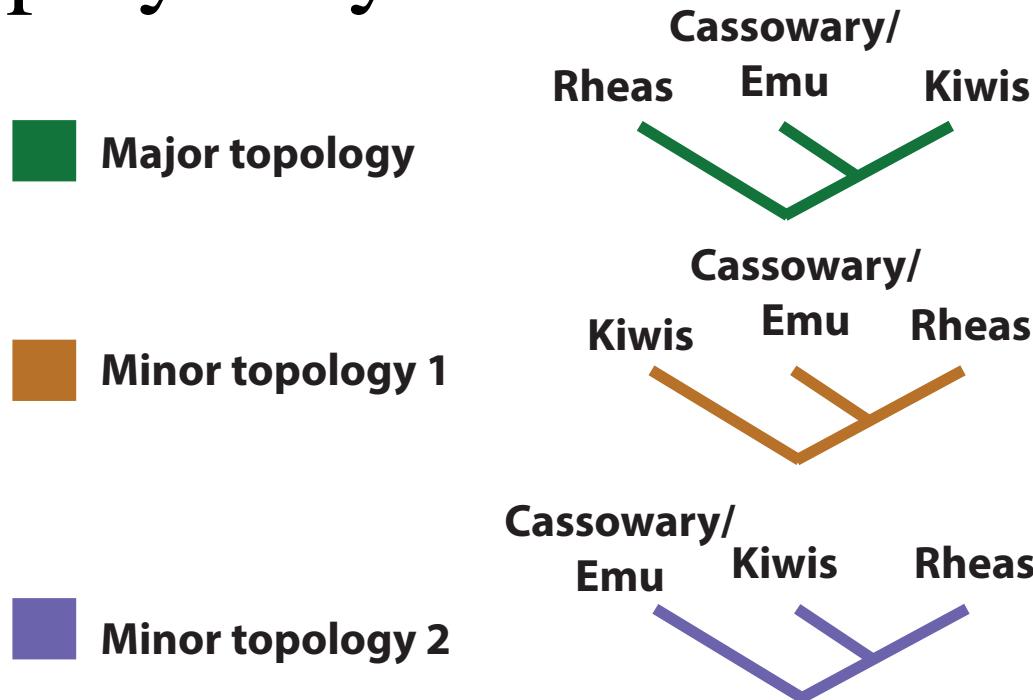
Rheas sister to:

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

- Moa + Tinamous
- Other non-ostrich palaeognaths

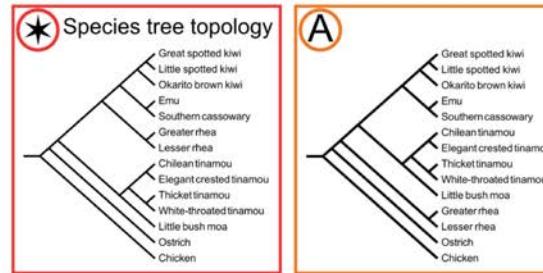
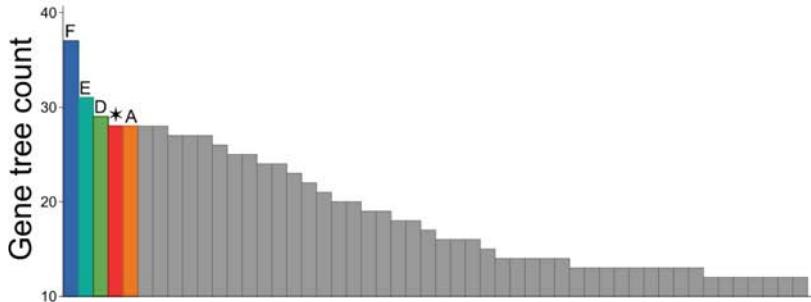


Gene tree distribution suggests a near polytomy at base of ratites

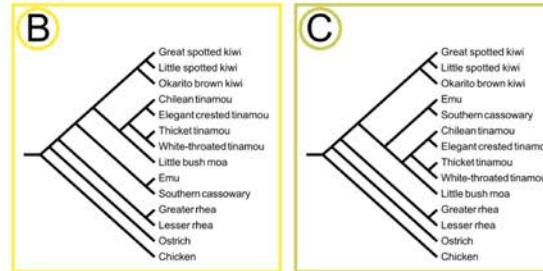
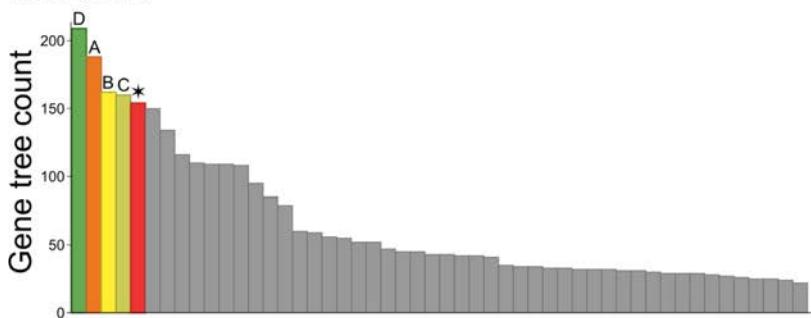


Anomaly zone: most common gene tree does not match the species tree

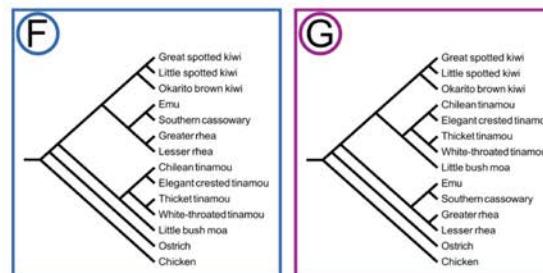
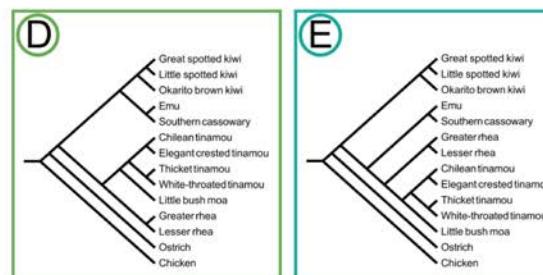
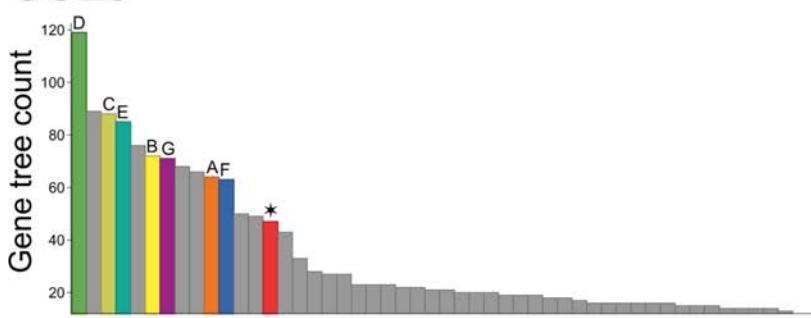
CNEEs



Introns

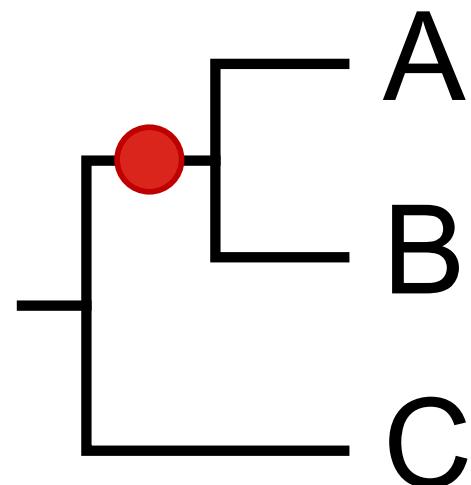
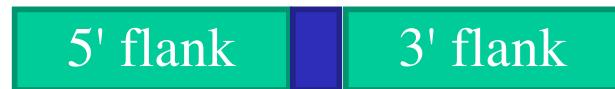


