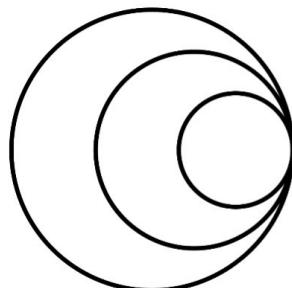


# The genetic basis of “Cancer-resistant” species

Daniel Chavez Ph.D

Pontificia Universidad Católica del Ecuador

[dechavezv@puce.edu.ec](mailto:dechavezv@puce.edu.ec)



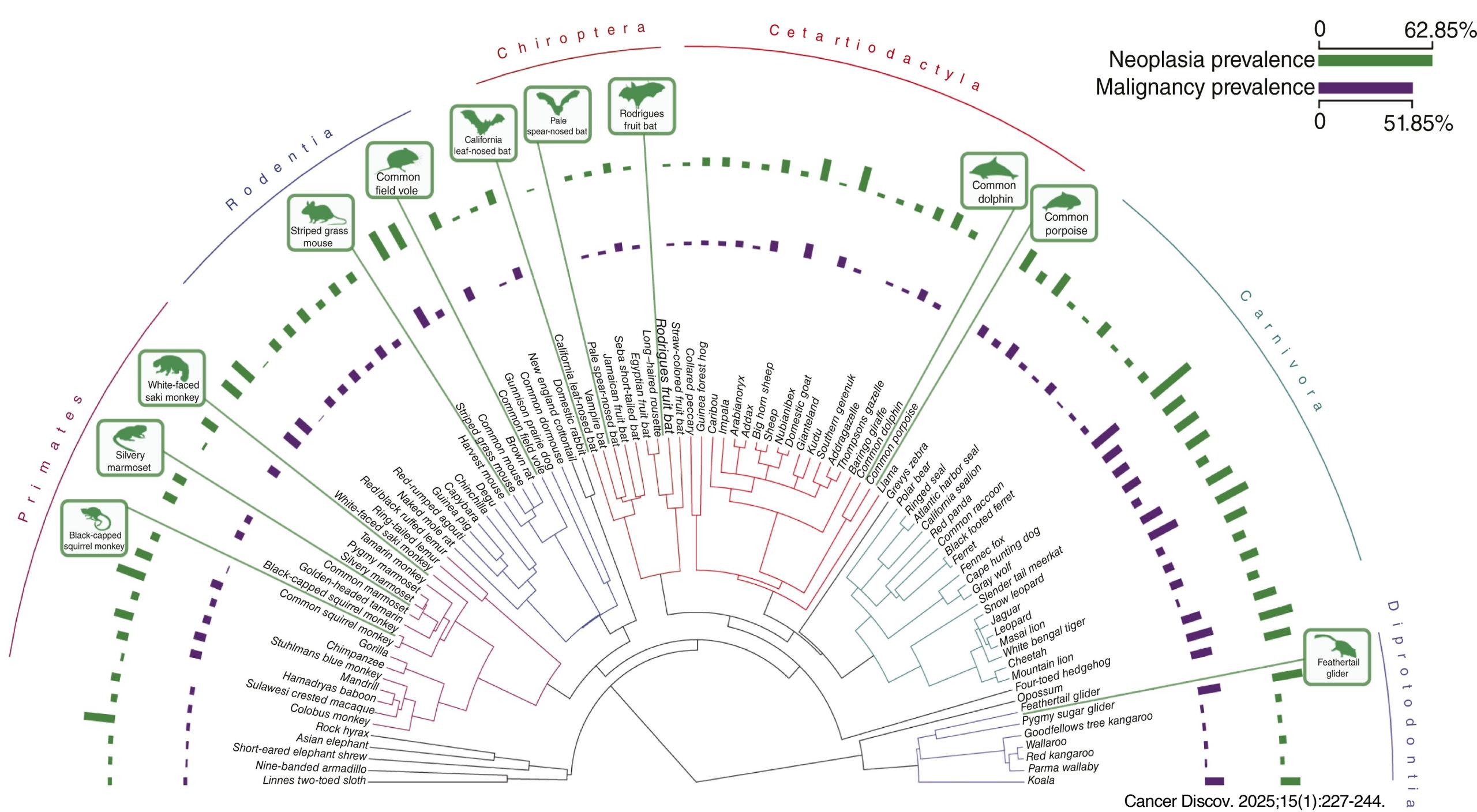
**welcome  
connecting  
science**

# Outline

- Existing time for comparative genomics and Cancer research
  - Necropsy Data
  - Whole genome alignments
- Major projects
  - TP53 evolution in elephants
  - Consequence of genetic loss in cancer prevalence
  - Positive selection on “cancer resistant species”
- Bioinformatic workflow tutorial

# Outline

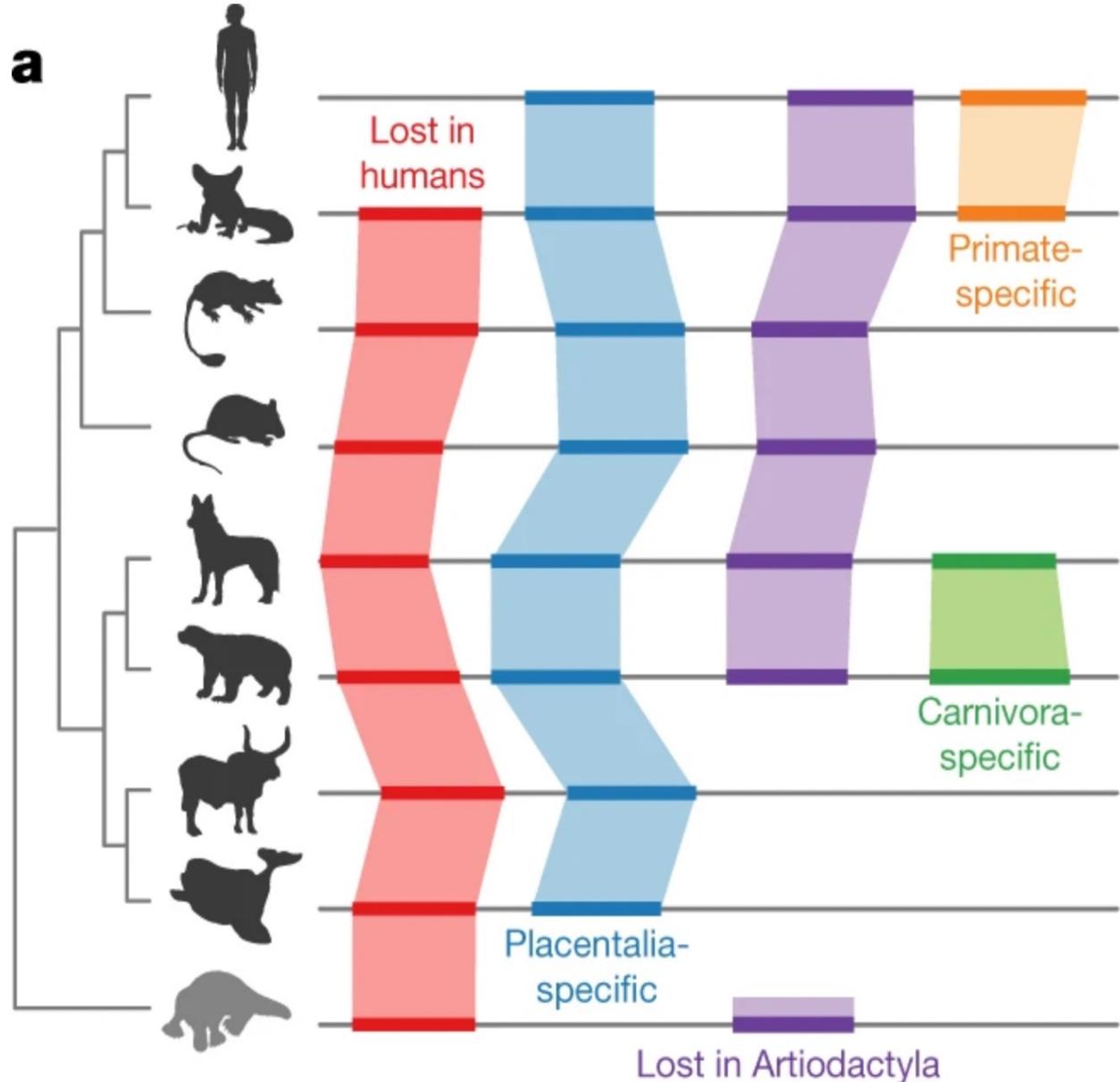
- Existing time for comparative genomics and Cancer research
  - **Necropsy Data**
  - Whole genome alignments
- Major projects
  - TP53 evolution in elephants
  - Consequence of genetic loss in cancer prevalence
  - Positive selection on “cancer resistant species”
- Bioinformatic workflow tutorial



