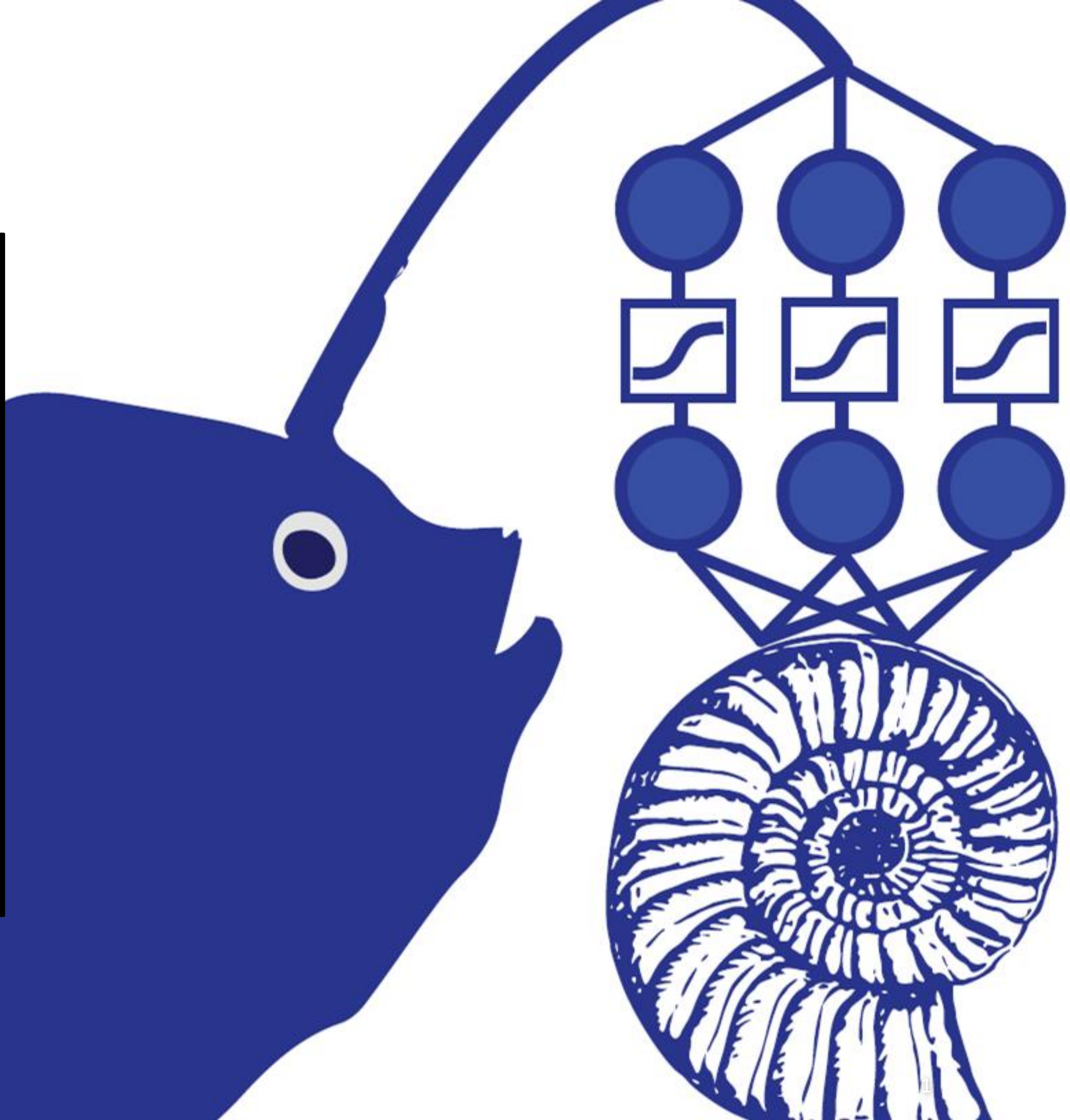
Introduction

DeepDive: Deep learning Diversity Estimation

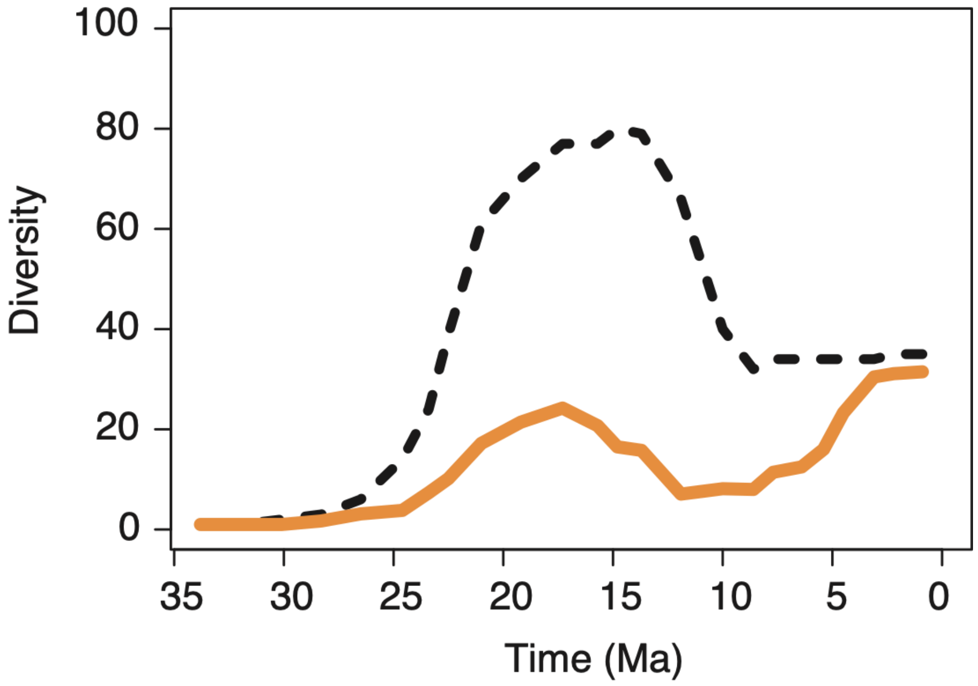
CPEG 2025



Estimating biodiversity from fossil data

- Fossils are our only direct evidence of past biodiversity.
- Non-random biases at spatial, taxonomic and temporal scales.
- Available methods are based on rarefaction or simple models assuming random sampling.

True vs observed diversity



Alroy 2010 Science; Starrfelt & Liow 2016 Phil Trans B; Flannery-Sutherland et al. 2022 Nature Comm