UCEs

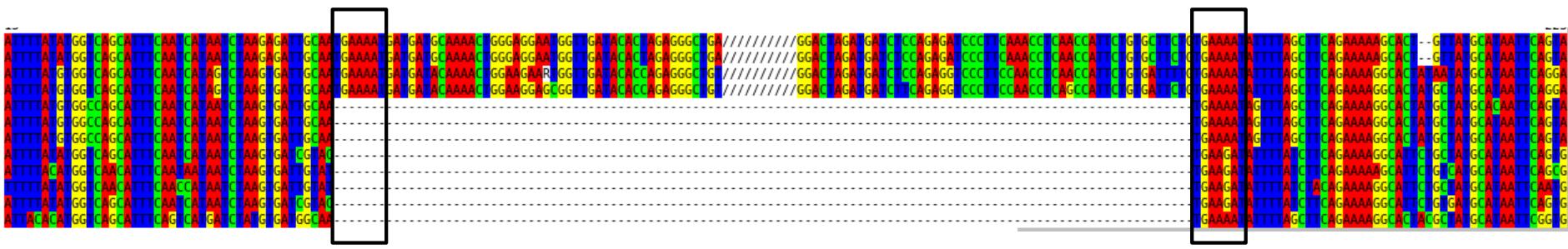


Corroboration of coalescent tree from transposable elements

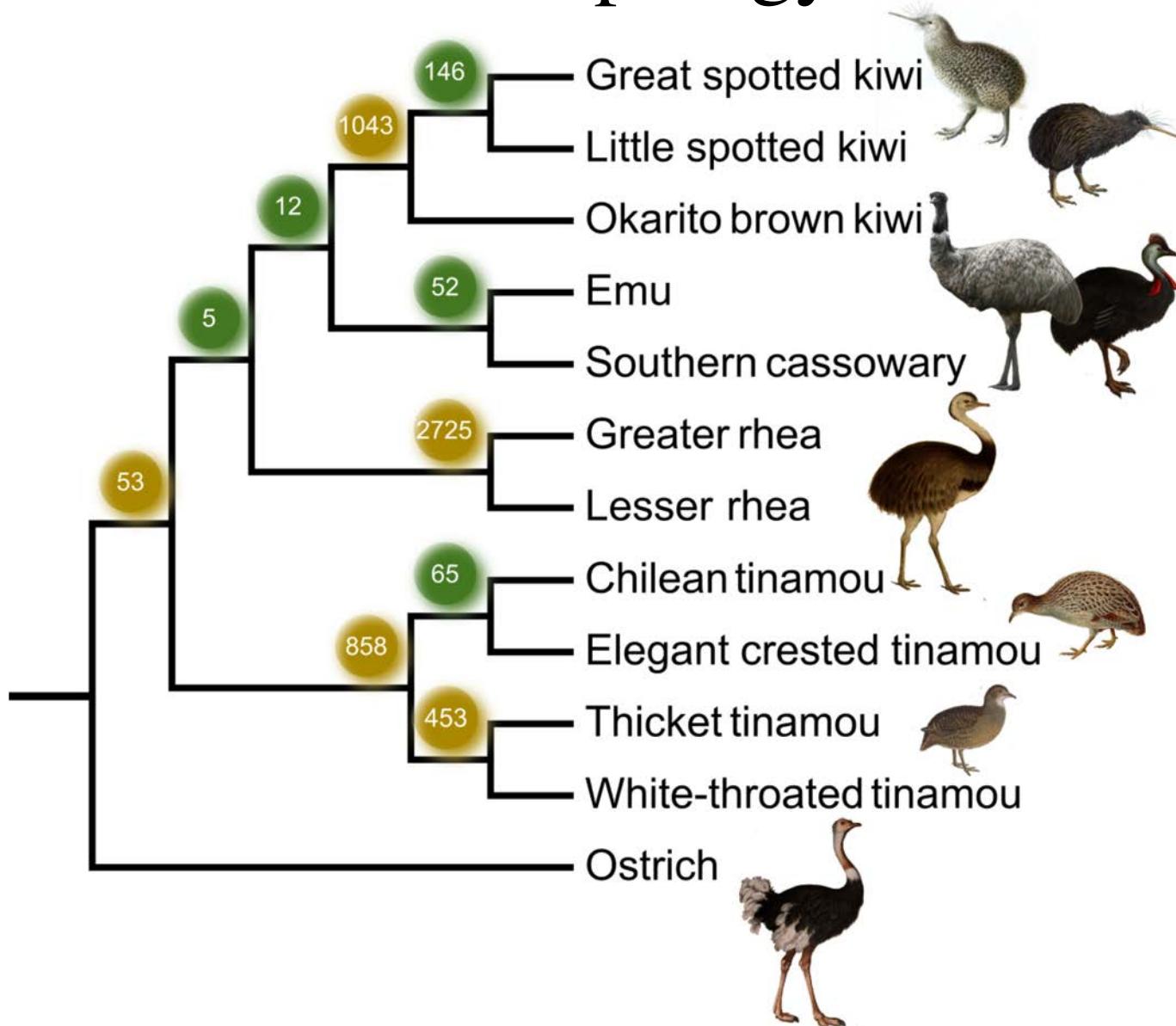
- Chicken Repeat1 (CR1) retroelement insertions
- (Virtually) homoplasy free
- Binary presence/absence
 - No model misspecification
 - No GC/rate variation bias
- BUT- subject to incomplete lineage sorting



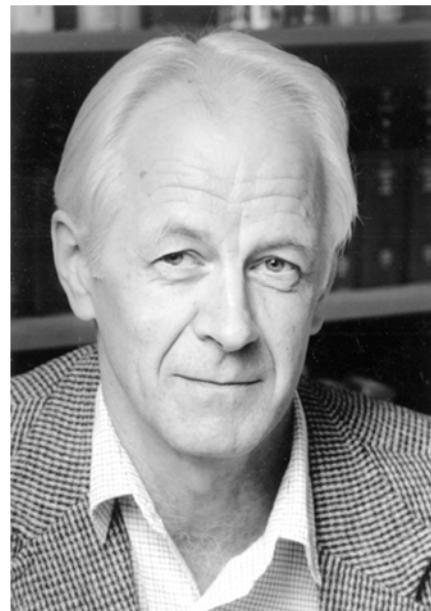
Example CR1 insertion



Most CR1s support the coalescent species tree topology



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.

