

Cesky Krumlov

Phylogenomics Workshop 26/01/2024



UPPSALA
UNIVERSITET

Insights into mammalian evolution from large-scale comparative genomics

Matt Christmas

Institute of medical biochemistry and microbiology

Uppsala Universitet

matthew.christmas@imbim.uu.se



Post doc/researcher/
Trying to make a
career out of it!



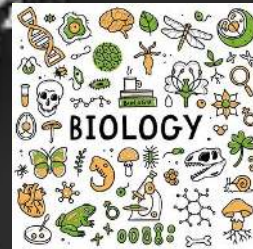
Prof. Matthew
Webster



Prof. Kerstin
Lindblad-Toh



Undergrad



Prof. Andrew Lowe



PhD



Zoonomia

Using the largest alignment of mammal genomes to date to explore mammalian evolution, disease, and conservation

RESEARCH ARTICLE SUMMARY

ZOONOMIA

Evolutionary constraint and innovation across hundreds of placental mammals

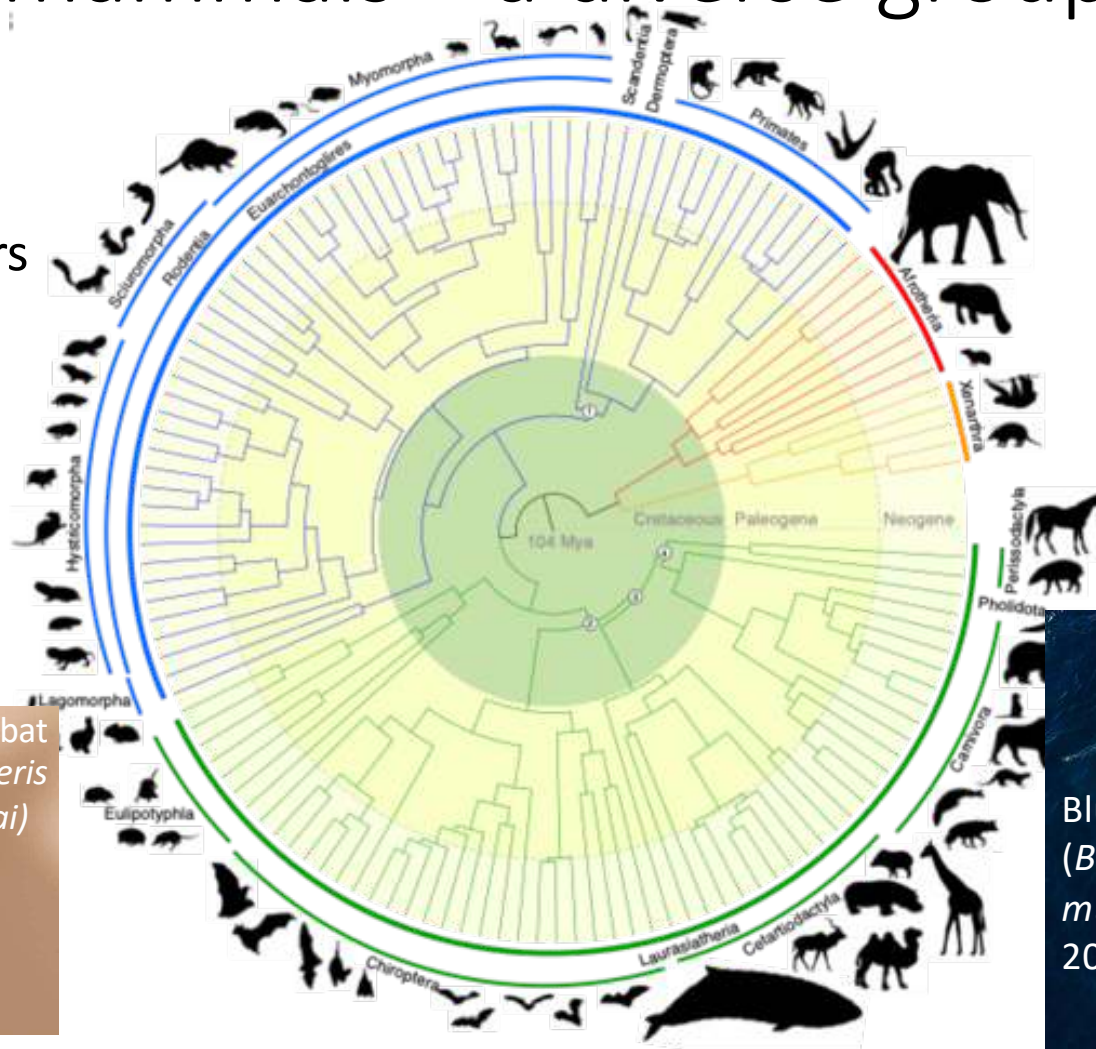
Matthew J. Christmas† and Irene M. Kaplow† et al.





Placental mammals – a diverse group

- 4 major subdivisions
- 19 extant orders
- >6,000 species
- ~100 million years of evolution



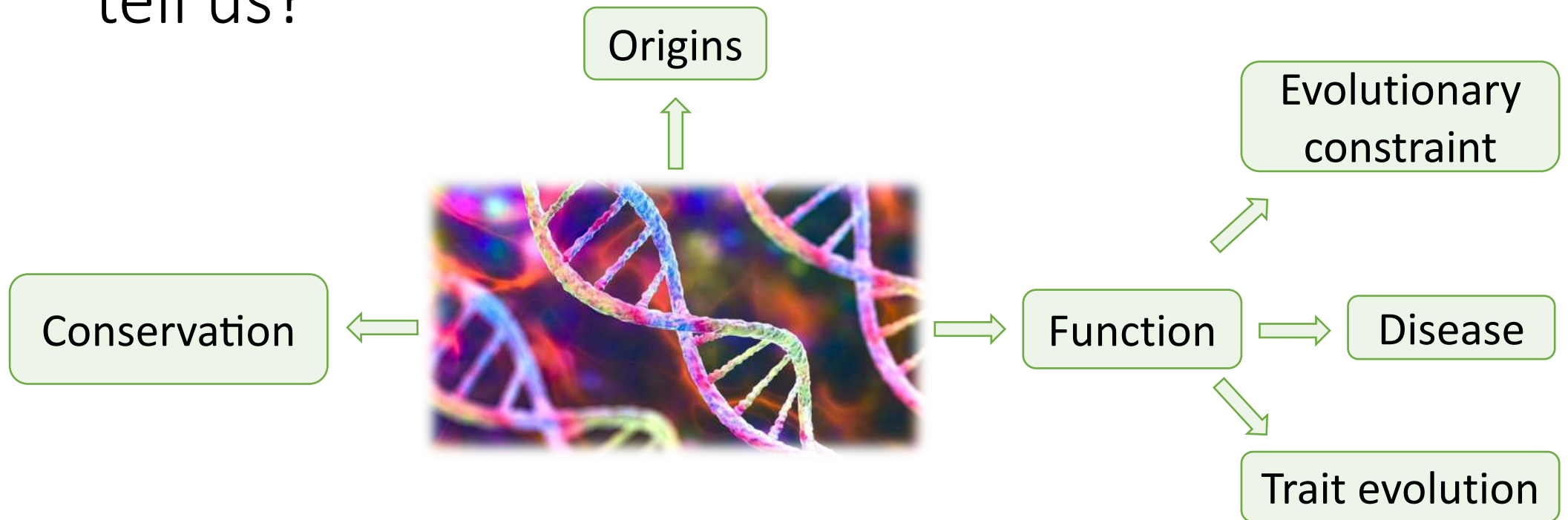
Bumblebee bat
(*Craseonycteris
thonglongyai*)

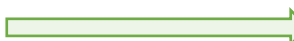
2 grams

Blue
(*Balaeniceps
mus*)
200 tons



What can a comparative genomics approach tell us?



Large multi-species genome alignments can be used to reveal answers to these questions  **The Zoonomia Project**

The Zoonomia project



UPPSALA
UNIVERSITET

Analysis

A comparative genomics multitool for scientific discovery and conservation

<https://doi.org/10.1038/s41586-020-2876-6>

Zoonomia Consortium*

Received: 17 April 2019

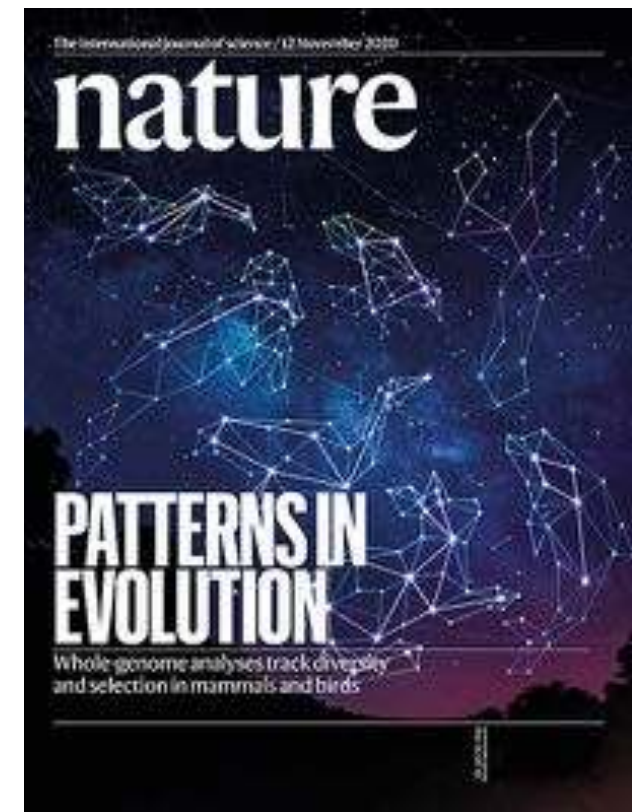
Accepted: 27 July 2020

Published online: 11 November 2020

Open access

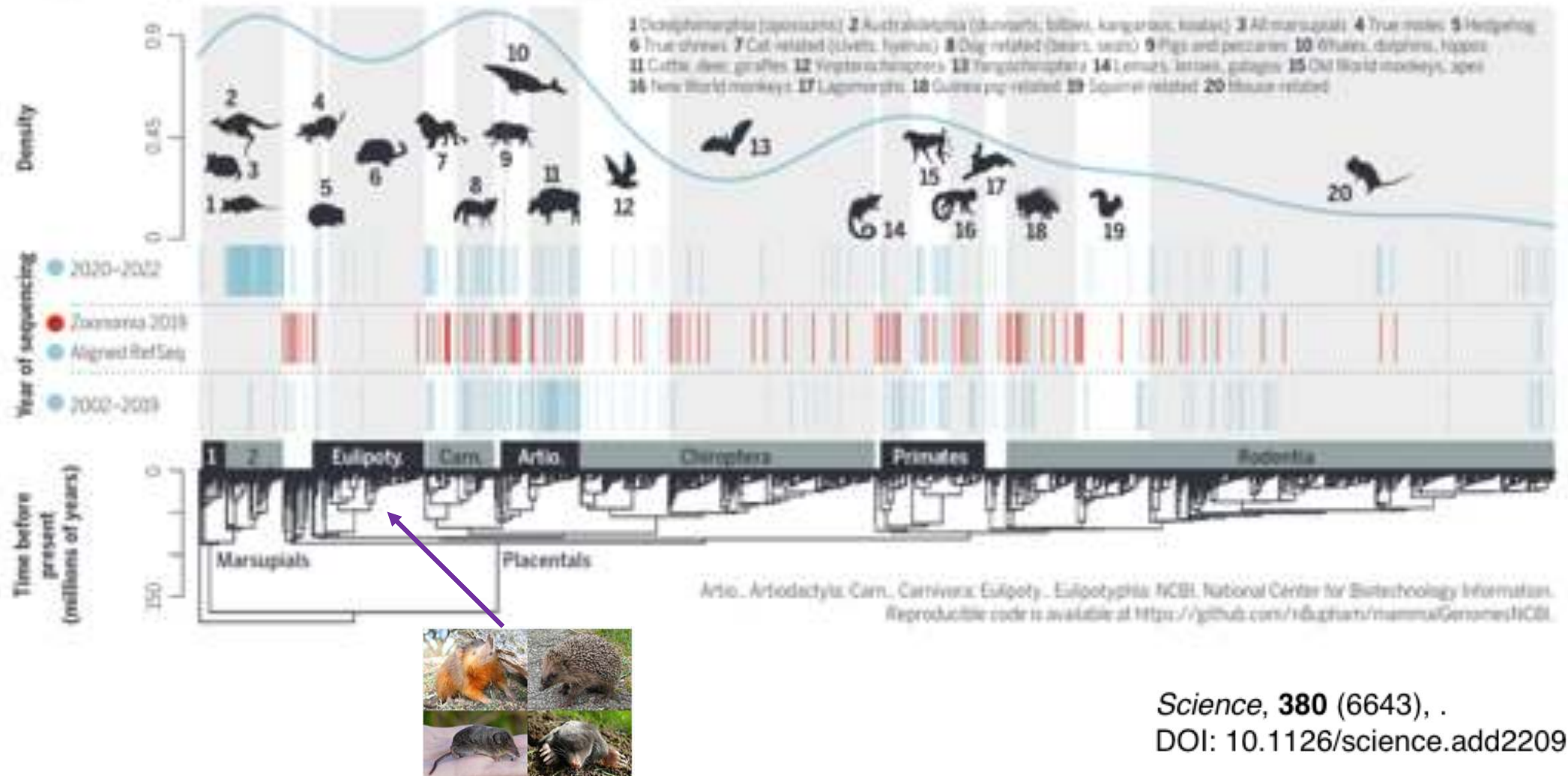
Check for updates

The Zoonomia Project is investigating the genomics of shared and specialized traits in eutherian mammals. Here we provide genome assemblies for 131 species, of which all but 9 are previously uncharacterized, and describe a whole-genome alignment of 240 species of considerable phylogenetic diversity, comprising representatives from more than 80% of mammalian families. We find that regions of reduced genetic diversity are more abundant in species at a high risk of extinction, discern signals of evolutionary selection at high resolution and provide insights from individual reference genomes. By prioritizing phylogenetic diversity and making data available quickly and without restriction, the Zoonomia Project aims to support biological discovery, medical research and the conservation of biodiversity.



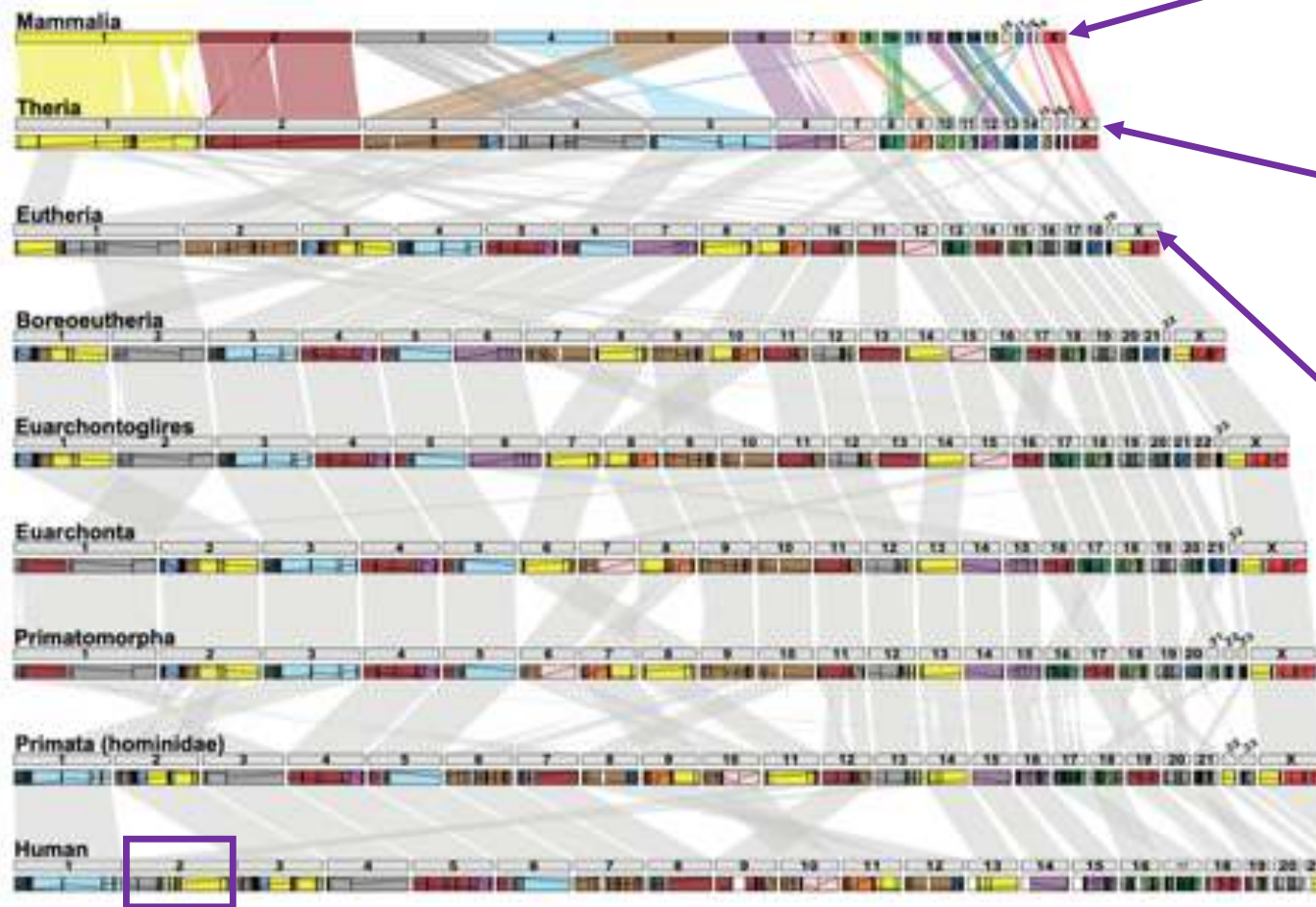
Genomes relative to year and phylogenetic relationships

Genomes for 675 mammal species relative to the Mammalia phylogenetic tree of 5911 living species shows the disproportionate representation of large-bodied and high-latitude species. Shown is the consensus timescaled phylogeny from Upham *et al.* (4) and genome data downloaded from NCBI on 9 February 2023.



Science, **380** (6643), .
DOI: 10.1126/science.add2209

Mammalian genome evolution



Ancestral mammal chromosomes
~180 Mya
19 autosomes + sex chromosomes

Ancestral therian (placentals +
marsupials) chromosomes 17
autosomes + sex chromosomes
96 chromosomal rearrangements
over 18 My

124 chromosomal
rearrangements over 53 My

Other great apes have 23
autosomes.
Fusion of two ancestral
chromosomes in human lineage

PNAS

RESEARCH ARTICLE | EVOLUTION

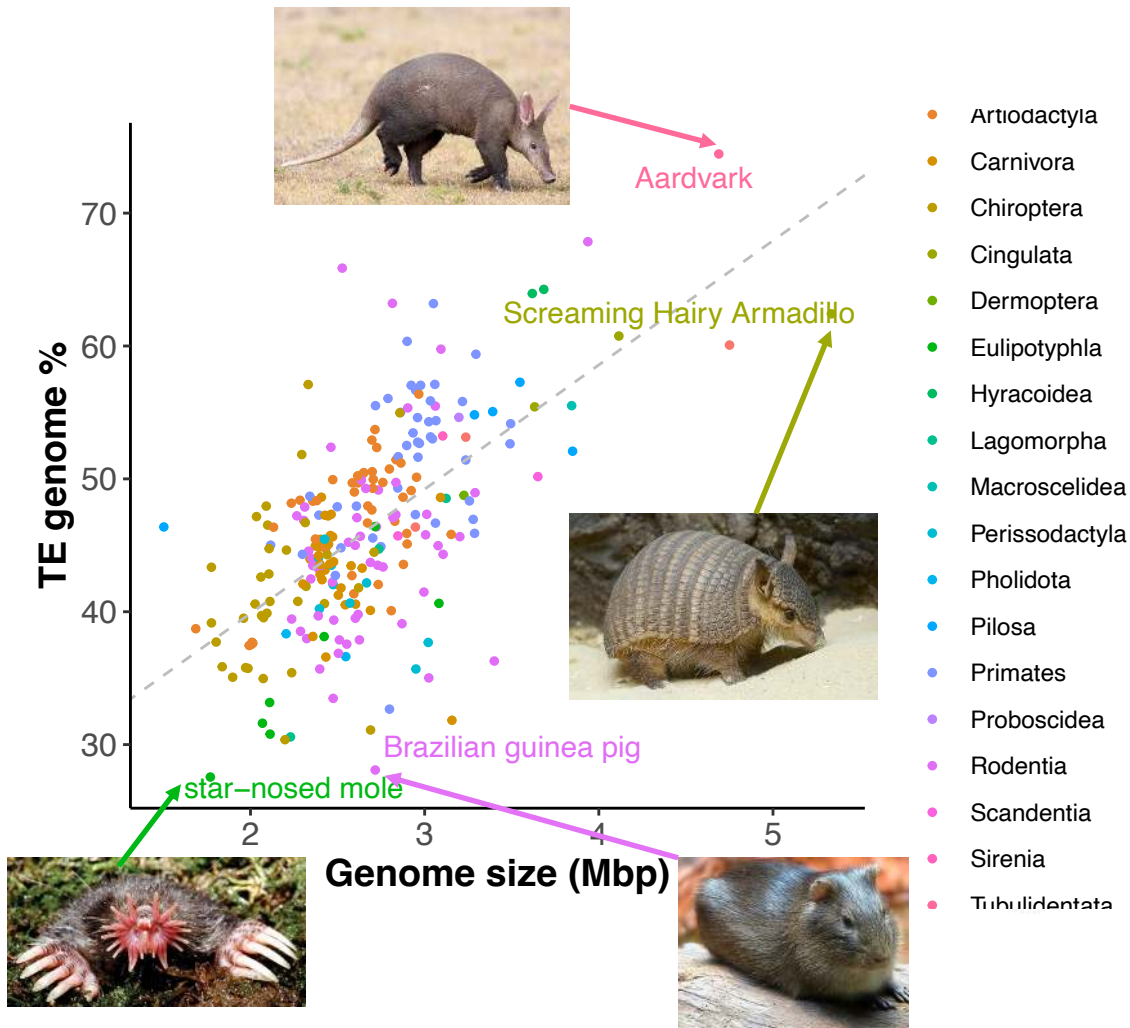
OPEN ACCESS



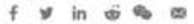
**Evolution of the ancestral mammalian karyotype and
syntenic regions**