Problems in estimating biodiversity from fossils

- Spatial biases make global biodiversity patterns unidentifiable using the current methods.

Science AAAS

PALEONTOLOGY

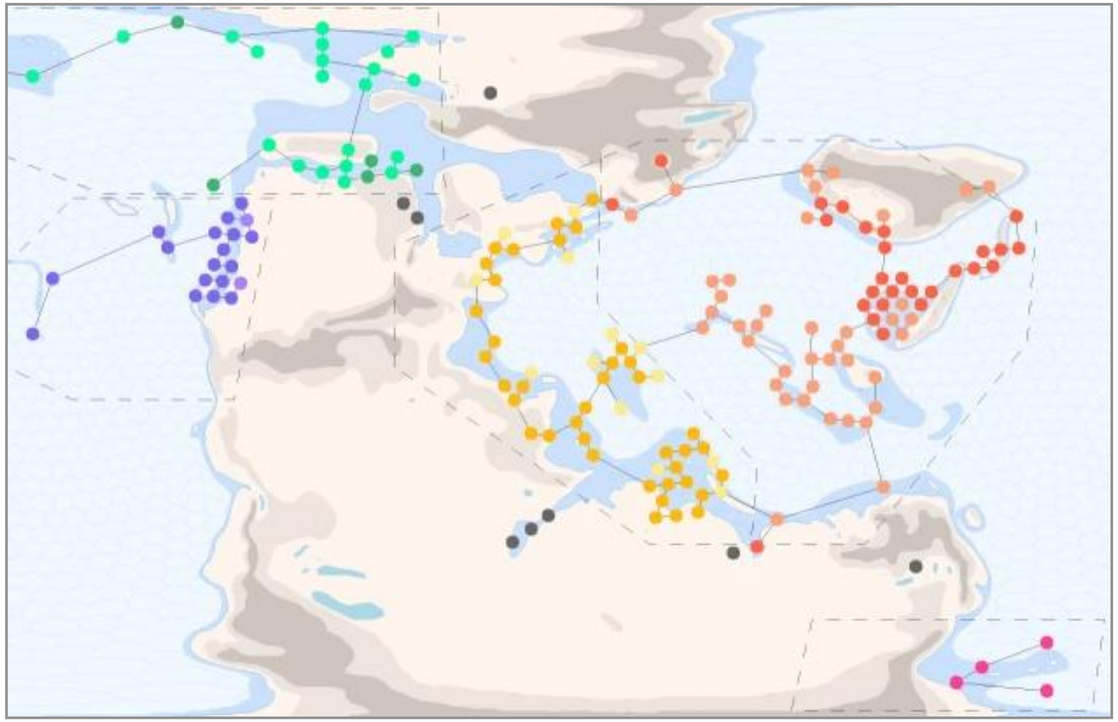
The spatial structure of Phanerozoic marine animal diversity

R. A. Close[1]*, R. B. J. Benson[2], E. E. Saupe[2], M. E. Clapham[3], R. J. Butler[1]

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

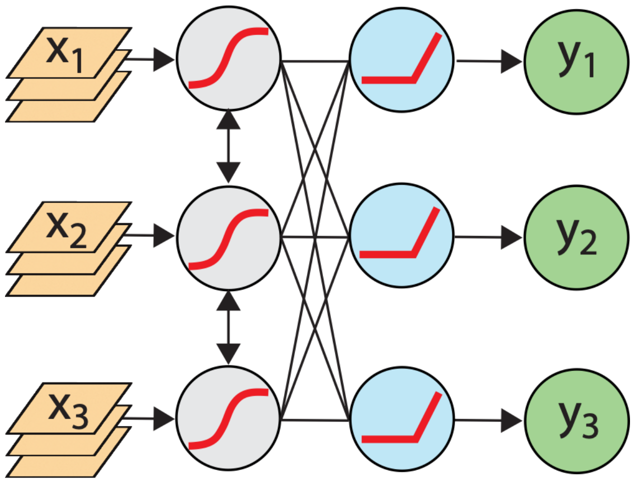
The apparent exponential radiation of Phanerozoic land vertebrates is an artefact of spatial sampling biases