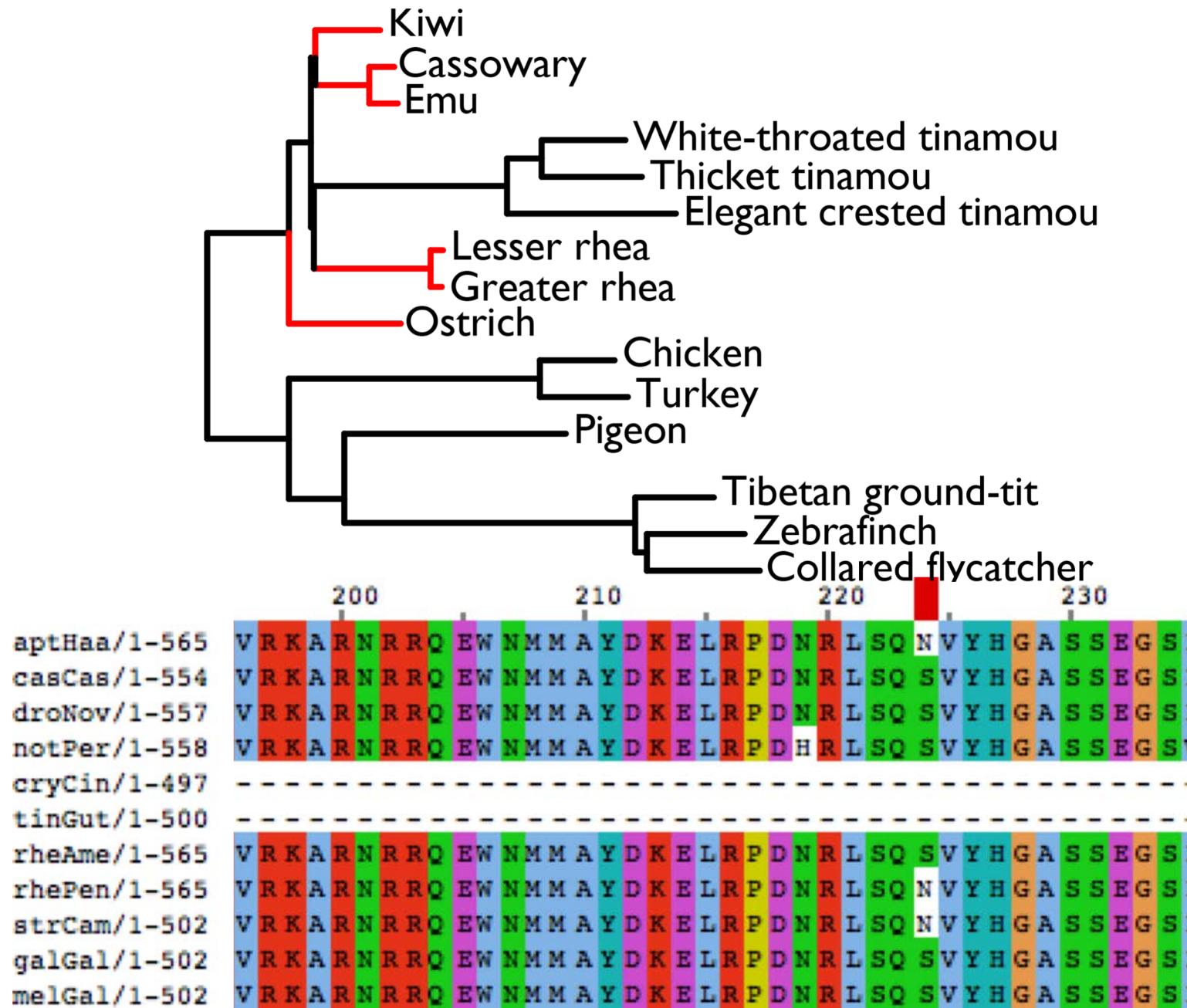


Mary-Claire King and A. C. Wilson

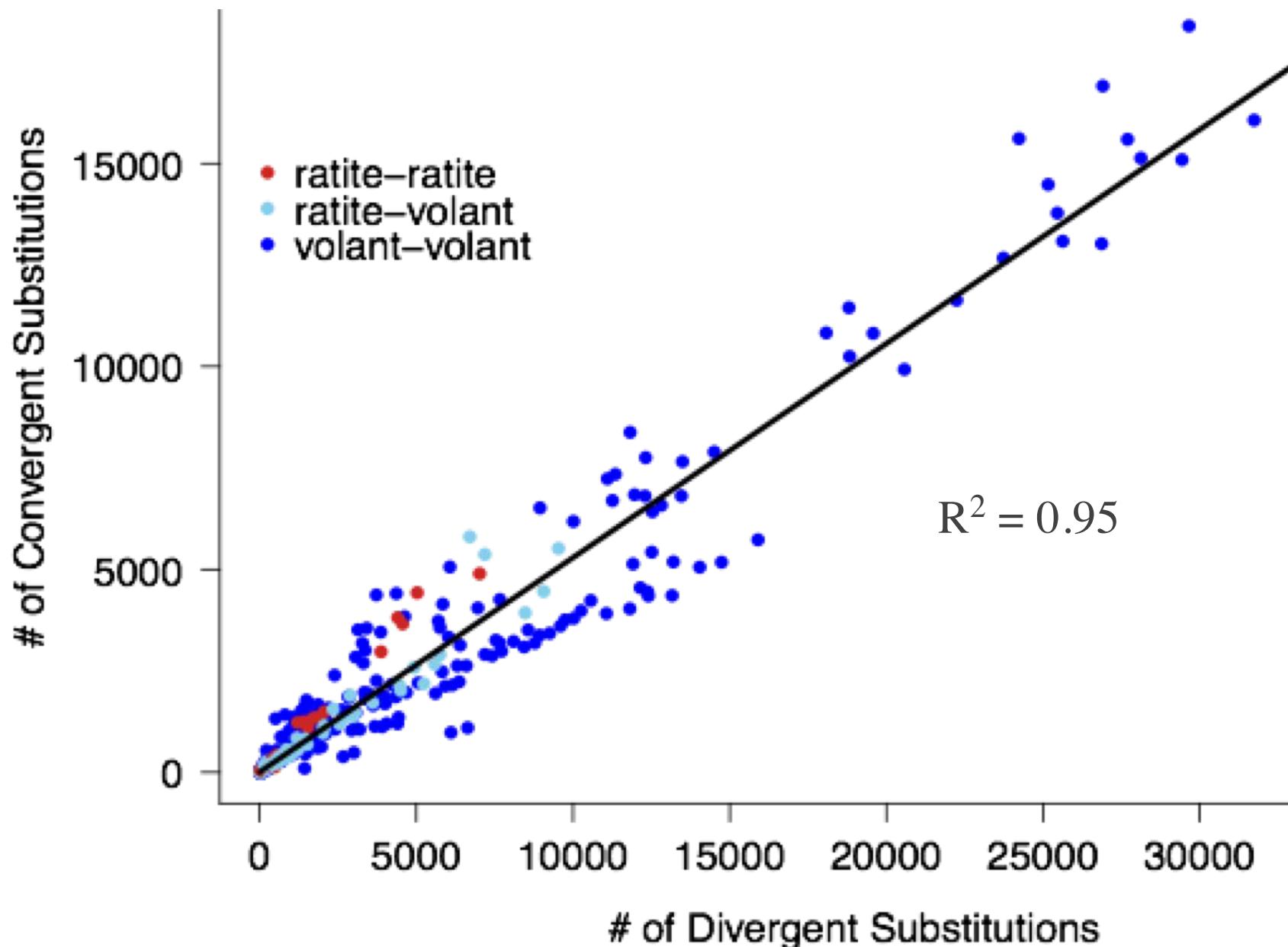
SCIENCE

11 April 1975, Volume 188, Number 4184

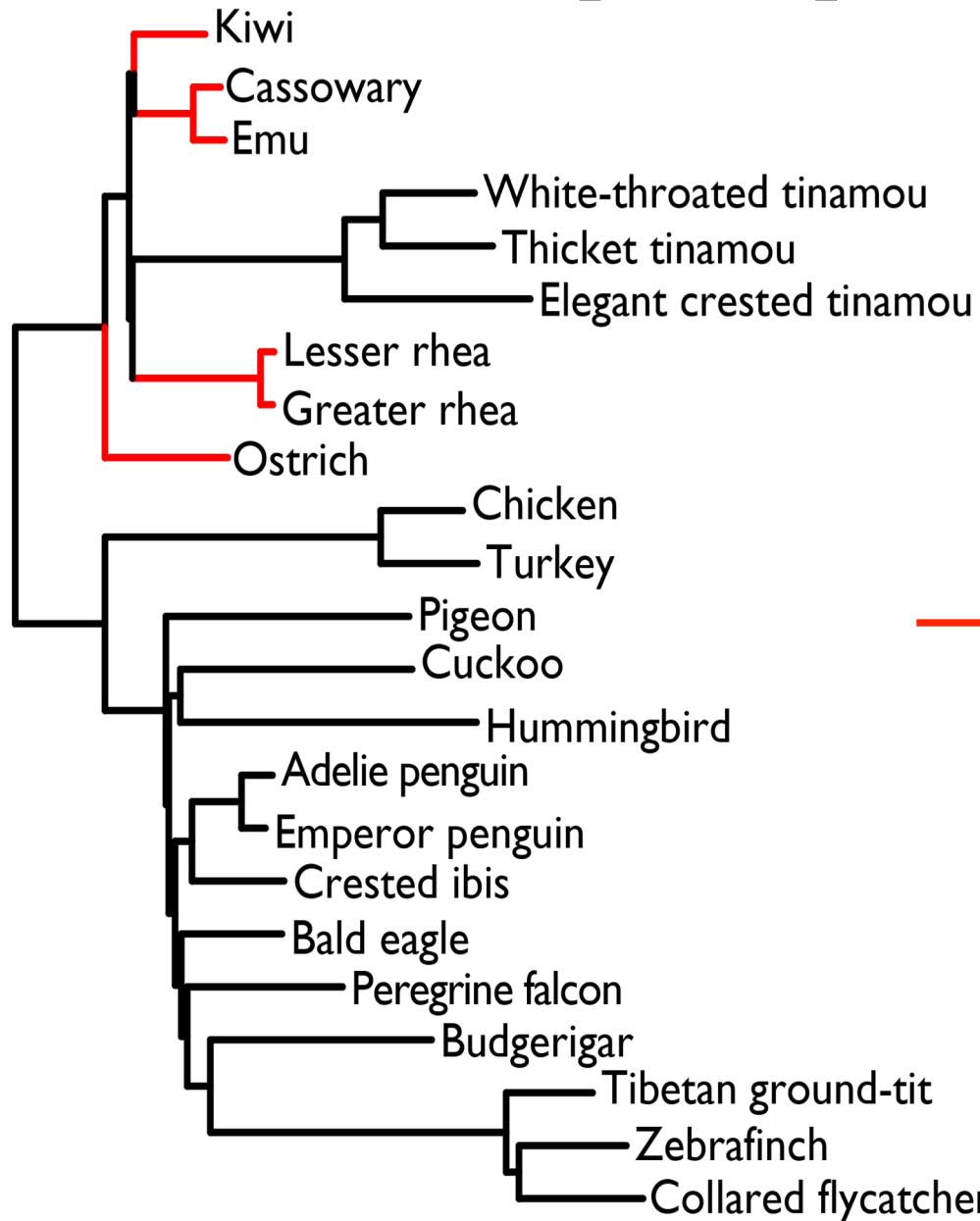
Searching for evidence of genome-wide amino acid convergence in ratites