Joana Deman¹, Marco Corbelli², Jaebum Kim³, Jason Turner-Maier⁴, Marta Farré⁵, Denis M. Larkin⁶, Oliver A. Ryder^{7,8}, Cynthia Steiner⁹,
Marlys L. Houck¹⁰, Shaune Hall¹¹, Lily Shuai¹², Stephen Thomas¹³, Thomas Swale¹⁴, Mark Daly¹⁵, Jonas Koriach¹⁶, Marcela Ullano-Salv^{17,18}, Camilla J. Mazzoni¹⁹,
Bruce W. Birren²⁰, Diane P. Genereux²¹, Jeremy Johnson²², Kerstin Lindblad-Toh²³, Jonas Koriach²⁴, Elinor K. Karlsson²⁵, Martin T. Nweeia²⁶,
Rebecca N. Johnson²⁷, Zoonomia Consortium¹, and Harris A. Lewin^{28,29}

Genome size variation and TEs



SPECIAL ISSUE RESEARCH ARTICLE ZOONOMIA



Insights into mammalian TE diversity through the curation of 248 genome assemblies

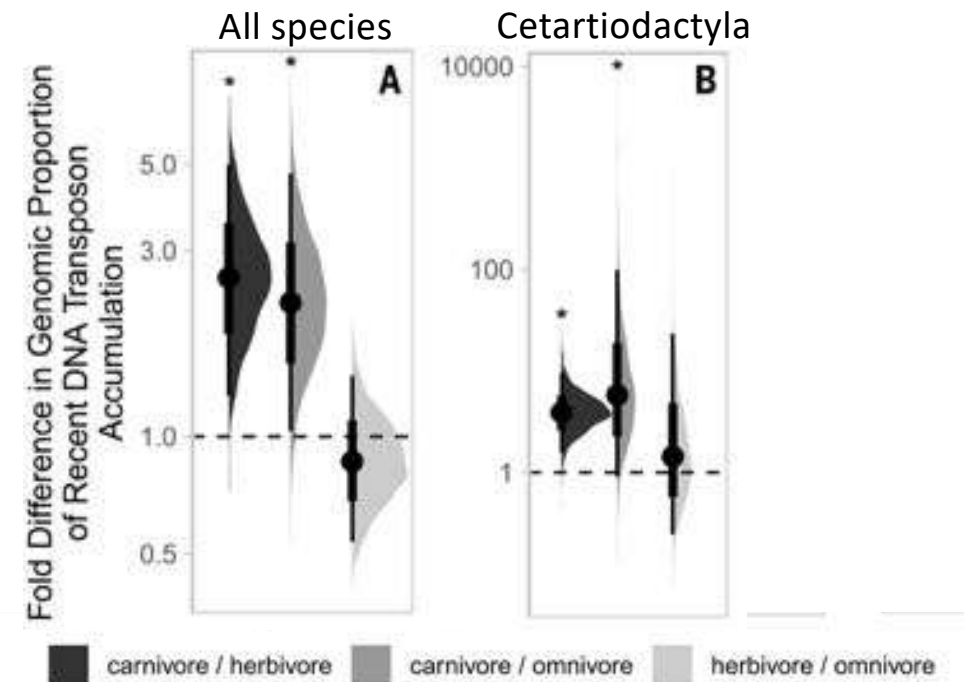
AUSTIN B. DESMARCK, NICOLE S. PALMAT, JENNY KORTSIAN, JENNA B. GRIMSHAW, MICHAELA HALSEY, KEVIN A. M. SULLIVAN

DIANA D. MORENO-SANTILLAN, CLAUDIA CROOKSHANKS, JACQUELYN ROBERTS, J. L. AND DAVID A. RAY +12 authors [Authors Info & Affiliations](#)

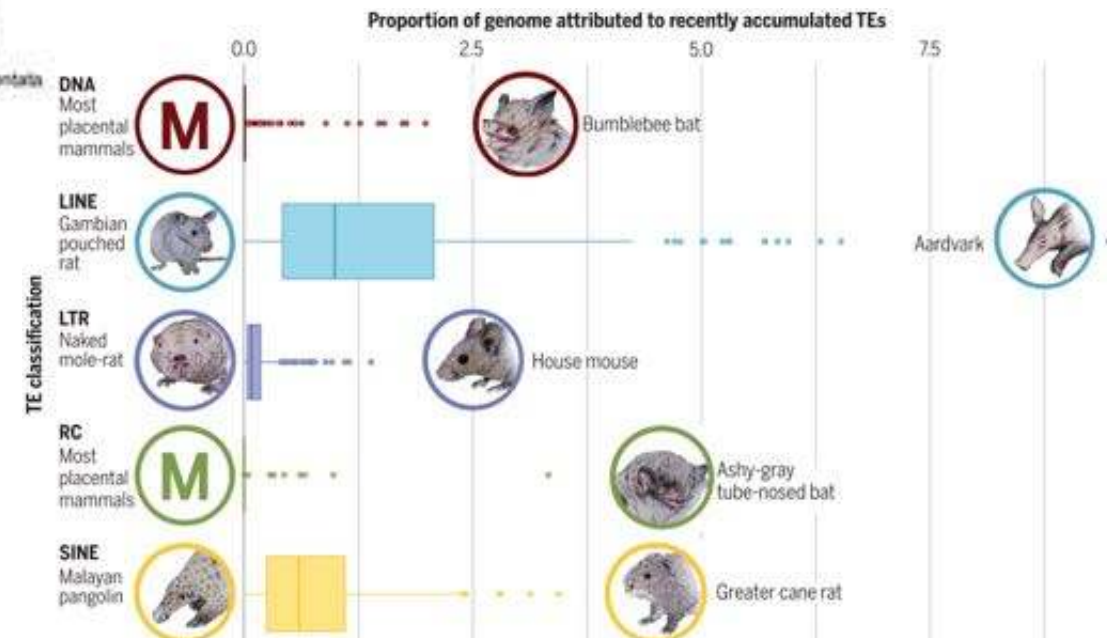
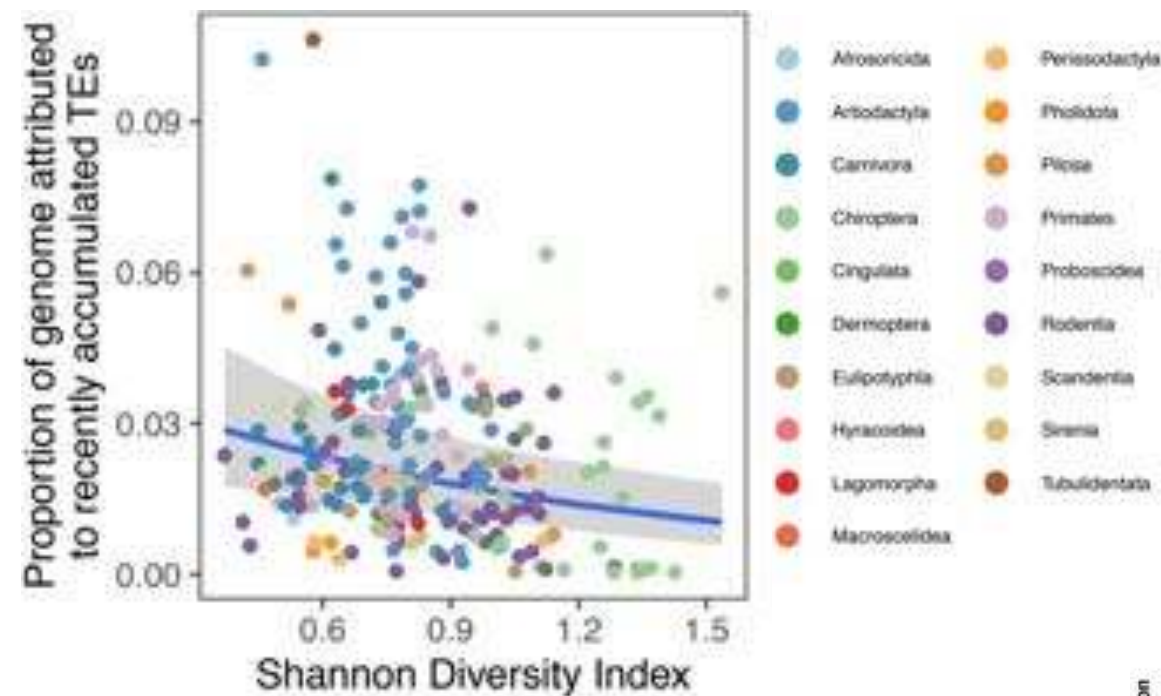
SCIENCE • 28 Apr 2023 • Vol 380, Issue 6643 • DOI: 10.1126/science.abn1430

Mean genome size = 2.67 Mb

Mean % TE = 46%



Genome size variation and TEs



SPECIAL ISSUE RESEARCH ARTICLE ZOOLOGIA

f t in d y

Insights into mammalian TE diversity through the curation of 248 genome assemblies

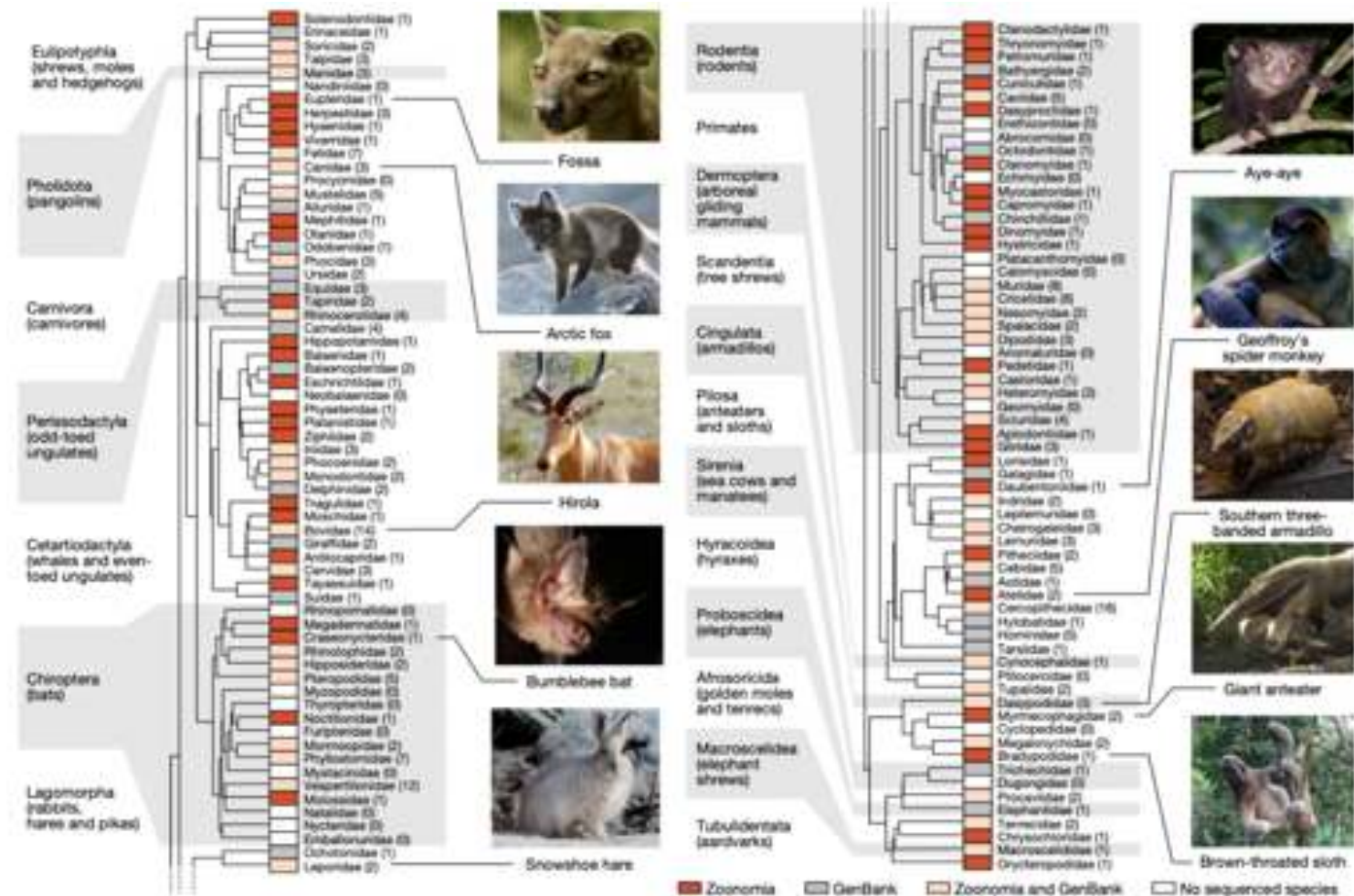
AUSTIN B. OSMAWSKI, NICOLE S. PAULAT, JENNY KOSTIAN, JENNIFER GRIMSHAW, MICHAELA HALSEY, KEVIN A. M. SULLIVAN

DIANA D. MORENO-SANTILLAN, CLAUDIA CROOKSHANKS, JACQUELYN ROBERTS [1-] AND DAVID A. RAY +12 authors [Authors Info & Affiliations](#)

SCIENCE • 29 Apr 2023 • Vol 380, Issue 6643 • DOI: 10.1126/science.abn1430

Zoonomia genomes

- 241 placental mammal genomes (240 species), representing >80% of mammalian families
- At least one long-range assembly (contig N50 > 20 kb and scaffold N50 > 10 Mb) per order
- Reference-free Cactus alignment
- Single-base measures of evolutionary constraint (phyloP)



Avoiding reference bias

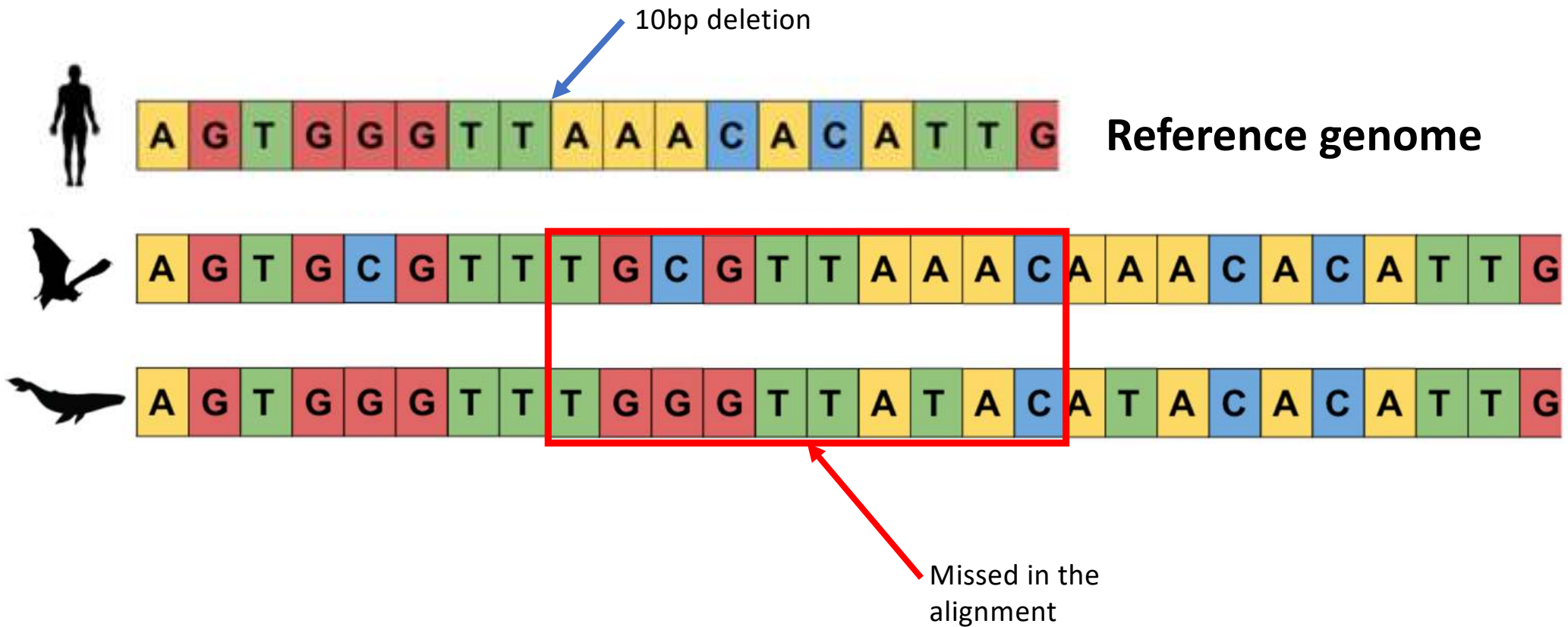
Progressive Cactus is a multiple-genome aligner for the thousand-genome era

Joel Armstrong, Glenn Hickey, Mark Diekhans, Ian T. Fiddes, Adam M. Novak, Alden Deran, Qi Fang, Duo Xie, Shaohong Feng, Josefin Stiller, Diane Genereux, Jeremy Johnson, Voichita Dana Marinescu, Jessica Alfoldi, Robert S. Harris, Kerstin Lindblad-Toh, David Haussler, Elinor Karlsson, Erich D. Jarvis, Guojie Zhang & Benedict Paten

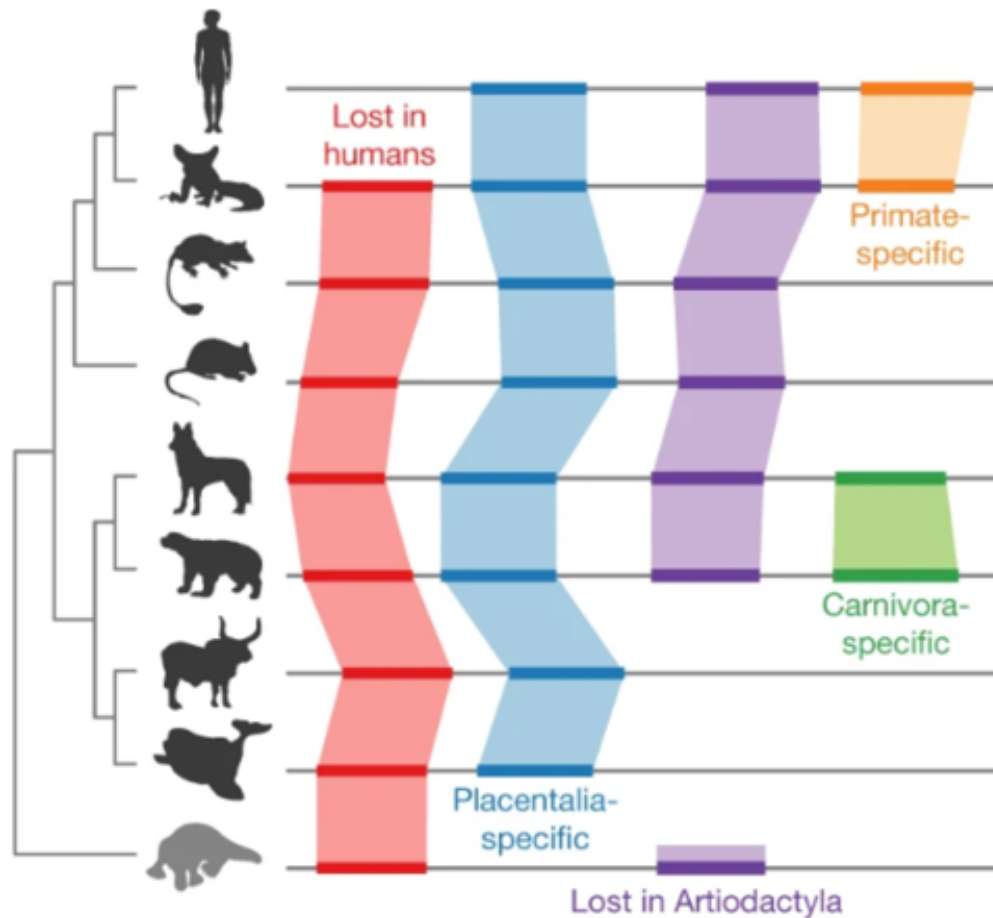
Nature 587, 246–251 (2020) | [Cite this article](#)