Alroy 2010 Science; Starrfelt & Liow 2016 Phil Trans B; Flannery-Sutherland et al. 2022 Nature Comm

DeepDive: deep learning estimation of biodiversity through time

Input: fossil distribution in space and time



Output: diversity estimates through time

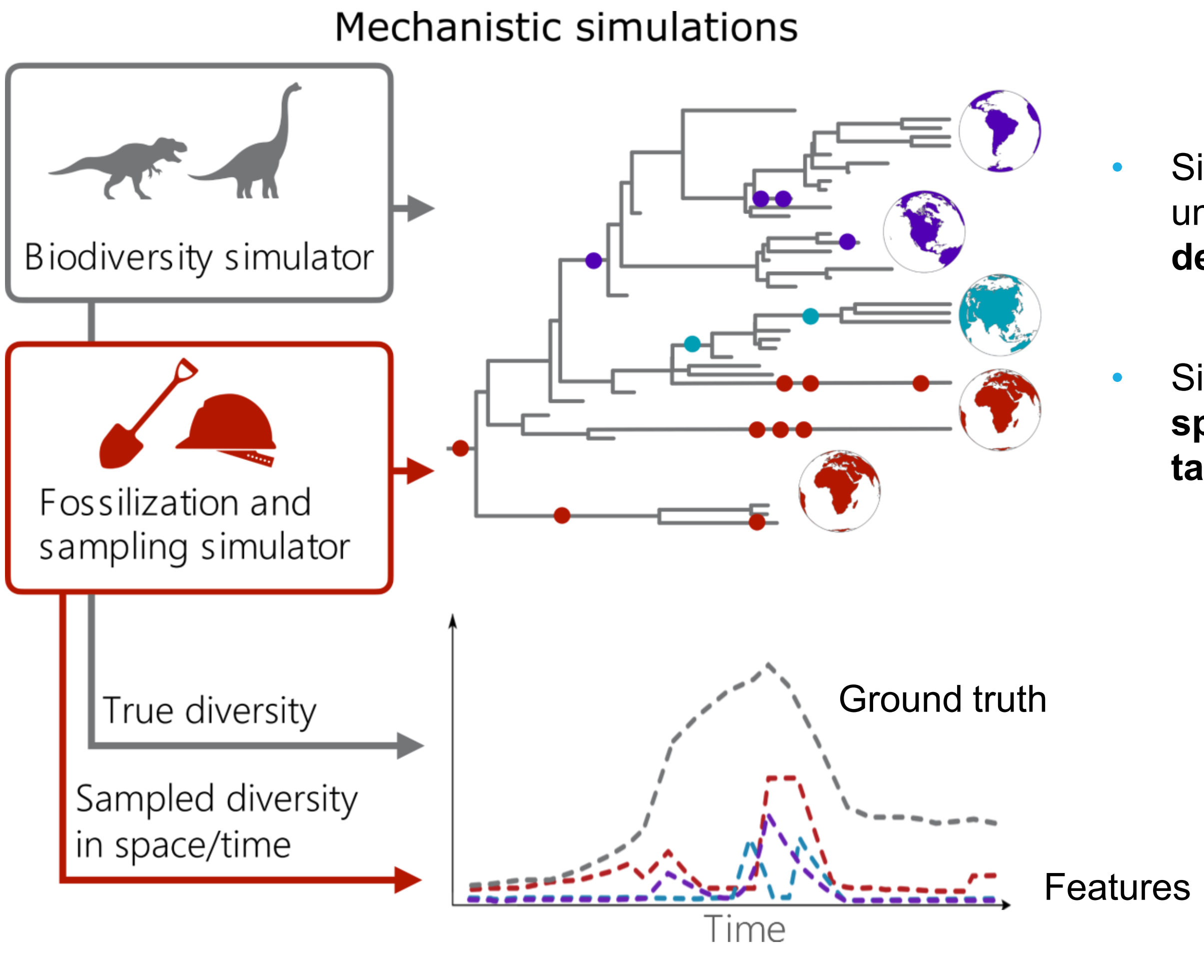

nature communications

Article
https://doi.org/10.1038/s41467-024-48434-7

DeepDive: estimating global biodiversity patterns through time using deep learning