No evidence for genome-wide convergent amino acid substitutions in ratites

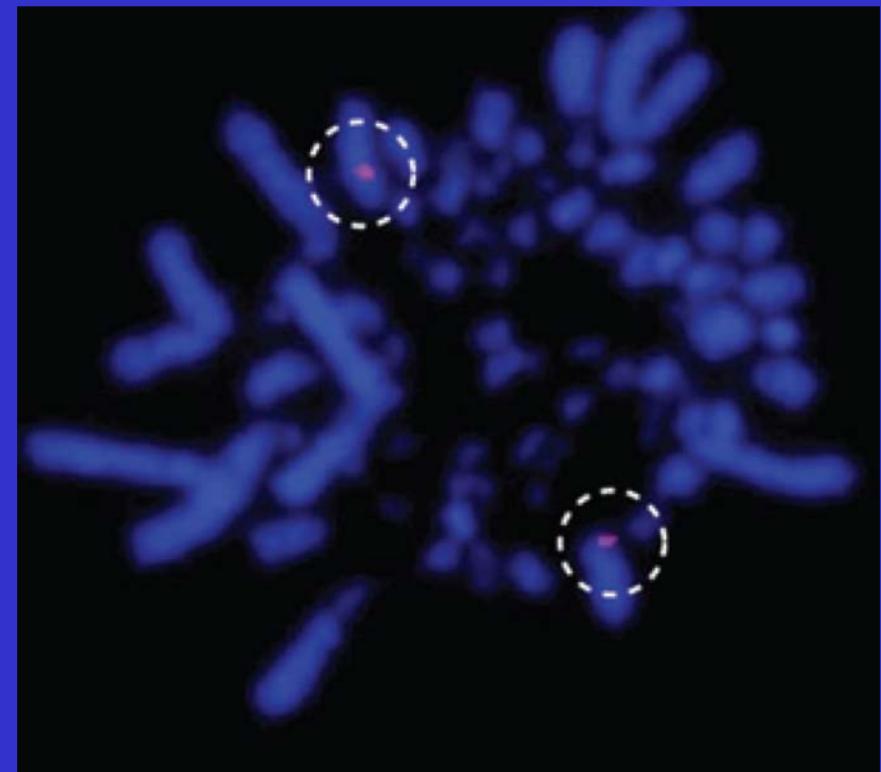


\sim 1-2% of protein-coding genes show evidence of ratite-specific positive selection



271 genes (10% FDR)
104 genes (1% FDR)

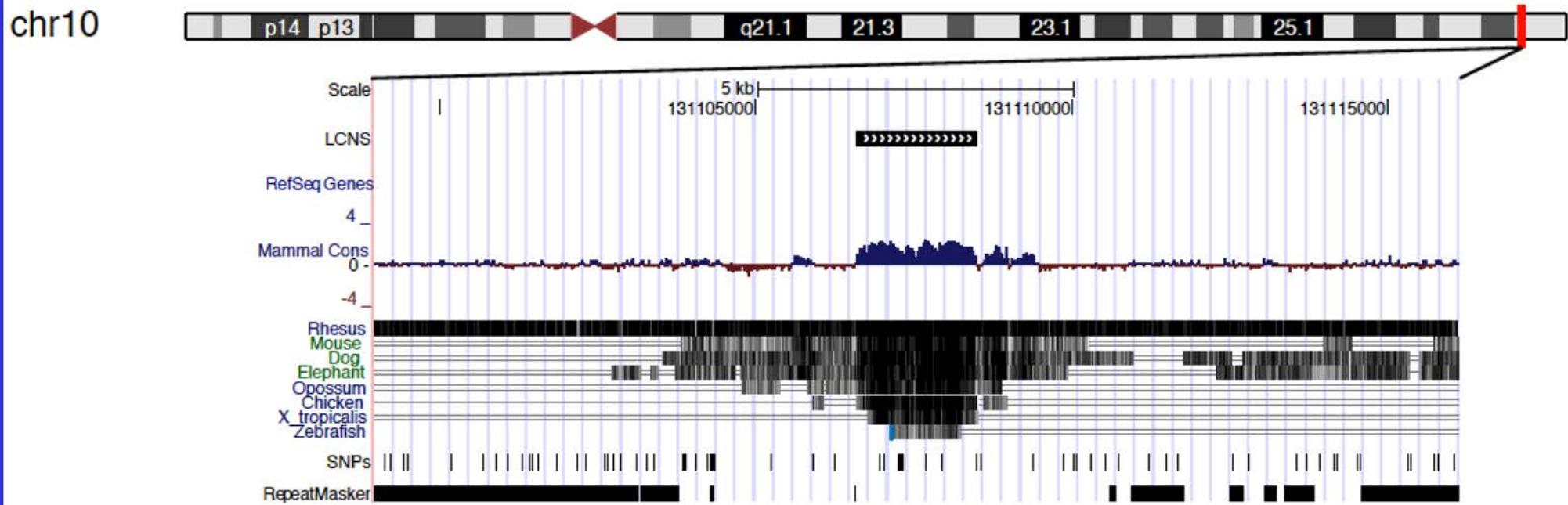
Non-coding ‘Dark matter’ of the genome: a regulatory network?



Karyotype of an Emu

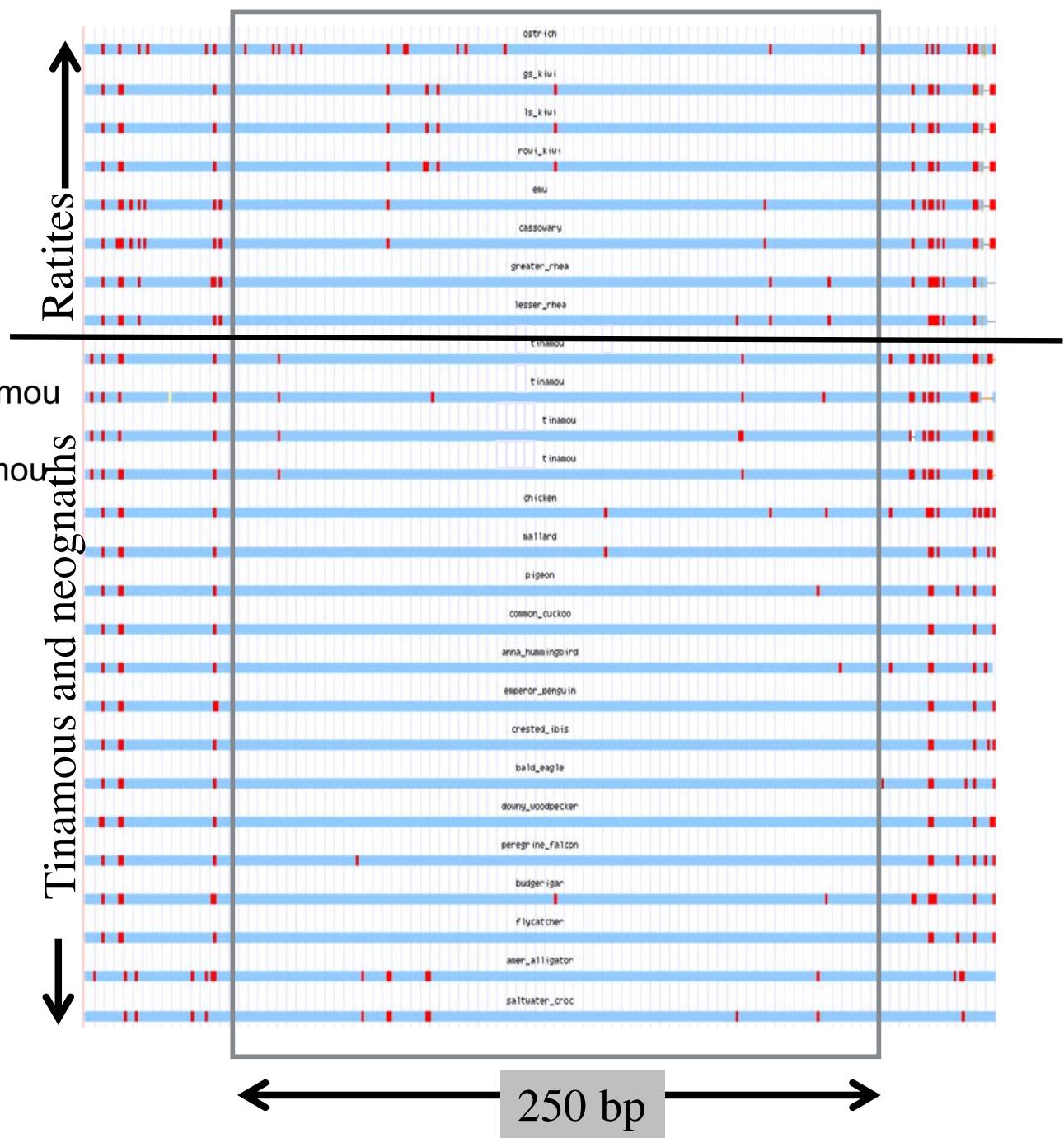
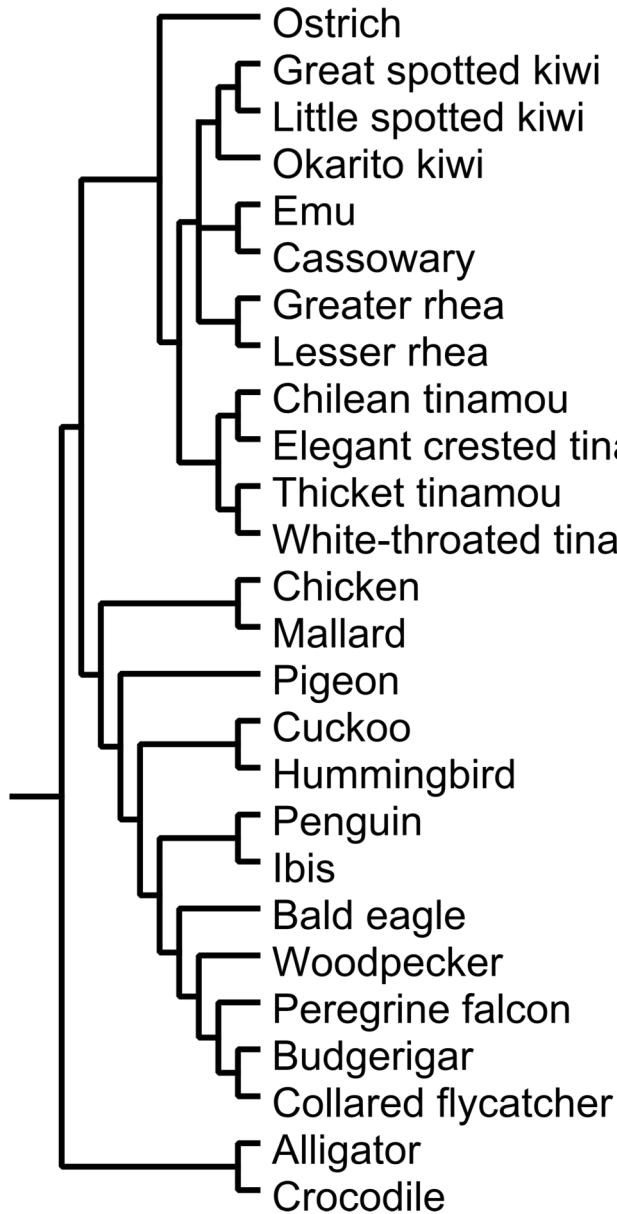
CNEEs: evolutionarily conserved non-coding enhancer regions