UPPSALA
UNIVERSITET



Reference-free alignment



Progressive Cactus is a multiple-genome aligner for the thousand-genome era

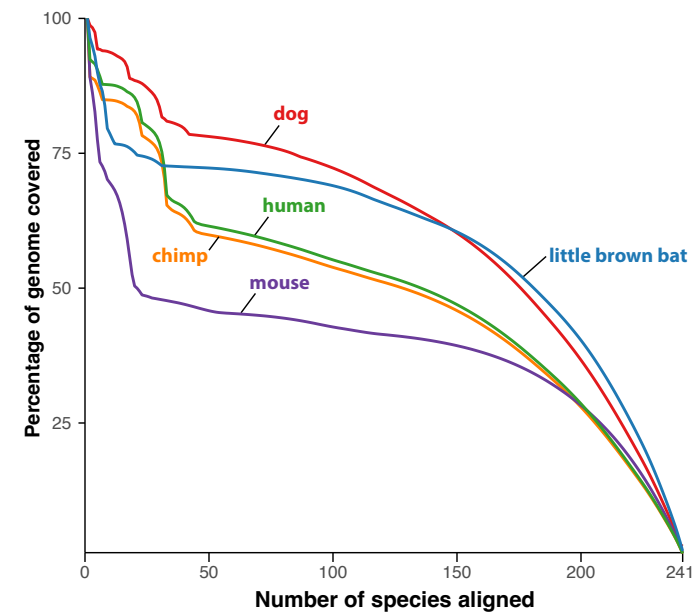
Joel Armstrong, Glenn Hickey, Mark Diekhans, Ian T. Fiddes, Adam M. Novak, Alden Deran, Qi Fang, Duo Xie, Shaohong Feng, Josefin Stiller, Diane Genereux, Jeremy Johnson, Voichita Dana Marinescu, Jessica Alföldi, Robert S. Harris, Kerstin Lindblad-Toh, David Haussler, Elinor Karlsson, Erich D. Jarvis, Guojie Zhang & Benedict Paten

Nature 587, 246–251 (2020) | [Cite this article](#)



UPPSALA
UNIVERSITET

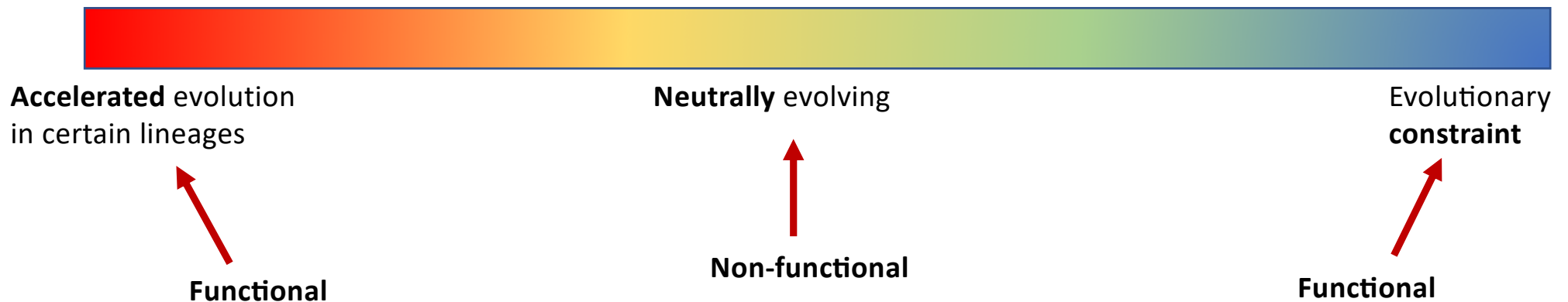
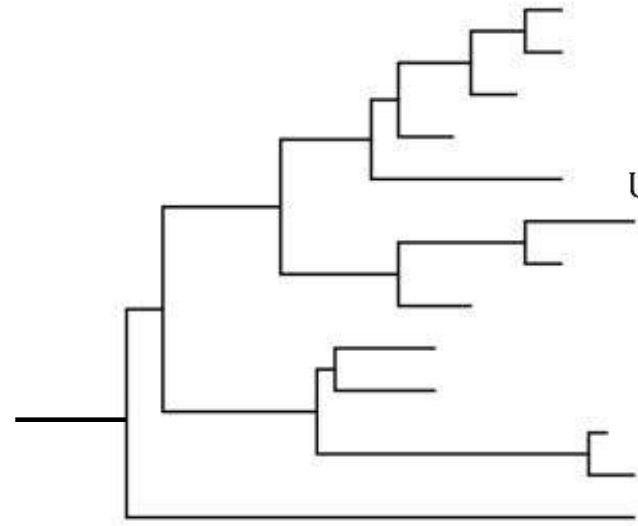
- All against all
- Can identify lineage-specific variation



Using comparative genomics to study genome evolution



UPPSALA
UNIVERSITET



Consequences of this can be utilized to identify functional genomic regions via **multi-species alignments**

A Constraint

B Neutral

C Acceleration



A	T	C	T	A	A	G	G	A	C	C	C	G	G	A	T	T	A	A	C	A
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---



A	T	C	T	A	A	G	G	T	C	G	C	G	G	A	A	A	C	C	C	A
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---



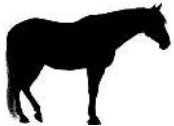
A	T	C	T	A	A	G	G	A	C	C	G	G	G	C	G	T	G	G	T	T
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---



A	T	C	T	A	A	G	C	A	A	C	C	G	G	A	T	C	C	A	G	G
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

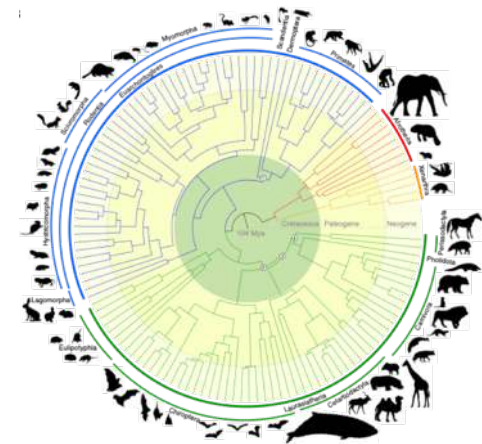
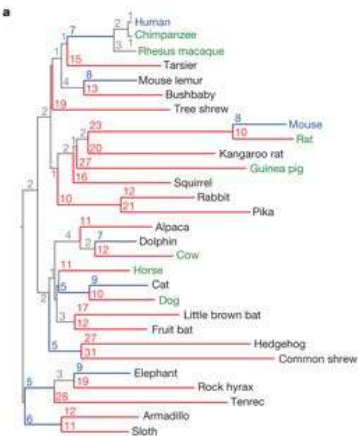


A	T	C	T	A	A	G	C	A	A	C	C	G	T	A	C	T	T	T	A	A
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---



A	T	C	T	A	A	G	G	A	C	G	C	T	G	A	G	G	G	A	T	T
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

More species, more power



	29 mammals (2011)	240 mammals (2023)
Total branch length (substitutions/site)	4.5	16.6
Probability of identical site (p)	0.02	6x10 ⁻¹⁸
Expected identical sites	62 Mbp	191 bp
Constraint resolution	12 bp (p = 1 x 10 ⁻²⁵)	Single base pair
Human genome constraint	>5.5%	>10.7%

Open Access | Published: 12 October 2011

A high-resolution map of human evolutionary constraint using 29 mammals

Kerstin Lindblad-Toh , Manuel Garber, Or Zuk, Michael F. Lin, Brian J. Parker, Stefan Washietl, Pouya Kheradpour, Jason Ernst, Gregory Jordan, Ewan Mauceli, Lucas D. Ward, Craig B. Lowe, Alisha K. Holloway, Michele Clamp, Sante Gnerre, Jessica Alföldi, Kathryn Beal, Jean Chang, Hiram Clawson, James Cuff, Federica Di Palma, Stephen Fitzgerald, Paul Flicek, Mitchell Guttman, Broad Institute Sequencing Platform and Whole Genome Assembly Team, Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, Genome Institute at Washington University, ... Manolis Kellis 

[Show authors](#)

Nature **478**, 476–482 (2011) | [Cite this article](#)

RESEARCH ARTICLE SUMMARY

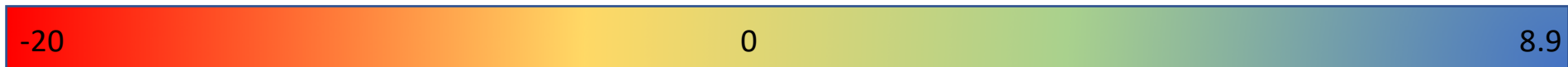
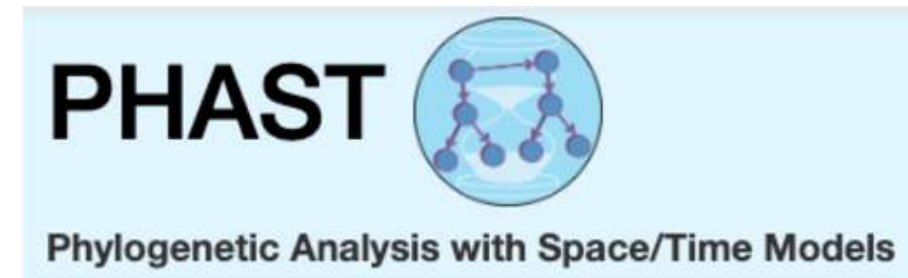
ZOONOMIA

Evolutionary constraint and innovation across hundreds of placental mammals

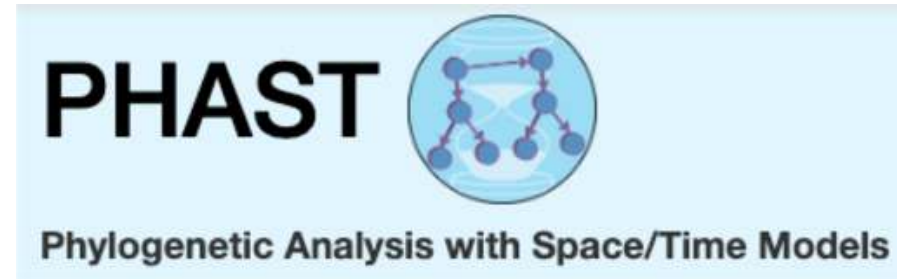
Matthew J. Christmas[†] and Irene M. Kaplow[†] et al.

Estimating evolutionary constraint and acceleration - phyloP

- **PhyloP** calculates constraint and acceleration p-values based on a multi-species alignment and **a model of neutral evolution** at individual nucleotides
- Each base position in the alignment has a phyloP score: 0 = neutral, >0 = constraint, <0 = acceleration



Running phyloP



- Generate neutral model file using phyloFit
 - Inputs: alignment file for neutral sites (.maf format from Cactus), tree topology
- Run phyloP
 - Inputs: alignment file and phylogenetic model from above
 - --method LRT – likelihood ratio test. How likely is the observed divergence given neutral model?
 - --mode CONACC gives positive p values/scores to indicate constraint and negative scores to indicate acceleration

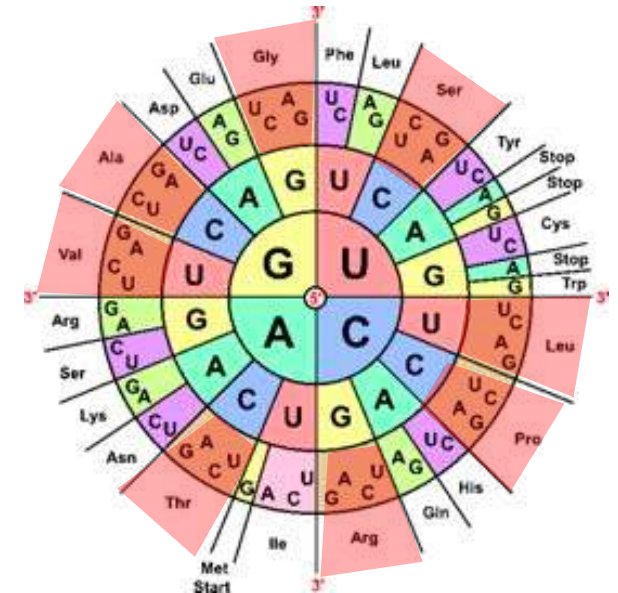
-20

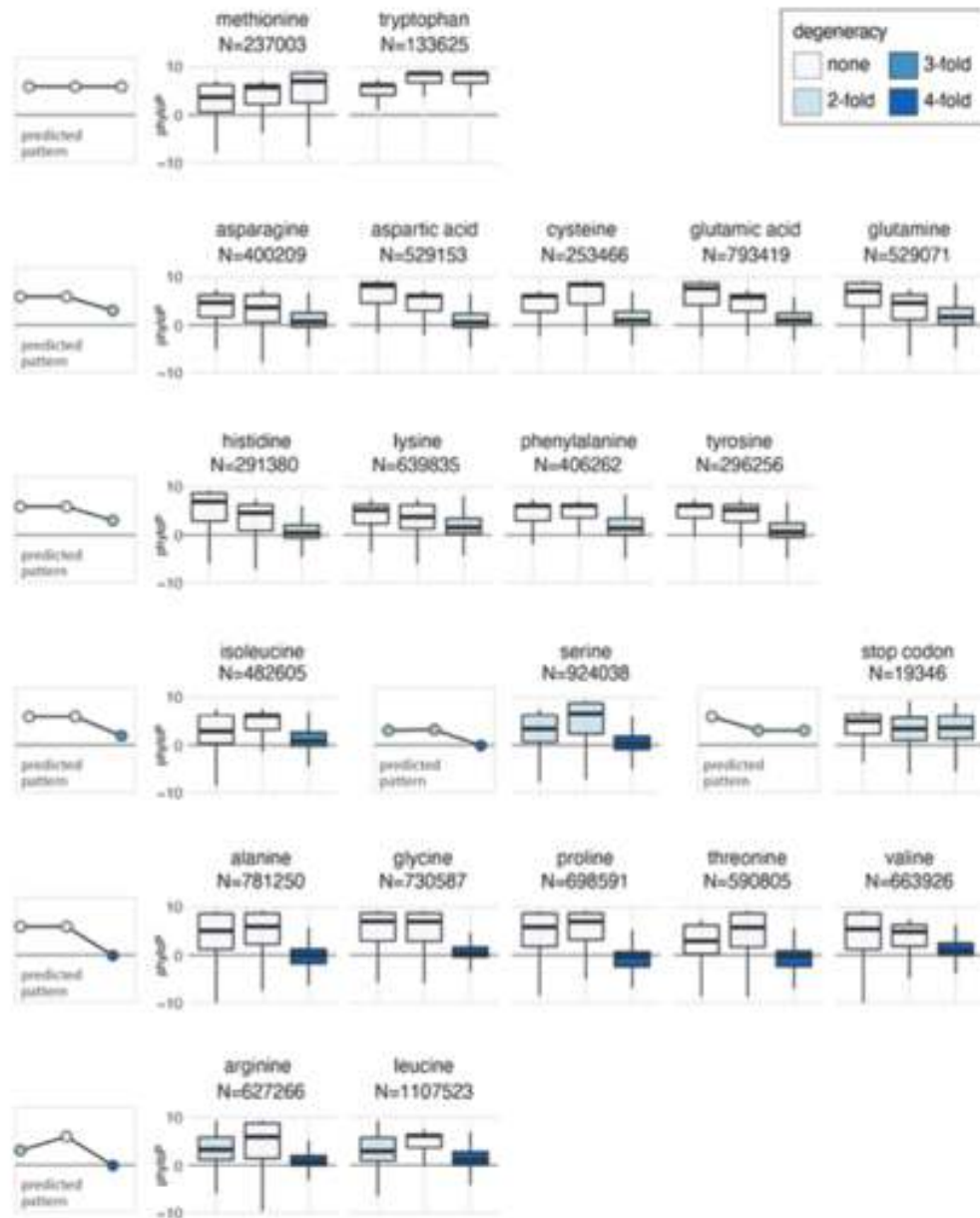
0

8.9

Models of neutral evolution

- To understand sequence evolution over time we need **accurate models of neutral evolution**, i.e. sequence changes over time in the absence of selection
- Important to have accurate models of neutral evolution for
 - Inferring phylogeny (model assumption)
 - Divergence time estimates
 - Detecting constraint and acceleration
- Models of neutral evolution:
 - Four-fold degenerate sites
 - **Ancestral repeats**

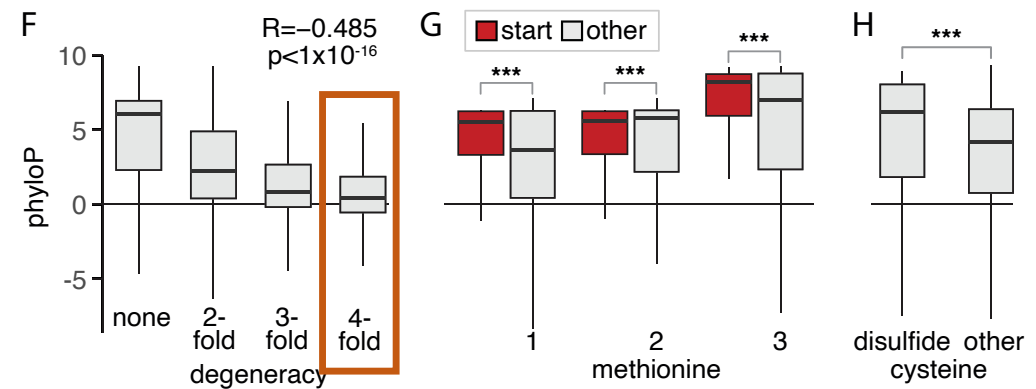
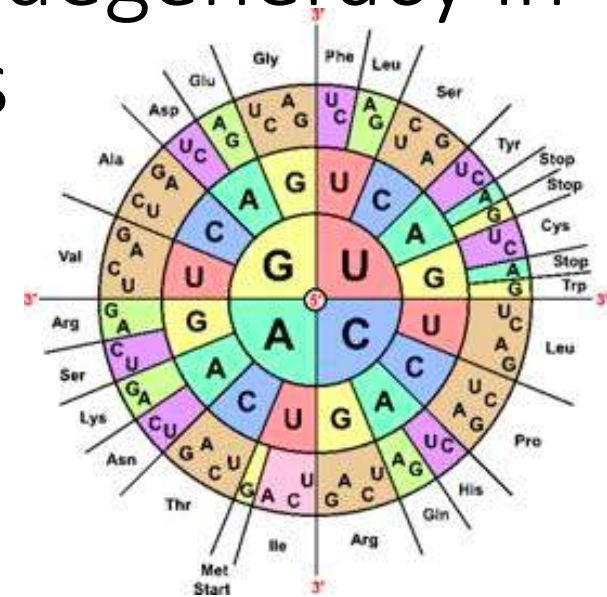




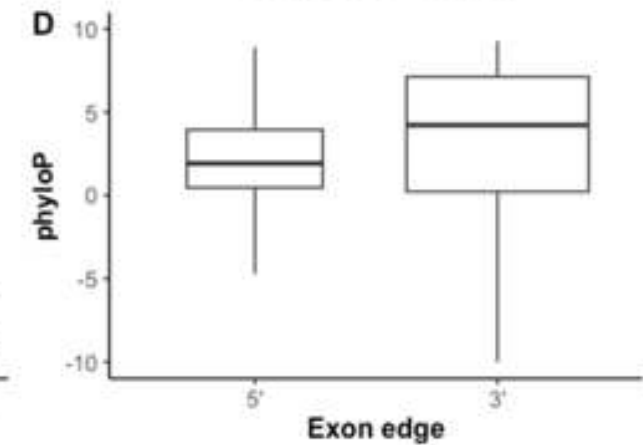
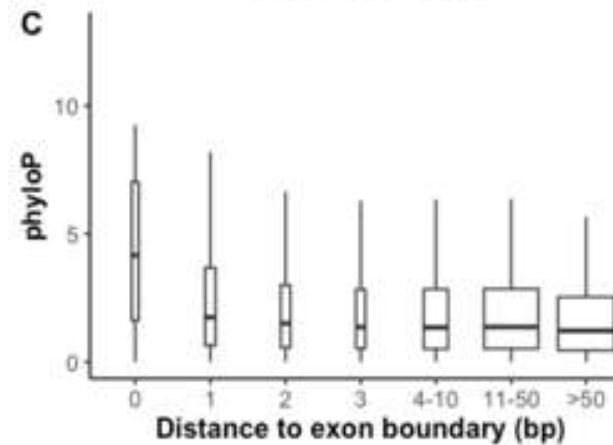
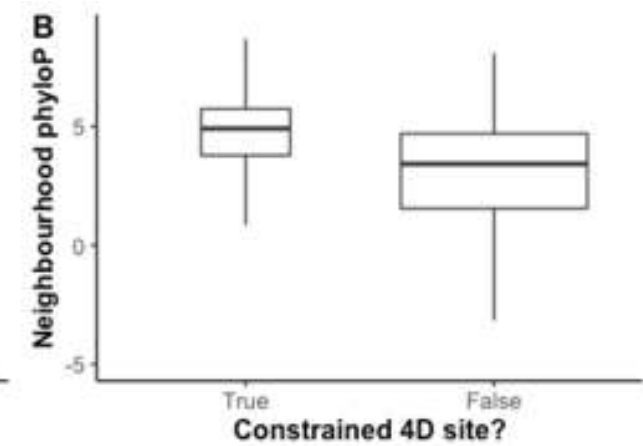
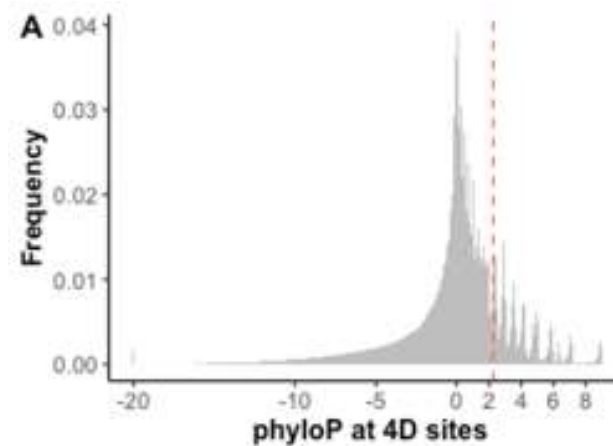
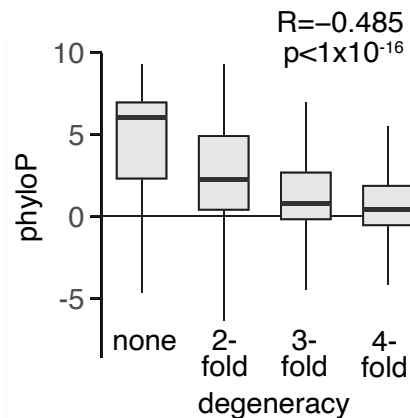
Constraint scores reflect degeneracy in codons



UPPSALA
UNIVERSITET



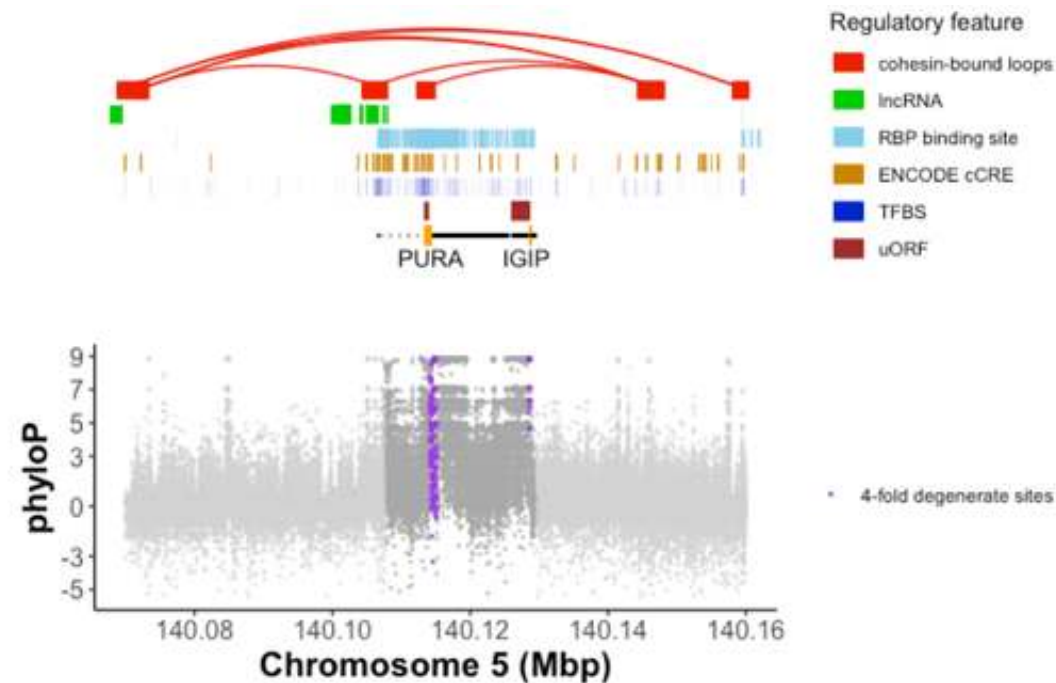
Four-fold degenerate sites generally evolve neutrally, but it's not all neutral...



