



Latent Space Phenotyping for Measuring Complex Evolutionary Traits

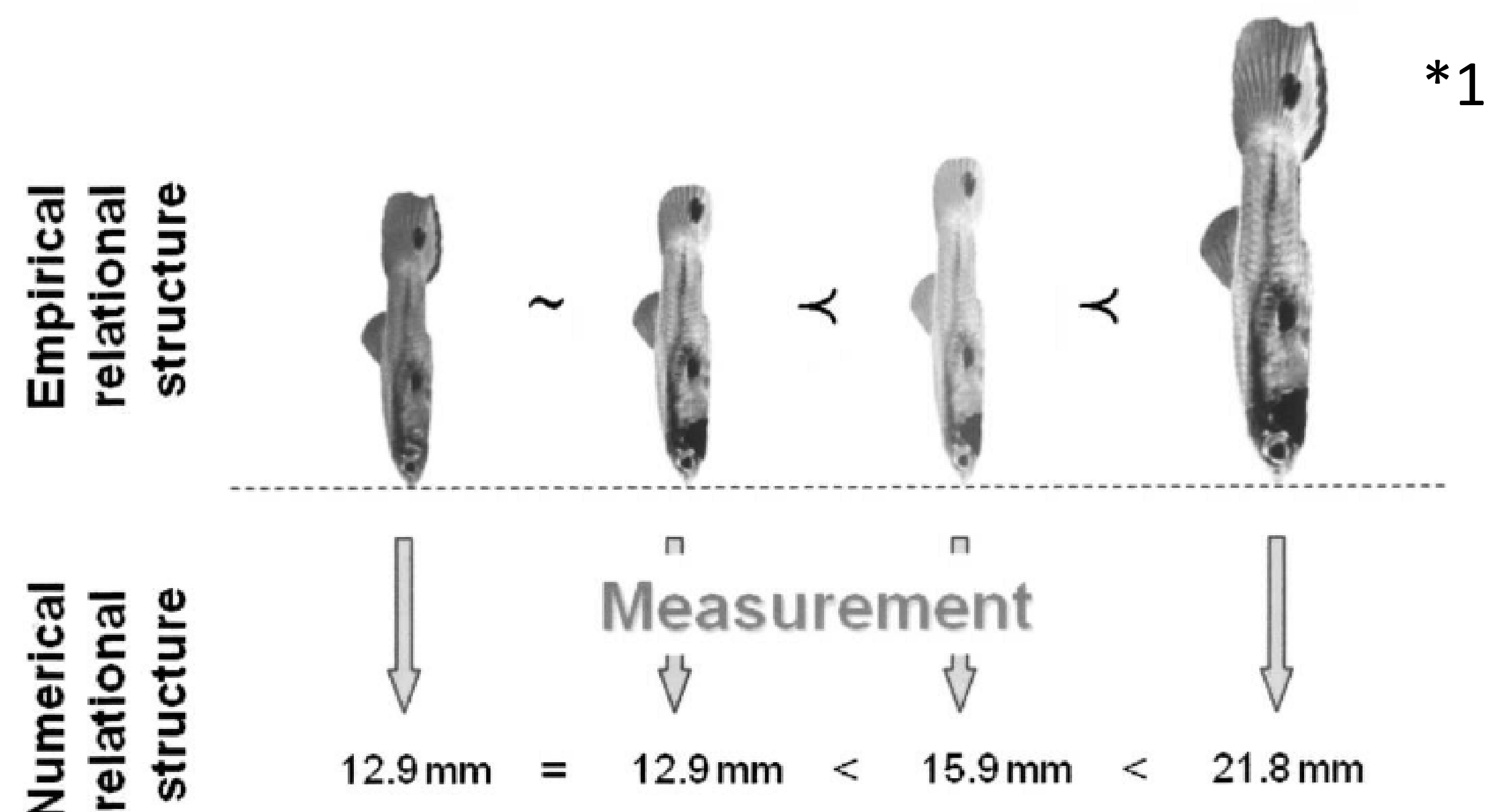
Caleb Charpentier, Josef Uyeda

Virginia Tech Department of Biological Sciences