# Outline

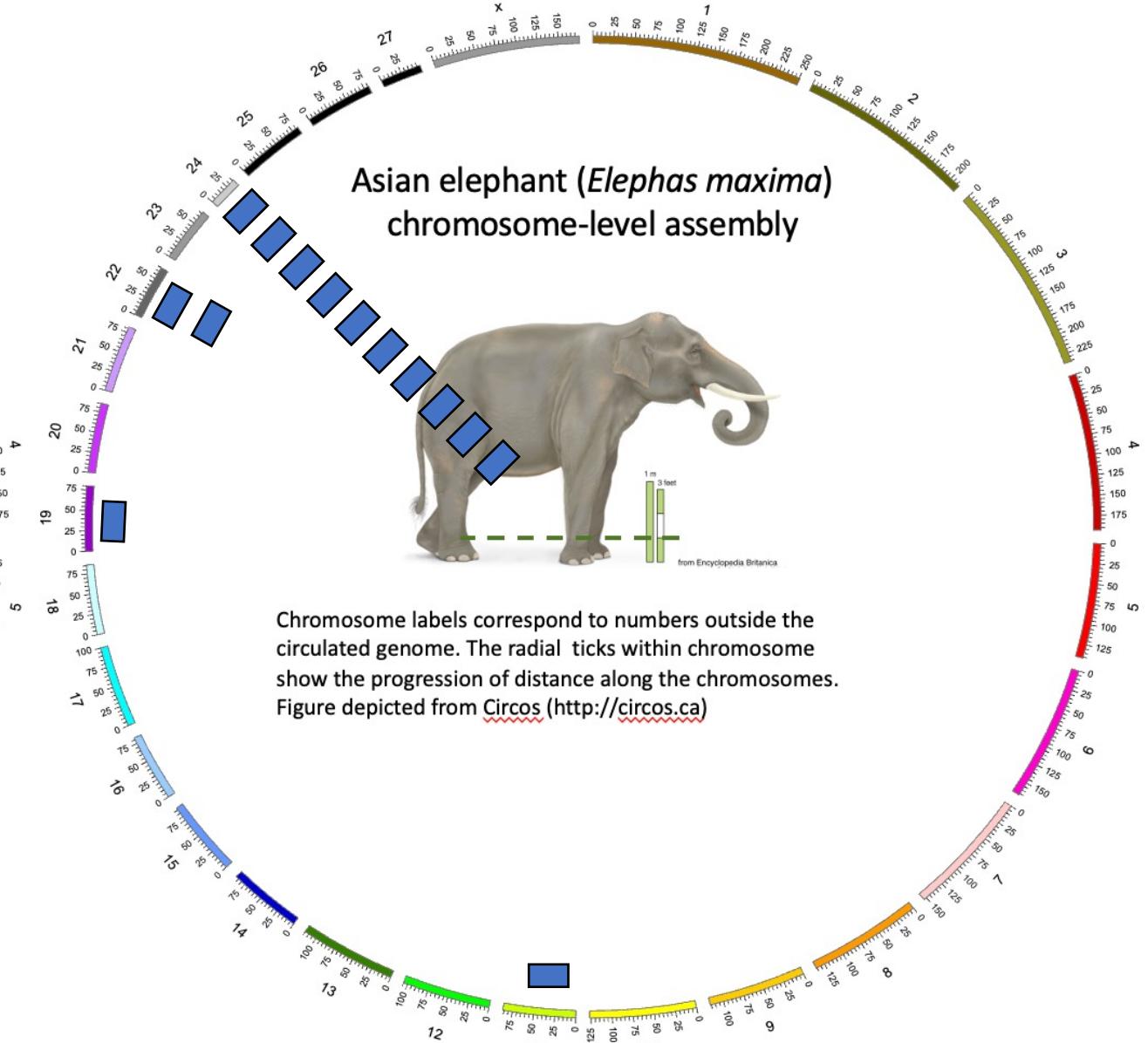
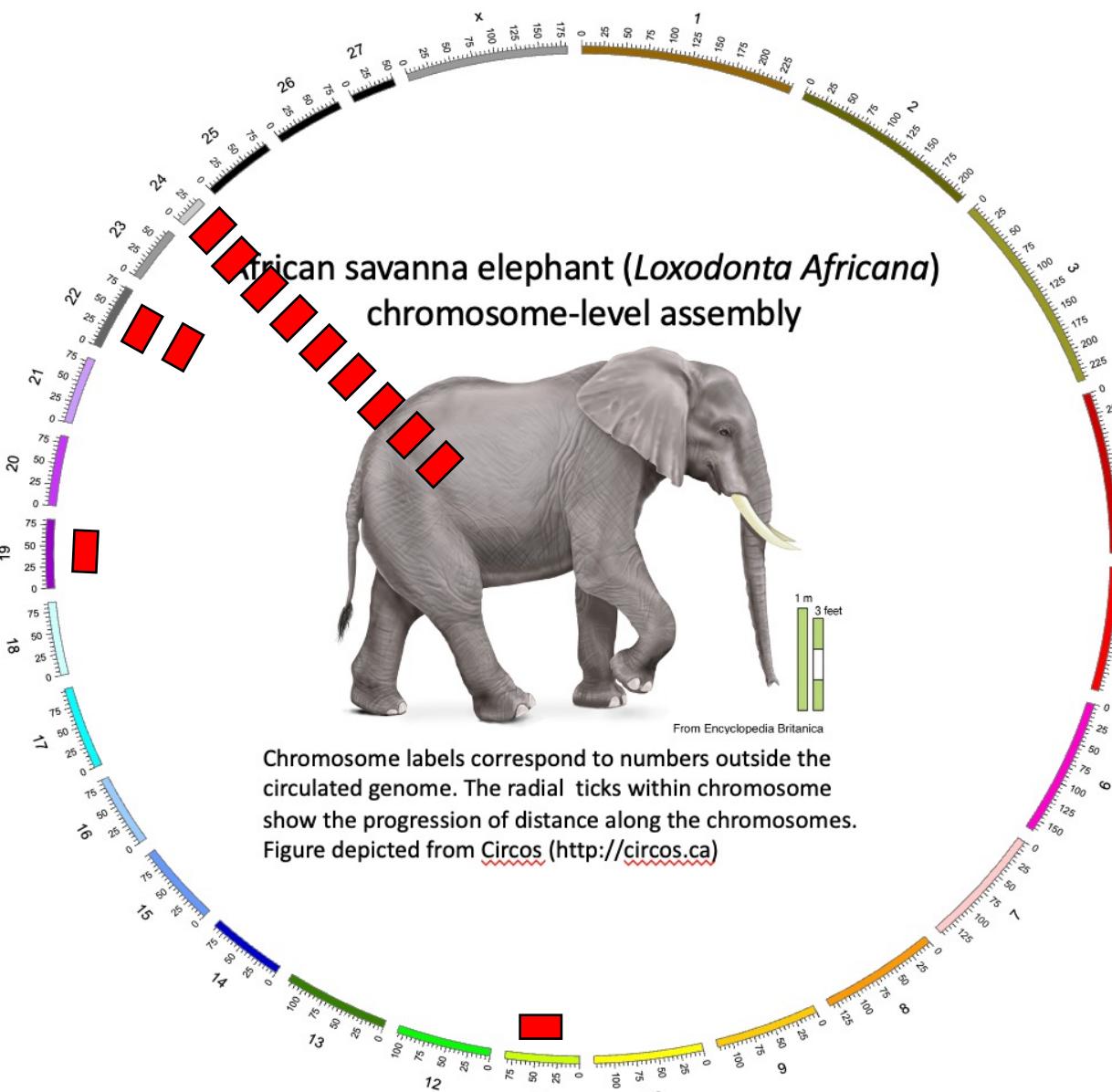
- Existing time for comparative genomics and Cancer research
  - Necropsy Data
  - **Whole genome alignments**
- Major projects
  - TP53 evolution in elephants
  - Consequence of genetic loss in cancer prevalence
  - Positive selection on “cancer resistant species”
- Bioinformatic workflow tutorial