Cooper et al. 2024 Nature Comms
Cooper et al. 2024 biorXiv

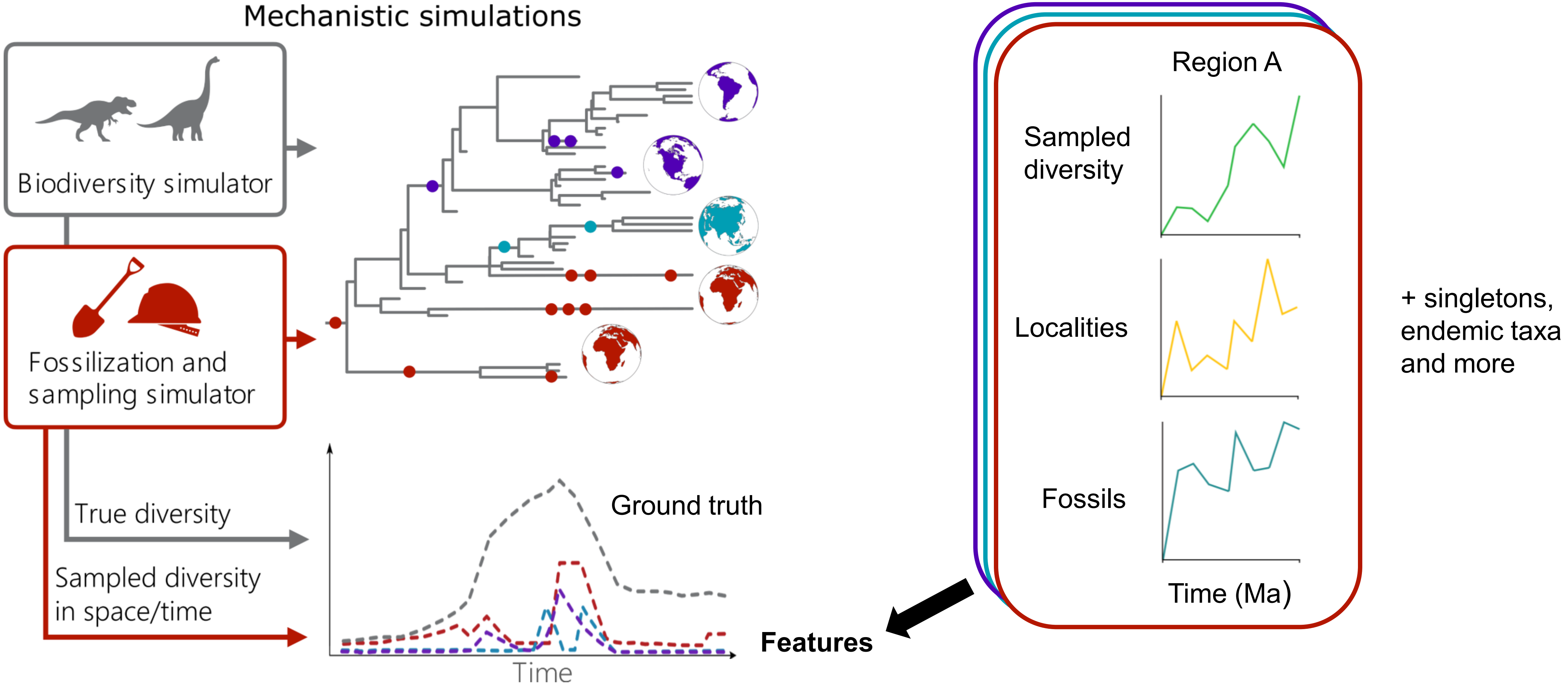
DeepDive



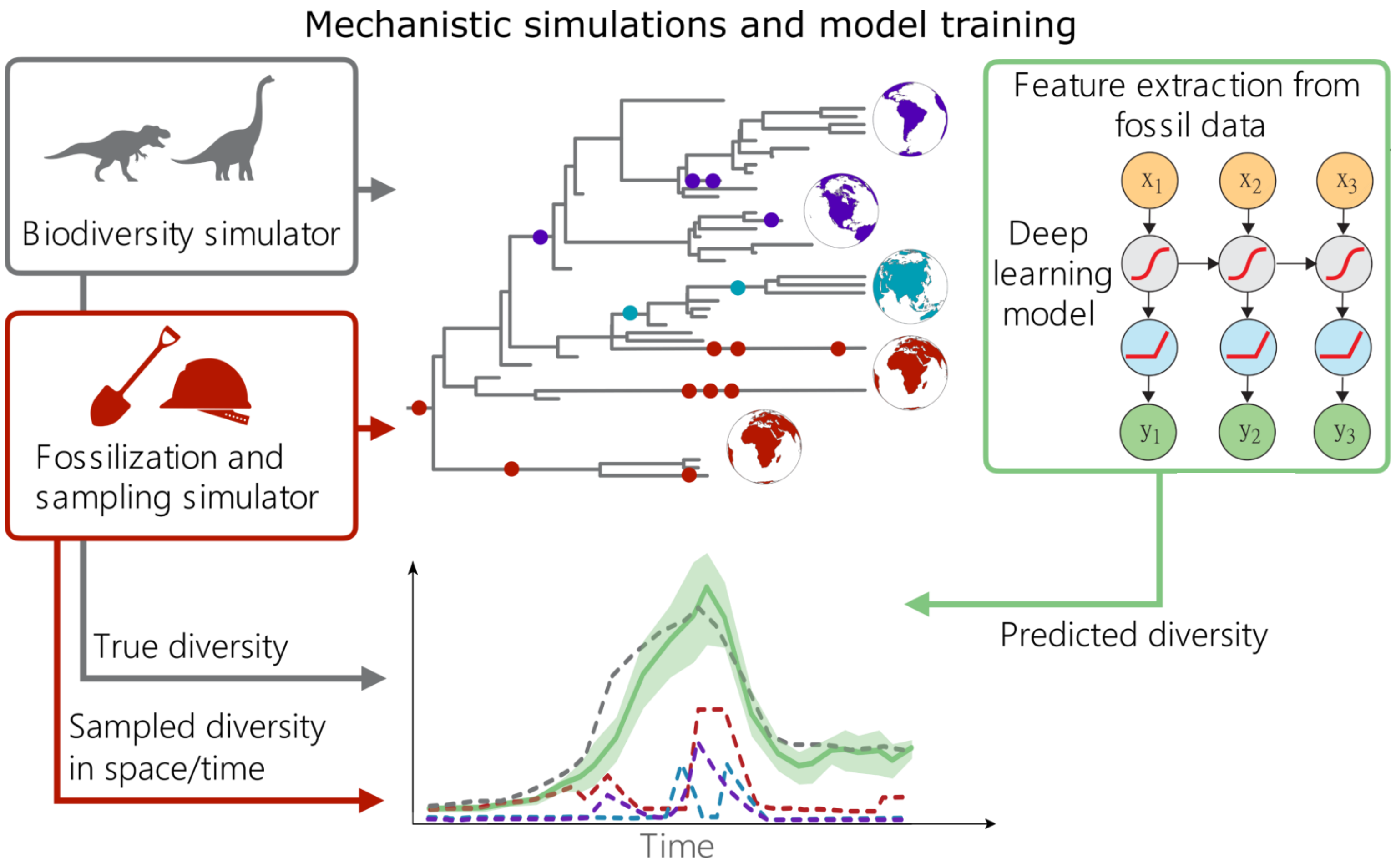
- Simulating clade evolution under spatial-explicit **birth-death processes**
- Simulating fossil data with **spatial, temporal, and taxonomic biases**