18.6% of 4D sites are under constraint (FDR=0.05)

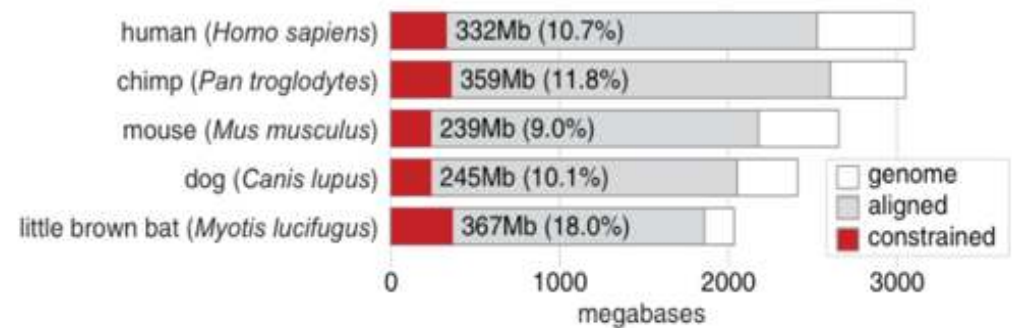
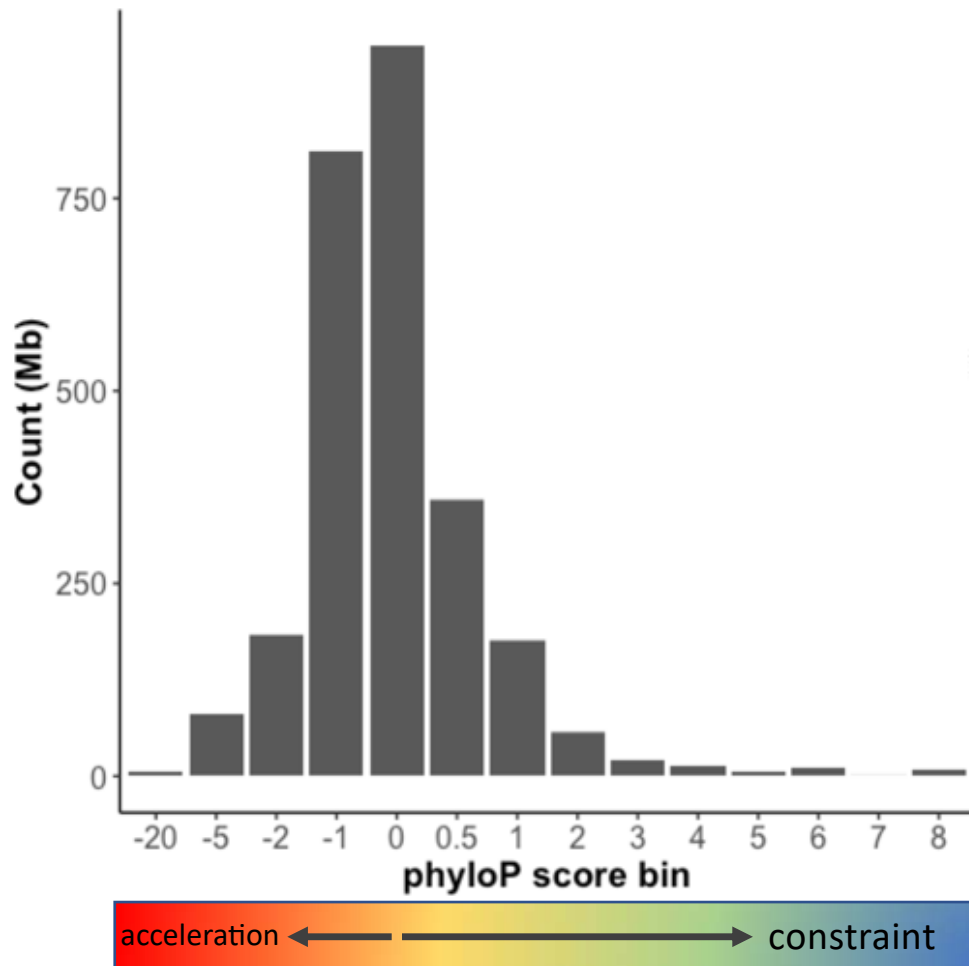
Christmas *et al.* In Prep.

Constraint at 4D sites relates to gene regulation



Christmas *et al.* In Prep.

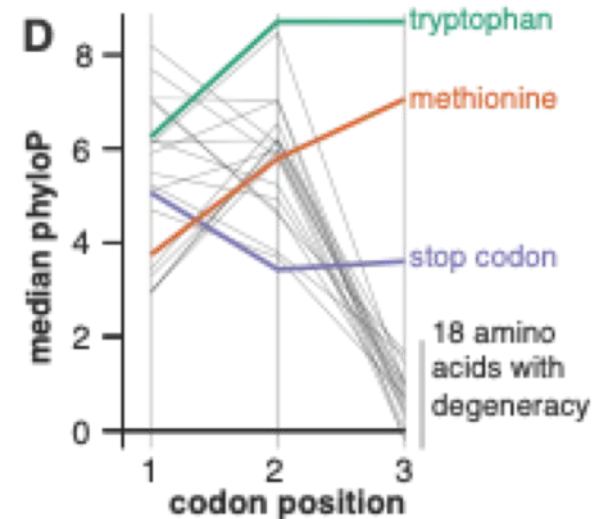
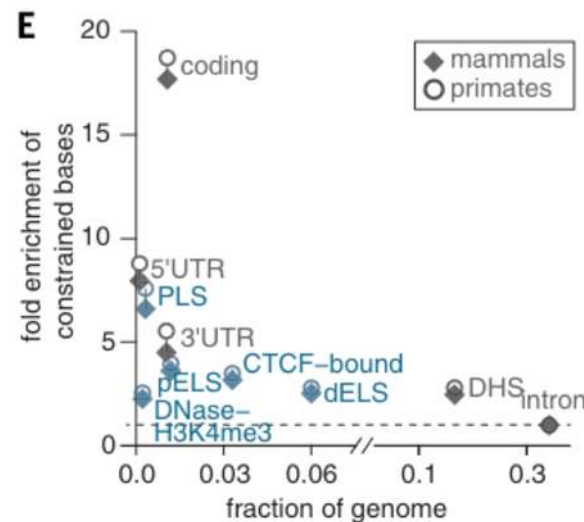
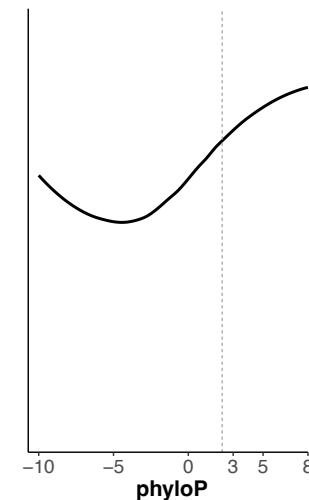
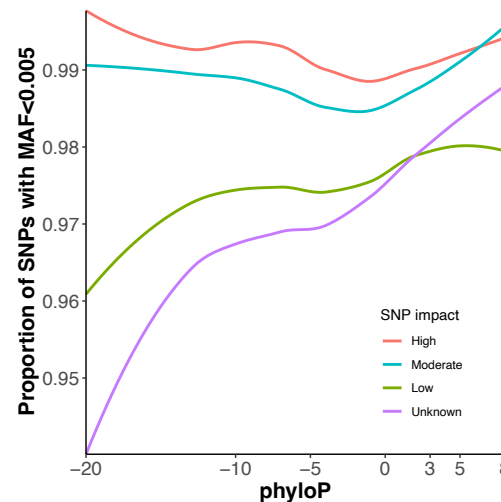
Constraint in mammals



- Max score: 8.903 – all species aligned and identical, **highly constrained**
- **3.6 million perfectly conserved positions**

Distribution of constraint

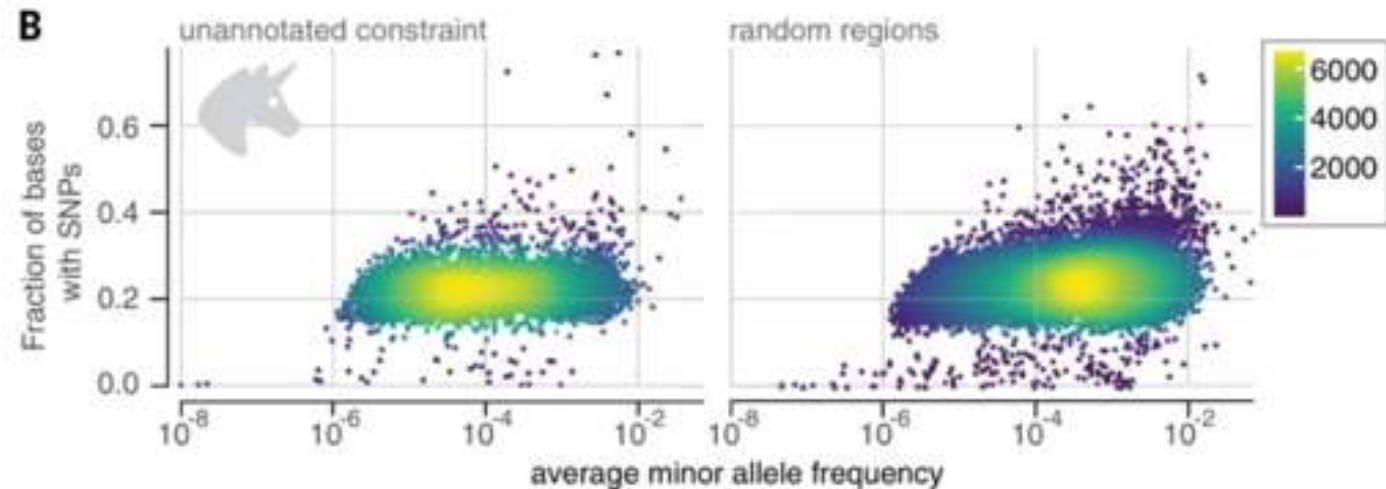
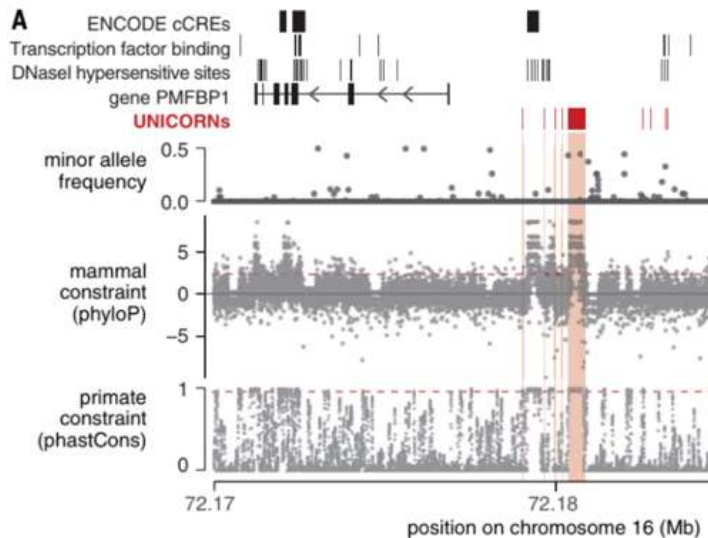
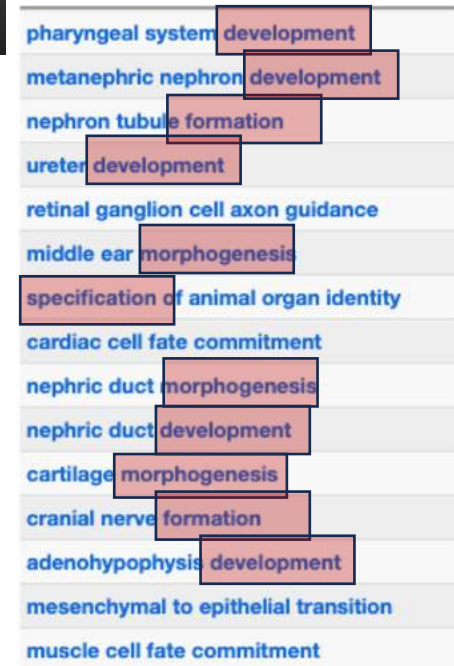
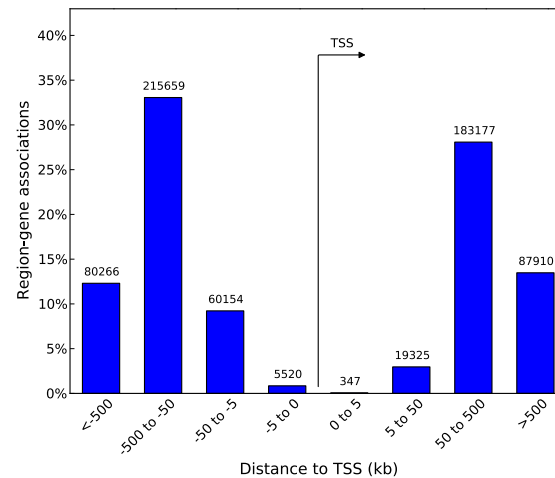
- 3.26% of bases in the human genome identified as under significant constraint at $FDR < 0.05$ ($phyloP > 2.270$)
- Form clusters – most (80%) are within 5bp of another constrained base
- Positions under constraint are less variable in humans (TOPMed SNPs)
- Constraint strongly enriched in coding sequence, moderately enriched in regulatory elements
- Single-base resolution of constraint scores evidenced by constraint in codons



Constraint in unannotated regions



- Nearly half of all constrained bases sit outside of annotations
- Identified >400,000 UNannotated Intergenic COnstrained RegioNS (UNICORNs)



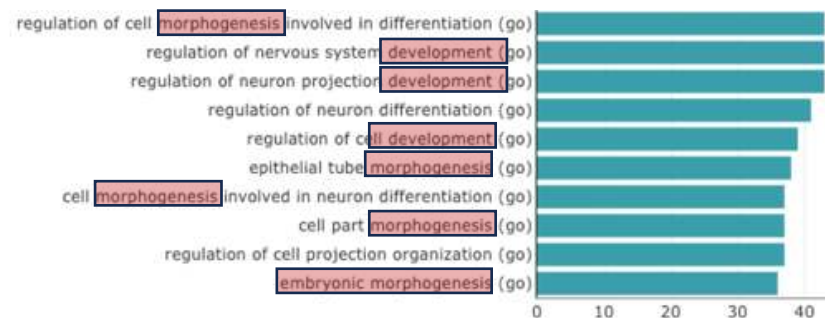
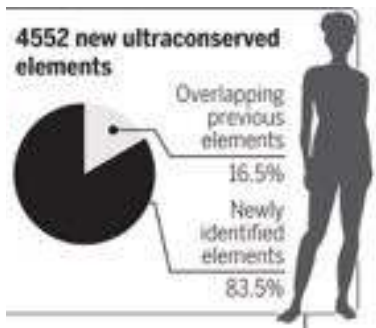
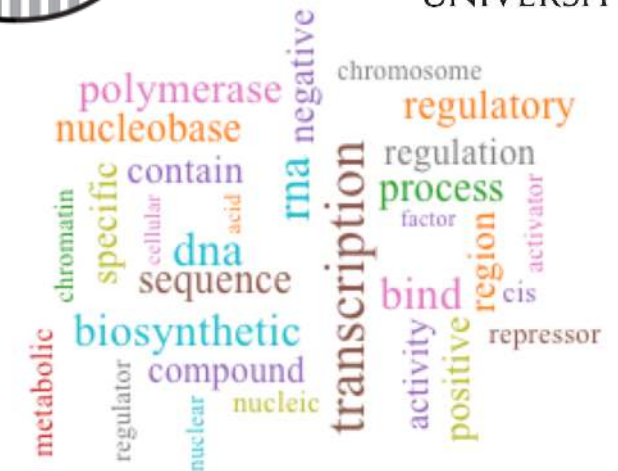


UPPSALA
UNIVERSITET

zooUCEs



- Original set (2004): 481 segments > 200bp 100% conserved between human, mouse, and rat
- zooUCEs: $\geq 98\%$ of species (235) are *aligned and identical*
- 4,552 zooUCEs ≥ 20 bp. 753 overlap 318 original UCEs. 27 ≥ 100 bp
- 69% are outside of protein coding exons



Genomic 'hotspots' of constraint

- Constraint measured in 100kb windows across human genome
- 53 bins with significant constraint ($q < 0.05$)
- Includes all HOX clusters
- Large gene deserts with high constraint

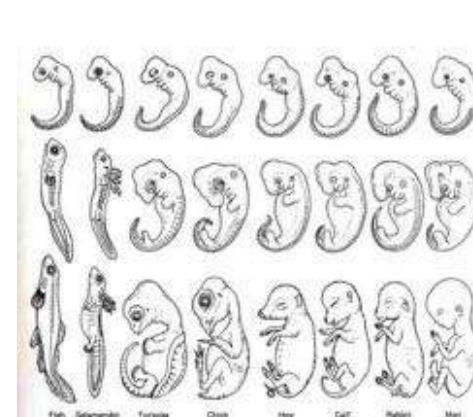
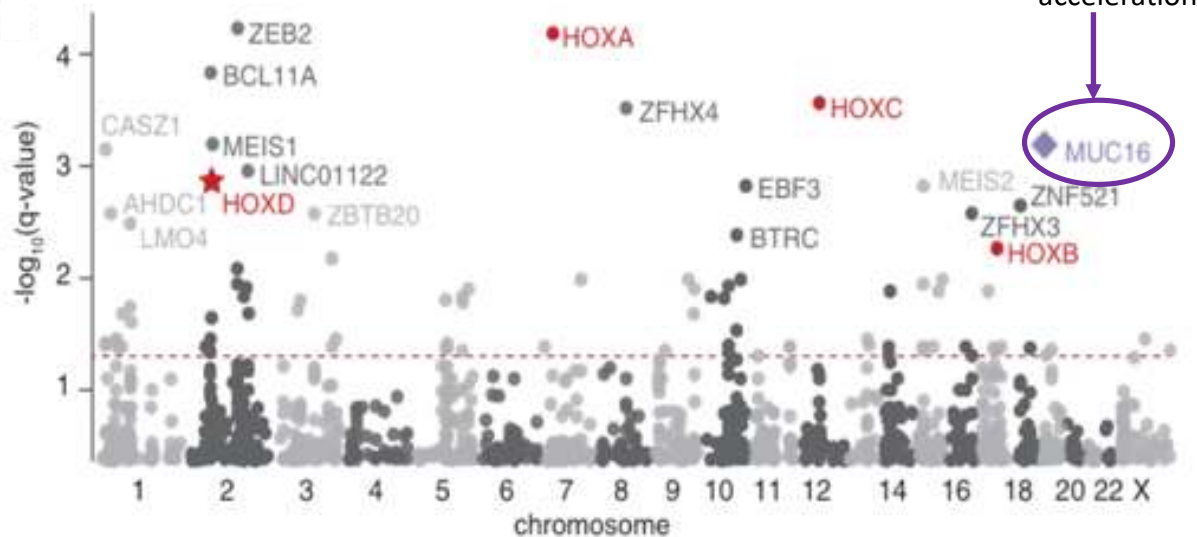
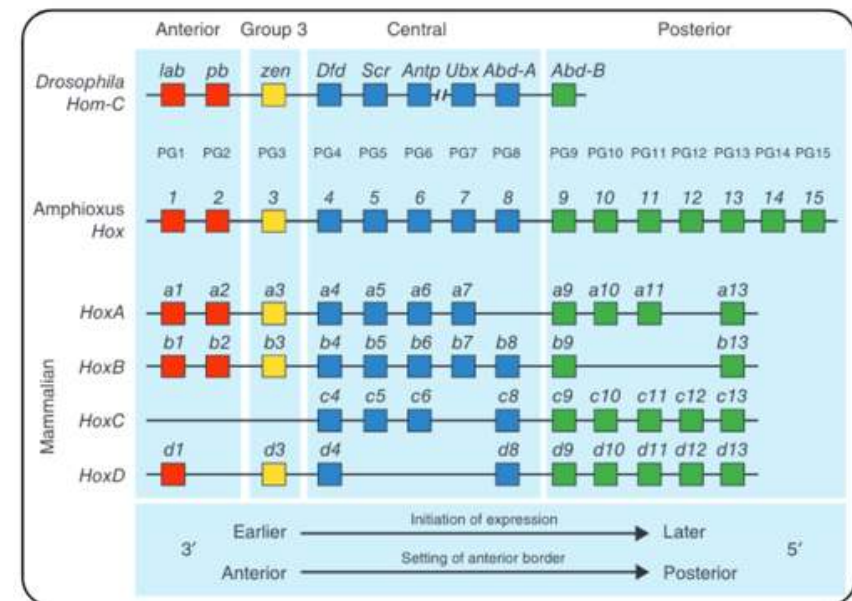


Figure 8.7. Comparison of vertebrate embryos at three different stages of embryonic development (early, mid, and late) across various species including Fish, Salamander, Turtle, Chick, Hry, Cat, Rabbit, and Man.