- Cactus alignment of 240 mammals
- Rather than aligning to a single anchor genome, Cactus infers an ancestral genome for each pair of assemblies.



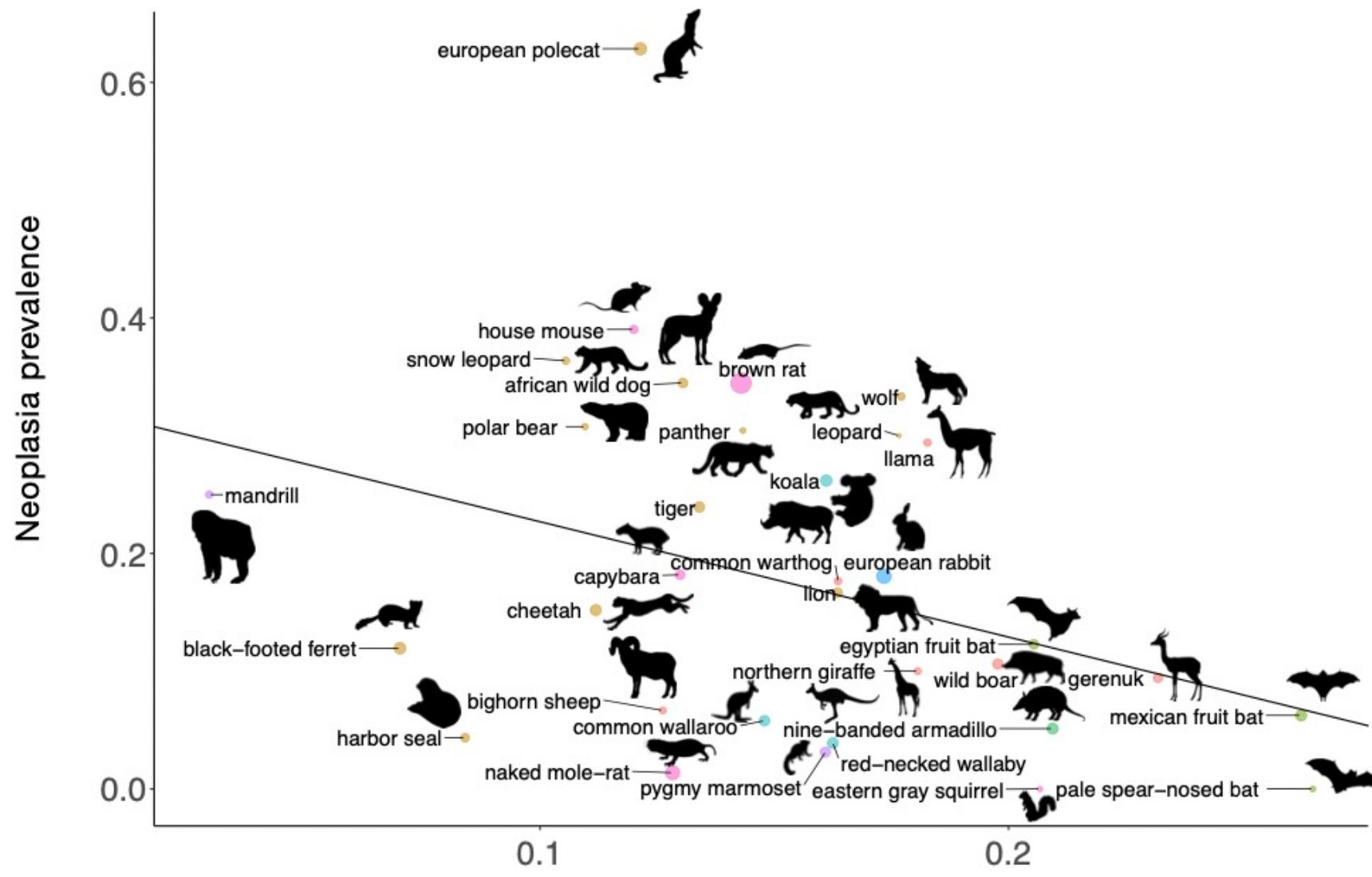
# Outline

- Existing time for comparative genomics and Cancer research
  - Necropsy Data
  - Whole genome alignments
- Major projects
  - **TP53 evolution in elephants**
  - Consequence of genetic loss in cancer prevalence
  - Positive selection on “cancer resistant species”
- Bioinformatic workflow tutorial



# Outline

- Existing time for comparative genomics and Cancer research
  - Necropsy Data
  - Whole genome alignments
- Major projects
  - TP53 evolution in elephants
  - **Consequence of genetic loss in cancer prevalence**
  - Positive selection on “cancer resistant species”
- Bioinformatic workflow tutorial



PGLS  
 $R^2 = 0.91$   
 P-value = 0.03  
 F-statistic = 4.64 on 1 and 31 DF  
 lambda = 0.01  
 33 species