CNEEs = conserved non-exonic elements
284,001 long (* > 50 bp) CNEEs in data set

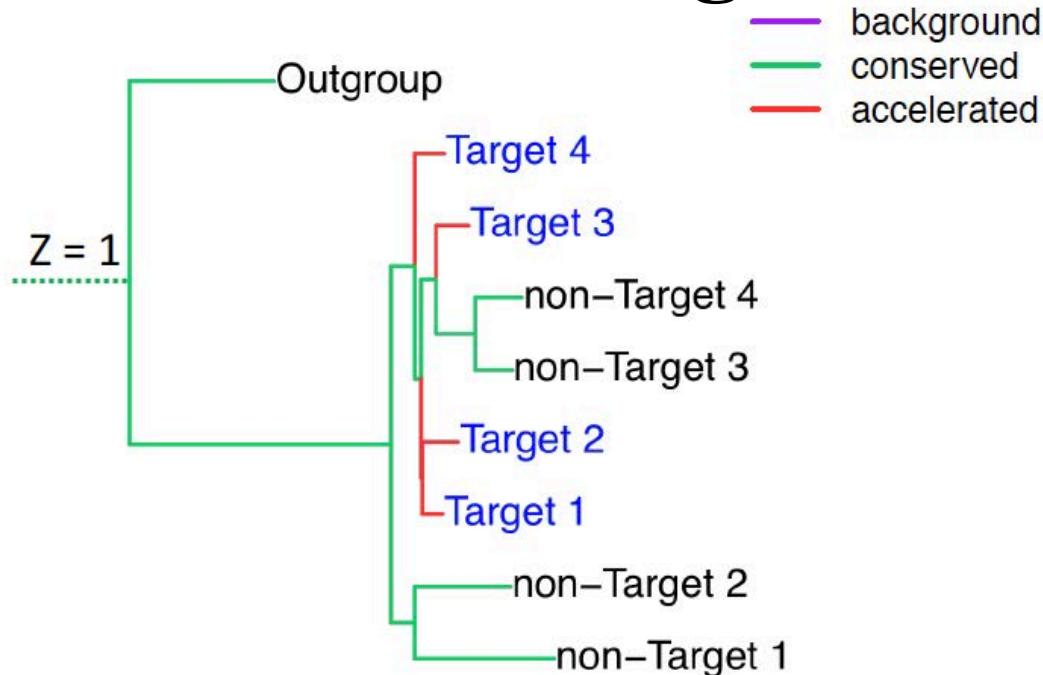


View of a segment of human chromosome 10 using UCSC Genome Browser

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

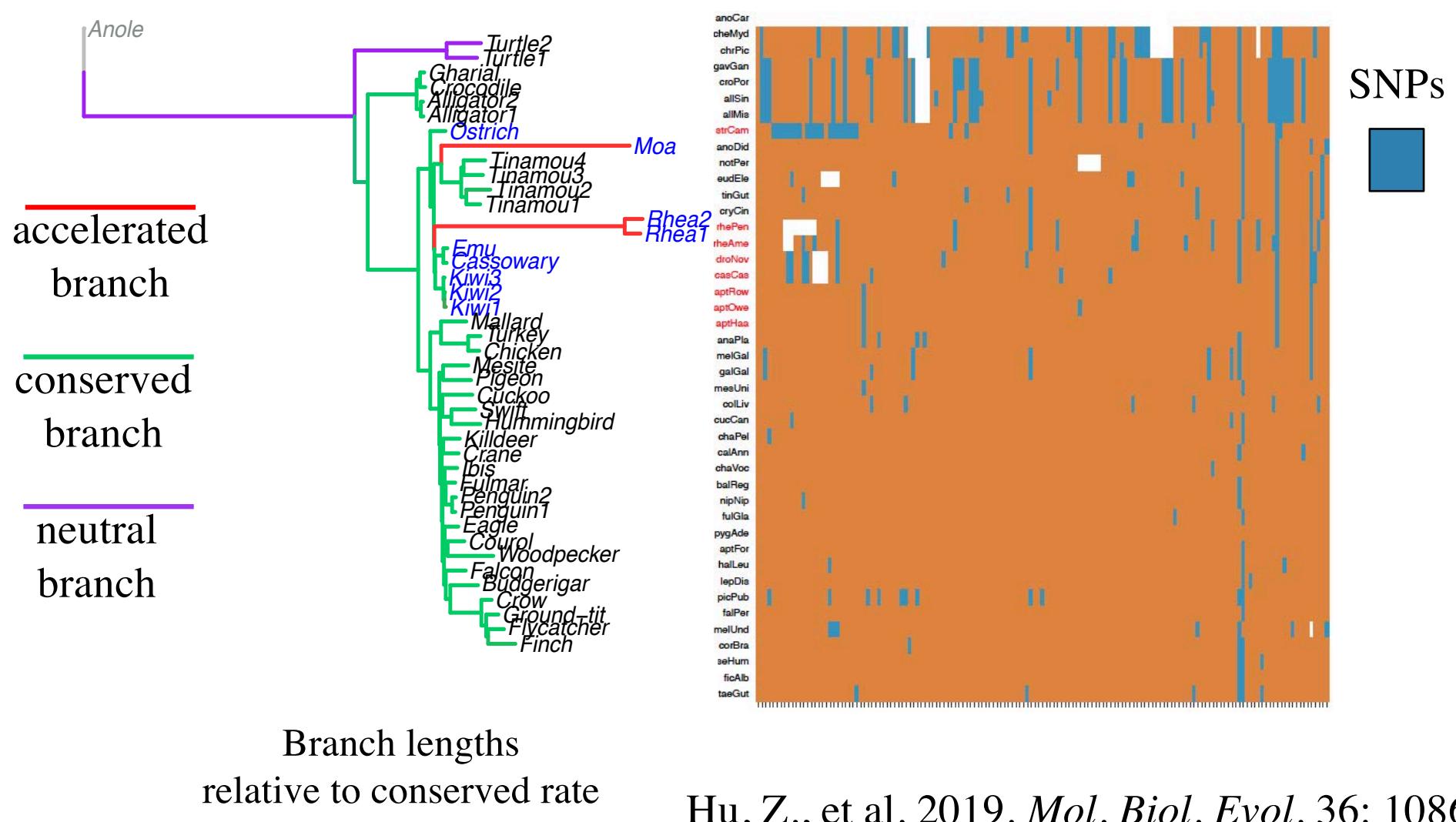
β = probability of loss of conserved state