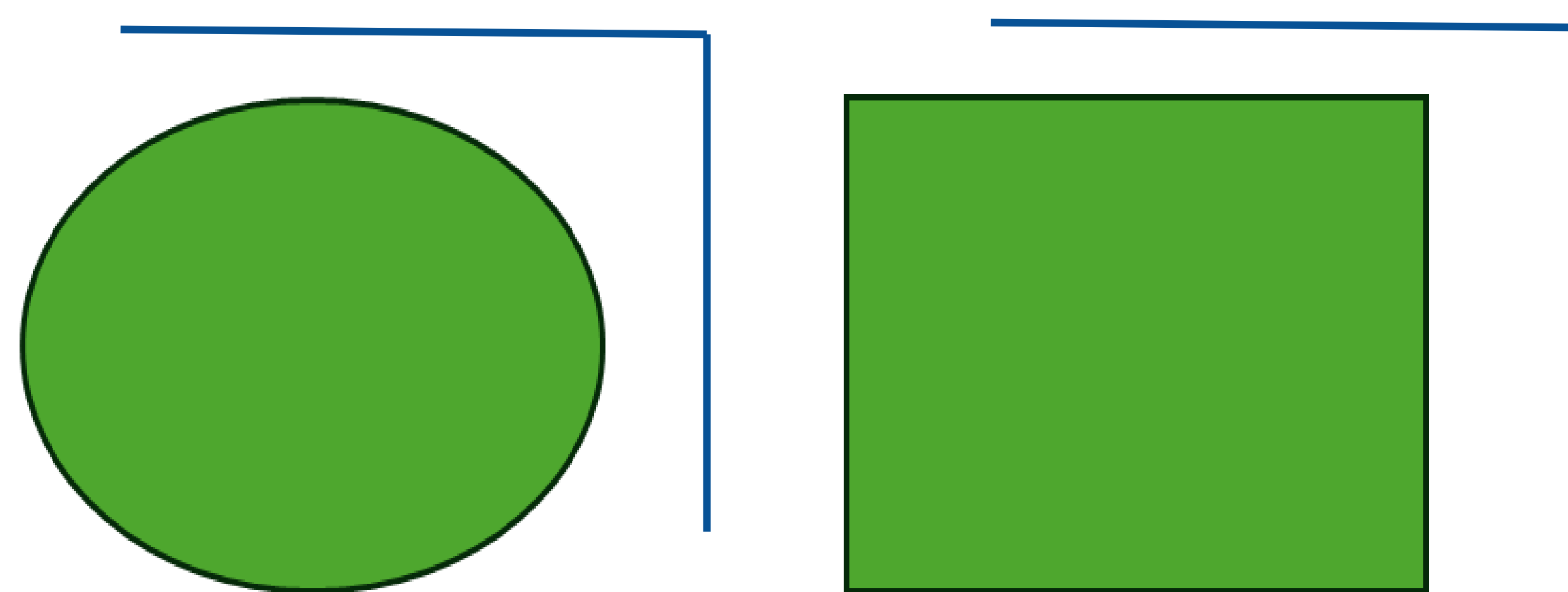


Why Does Measurement Matter?

Measurement generally requires defining scales for data, conditioned on a data-generating process