#### Necropsies

- 100
- 200
- 300

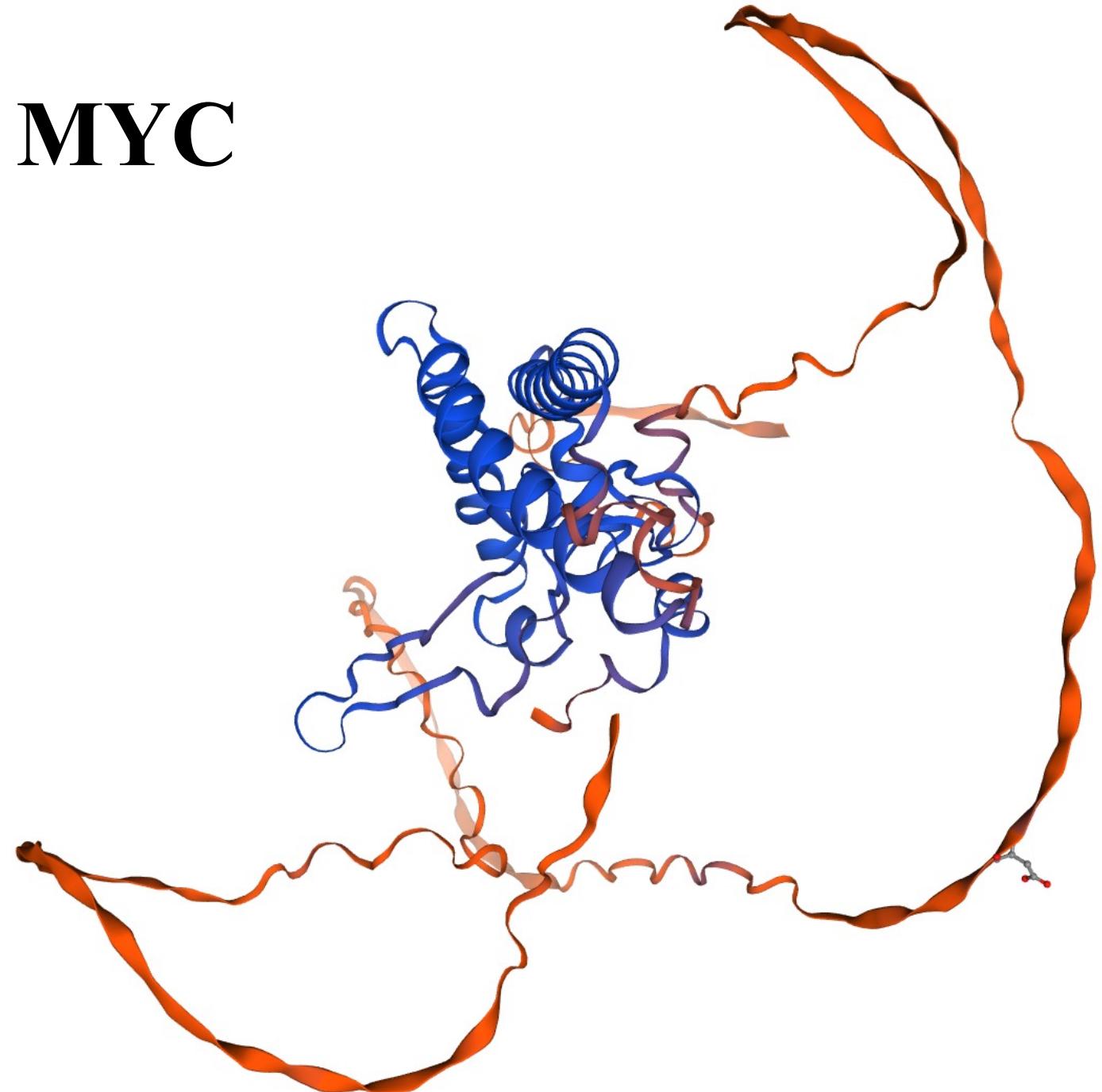
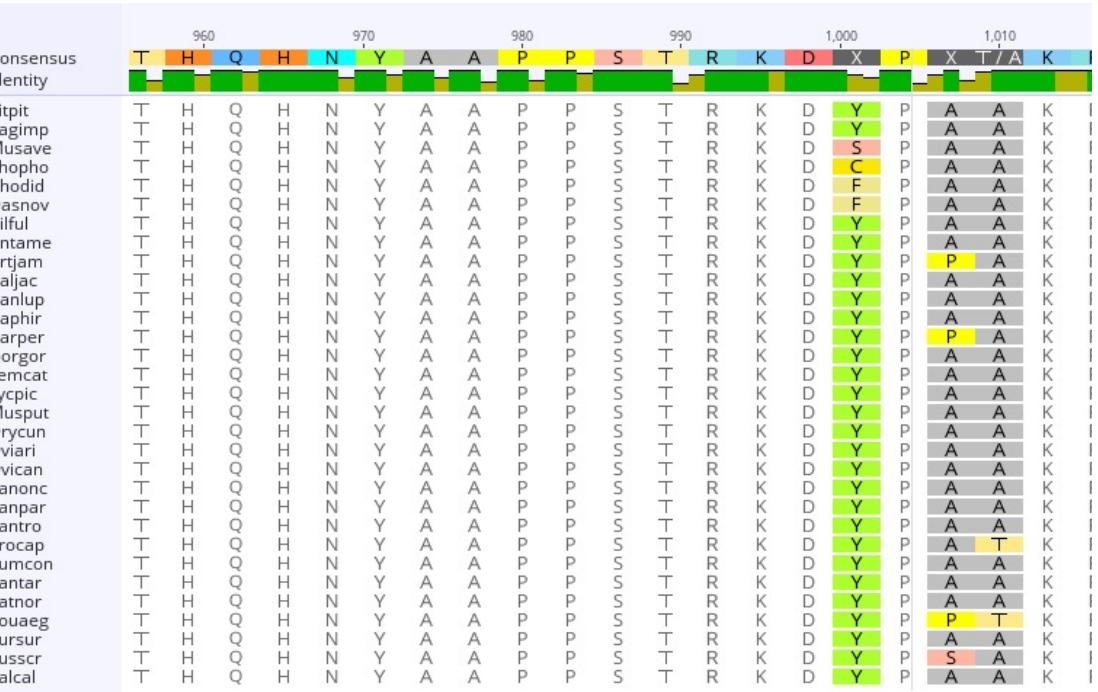
#### Orders

- Artiodactyla
- Carnivora
- Chiroptera
- Cingulata
- Diprotodontia
- Lagomorpha
- Primates
- Rodentia

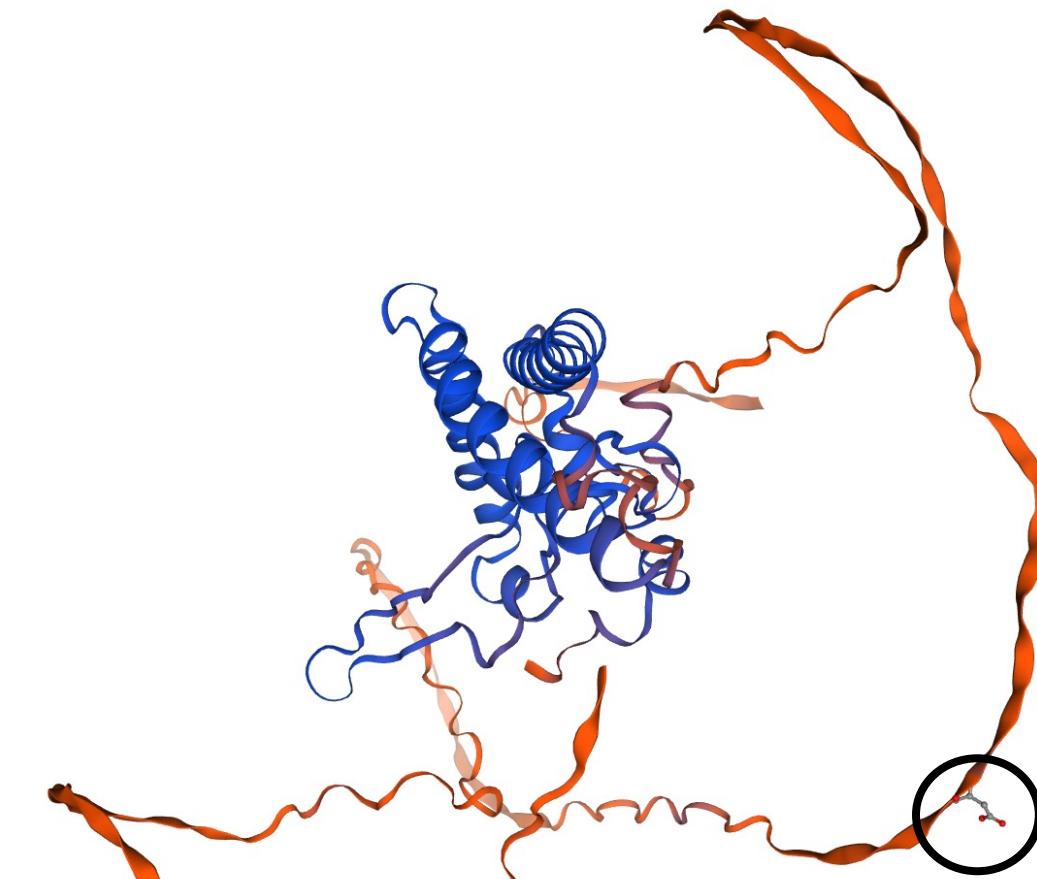
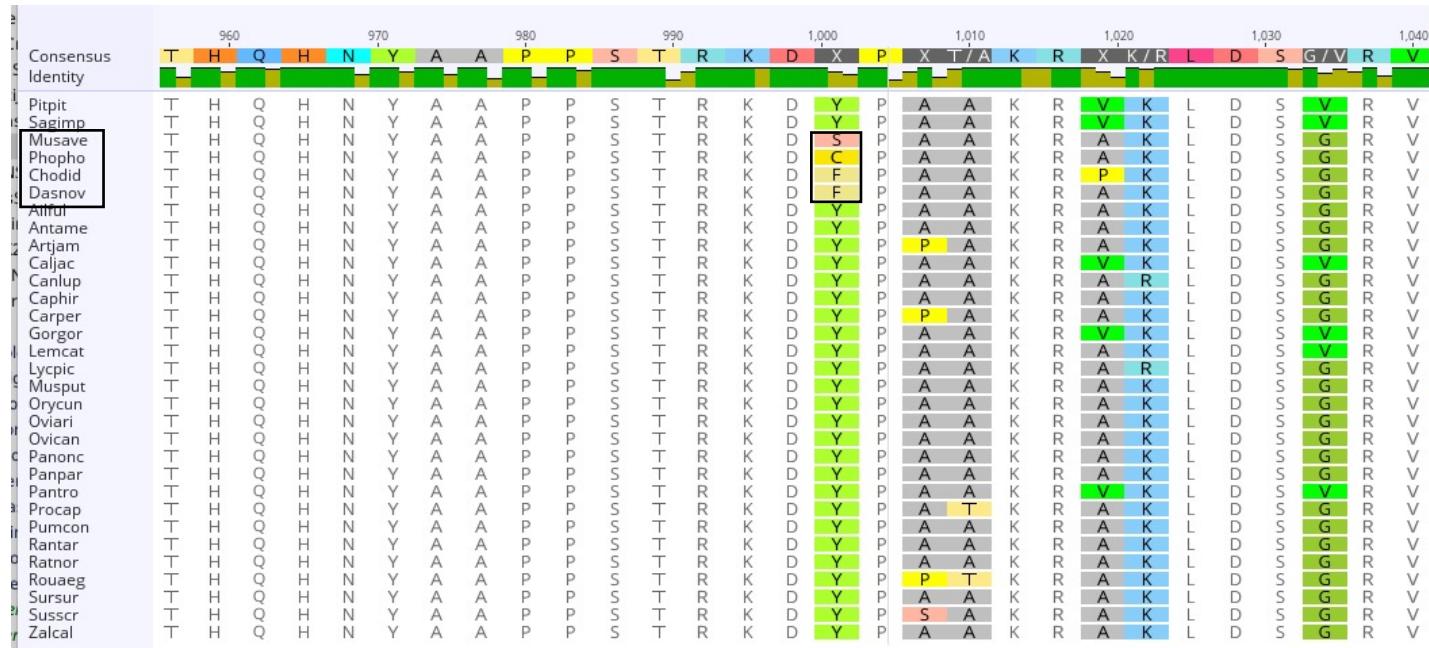
# Outline

- Existing time for comparative genomics and Cancer research
  - Necropsy Data
  - Whole genome alignments
- Major projects
  - TP53 evolution in elephants
  - Consequence of genetic loss in cancer prevalence
  - **Positive selection on “cancer resistant species”**
- Bioinformatic workflow tutorial

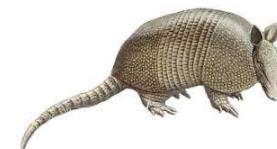
# MYC



# MYC



- Hazel dormouse (*Muscardinus avellanarius*)
- Nine-banded armadillo (*Dasypus novemcintus*)
- Harbour porpoise (*Phocoena phocoena*)
- Two-toed sloth (*Choloepus didactylus*)



Path	pval
cell_surface_receptor_linked_signal_transduction	0.00255041
<b>helperPathway</b>	0.00633378
tcytotoxicPathway	0.00638912
Matrix_Metalloproteinases	0.00694627
malatexPathway	0.00914194
HEMO_TF_LIST_JP	0.00995867
d4gdiPathway	0.01143078
<b>cell_cycle_checkpointII</b>	0.01198008
GLUCO	0.01289684
MAP00670_One_carbon_pool_by_folate	0.01362824
<b>ctla4Pathway</b>	0.01743544
<b>ST_Tumor_Necrosis_Factor_Pathway</b>	0.020149
MAP00620_Pyruvate_metabolism	0.02319838
<b>p53Pathway</b>	0.0237831
Complement_Activation_Classical	0.02731868
MYC_WT	0.03707732
HOX_LIST_JP	0.03864926
<b>dnafragmentPathway</b>	0.04117894