Cooper et al. 2024 Nature Communications

DeepDive


Mechanistic simulations
Biodiversity simulator
Fossilization and sampling simulator
True diversity
Sampled diversity in space/time
Ground truth
Time
Features
Region A
Sampled diversity
Localities
Fossils
Time (Ma)
+ singletons, endemic taxa and more

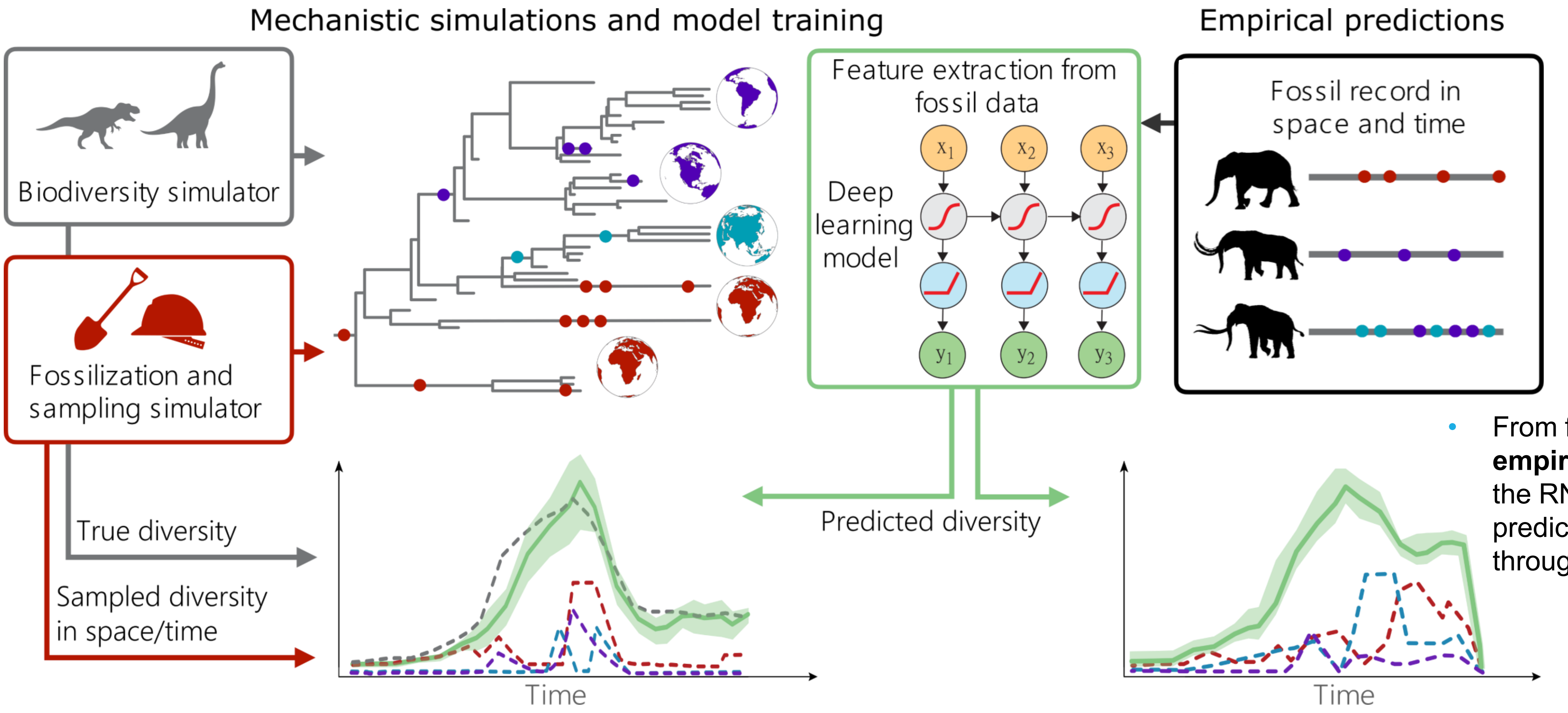
DeepDive



Deep learning models for diversity estimation

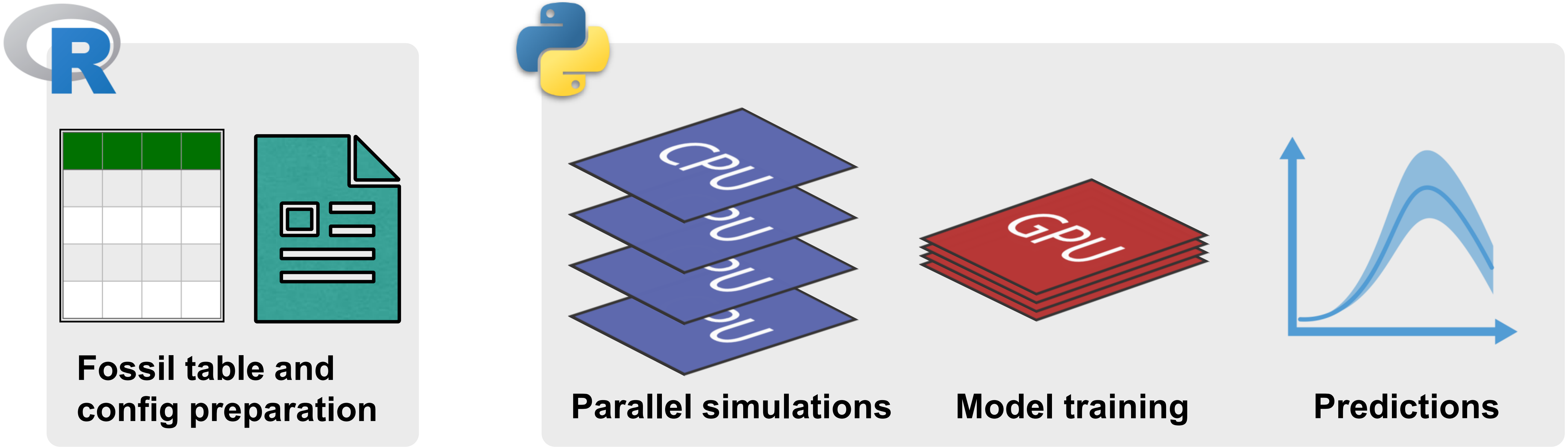
- Training a **Recurrent Neural Network** to predict true diversity from sparse fossil data

DeepDive



- From features of **empirical data**, the RNN can predict diversity through time

DeepDiveR



DeepDiveR

DeepDive

github.com/DeepDive-project

Data preparation

- Occurrence data e.g. from the Paleobiology Database can be formatted as input for DeepDive using the function prep_dd_input()

Taxon	Region	MinAge	MaxAge	Locality
Ailurus	Asia	0.6	1.3	1
Ailurus	Asia	9.5	10.35	2
Alopecocyon	Asia	9.5	10.35	2
Alopecocyon	Europe	9.7	11.11	3
…	…	…	…	…

	A	B	C	D	E	F
1	Replicate	Type	Region	t1	t2	t3
2	1	bin_start	NA	66.043	65	64
3	1	bin_mid	NA	65.5215	64.5	63.5
4	1	bin_dur	NA	1.043	1	1
5	1	locs	Africa	0	0	0
6	1	locs	Asia	1	0	0
7	1	locs	Europe	0	0	0
8	1	locs	NorthAme	0	2	1
9	1	locs	SouthAme	0	0	0
10	1	occs	Africa	0	0	0
11	1	occs	Africa	0	0	0
12	1	occs	Africa	0	0	0