For branch s ,

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

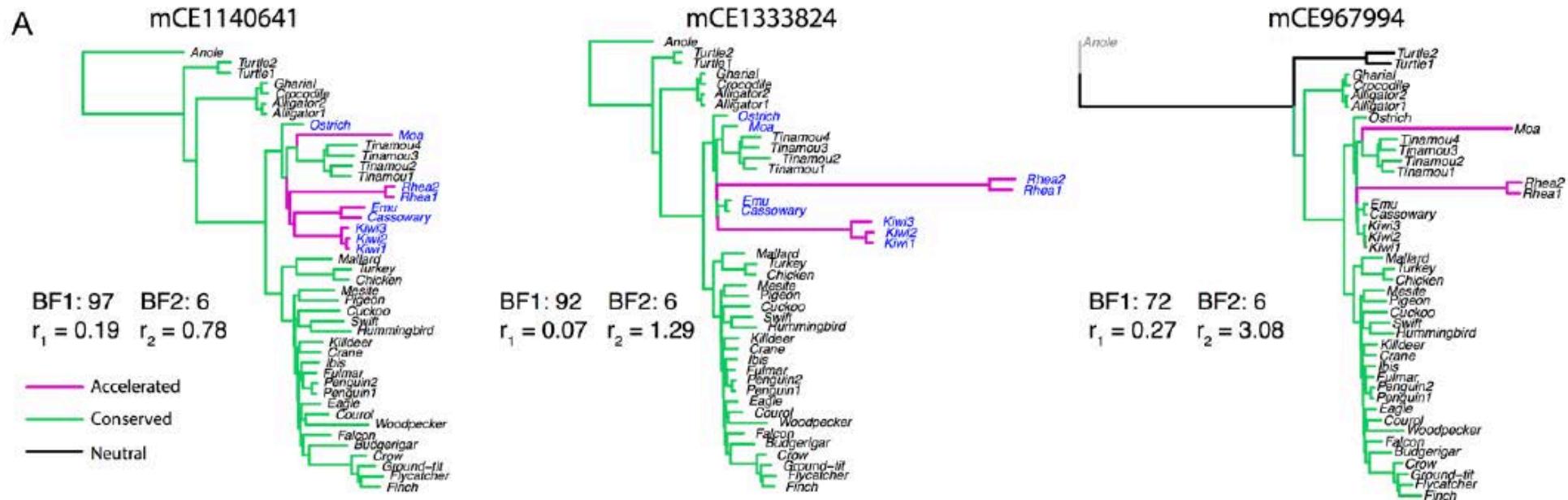
$$BF1 = \frac{P(Y|M_1)}{P(Y|M_0)} \text{ and } BF2 = \frac{P(Y|M_1)}{P(Y|M_2)}.$$

A convergently accelerated CNEE detected with a novel Bayesian method

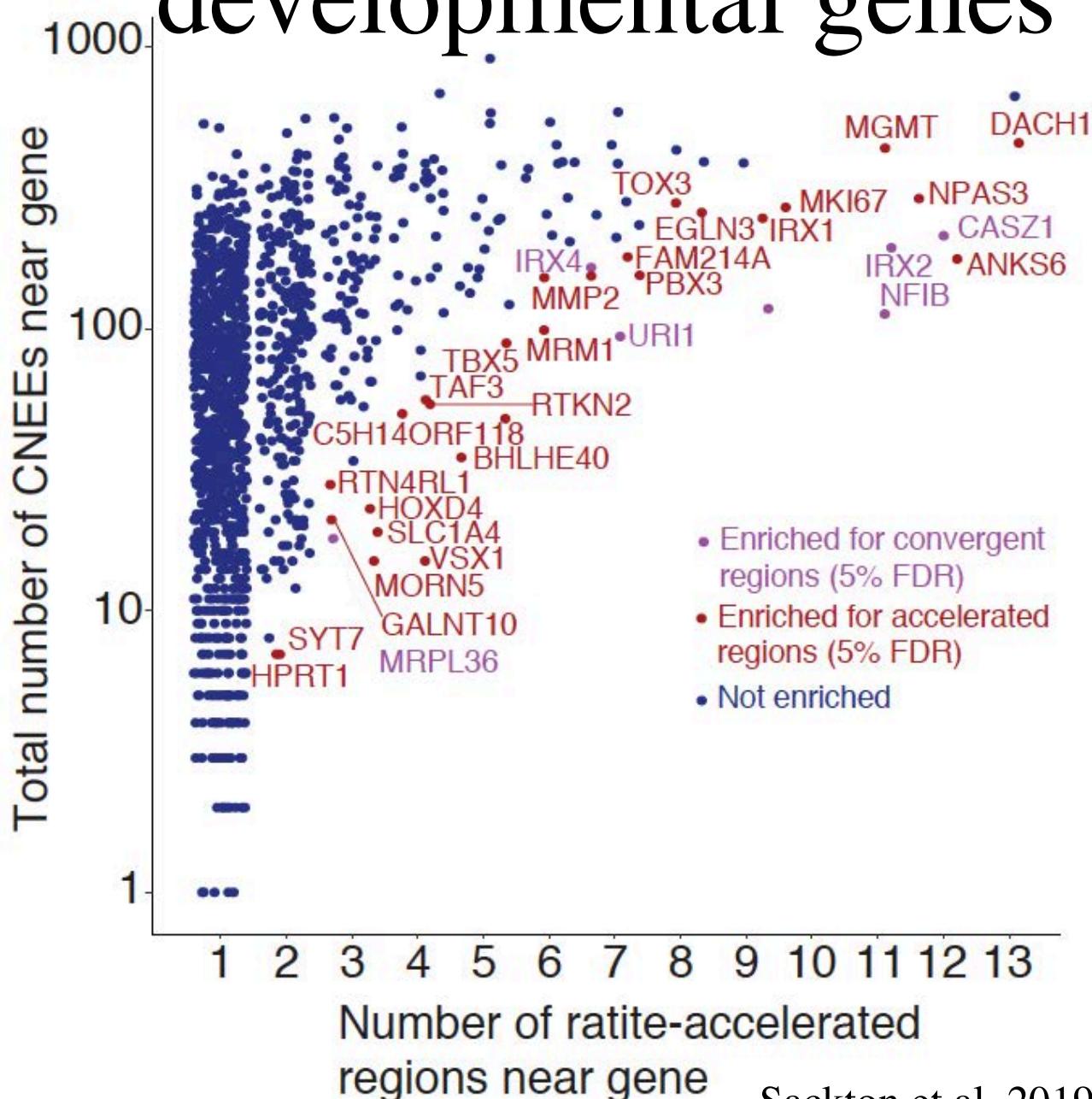


Hu, Z., et al. 2019. *Mol. Biol. Evol.* 36: 1086

Additional examples of convergently accelerated CNEEs



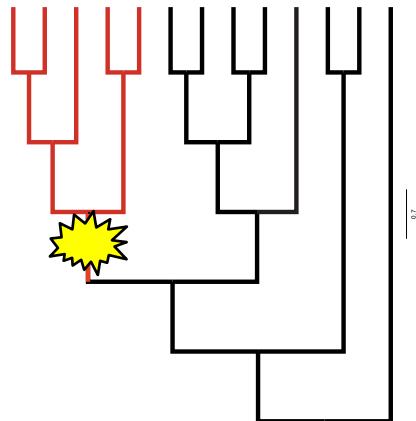
Rapid regulatory evolution near developmental genes



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

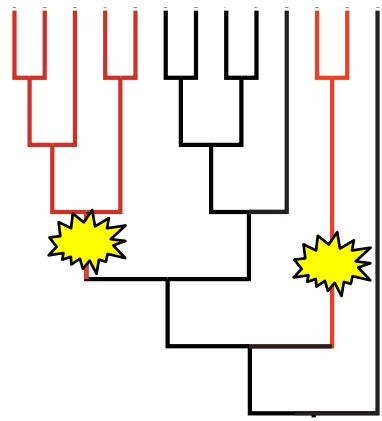
Homing in on regulators for flightlessness through convergence

4,260 CNEEs



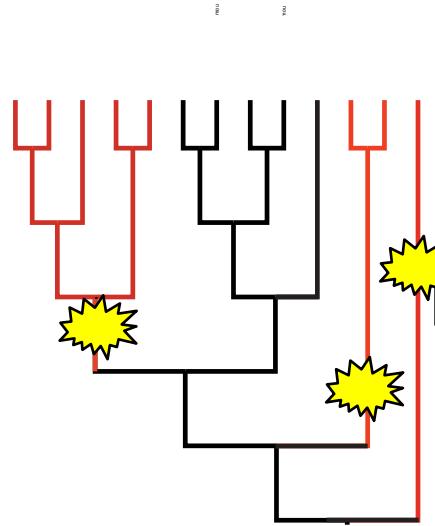
CNEE relaxed
on 1 lineage

1,270 CNEEs



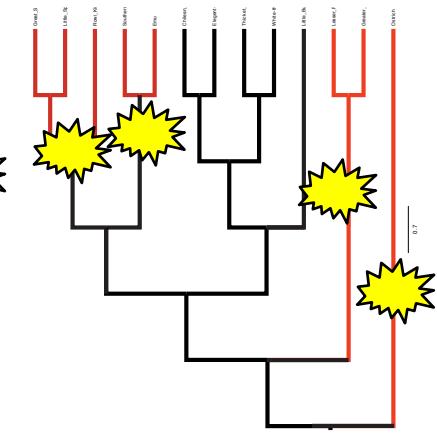
Same CNEE relaxed
in 2 lineages

252 CNEEs



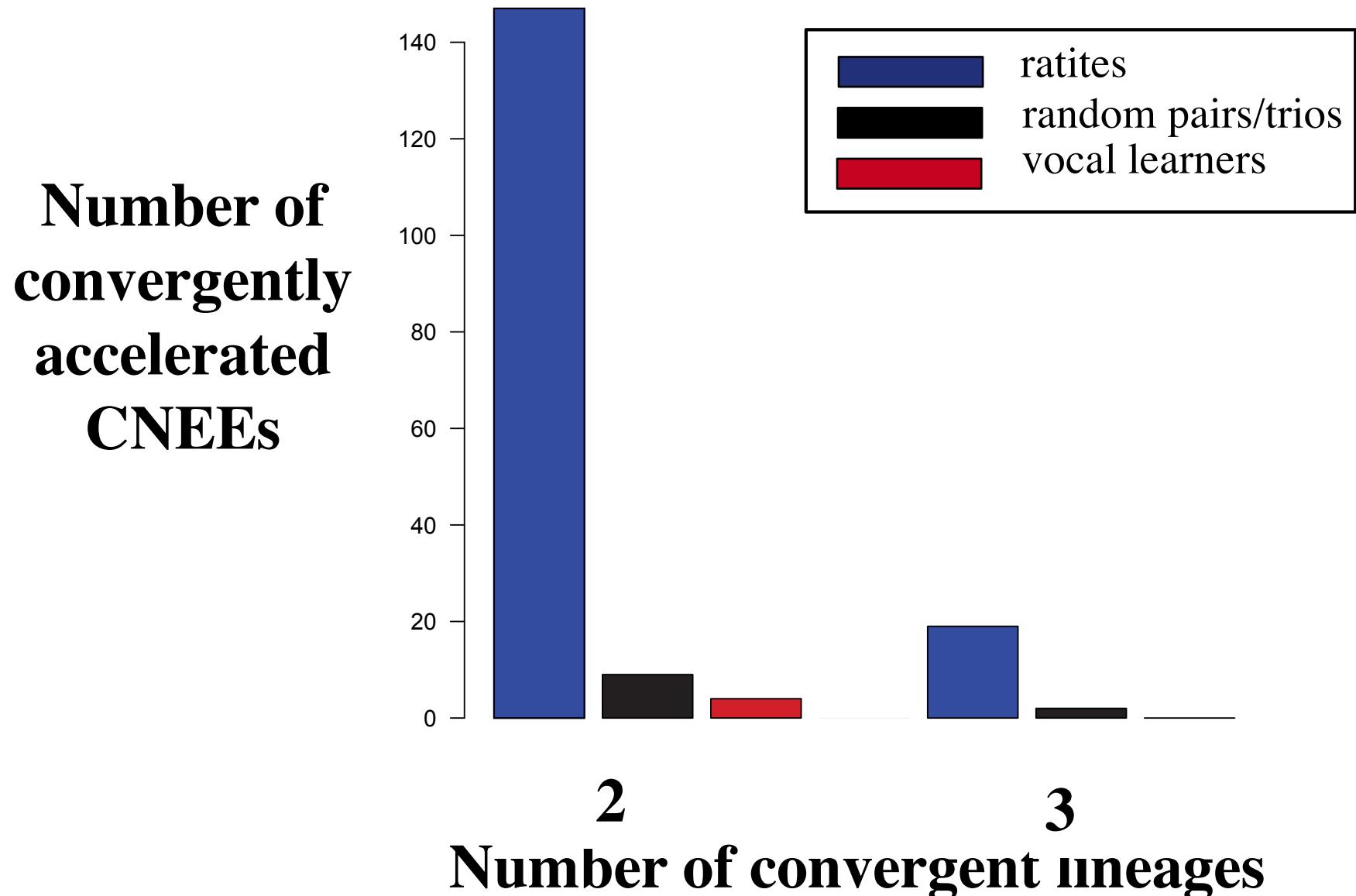
Same CNEE relaxed
in 3 lineages

66 CNEEs



Same CNEE relaxed
in 4 lineages

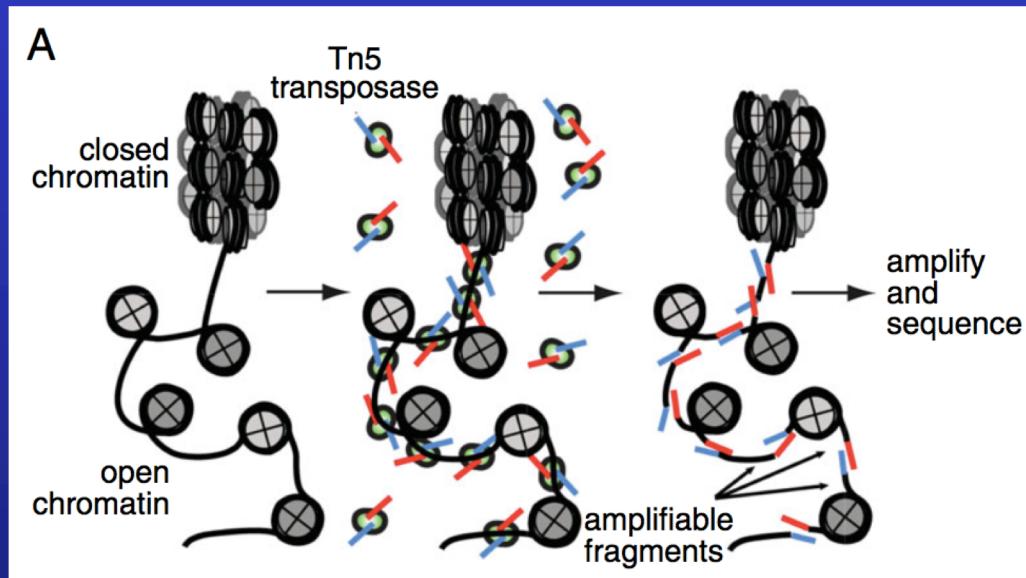
**Ratite genomes exhibit higher numbers of
convergently accelerating CNEEs
than do vocal learners or random trios of taxa**



Assay for Transposase-Accessible Chromatin

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

Stage HH24-25 chickens and rheas



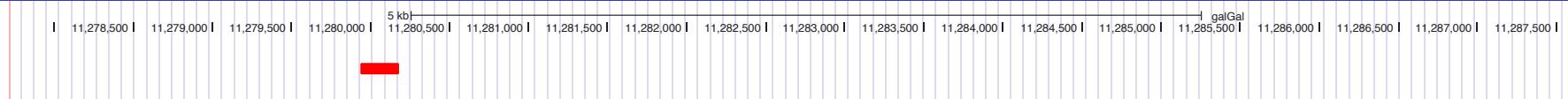
Buenrostro et al. 2015. Curr Protoc Biol. 2015;
109: 21.29.1–21.29.9.



Differences in ATAC-seq peaks between rhea and chicken suggest changes in limb gene regulation

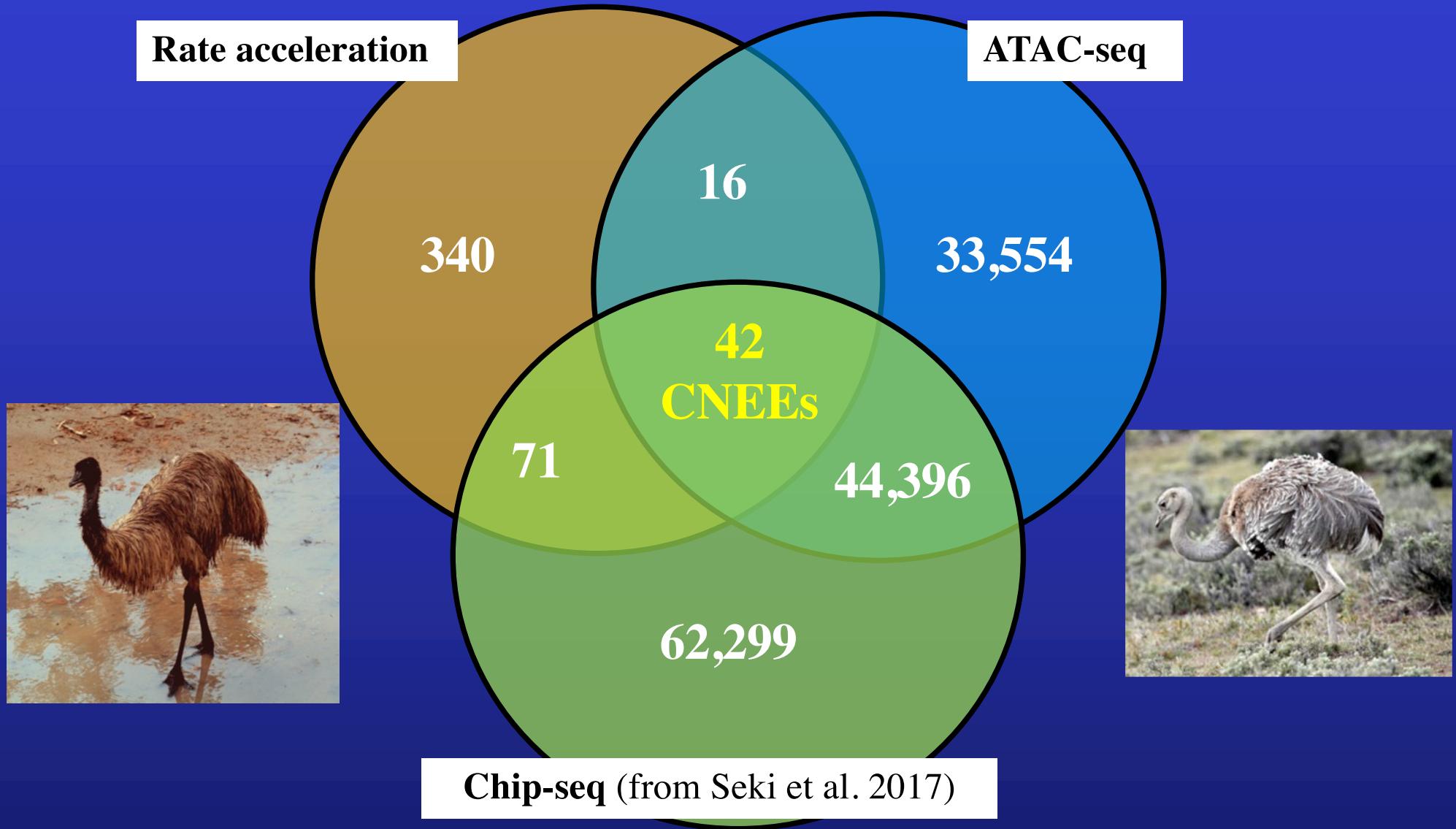
Ratite accelerated element 1317692 is contained under chicken ATAC peaks ...

CNEE

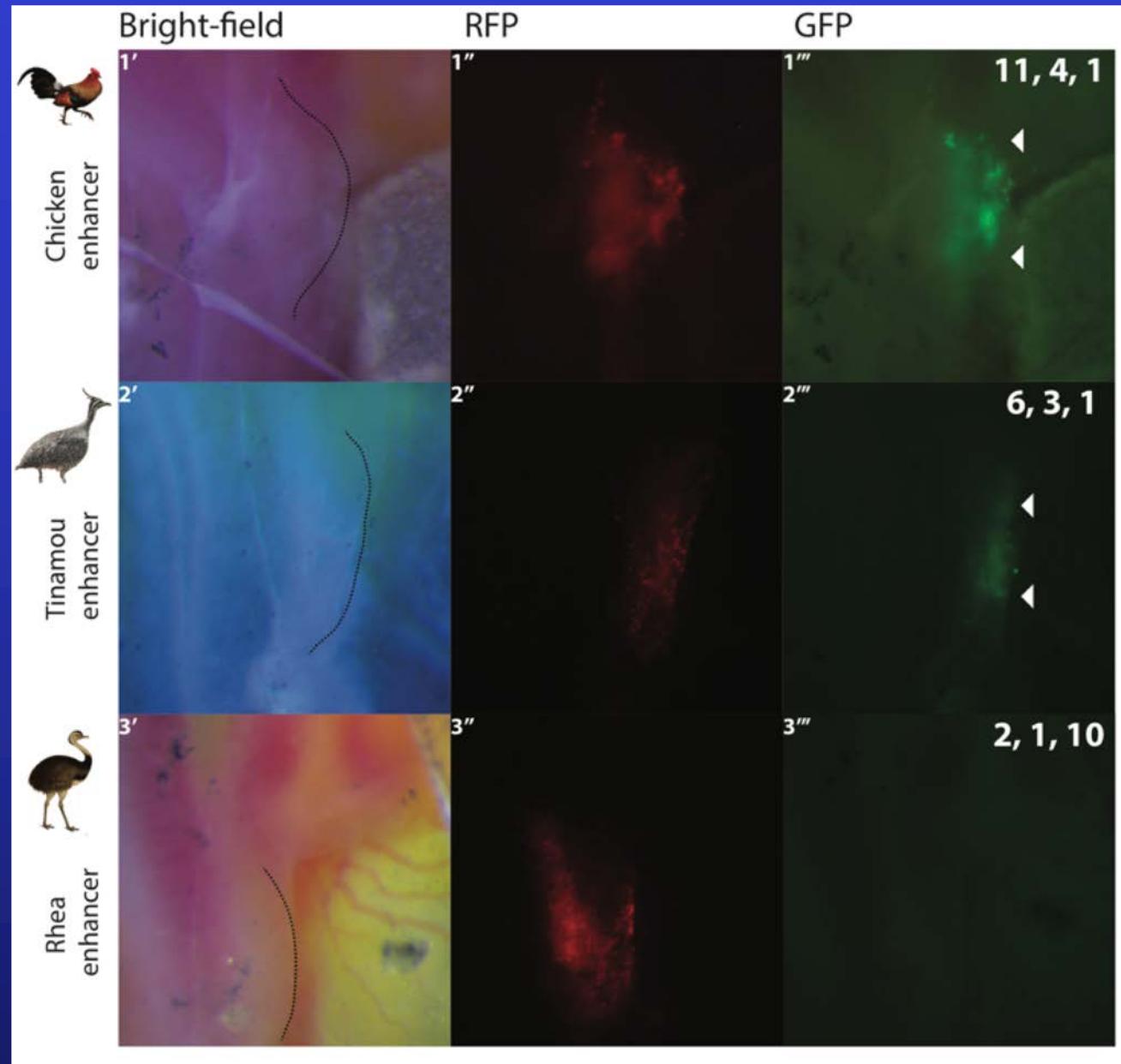


... but rhea is missing this peak

Combined information from multiple sources suggests candidate enhancers for flightlessness phenotypes



Volant version of CNEE drives gene expression in the developing forelimb of chicken but flightless version does not



Conclusions

- Ratite relationships require methods accommodating gene tree heterogeneity
- Noncoding CNEEs show more evidence for convergence than do coding regions
- Functional marks (ATAC-seq) suggests noncoding CNEEs are functional and related to species differences



Acknowledgements

Cliff Tabin Tim Sackton Phil Grayson Scott Edwards Chad Eliason

Statistics:
Zhirui Hu
Jun Liu



Grant numbers EAR-1355343/DEB-1355292

Michele Clamp Allan Baker Julia Clarke Alison Cloutier