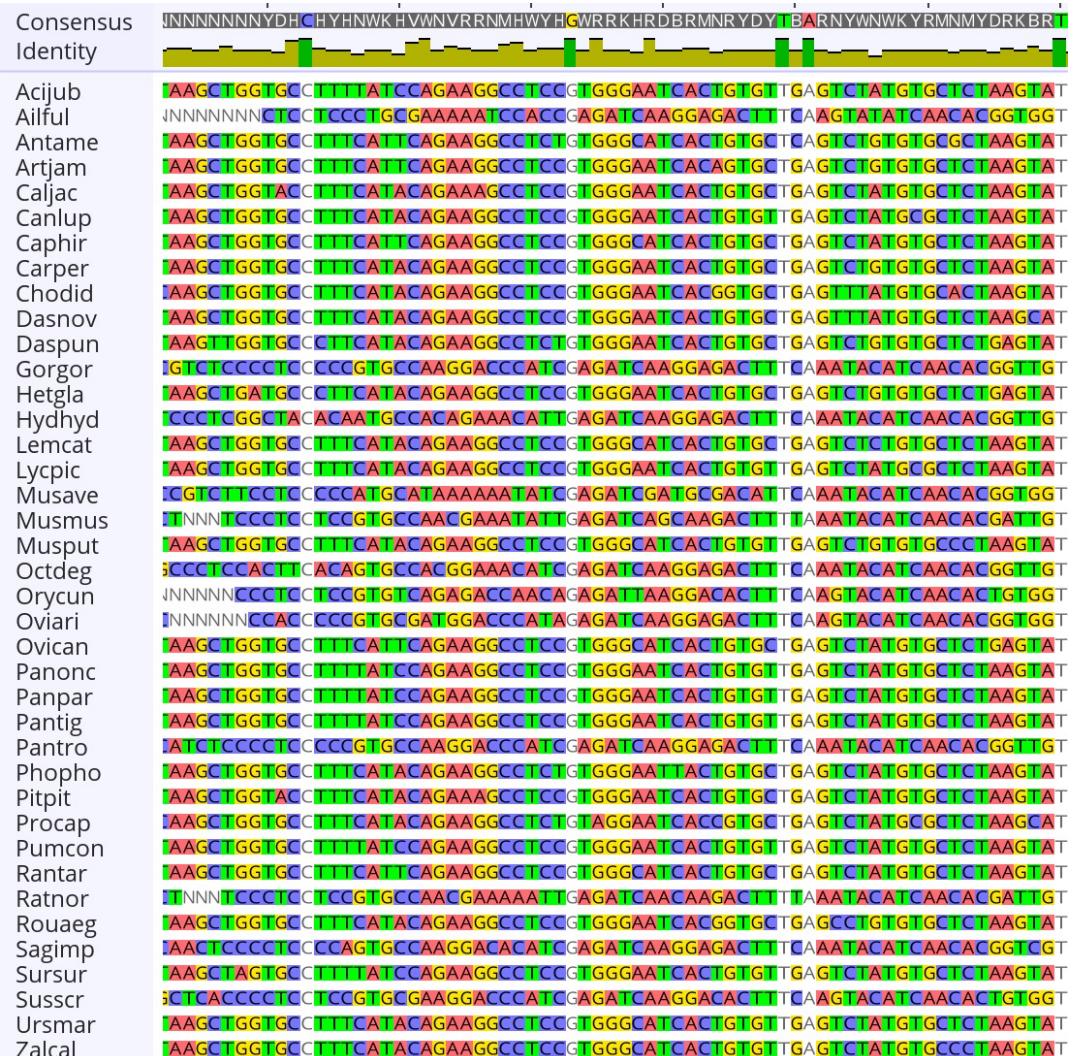
- Enrichment of most significant genes
- Several pathways that could be related to cancer resistance
  - Cell-cycle
  - DNA repair.
  - Immune system.

# Outline

- Existing time for comparative genomics and Cancer research
  - Necropsy Data
  - Whole genome alignments
- Major projects
  - TP53 evolution in elephants
  - Consequence of genetic loss in cancer prevalence
  - Positive selection on “cancer resistant species”
- **Bioinformatic workflow tutorial**

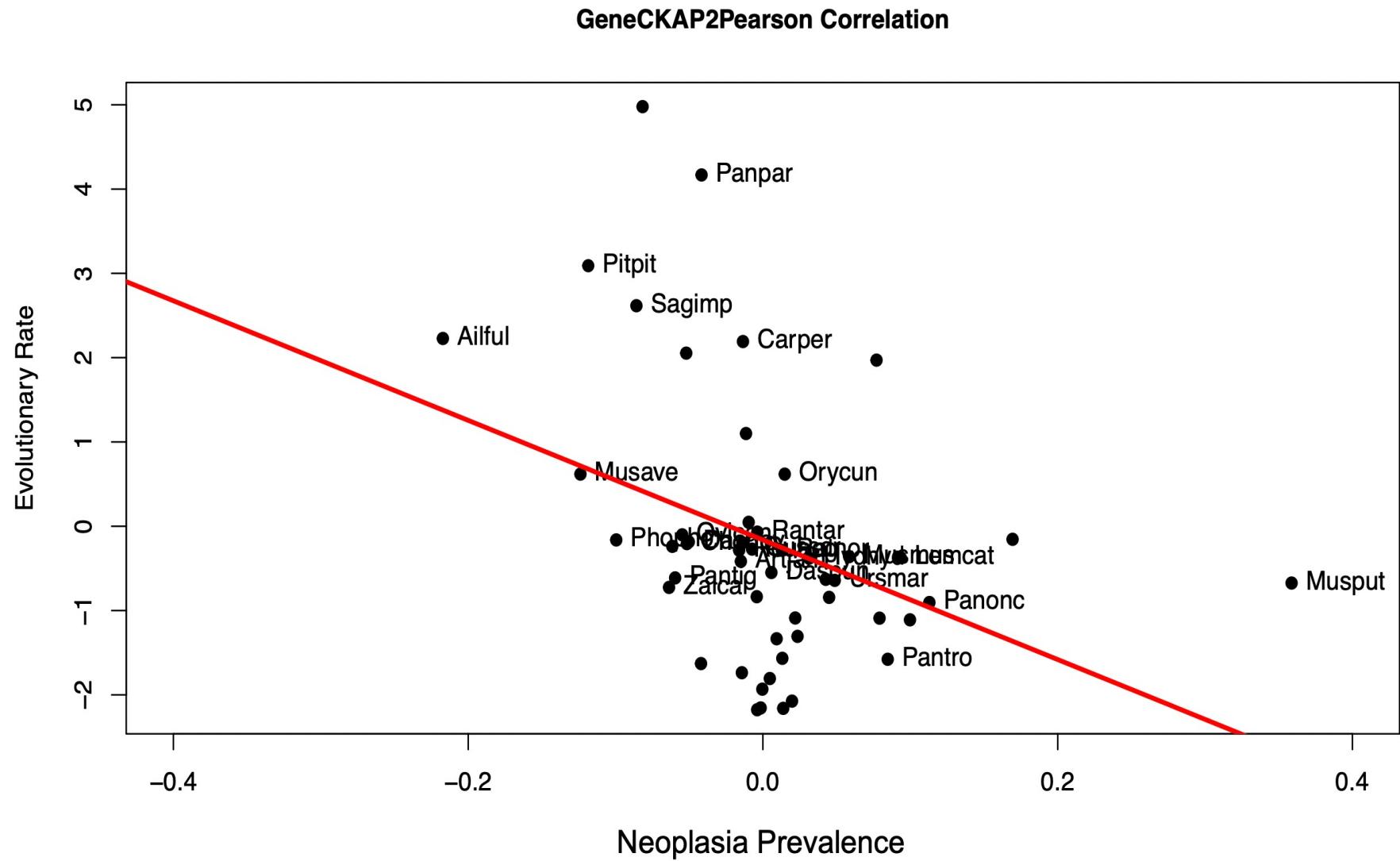
Cactus alignment is available in a very complex format called a hierarchical alignment (HAL)

- **Step1.** Get Cactus alignment in HAL Format
  - **Step2.** Get genomes in fasta format with hal2fasta
  - **Step3.** Liftover the hg38 bed to get CDS coordinates per species



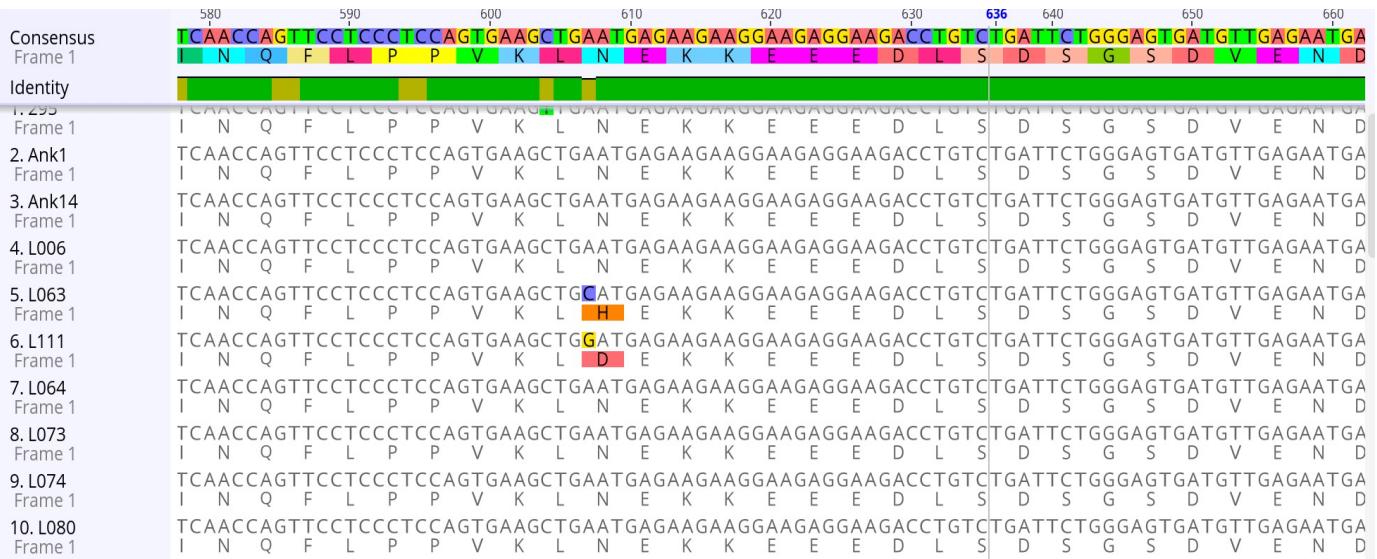


# RERconverge associates evolutionary rates with convergent traits

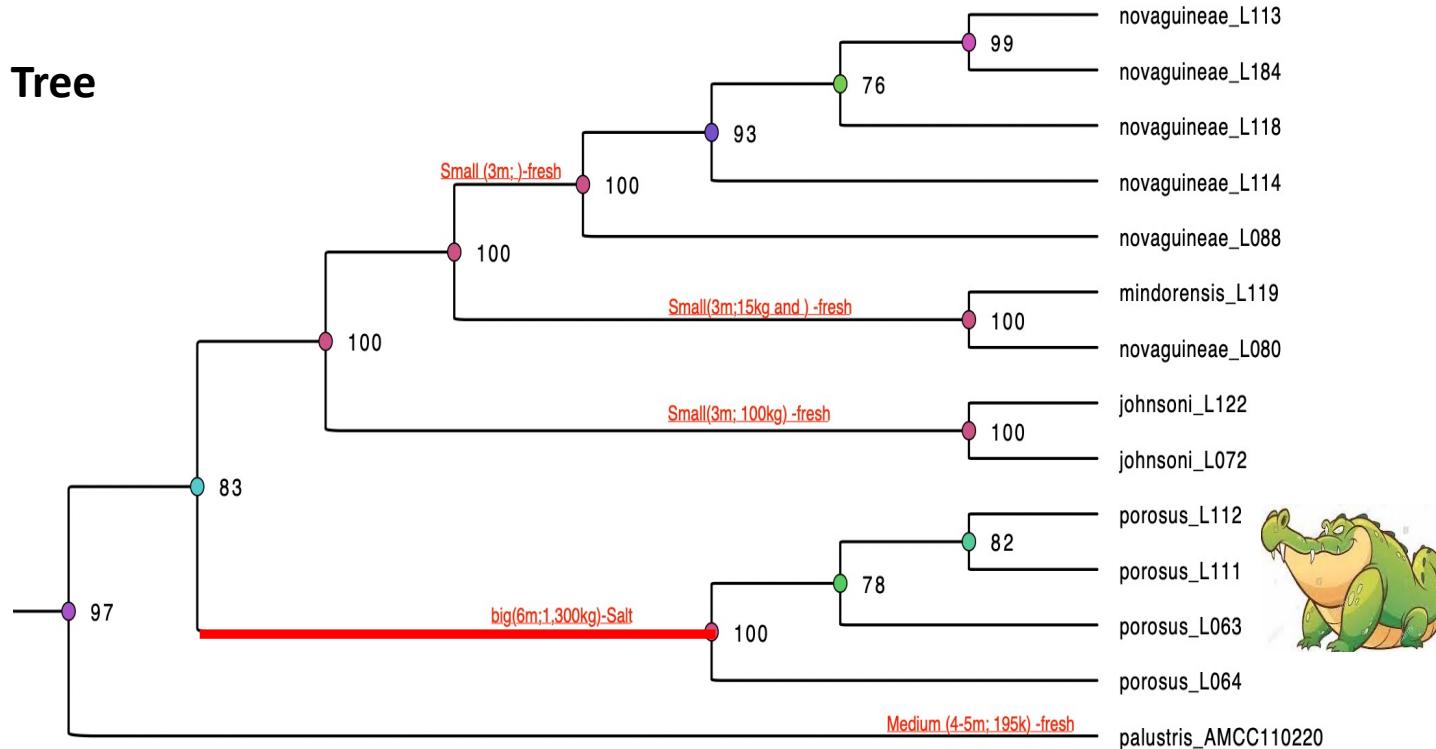


- Start with 241 genomes of mammals. (Cactus alignment from DAN zoo)
- A subset of 103 mammals containing necropsy data.
- For each species we analyzed ~ 24k genes (one gene is shown)
- RERconvergence to correlate Evolutionary rates and necropsy data.

## 1. Alignment



## 2. Tree



**Ko** : Conserved ( $0 < \omega_0 < 1$ ) on both branches

**K1:** neutral selection ( $\omega_1 = 1$ ) on both branches

**K2a: positive selection ( $\omega_2 \geq 1$ ) on forward and conserved ( $0 < \omega_0 < 1$ ) on background selection .**

**K2b: positive selection ( $\omega_2 \geq 1$ ) on forward  
and neutral selection ( $\omega_1 = 1$ ) on background selection.**

- $\omega < 1 \rightarrow$  purifying selection
- $\omega = 1 \rightarrow$  neutral evolution
- $\omega > 1 \rightarrow$  positive selection

$$\omega = dn/ds$$

## **Limitation about positive selection models:**

Problem: Multiple sites changes within a codon

<https://pubmed.ncbi.nlm.nih.gov/29967485/>

Solution: Masked for multiple mutations within a codon with tools like SWAMP  
(<https://github.com/peterwharrison/SWAMP>)

Problem: everything could be under positive selection.

<https://www.biorxiv.org/content/10.1101/2021.10.26.465984v1.full>

Solution: Run a site model without the candidate linages and check for selection

Problem: Ignoring synonyms rate variation across linages could inflate false positive

<https://academic.oup.com/mbe/article/37/8/2430/5739973?login=false>

Solution: Model synonym variation with Bused in hyphy (more flexible)

Problem: multiple hits in the same site could underestimate positive selection

Solution: Allow paml to estimate different omegas across linages

## **Other Readings:**

Review of positive selection with whole genomes

<https://www.sciencedirect.com/science/article/pii/S0168952520301852>

Email: dechavezv@puce.edu.ec

# Acknowledgments



Carlo Maley



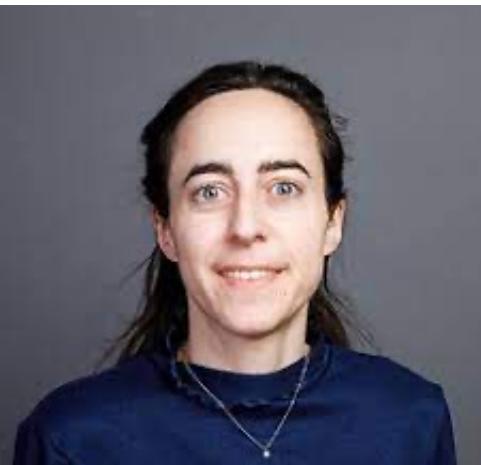
Amy Boddy



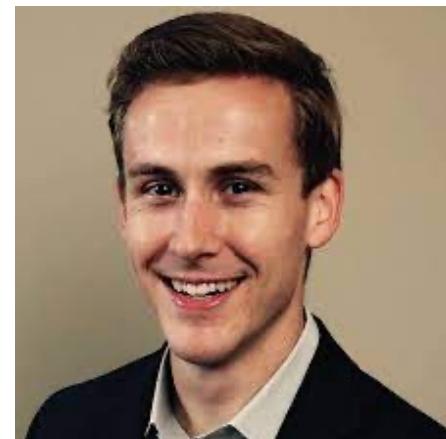
Lisa M. Abegglen



Marc Tollis



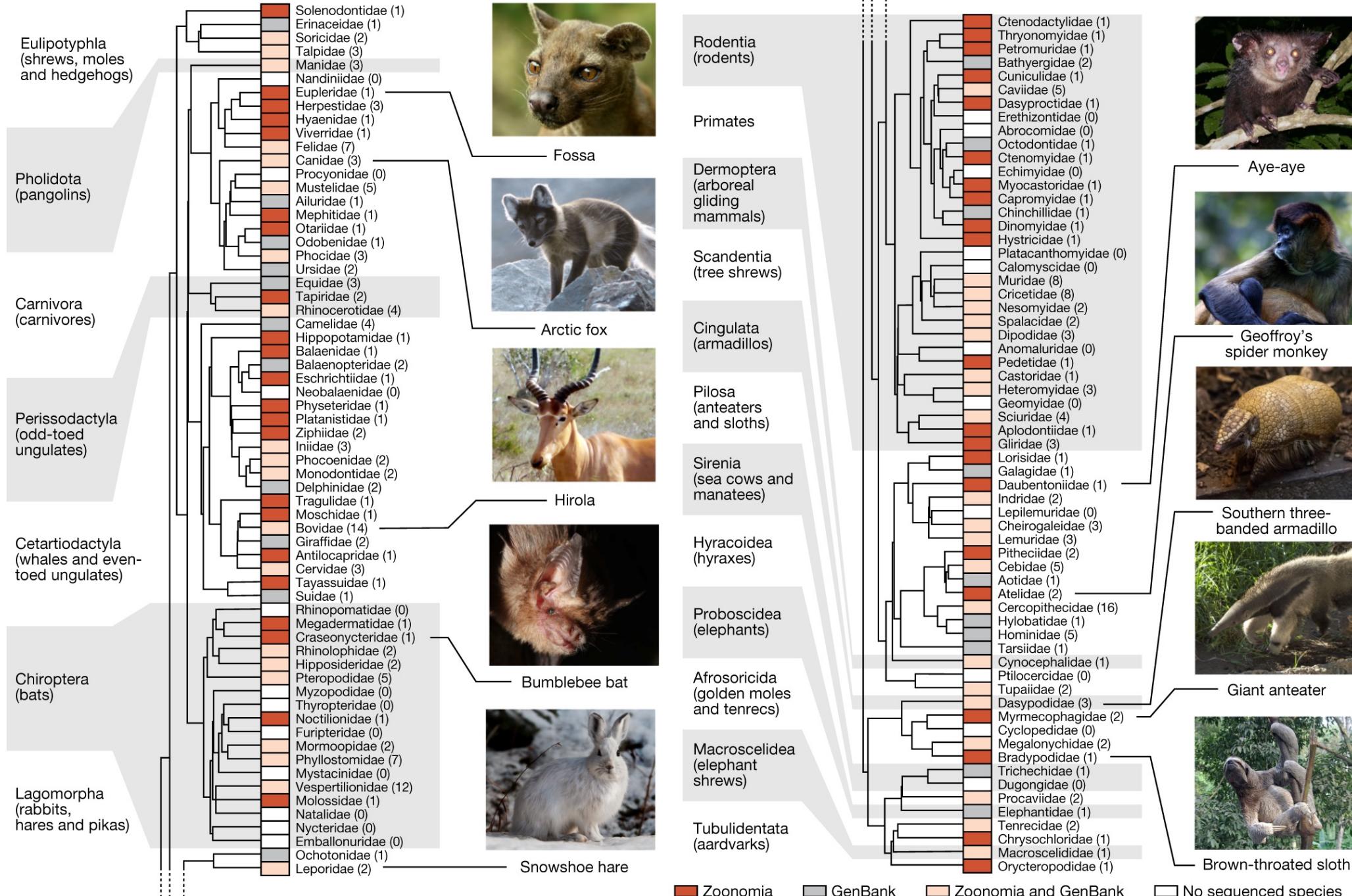
Steffi Kapsetaki



Zachary Compton







Zoonomia Consortium. 2020. Nature