UPPSALA
UNIVERSITET



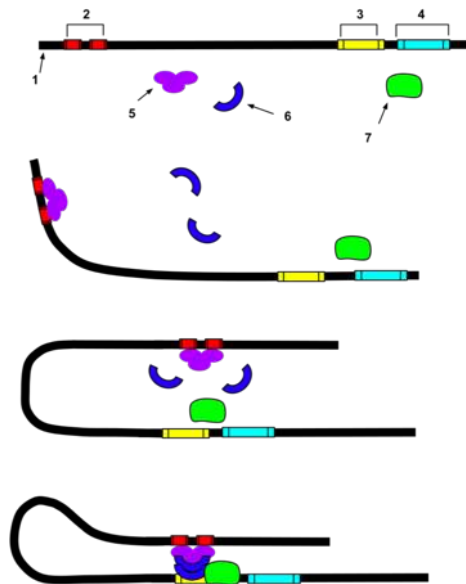
Mammalian Embryo: Hox Genes

Hugo J Parker, Stowers Institute for Medical Research, Kansas City, Missouri, USA

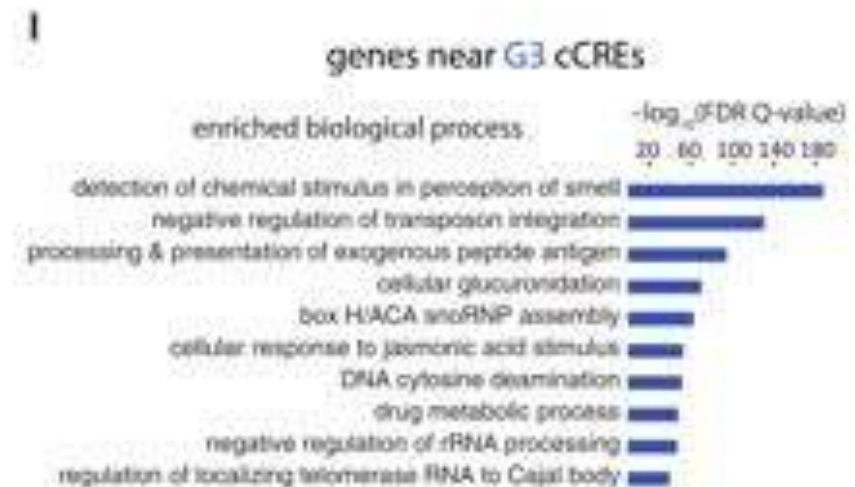
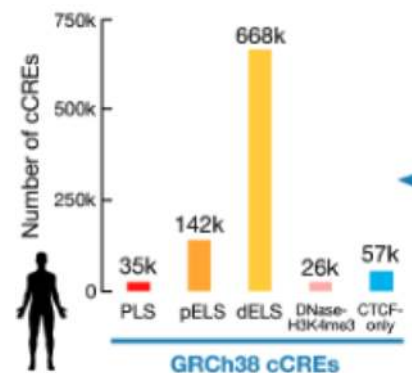
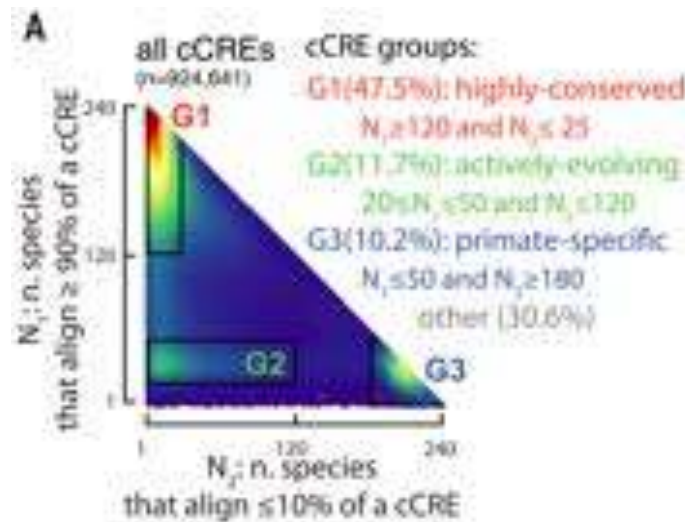
Based in part on the previous versions of this eLS article 'Mammalian Embryo: Hox Genes' (2009, 2015).



So what's all this constrained sequence doing if not coding for proteins?



1. DNA
2. Enhancer
3. Promoter
4. Gene
5. Transcription
- Activator Protein
6. Mediator Protein
7. RNA Polymerase



SPECIAL ISSUE RESEARCH ARTICLE | ZOOONOMIA

f t in d e

Mammalian evolution of human cis-regulatory elements and transcription factor binding sites

GREGORY ANDREWS, KALI FAN, HENRY E. PHATT, MOHESANHA PHALKE, ZOOONOMIA CONSORTIUM, ELINOR S. KARLSSON, KRISTIN LINDBLAD-TOH, STEVEN GAZAL, JILL E. MOORE, AND ZHIPING WENG | [Authors Info & Affiliations](#)

SCIENCE • 28 Apr 2023 • Vol 380, Issue 6643 • DOI:10.1126/science.abe7920



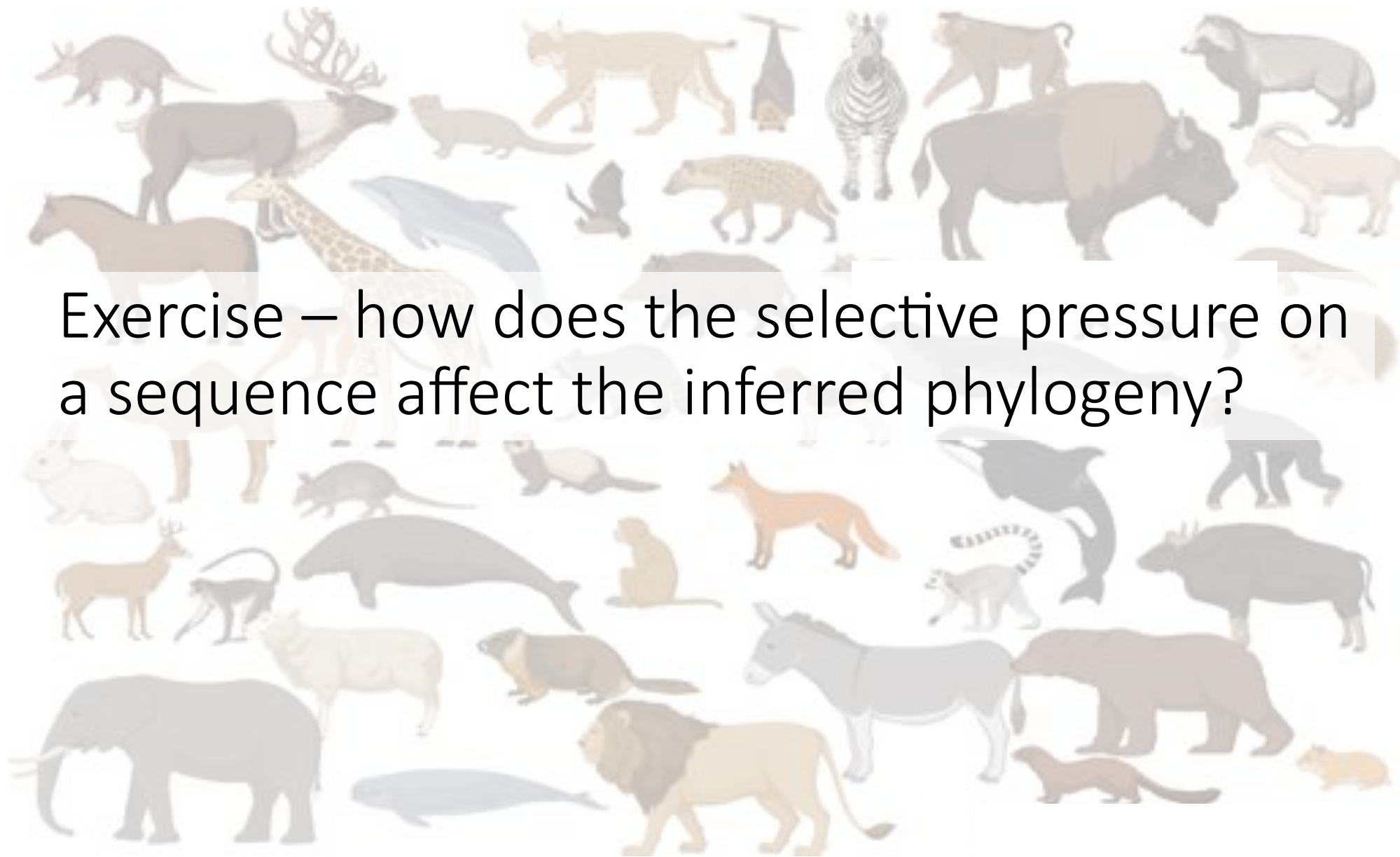
Questions?

Break

30 mins

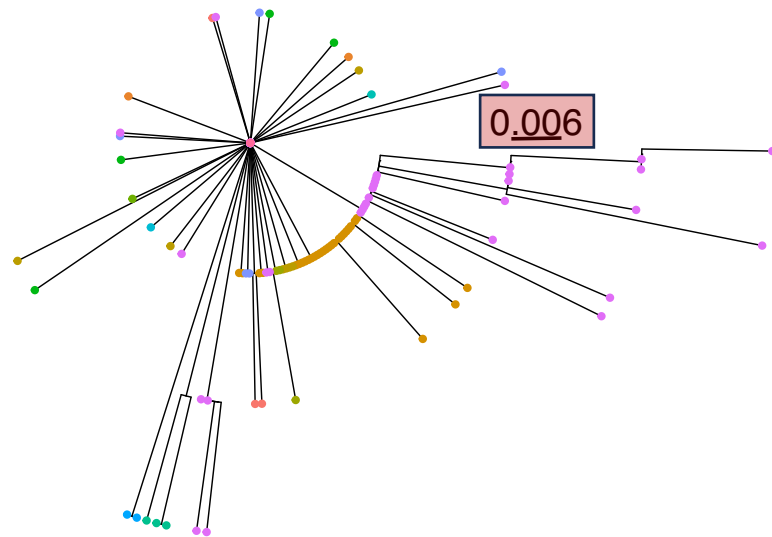
A large collection of various mammals, including deer, giraffe, dolphin, zebra, bison, goat, rabbit, mouse, cat, fox, orca, monkey, elephant, lion, donkey, bear, and walrus, arranged in a grid-like pattern. The text "Using estimates of constraint to explore mammalian evolution" is overlaid on the image.

Using estimates of constraint to explore
mammalian evolution



Exercise – how does the selective pressure on a sequence affect the inferred phylogeny?

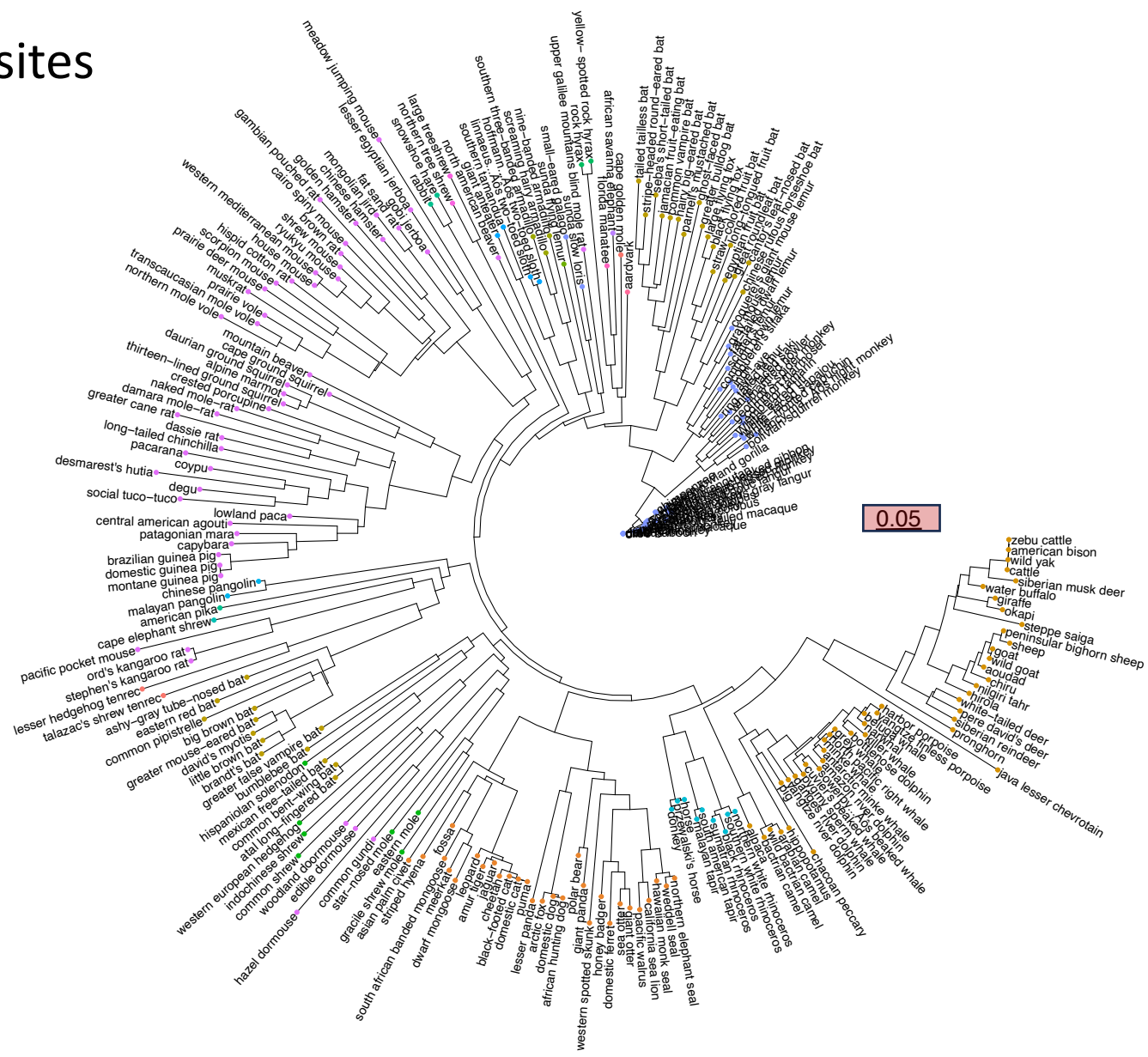
Constrained sites



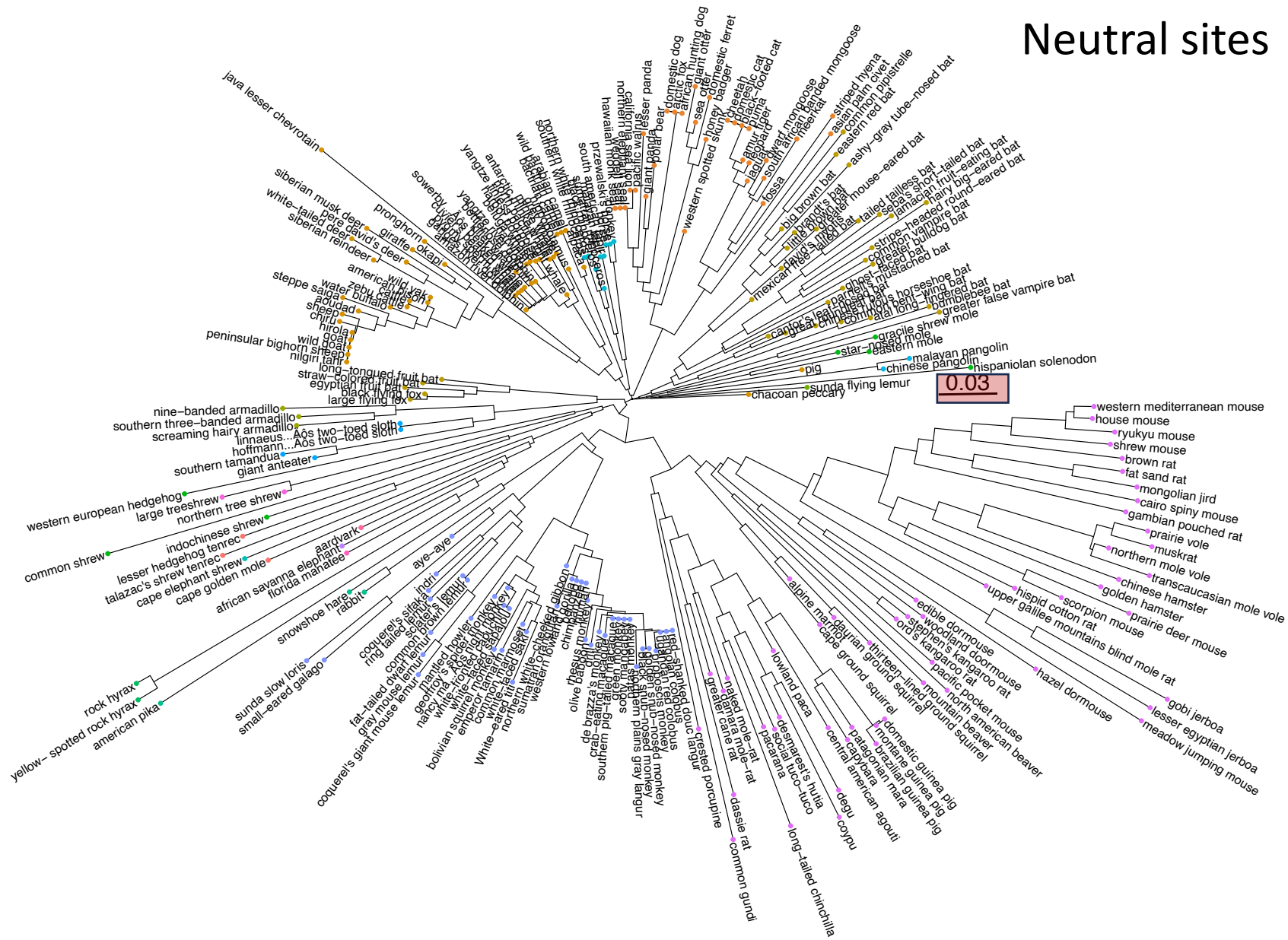
Order

- Afrosoricida
- Carnivora
- Cetartiodactyla
- Chiroptera
- Cingulata
- Dermoptera
- Eulipotyphla
- Hyracoidea
- Lagomorpha
- Macroscelidea
- Perissodactyla
- Pholidota
- Pilosa
- Primates
- Proboscidea
- Rodentia
- Scandentia
- Sirenia
- Tubulidentata

Accelerated sites

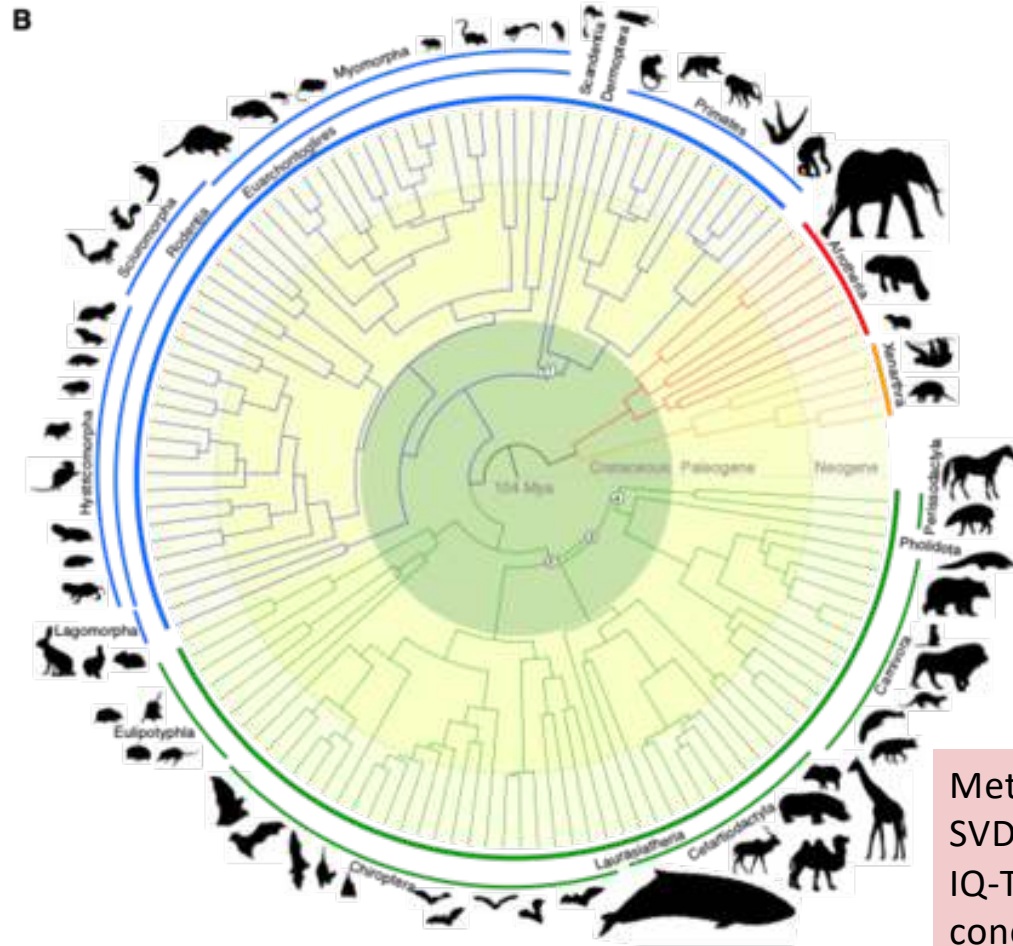
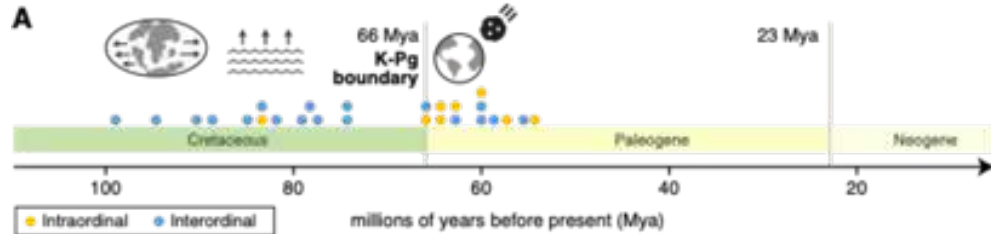


Neutral sites



Order

- Afrosoricida
- Carnivora
- Cetartiodactyla
- Chiroptera
- Cingulata
- Dermoptera
- Eulipotyphla
- Hyracoidea
- Lagomorpha
- Macroscelidea
- Perissodactyla
- Pholidota
- Pilosa
- Primates
- Proboscidea
- Rodentia
- Scandentia
- Sirenia
- Tubulidentata



New phylogeny resolves some long-standing debates in mammalian evolution

- Based on ~500k near-neutrally evolving sites
- Supports the 'long-fuse' model of mammalian diversification

Methods:
SVDquartets – coalescent trees
IQ-TREE – maximum likelihood concatenated trees

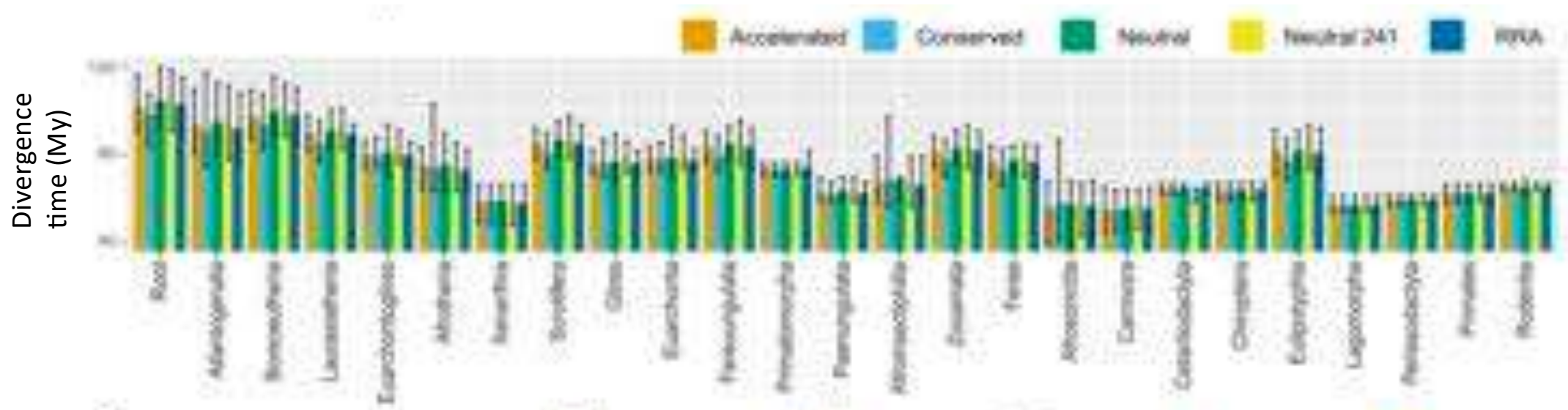
RESEARCH ARTICLE SUMMARY

ZOONOMIA

A genomic timescale for placental mammal evolution

Nicole M. Foley et al.

Effects of selection when estimating divergence time



Take-home messages:

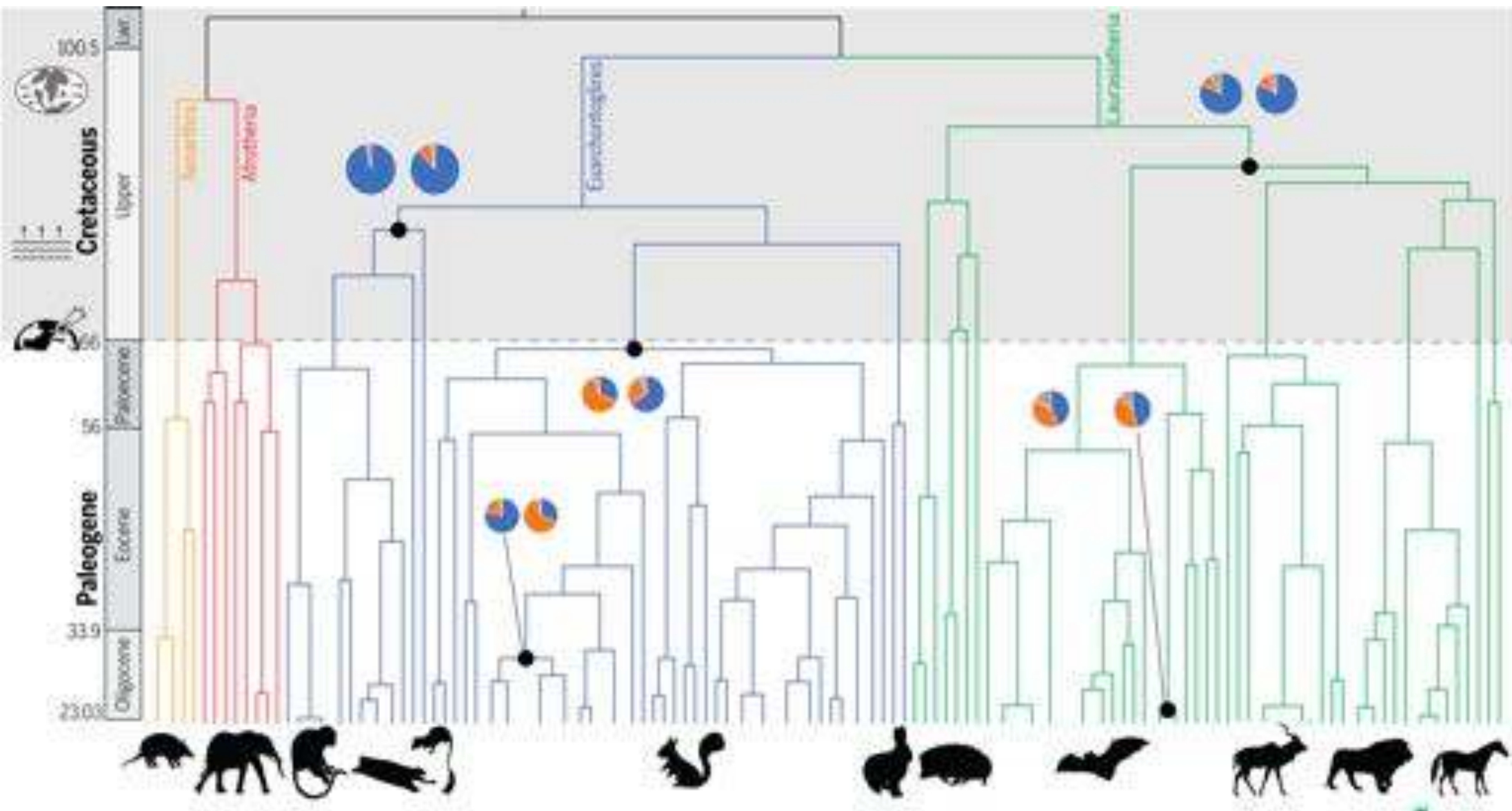
- Neutral is best – models assume this
- BUT with a massive amount of data (i.e. genome-wide), can still achieve accurate estimates
- Depends on the phylogenetic distance you are looking over

RESEARCH ARTICLE SUMMARY

ZOONOMIA

A genomic timescale for placental mammal evolution

Nicole M. Foley et al.



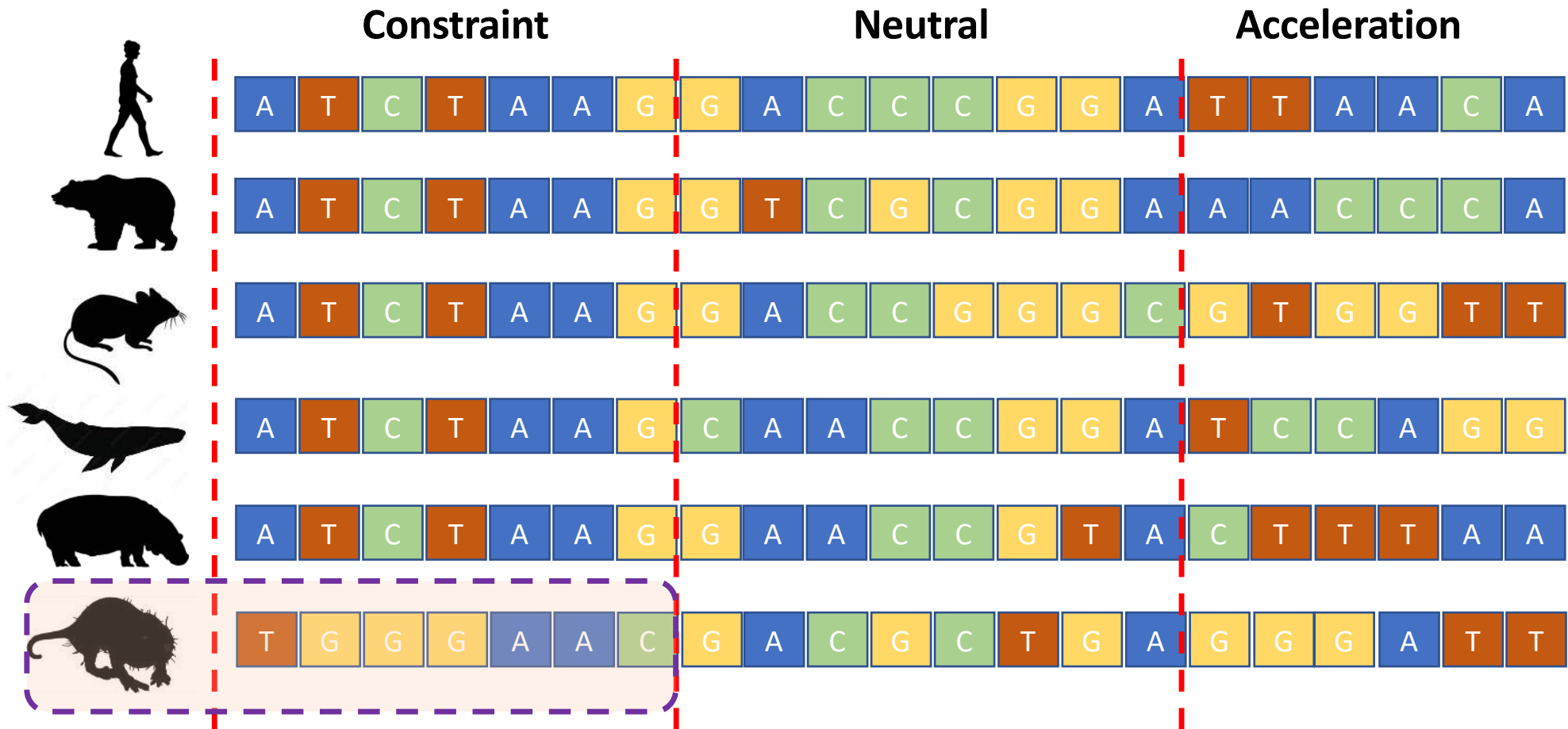


Linking genotype to phenotype

Identifying the genomic changes that underlie specific traits



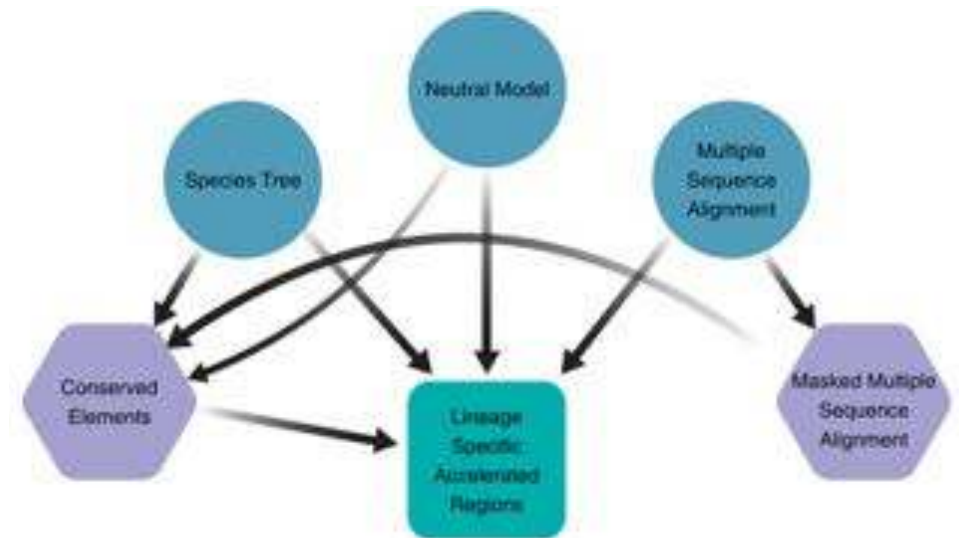
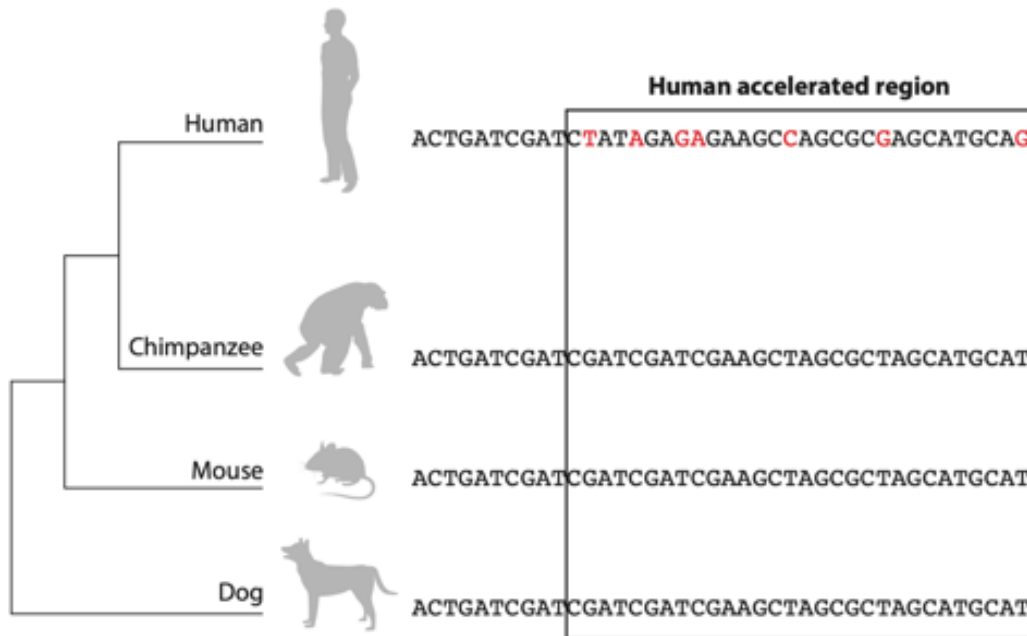
Lineage-specific acceleration



Identifying lineage-specific positive selection



UPPSALA
UNIVERSITET

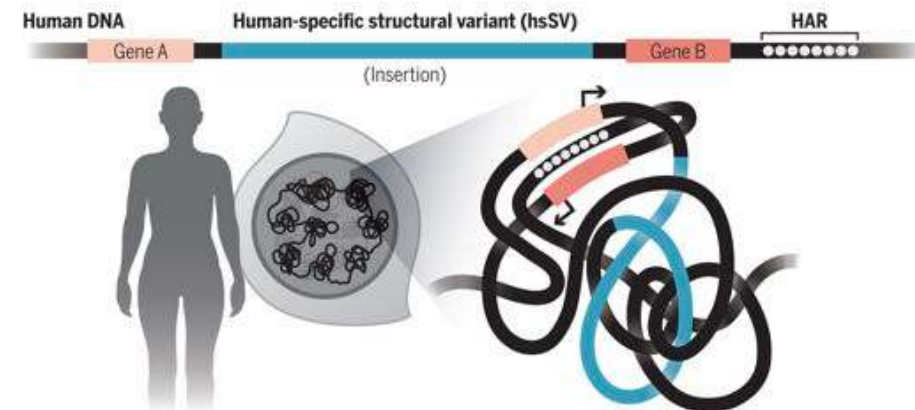
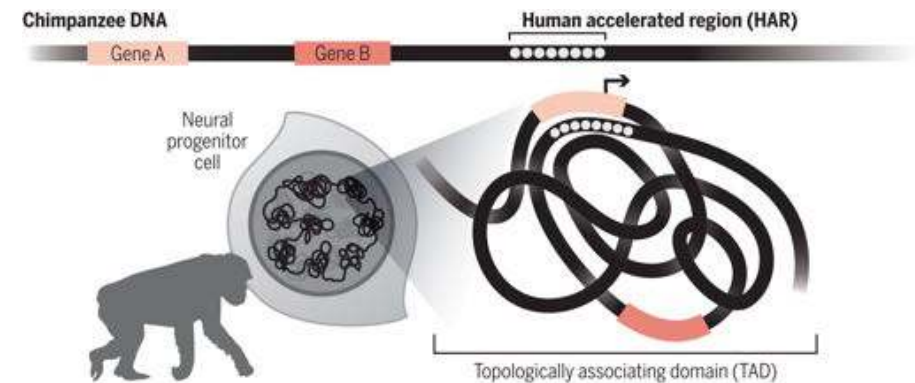


Whalen & Pollard, Annu. Rev. Genet. 2022. 56:423–39

Human accelerated regions

What/where might these be?

- 312 human accelerated regions (zooHARs) – many are neurodevelopmental enhancers with cell type-specific activity
- Overlap TADs containing human-specific structural variants
- Enrichment of hARs that affect 3D genome organisation
- ‘Rewiring’ of regulatory interactions between hARs and neurodevelopmental genes



RESEARCH ARTICLE

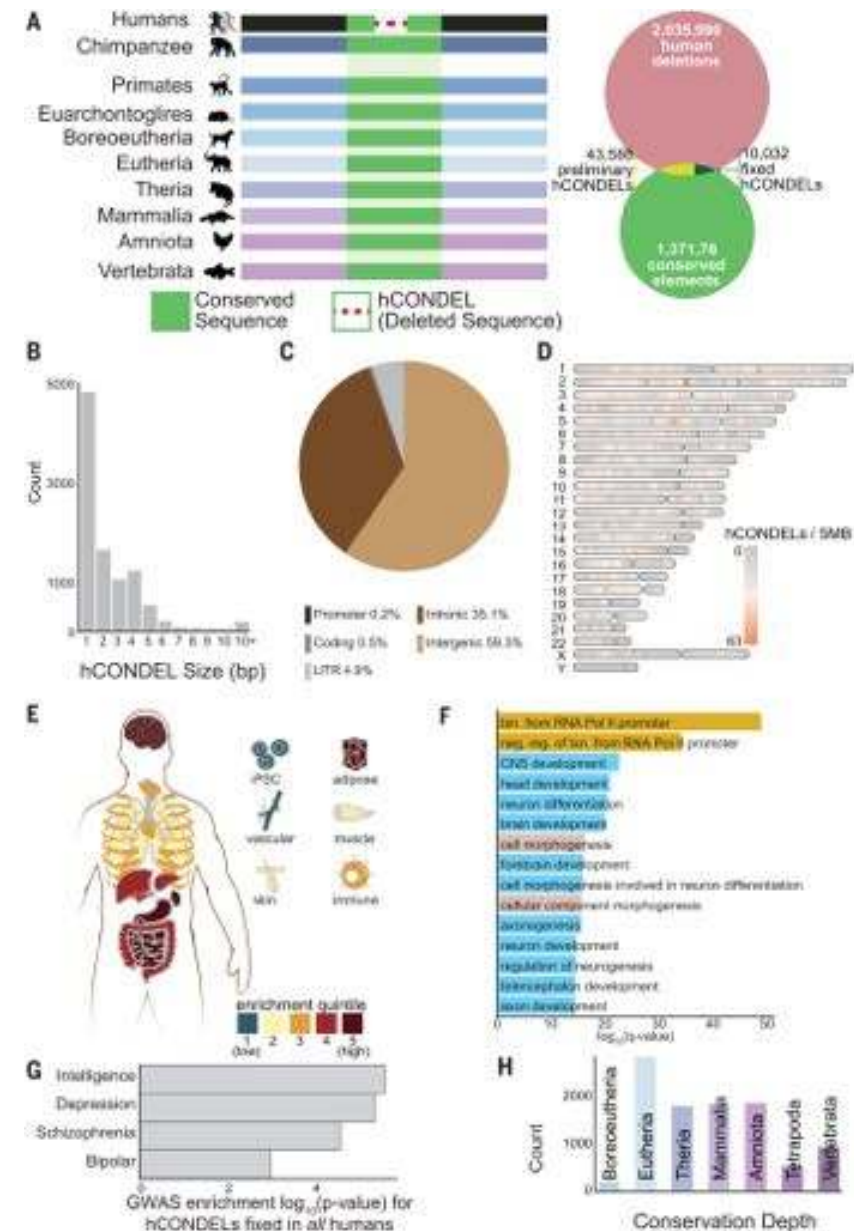
ZOONOMIA

Three-dimensional genome rewiring in loci with human accelerated regions

Kathleen C. Keough^{1,2,3}, Sean Whalen¹, Fumitaka Inoue^{2,3,†}, Pawel F. Przytycki^{1,†}, Tyler Fair^{4,5}, Chengyu Deng^{2,3}, Marilyn Steyert^{4,5,6,7}, Hane Ryu^{2,3}, Kerstin Lindblad-Toh^{8,9}, Elinor Karlsson^{9,10,11}, Zoonomia Consortium[§], Tomasz Nowakowski^{4,5,6,7}, Nadav Ahituv^{2,3}, Alex Pollen^{7,12}, Katherine S. Pollard^{1,3,13,14,*}

- Use the alignment to identify sites/sequences deleted in a single species
- Human-specific conserved deletions (hCONDELs)

- Use the alignment to identify sites/sequences deleted in a single species
- Human-specific conserved deletions (hCONDELs)



Detecting smells – a universal animal trait



Moles Smell in Stereo to Find Food, Dodge Predators



Research articles

Major histocompatibility complex-associated odour preferences and human mate choice: near and far horizons

Jan Havlíček¹, Jamie Winternitz² and S. Craig Roberts
Published: 20 April 2020 <https://doi.org/10.1098/rstb.2019.0260>

RESEARCH ARTICLE | BIOLOGICAL SCIENCES

Elephants have a nose for quantity

Joshua M. Plotnik¹, Daniel L. Brubaker, Rachel Dale², and Nicola S. Clayton¹ [Authors Info & Affiliations](#)

Edited by Raghavendra Gadagkar, Indian Institute of Science, Bangalore, India, and approved April 11, 2019 (received for review November 16, 2018)

June 3, 2019 | 116 (25) 12566–12571 | <https://doi.org/10.1073/pnas.1816284116>

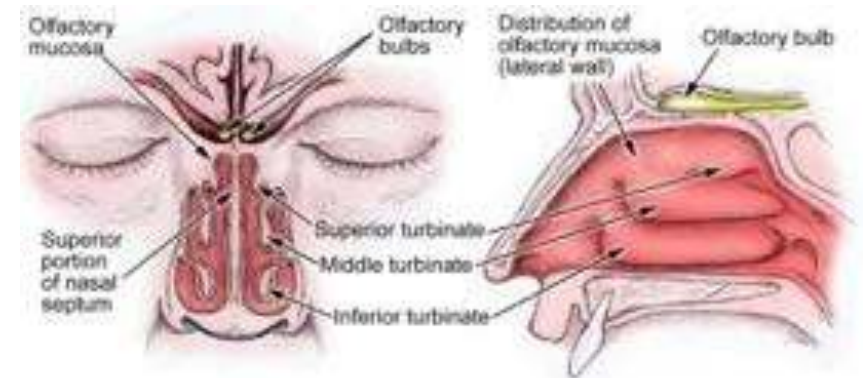
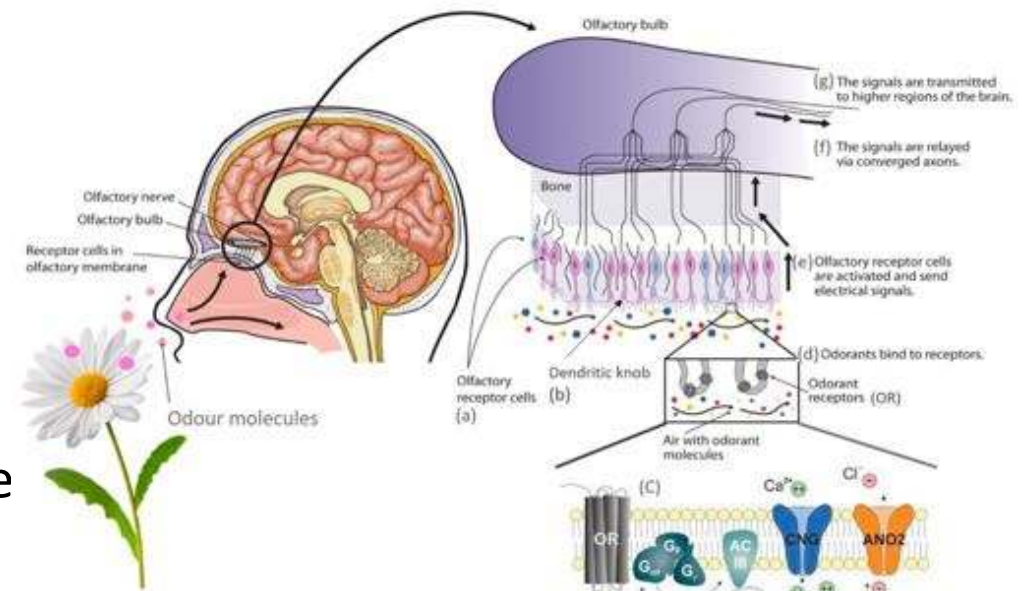


Published 2022



Olfaction in mammals

- **Olfactory receptor (OR) genes:** G-protein-coupled receptors, contain several transmembrane α -helical domains. Detect odour molecules in the environment
- Largest gene superfamily in vertebrate genomes
- Olfactory turbinates - extensive framework of delicate bones in nasal cavity that greatly enlarge the surface area available for conditioning inspired air, reducing water loss, and improving olfaction

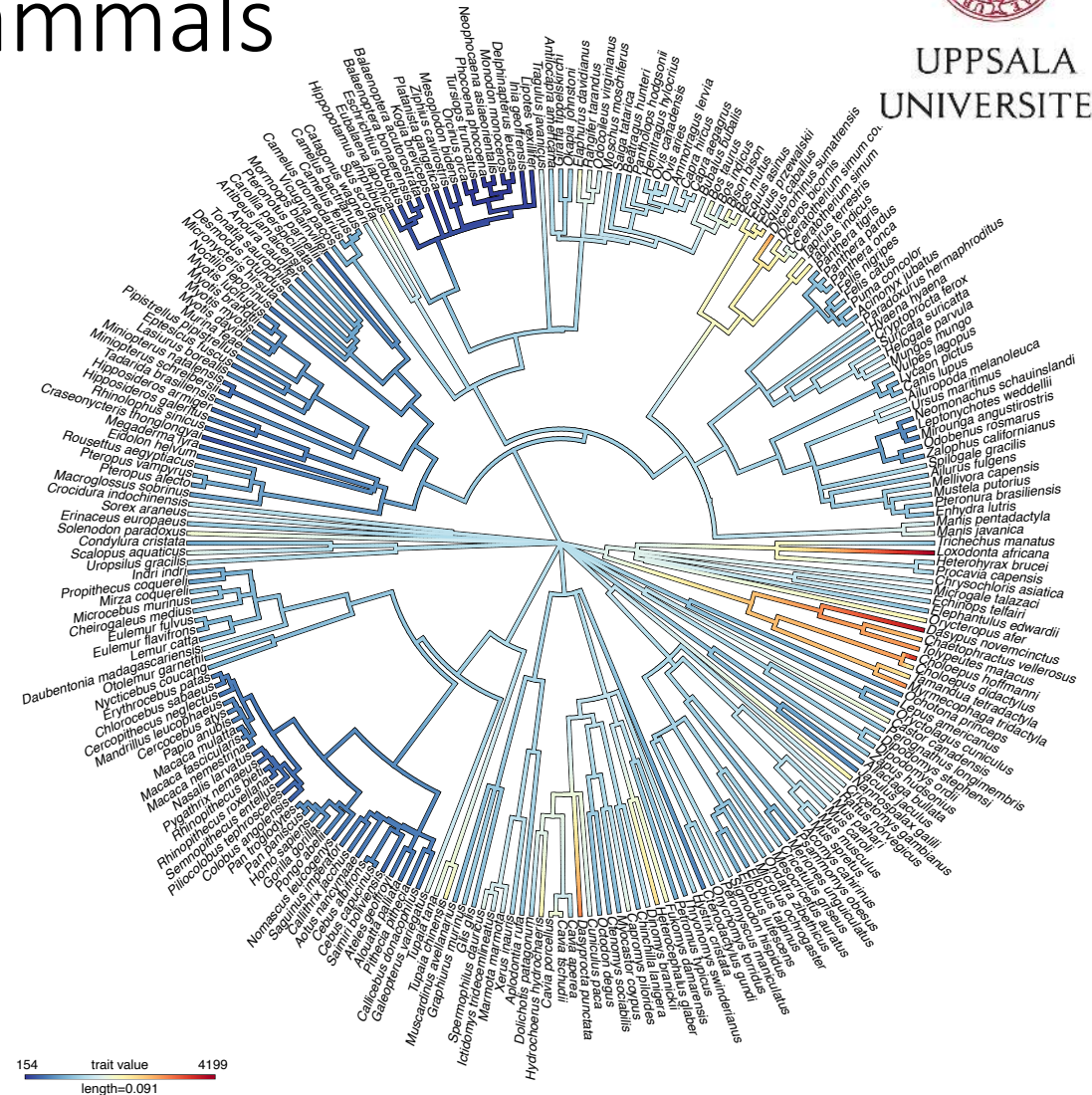




UPPSALA
UNIVERSITET

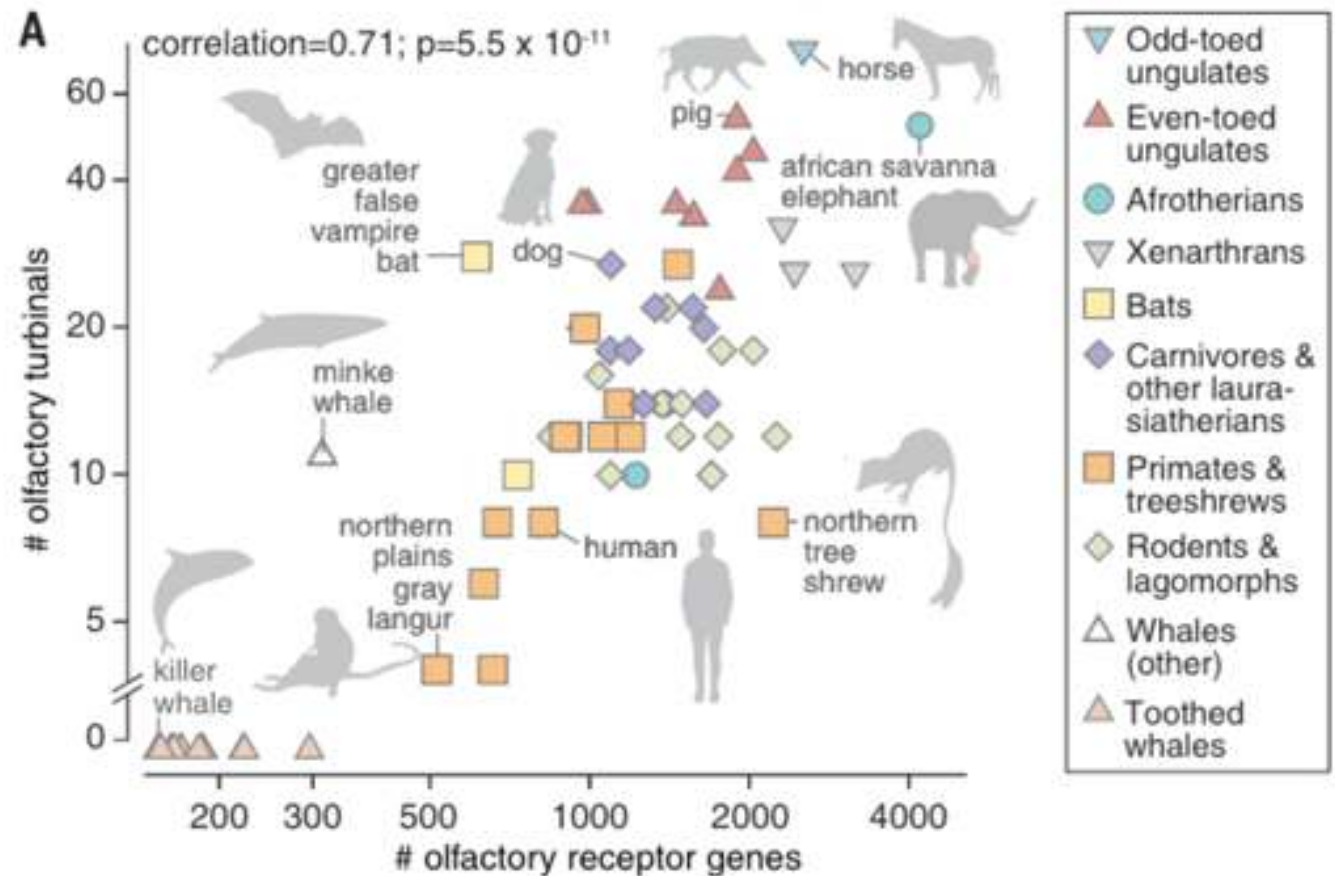
Counting OR genes in mammals

- Olfactory receptor gene (OR) detection in 249 species
- African elephant has largest repertoire (1,765 functional genes) 
- Killer whale smallest (24 functional genes) 
- Bumblebee bat and North Pacific right whale – only 332 and 392 respectively



Correlating genotype and phenotype

- OR counts strongly correlated with olfactory turbinal counts
- Highlights potential for investigating multifactorial evolutionary responses to a selection pressure



UPPSALA
UNIVERSITET

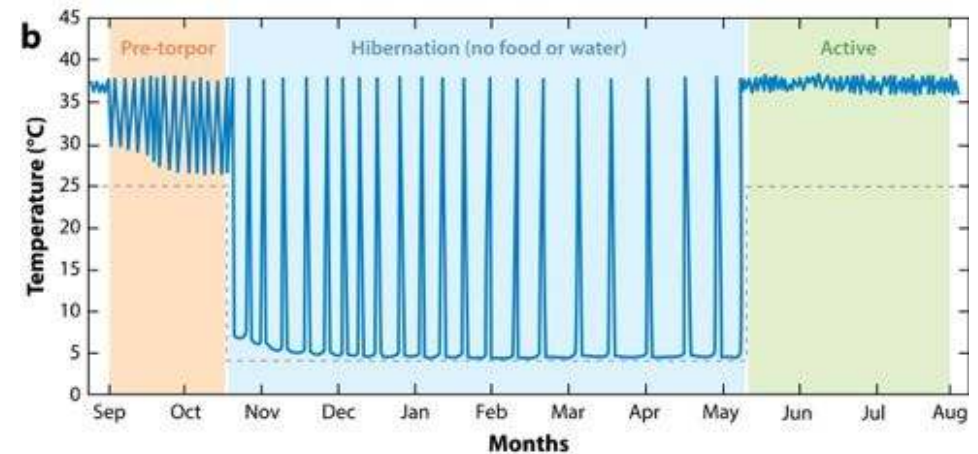
Hibernation in mammals

- Physiological state of metabolic depression (torpor). species capable of core temperature depression below 18°C for >24h
- Ancestral trait in mammals?
Found in all deep lineages



	Sleep	Daily heterothermy	Hibernation	
			Shallow torpor	Deep torpor
Minimum body temperature (°C)	18–37°C	5–30°C	10–30°C	~3–10°C
Respiration rate (% of active)	100–300%	5–20%	84%	2–3%
Heart rate (% of active)	70–90%	10–30%	10–50%	1–4%
Metabolic rate (% of BMR)	70–90%	5–70%	10–20%	1–10%
Blood pressure (% of active)	50%	60–70%	20–60%	
Daily activity level	High	High	Low	None
Average energy savings	5–15%	60–70%	Up to 90%	
Duration	<24 h	<24 h	>24 h	>24 h
Occurrence	Year-round	Year-round	Season inducible	Seasonal

*Shading refers to magnitude of parameter (i.e., more intense color represents higher body temperature or longer duration of time).
Abbreviation: BMR, basal metabolic rate.

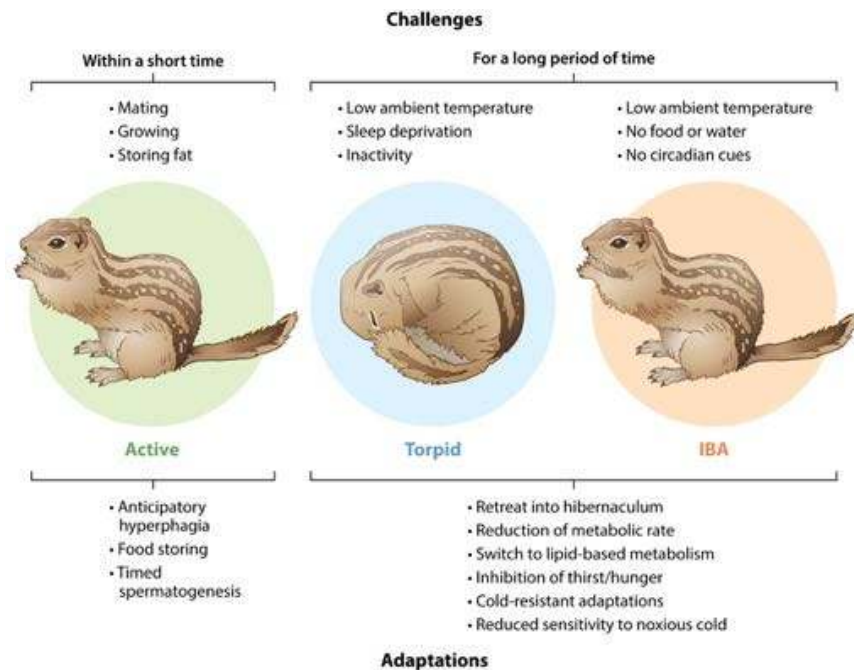


Hibernation in mammals

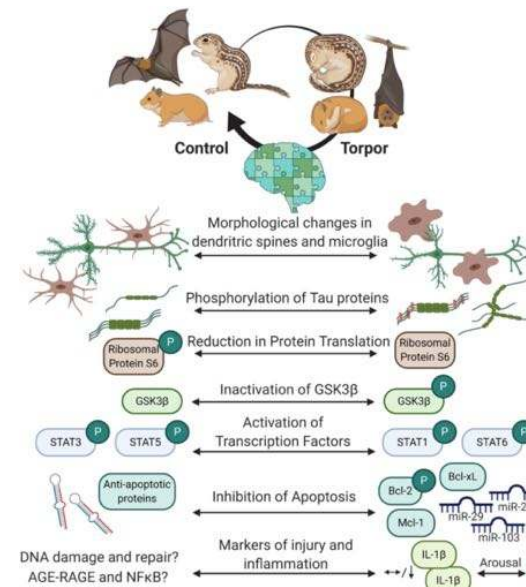
- Effects of the cold on the function of cellular and molecular processes requires specific adaptations to survive at the limitations of basic biology and physiology, e.g. transcription and translation can be halted



Arctic ground squirrel (*Spermophilus parryii*) can drop its core body temperature below 0°C – the lowest of any mammal



Mohr SM, et al. 2020. *Annu. Rev. Cell Dev. Biol.* 36:315–38



Protective mechanisms in the hibernating brain

REVIEW article
Front. Physiol., 20 January 2021
Sec. Integrative Physiology
Volume 11 – 2020 | <https://doi.org/10.3389/fphys.2020.623665>