Configuration

1. General:
Settings called at every stage of the analysis.

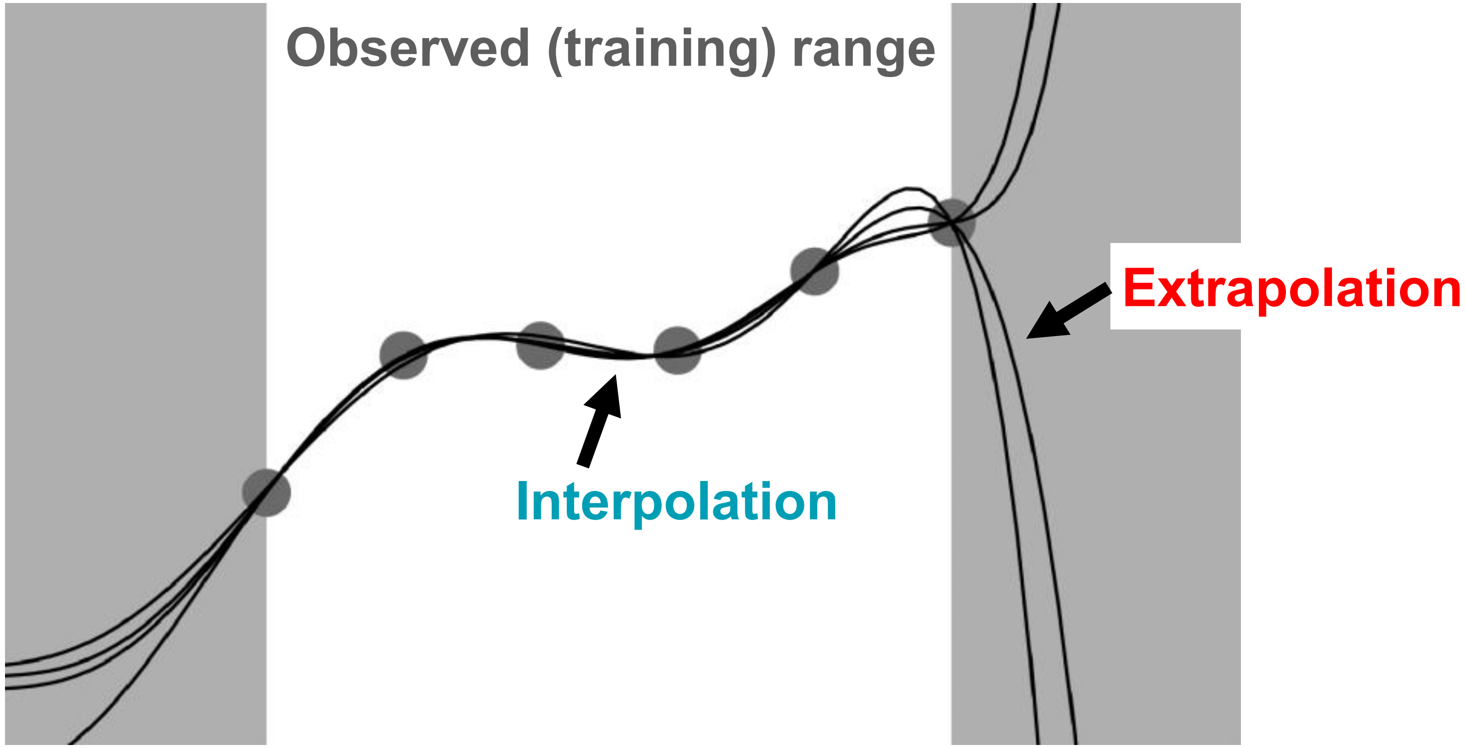
2. Simulations:
Settings for generating training and test data sets.

3. Regional constraints:
The ages at which regions become available or unavailable.

4. Model training:
Specification of deep learning models to be trained.

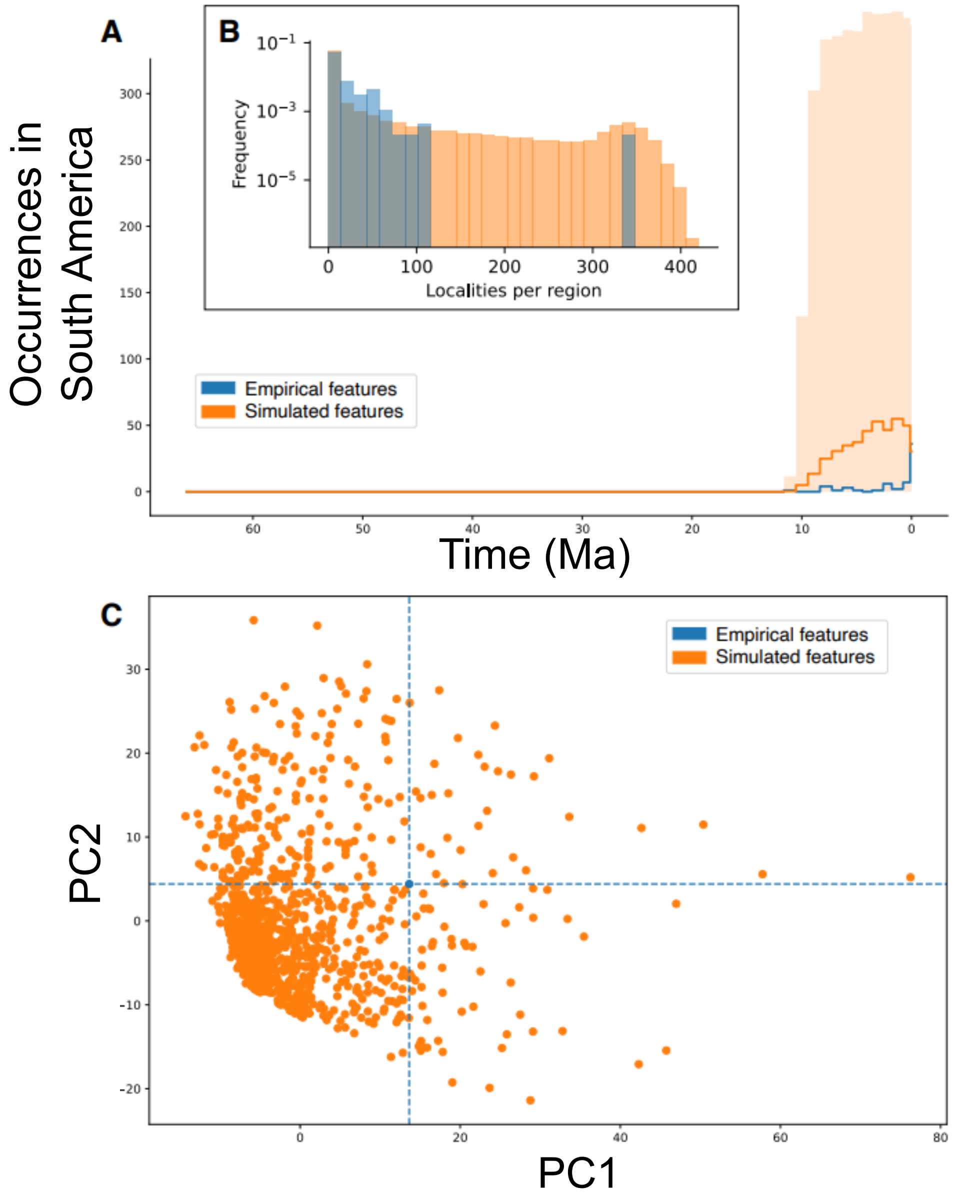
5. Empirical predictions:
Settings for generating predictions from empirical data.