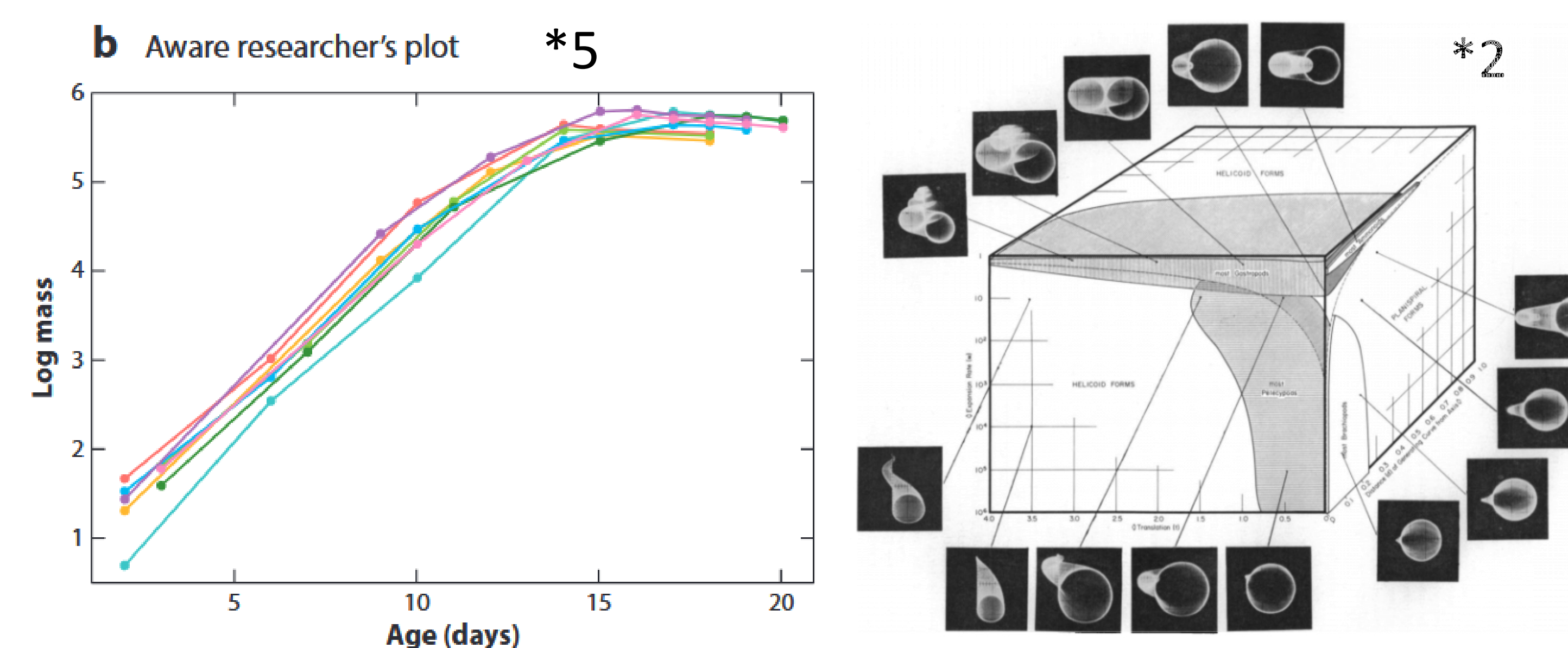
Measurement is a process of transferring observations to a symbolic space subject to isomorphic transformations



Human-devised scales are often inadequate for measurement

Traits on Complex Scales

Biological patterns can be represented in diverse mathematical ways



Traits can take numerical, functional, or other complex values

KGML

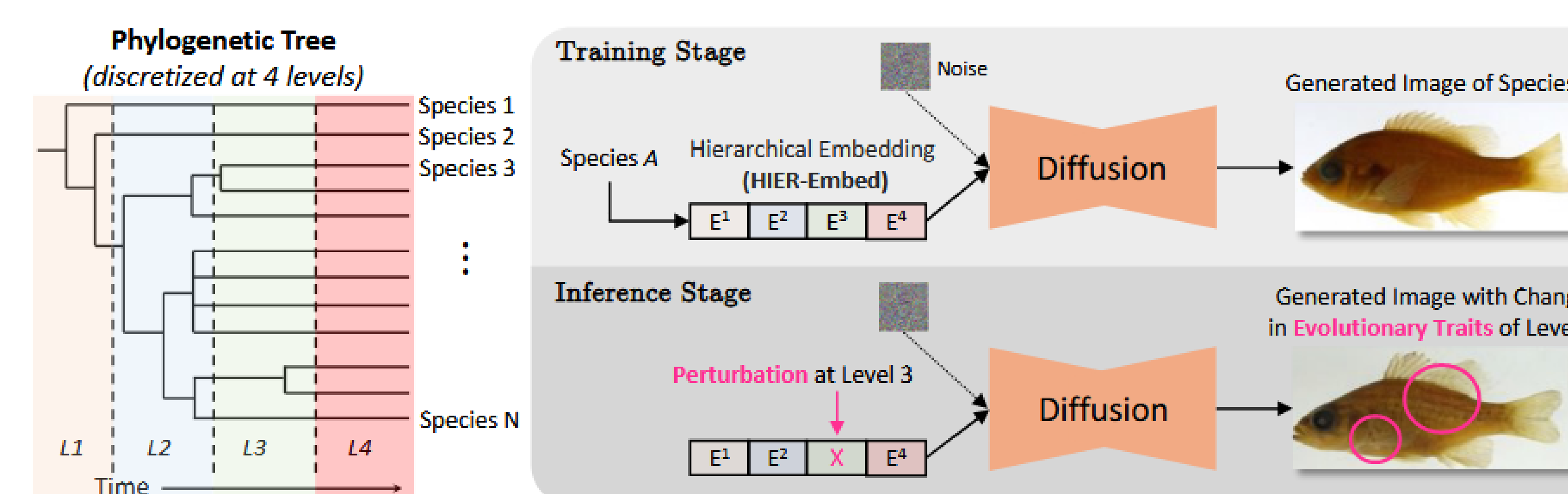
Performance \propto Accuracy + Simplicity + Consistency

Rapid progress on any biological problem rests on the hope that there is at least one viewpoint to each problem that makes causation relatively simple.

-- David Houle

For every complex problem there is a solution which is clear, simple and wrong.

-- Often attributed to H.L. Mencken



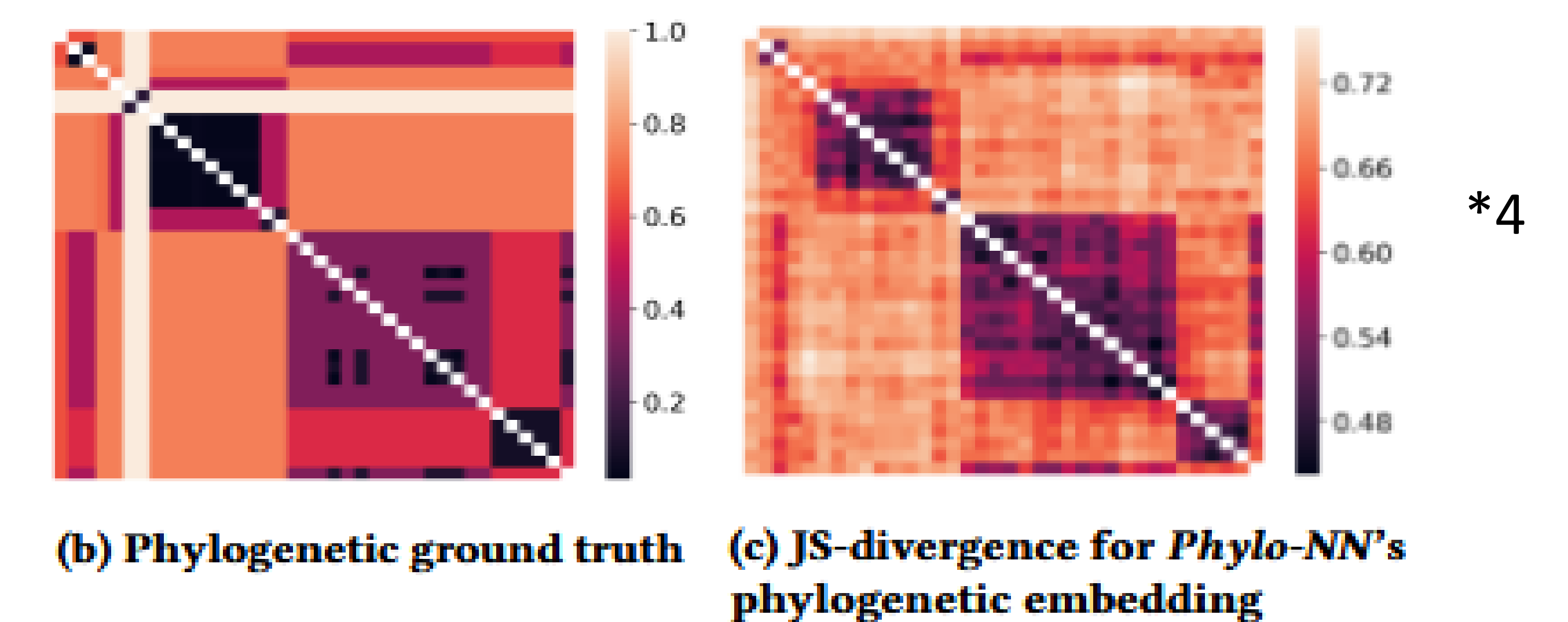
Phylogenetic grounding can make trait discovery more interpretable

Tests of Performance

A Generalized K Statistic for Estimating Phylogenetic Signal from Shape and Other High-Dimensional Multivariate Data *6

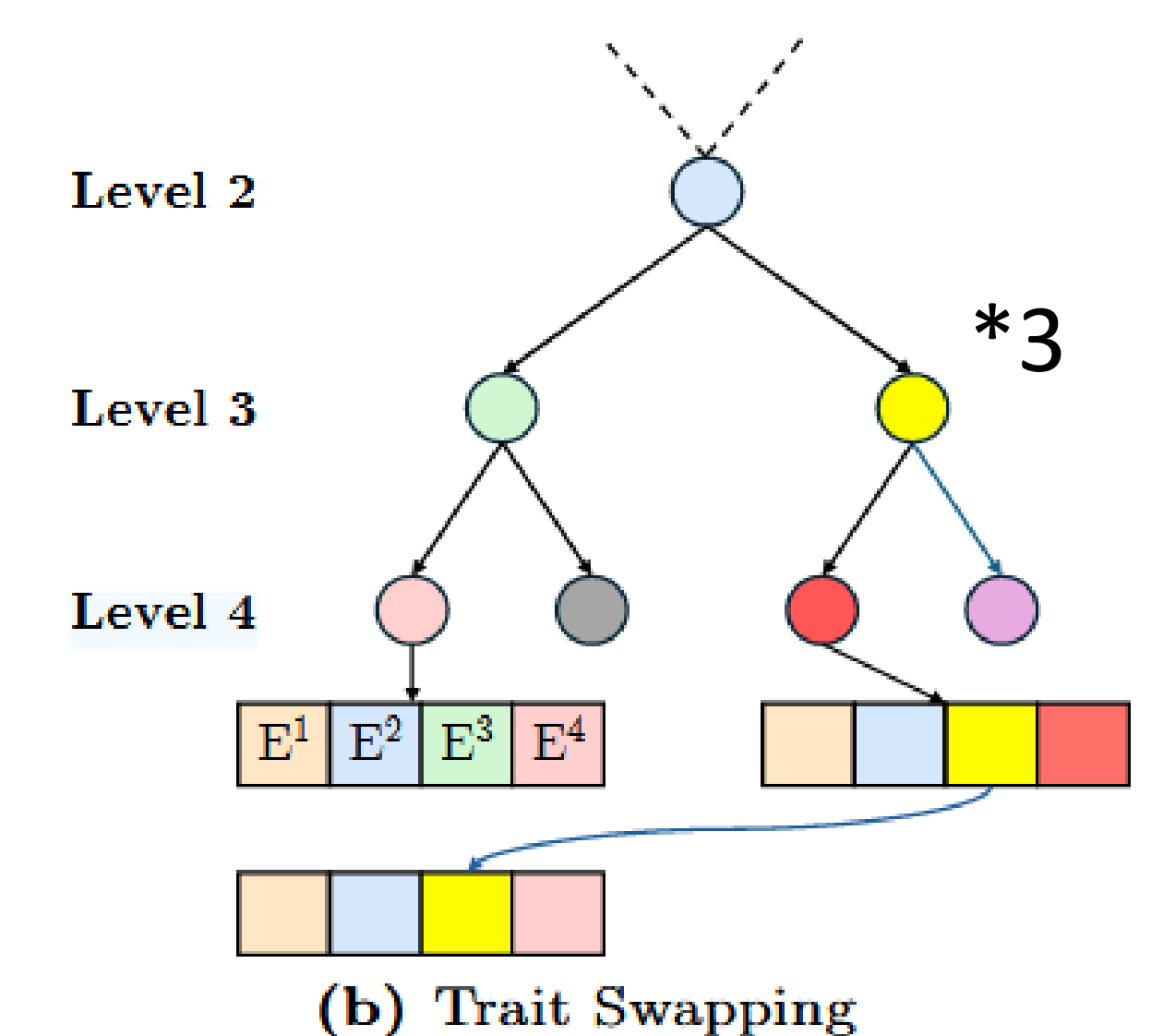
DEAN C. ADAMS*

Department of Ecology, Evolution, and Organismal Biology, and Department of Statistics, Iowa State University, Ames IA, 50011, USA; *Correspondence to be sent to: Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA; E-mail: dcadams@iastate.edu.



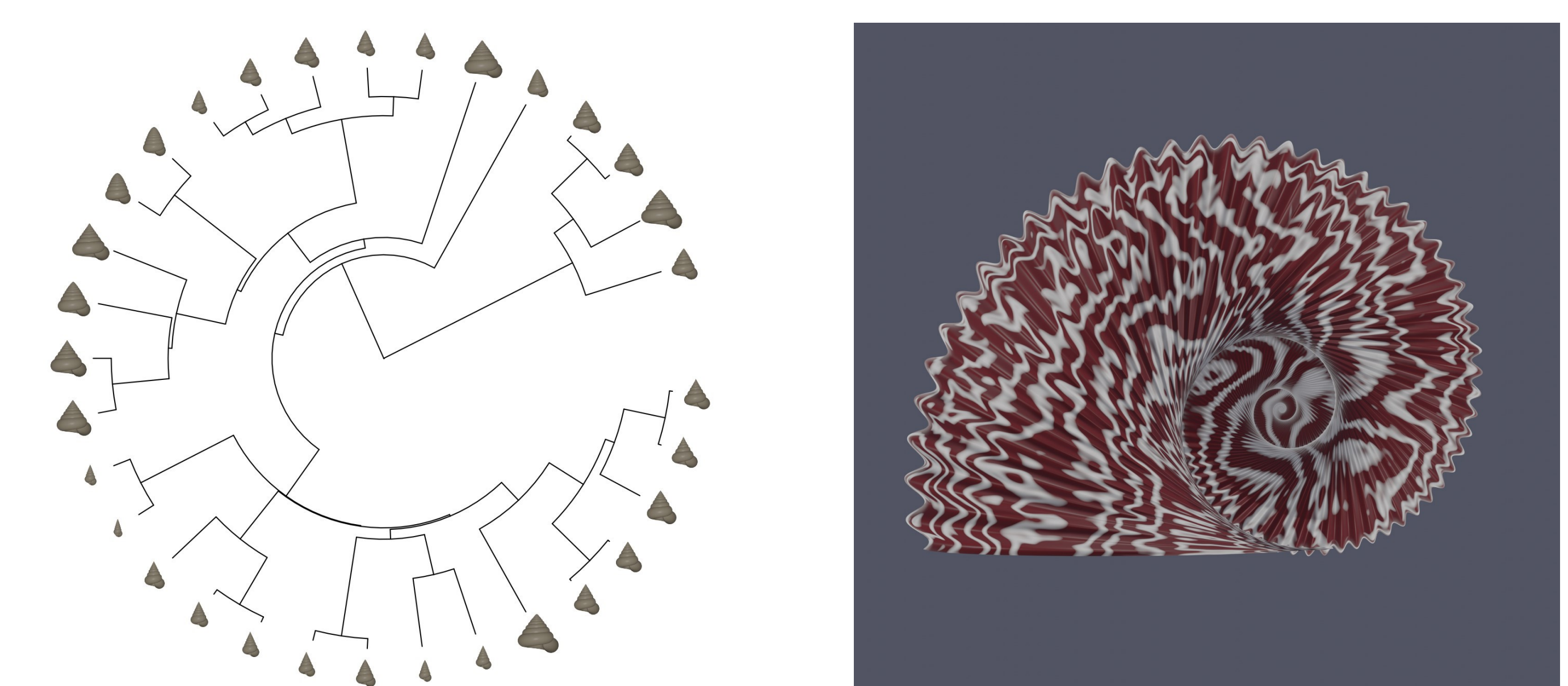
Level 4	E^1	E^2	E^3	E^4
Level 3	E^1	E^2	E^3	X
Level 2	E^1	E^2	X	X
Level 1	E^1	X	X	X

(a) Trait Masking

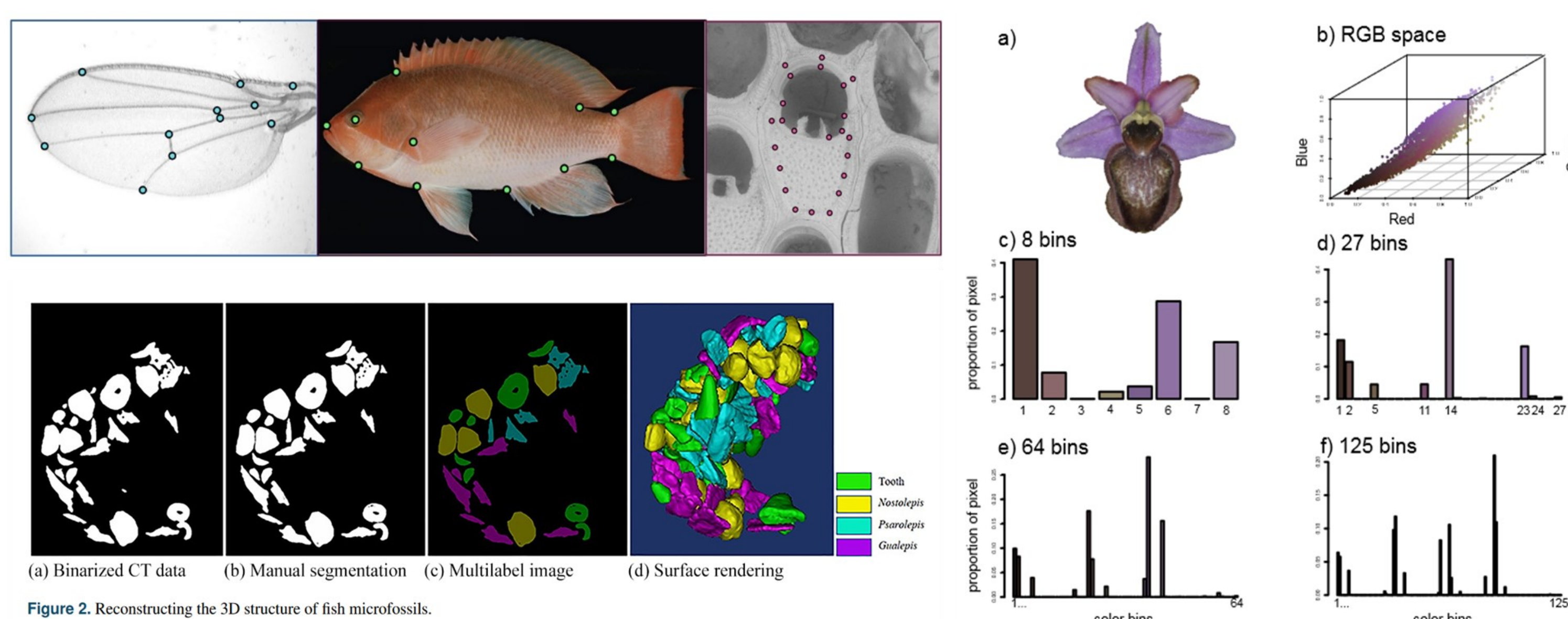


Various cross validation and summary statistic-based tests can be used to evaluate model performance

Simulations



Simulated images from the TraitBlender pipeline



Evolutionary processes are high-dimensional and have many potential confounders