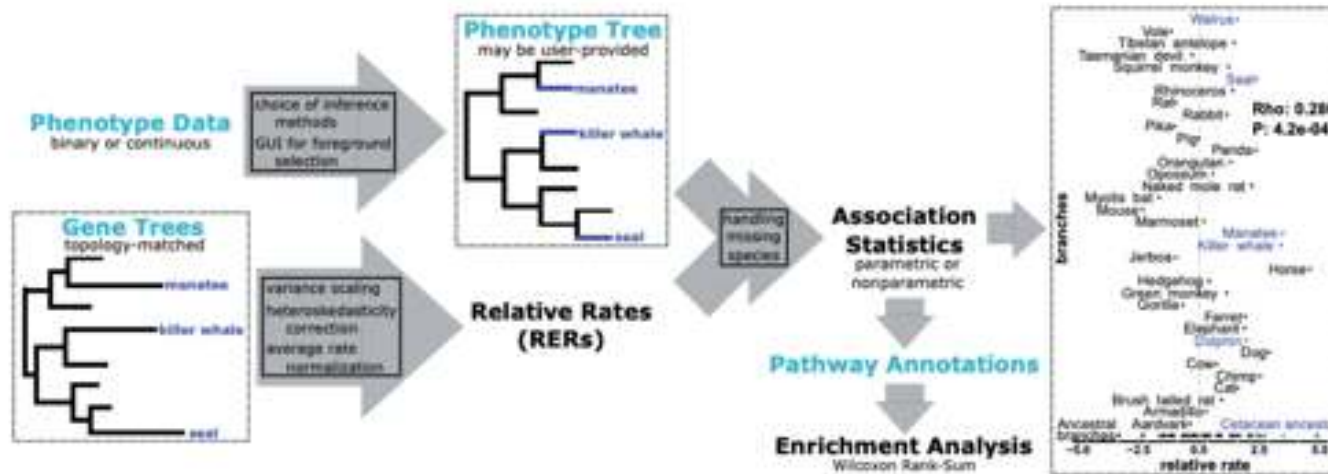
Testing for differential evolutionary rates associated with hibernation



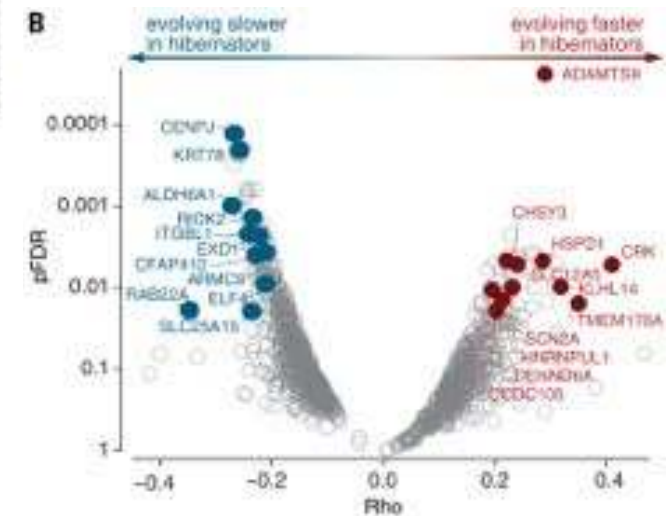
Phylogenetics

RERconverge: an R package for associating evolutionary rates with convergent traits

Amanda Kowalczyk^{1,2,†}, Wynn K. Meyer^{1,†}, Raghavendran Partha^{1,2,†}, Weiguang Mao^{1,2}, Nathan L. Clark^{1,2} and Maria Chikina^{1,2,*}

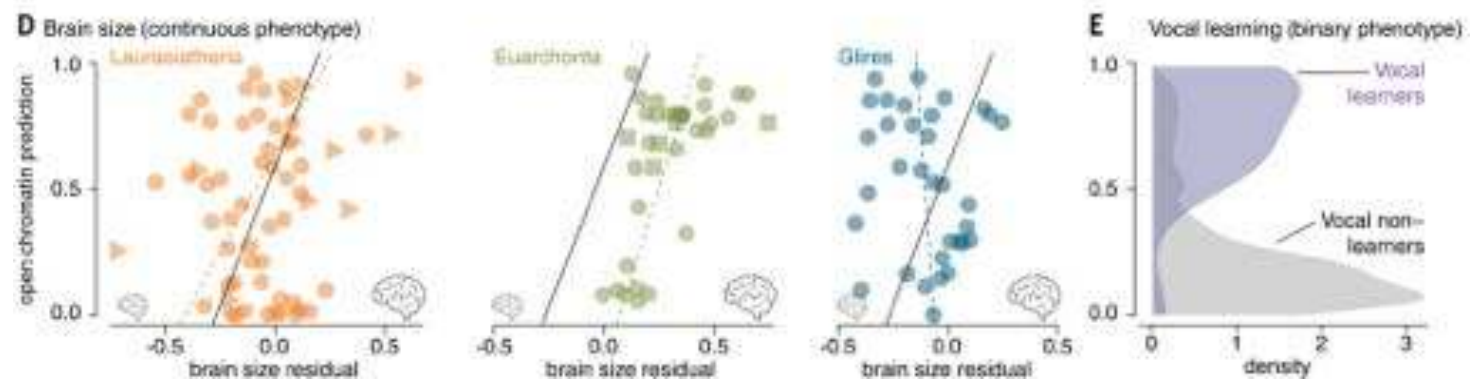
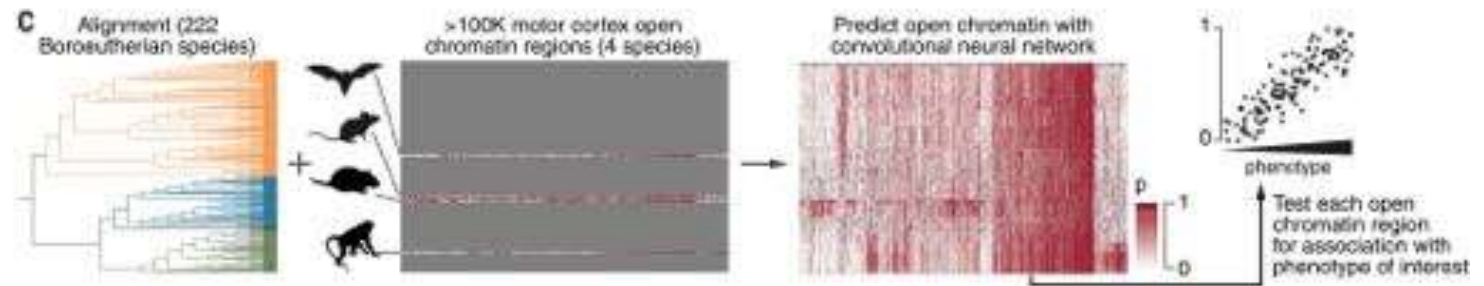
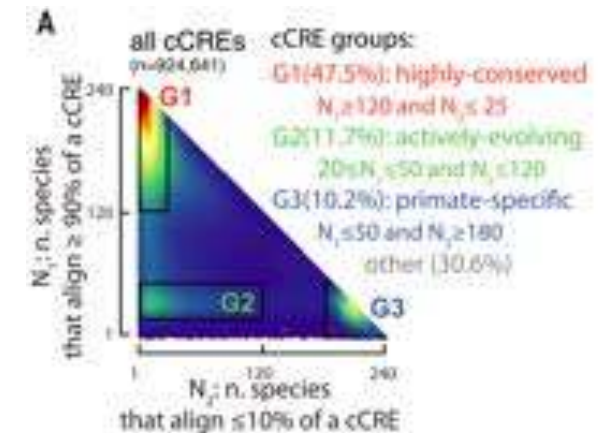
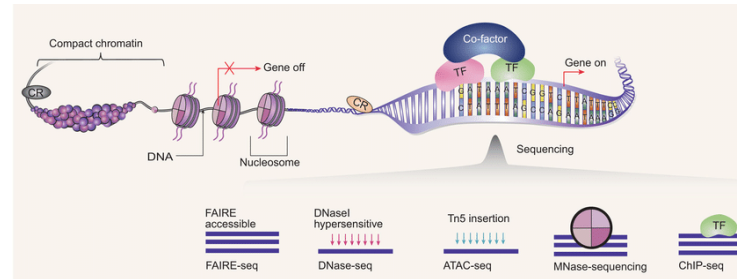


- RERconverge identified 511 genes slower-evolving and 253 faster-evolving in hibernators
- Faster evolving genes enriched in GO pathways relating to synaptic transmission
- Slower evolving genes enriched in GO pathways relating to DNA repair – under *higher constraint* in hibernators



What about regulatory sequences?

- Tissue-Aware Conservation Inference Toolkit (TACIT) for associating *cis*-regulatory elements with specific phenotypes
- Associates open chromatin predictions with phenotype annotations

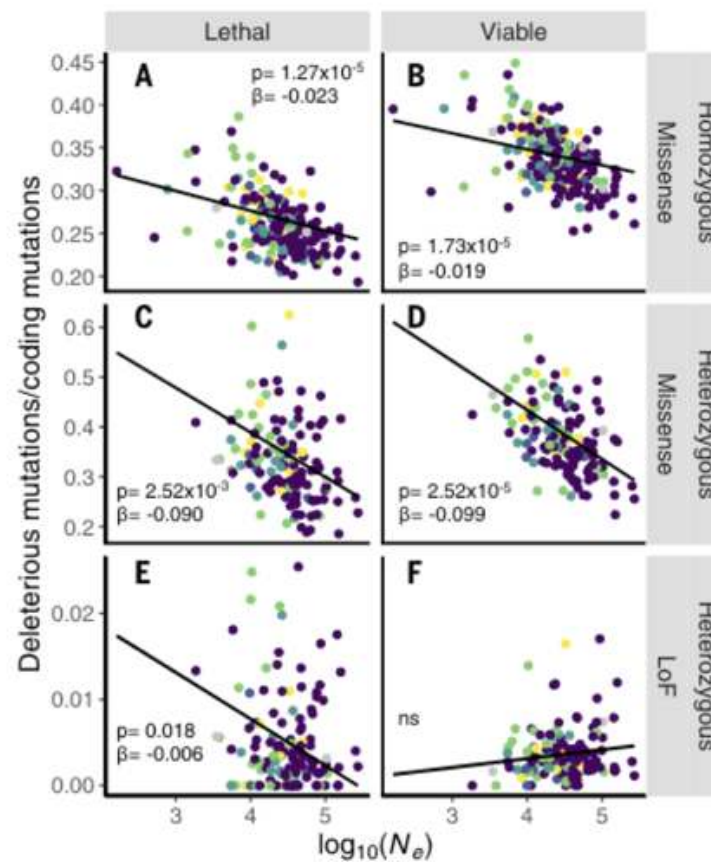
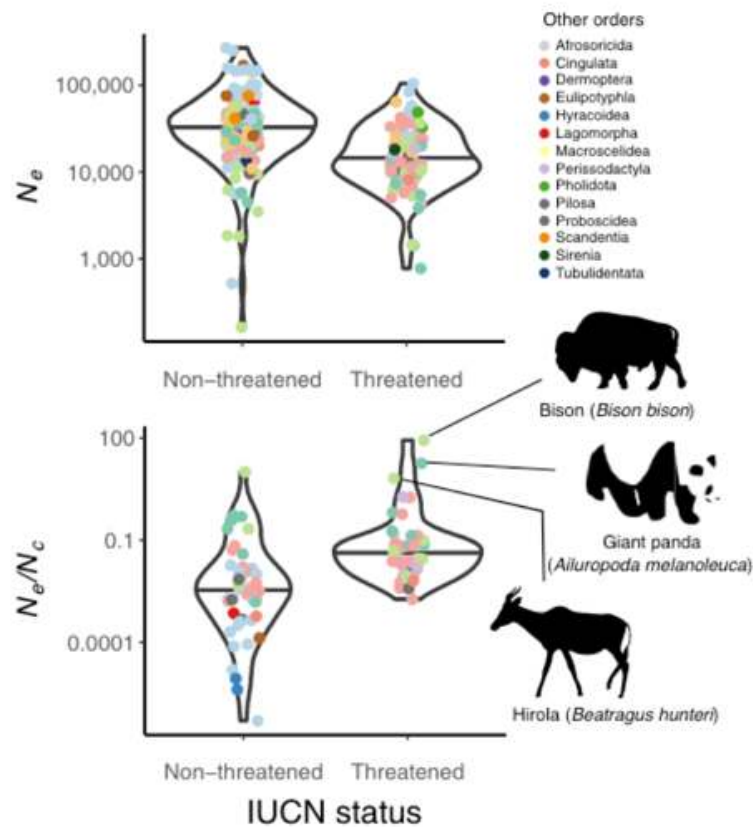


ZOONOMIA

Relating enhancer genetic variation across mammals to complex phenotypes using machine learning

Irene M. Kaplow^{1,2}, Alyssa J. Lawler^{1,2}, Daniel E. Schäffer^{1,2}, Chaitanya Srinivasan, Heather H. Sestili, Morgan E. Wirthlin, BaDoi N. Phan, Kavya Prasad, Ashley R. Brown, Xiaomeng Zhang, Kathleen Foley, Diane P. Genereux, Zoonomia Consortium, Elinor K. Karlsson, Kerstin Lindblad-Toh, Wynn K. Meyer, Andreas R. Pfenning^{1,2}

Using genomes to inform conservation



RESEARCH ARTICLE SUMMARY

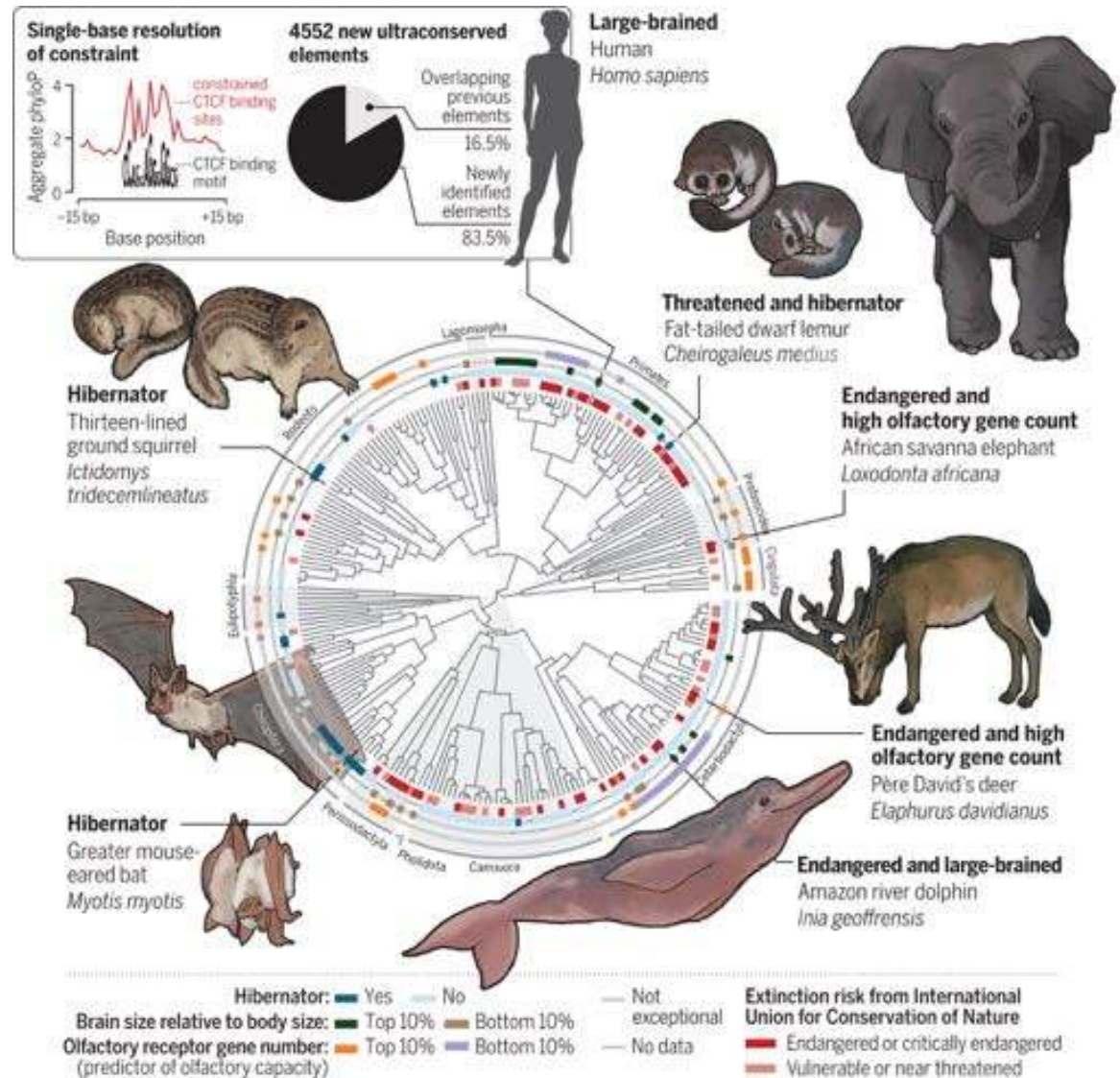
ZOONOMIA

The contribution of historical processes to contemporary extinction risk in placental mammals

Aryn P. Wilder^{*,†}, Megan A. Supple^{*,†}, Ayshwarya Subramanian, Anish Mudide, Ross Swofford, Aitor Serres-Armero, Cynthia Steiner, Klaus-Peter Koepfli, Diane P. Genereux, Elinor K. Karlsson, Kerstin Lindblad-Toh, Tomas Marques-Bonet, Violeta Munoz Fuentes, Kathleen Foley, Wynn K. Meyer, Zoonomia Consortium, Oliver A. Ryder^{*,†}, Beth Shapiro^{*,†}

Conclusions

- Massive, multi-way whole genome alignments are now possible!
- Informative for identifying effects of selection on the genome
- Single-base measures of constraint and acceleration pinpoint functional sequences
- Identification of neutrally evolving sequences important for inferring accurate phylogenetic relationships and divergence times
- Comparative genomics can reveal the genetic underpinnings of traits
- Measures of genomic variation informative for extinction risk and conservation



Uppsala University*Kerstin Lindblad-Toh*

Michael Dong

Matteo Bianchi

UPPSALA
UNIVERSITET

Jennifer Meadows

Ola Wallerman

Sergey Kozyrev

Voichita Marinescu

Broad Institute*Diane Genereux*

Xue Li

Lucas Moreira

Ross Swofford

Jeremy Johnson

Bruce Birren

James Xue

Elinor K. Karlsson**Umass Chan**

Gregory Andrews

Kaili Fan

Zhiping Weng

University of Nevada

Allyson Hindle

Cornelia Fanter

Carnegie Mellon University*Irene Kaplow*

Amanda Kowalczyk

Alyssa Lawler

BaDoi Phan

Daniel Schäffer

Morgan Wirthlin

Andreas Pfenning

University College Dublin

Graham Hughes

Colleen Lawless

Louise Ryan

Emma Teeling

University of Maine

Ana Breit

Danielle Levesque

UC Davis

Joana Damas

Harris Lewin

Gladstone Institute, UCal

Kathleen Keough

Abigail Lind

Katherine Pollard

Max Planck Institute

Bogdan Kirilenko

Michael Hiller

San Diego Zoo

Cynthia Steiner

Aryn Wilder

Oliver Ryder

UC Santa Cruz

Joel Armstrong

Mark Diekhans

Megan Supple

Beth Shapiro

Benedict Paten

Texas A&M

Nicole M. Foley

Jenna Grimshaw

Austin Osmanski

Nicole Paulat

Bill Murphy

Institute for Systems Biology, Seattle

Robert Hubley

Arian Smit

Universitat Pompeu Fabra, Barcelona

Alejandro Valenzuela

Tomas Marques-Bonet

Arcadi Navarro

Patrick Sullivan: KI, UNC

Linda Goodman: Fauna Bio Inc.

Franziska Wagner: Museum of zoology, Dresden

Steven Reilly: Yale school of medicine

Steven Gazal: Uni. Southern California

Klaus-Pieter Koepfli: Smithsonian

Wynn Meyer: Lehigh Uni.

Martin Nweeia: Narwhal genome initiative

Mark Springer: Uni. California

David Ray: Texas Tech Uni.

Irina Ruf: Senckenberg Research Institute

UPPSALA
UNIVERSITET

Vetenskapsrådet

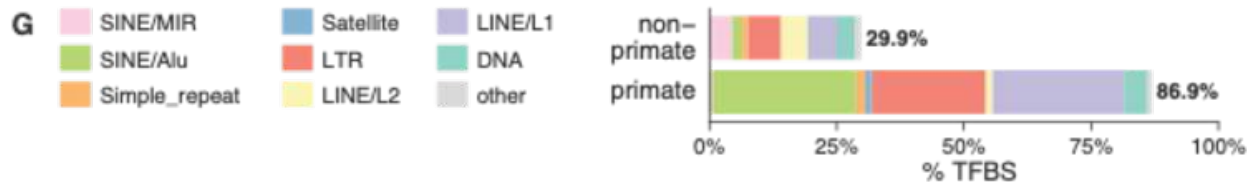


SciLifeLab

Knut and Alice
Wallenberg
Foundation

Constraint in human TEs

- ~11% of constrained bases are within repeats in the human genome
- Enriched in DNA transposons and simple repeats
- Repeat classes depleted in constraint have been more recently active
- Most primate-specific TFBS overlap TEs



Greg Andrews, Zhiping Weng, UMass



UPPSALA
UNIVERSITET