Training simulations must reflect the data

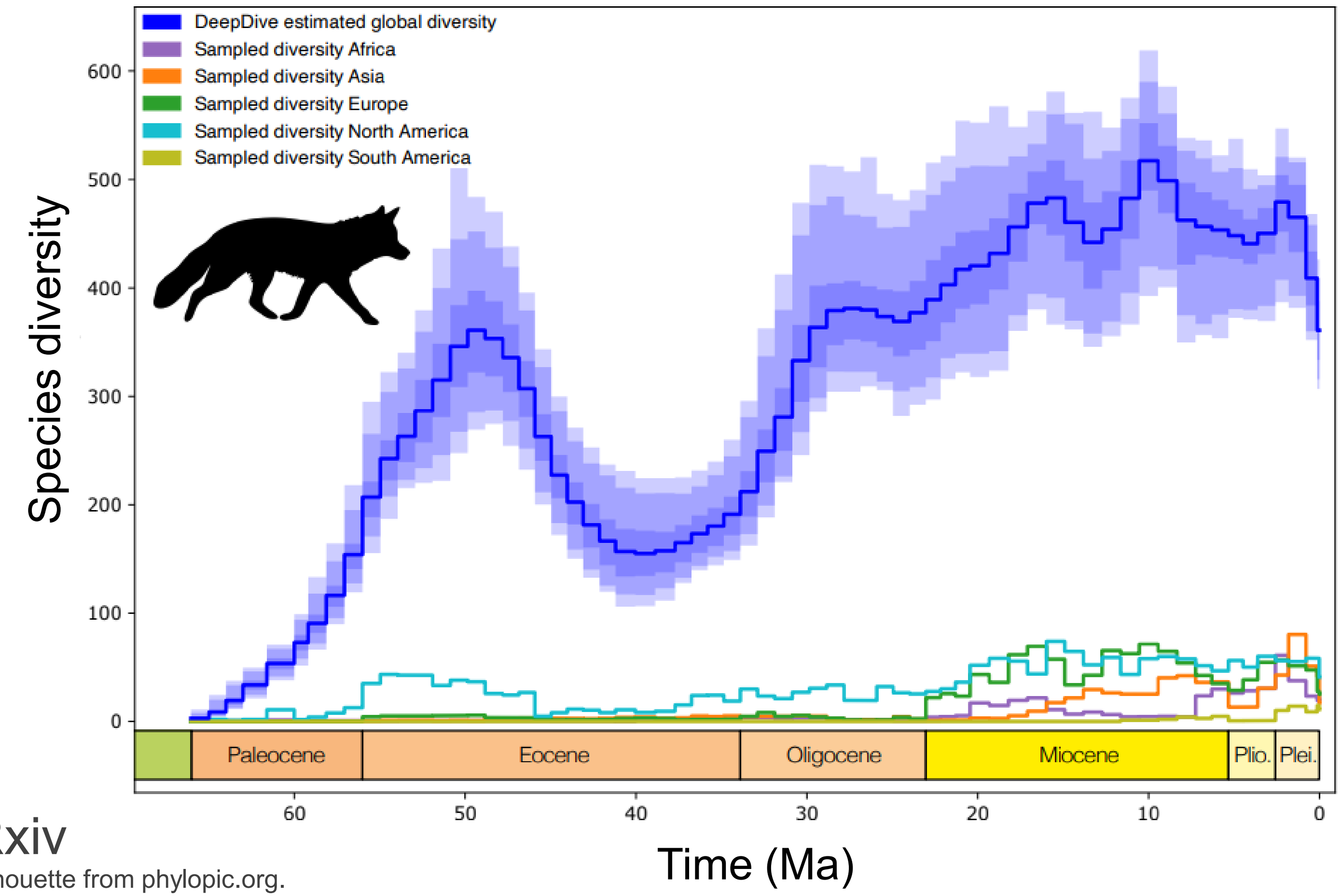


Autotune adjusts simulations to reflect empirical data

- Parameters with default NA are adjusted automatically and a new configuration saved unless user specified
- E.g. Across each region:
  - Sampled species
  - Longevity
  - Singletons
  - Gaps in data
  - Occurrences



Inferred diversity curve



Cooper et al. 2024 bioRxiv

Data from Faurby et al 2024 Proceedings B, silhouette from phylopic.org.

Let’s get started!

You can access the next tutorial at:

https://github.com/thauffe/cpeg25/tree/main/deepdive_analysis

or by following the QR code

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

SIB Swiss Institute of Bioinformatics

Swiss National Science Foundation

CPEG 2025

Practical

DeepDive and DeepDiveR

Rebecca B. Cooper

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG