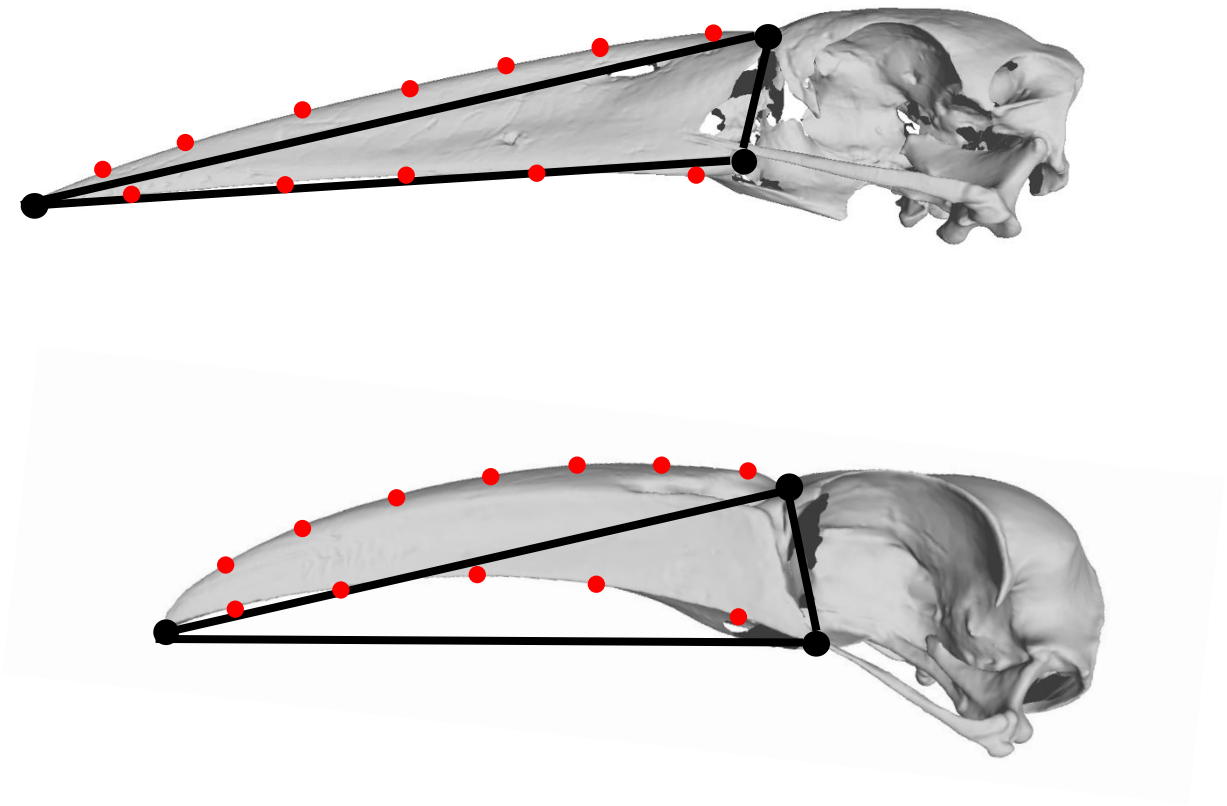
Advanced Shape Analysis

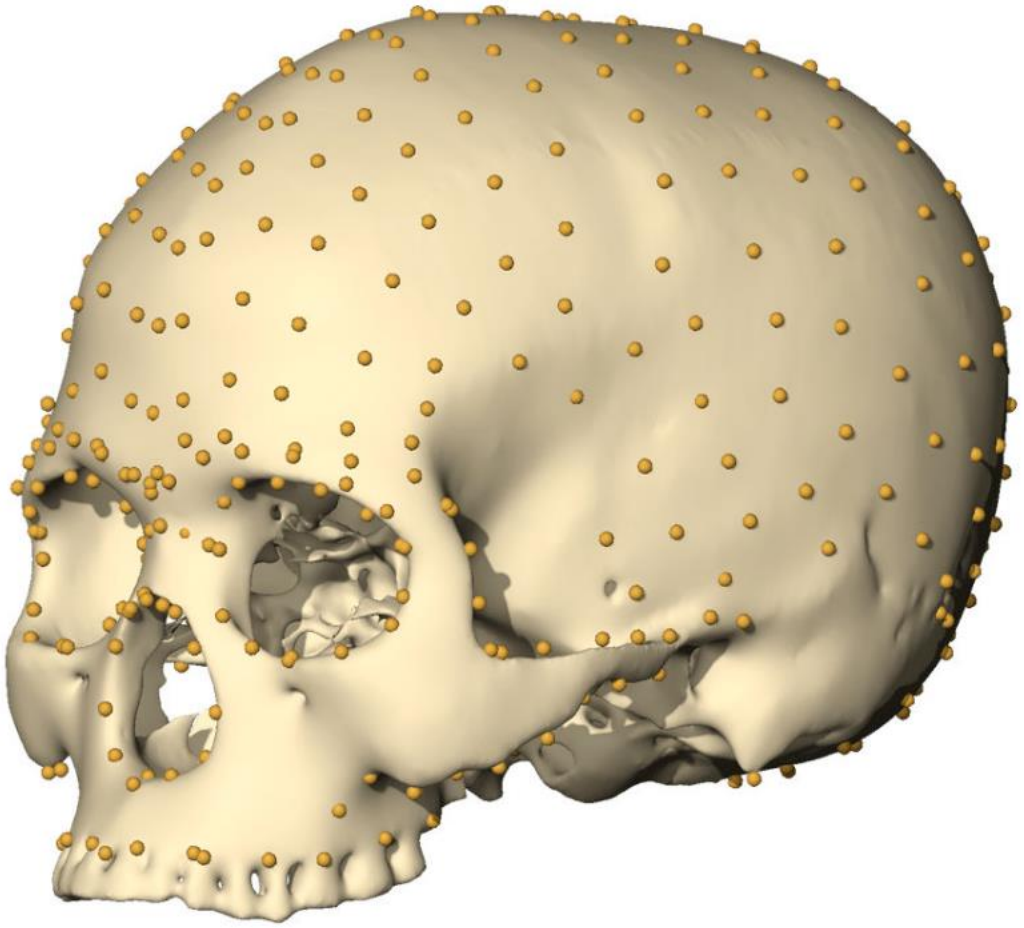
Ryan N. Felice



Sliding Semilandmarks



Surface Semi-landmarks



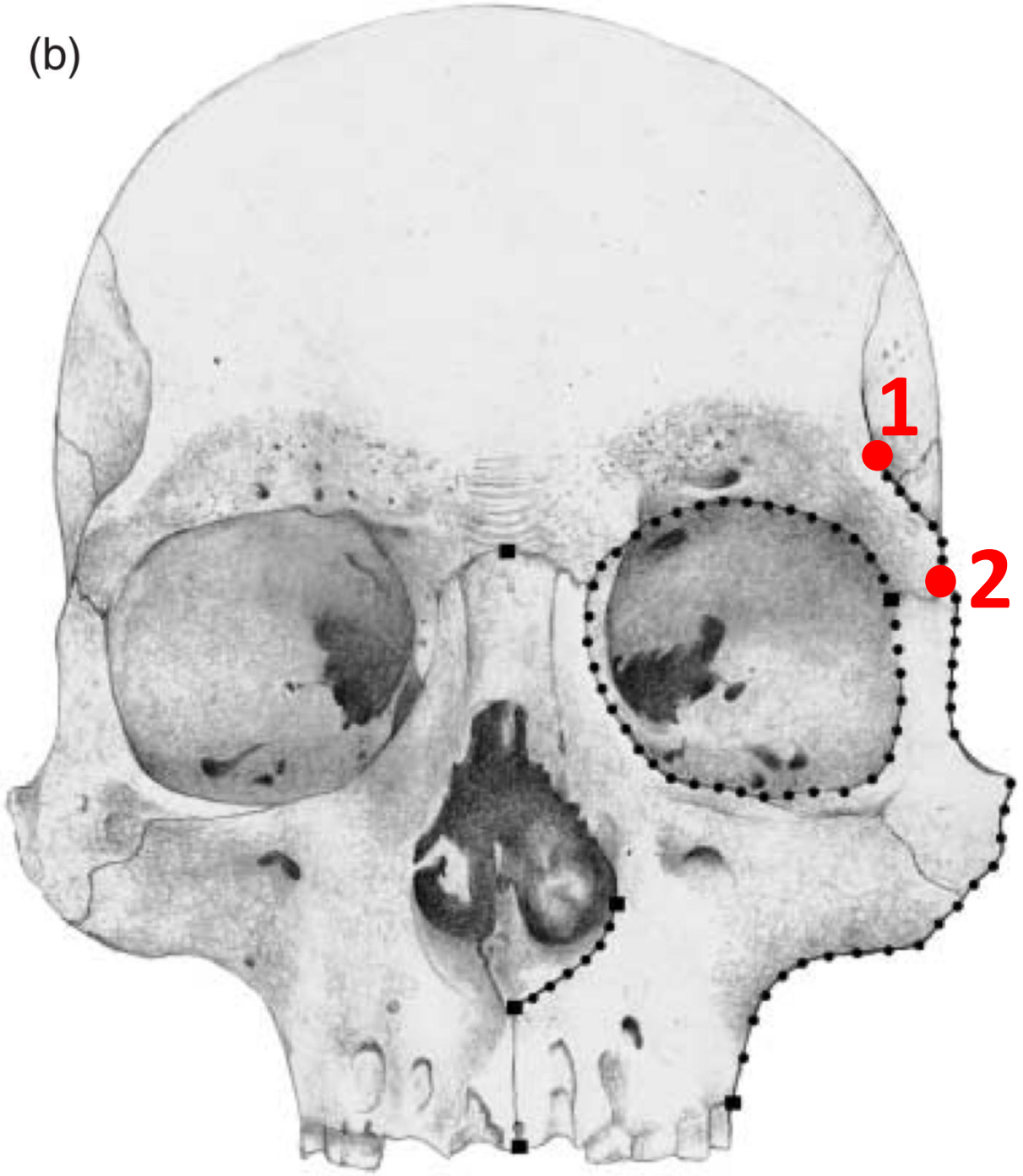
Gunz and Mitteroecker 2013

Sliding Semilandmarks



Perez et al 2006

Sliding Semilandmarks

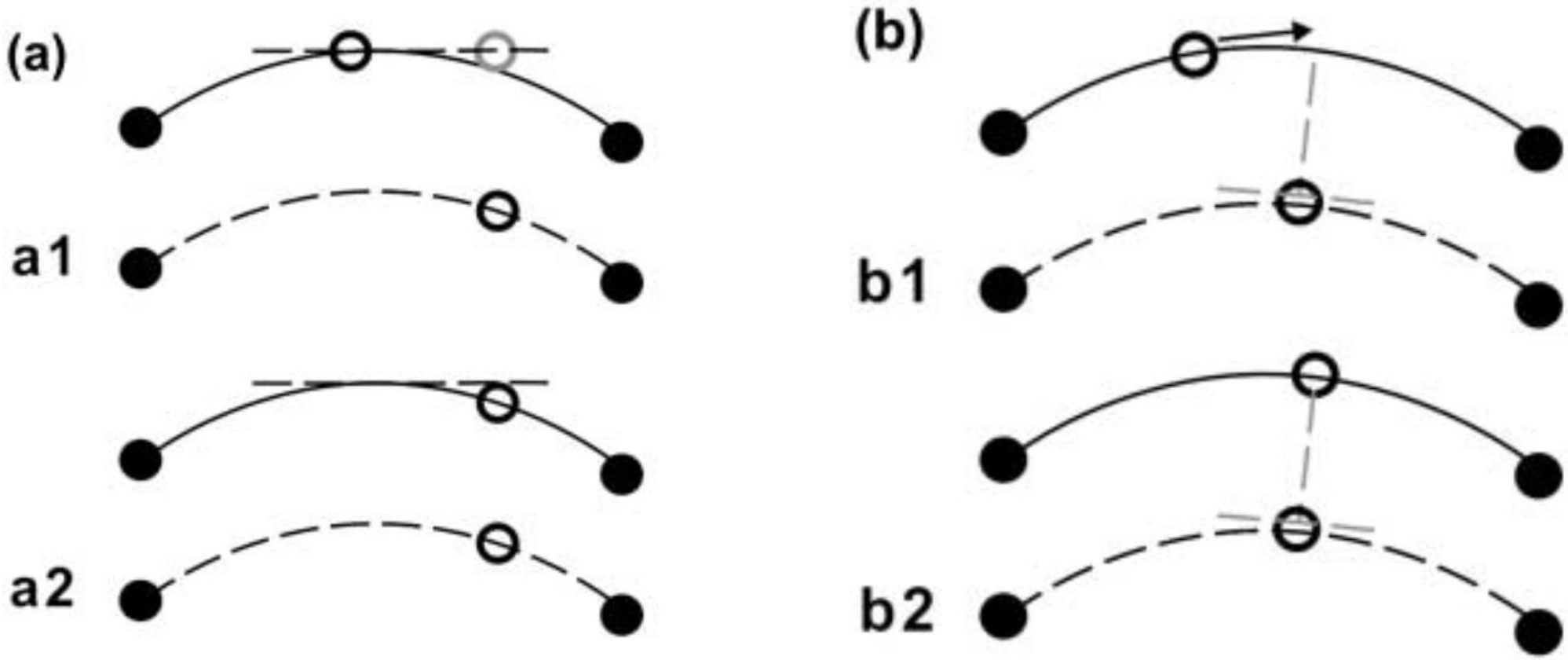


Perez et al 2006

Sliding Semilandmarks

Bending Energy

Procrustes Distance

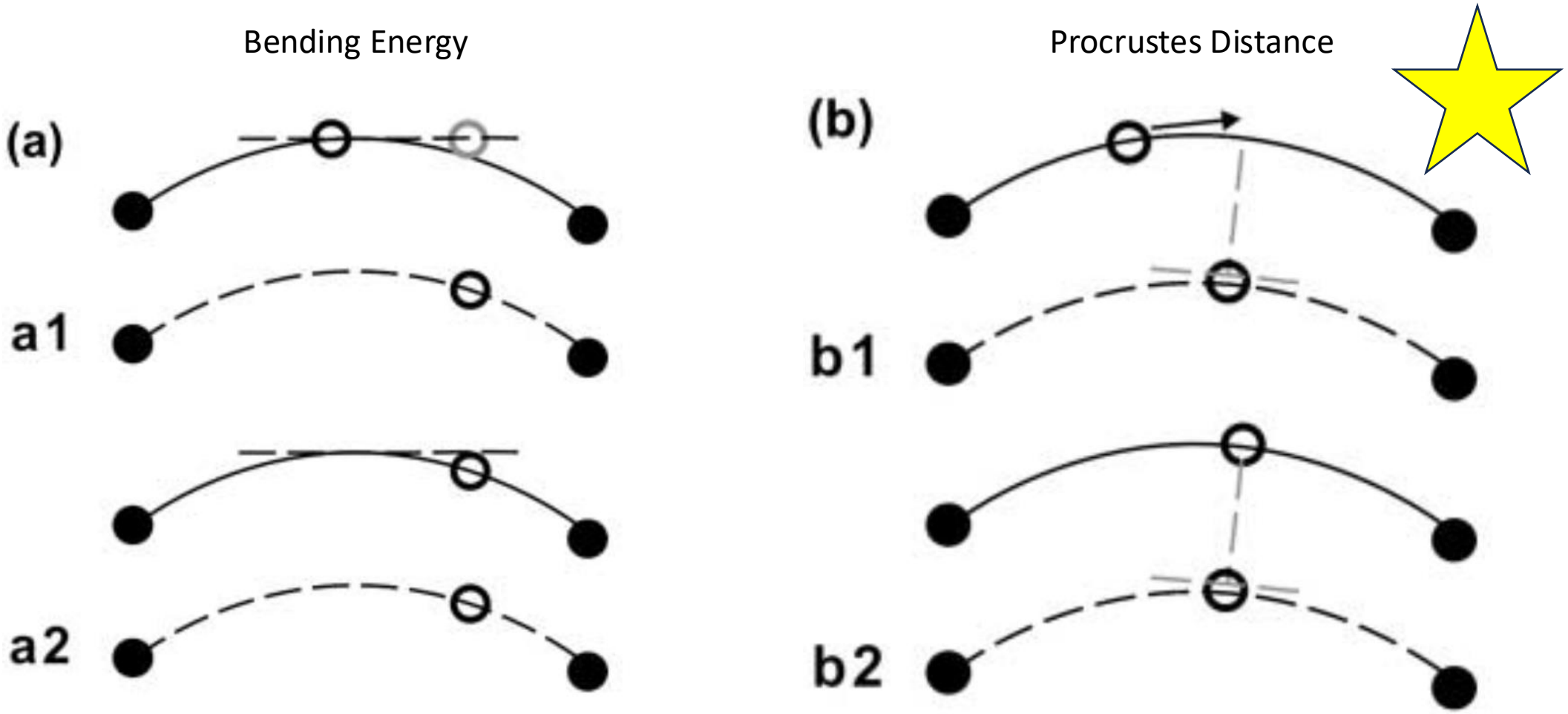


Perez et al 2006

Sliding Semilandmarks

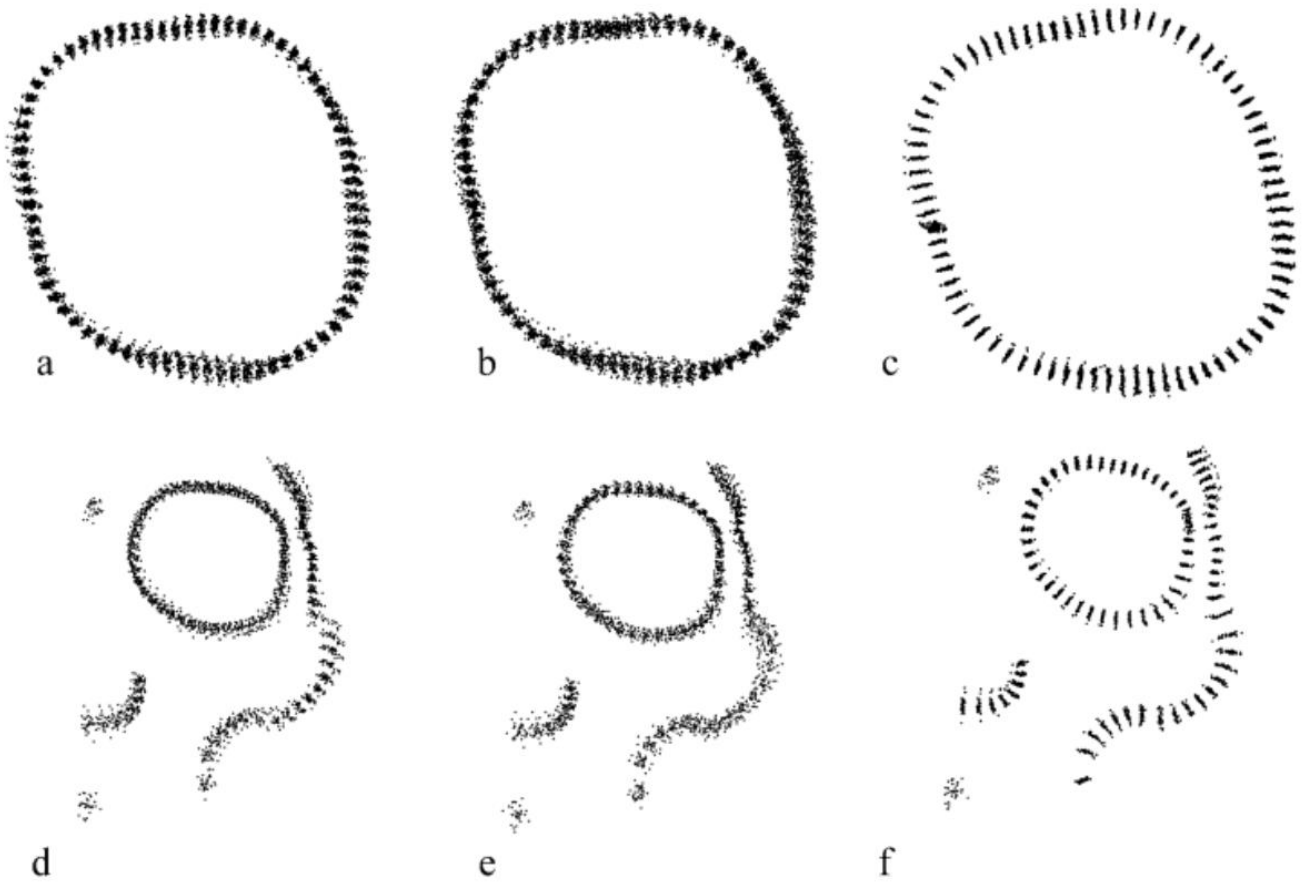
Bending Energy

Procrustes Distance

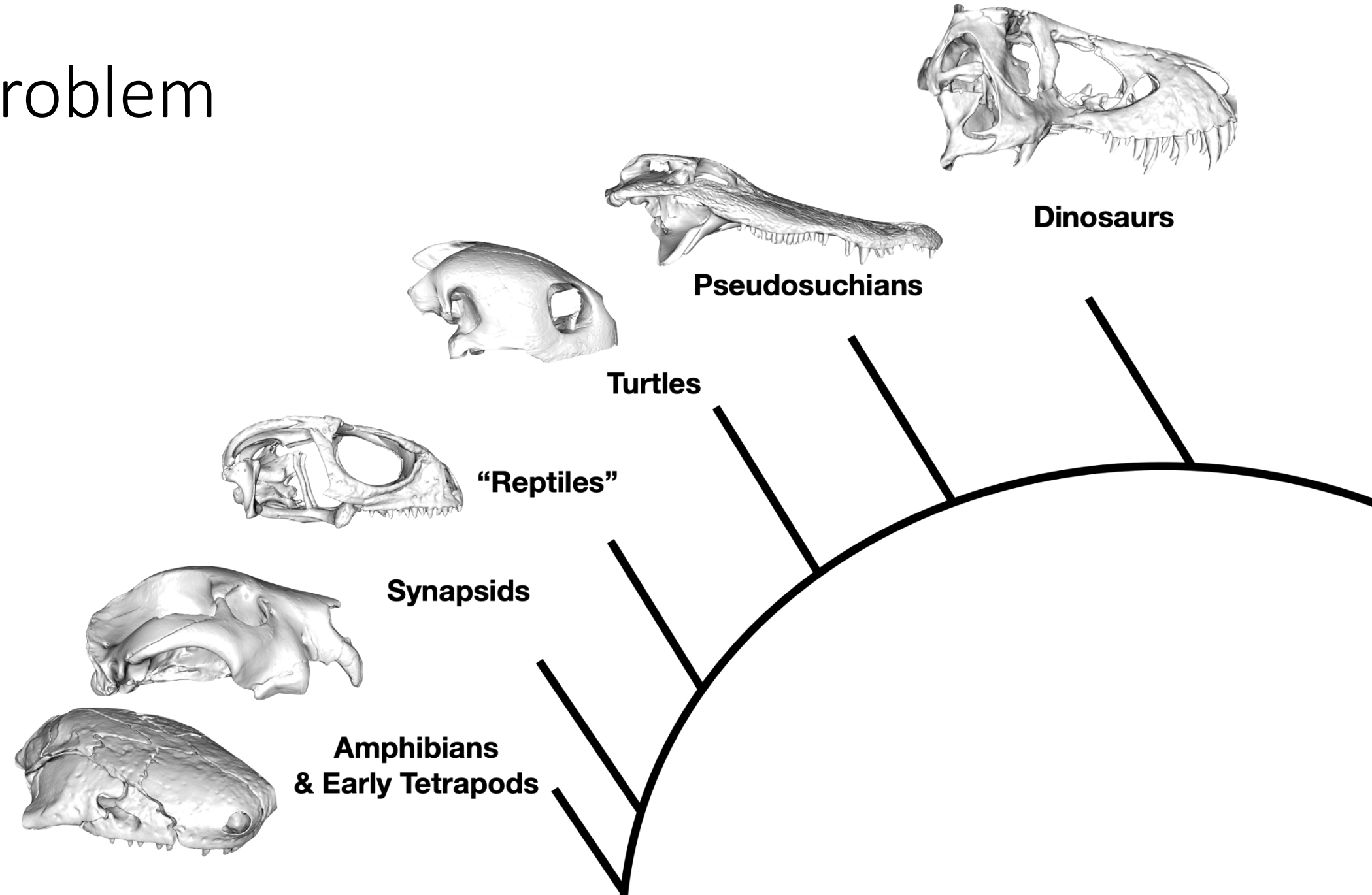


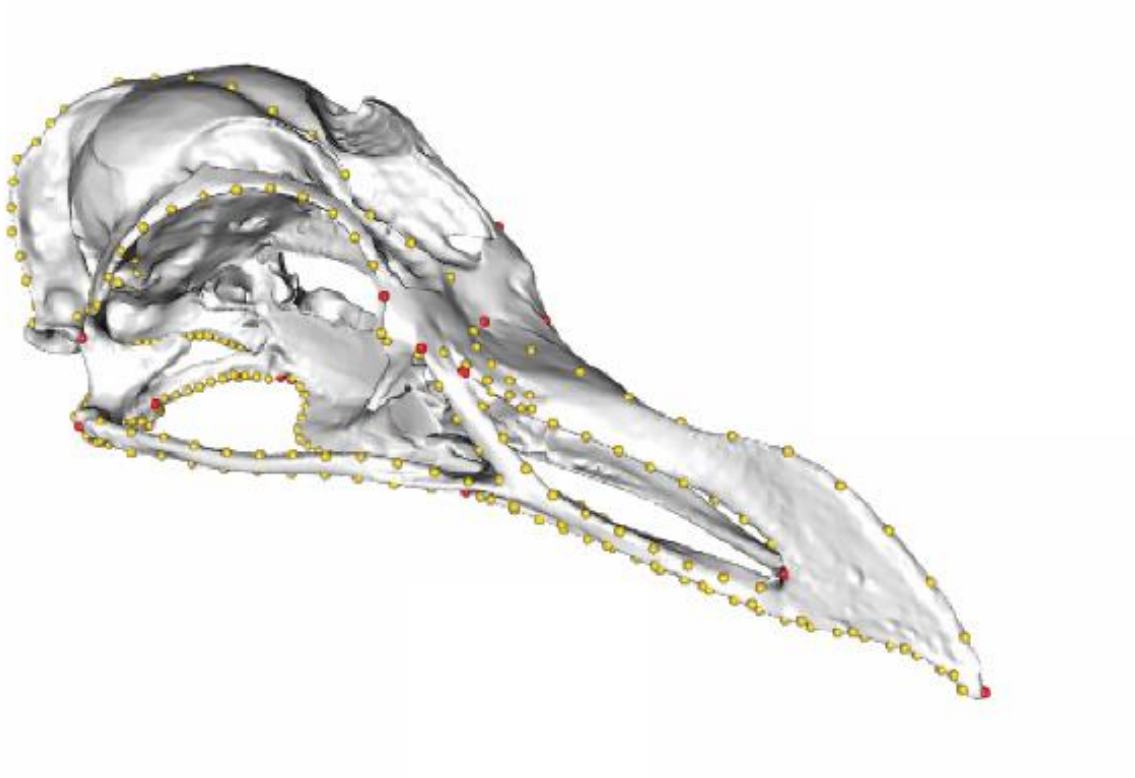
Perez et al 2006

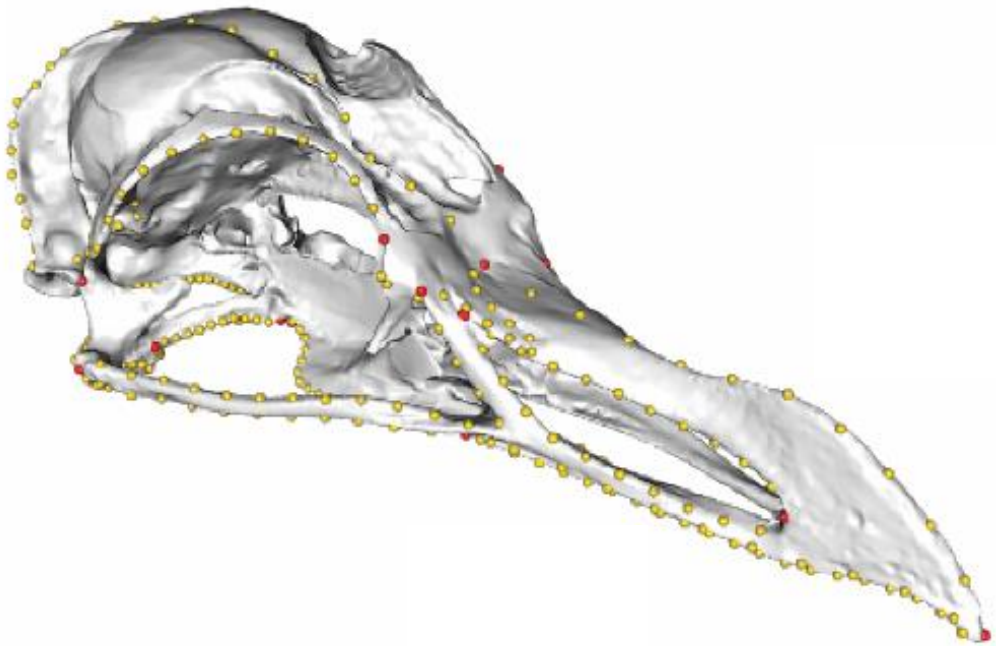
Sliding Semilandmarks

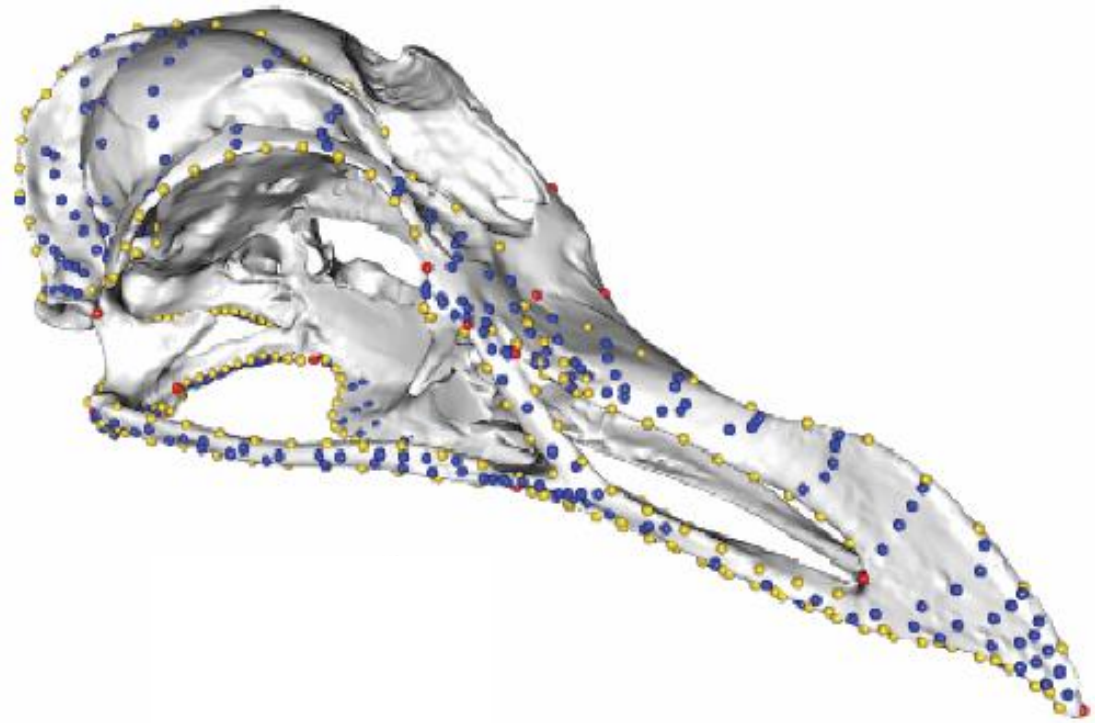


The Problem

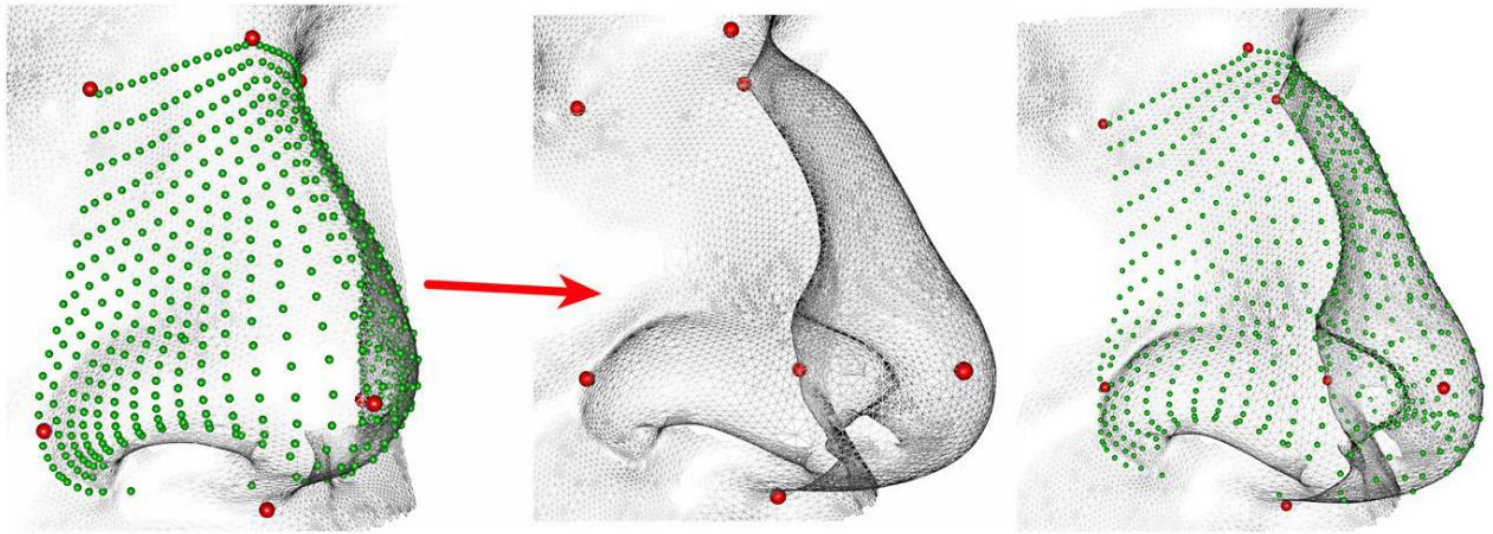

Dinosaurs
Pseudosuchians
Turtles
“Reptiles”
Synapsids
Amphibians
& Early Tetrapods





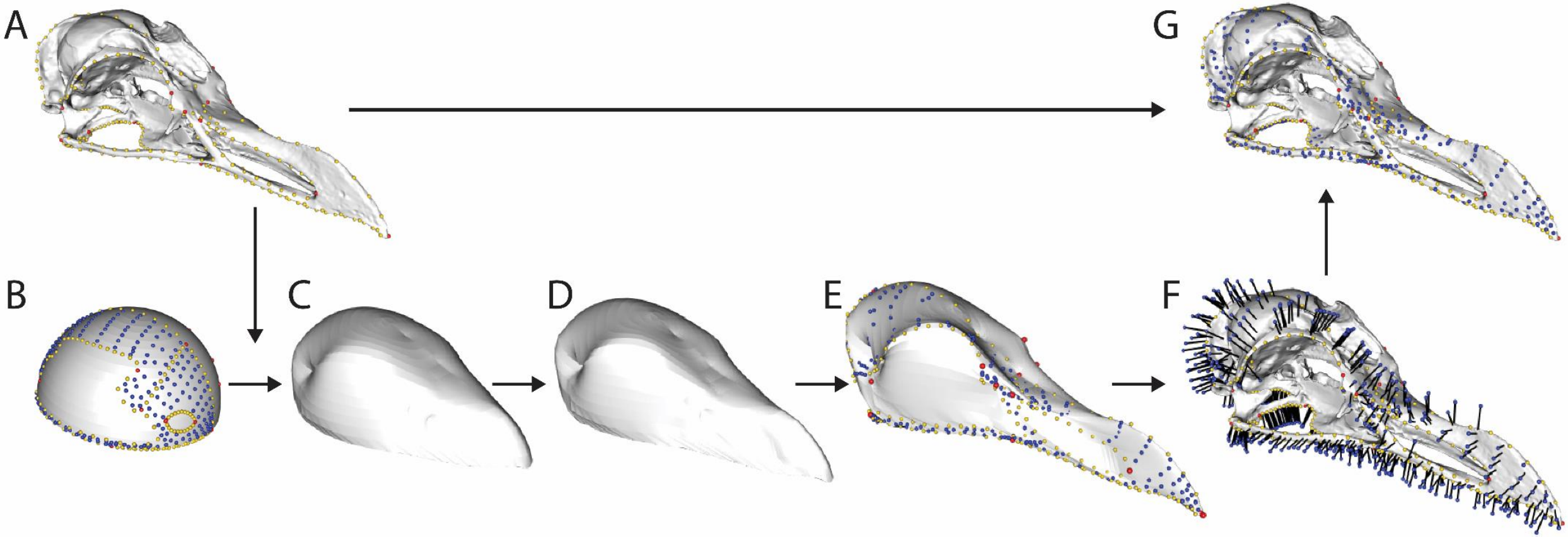


Semi-automated patching

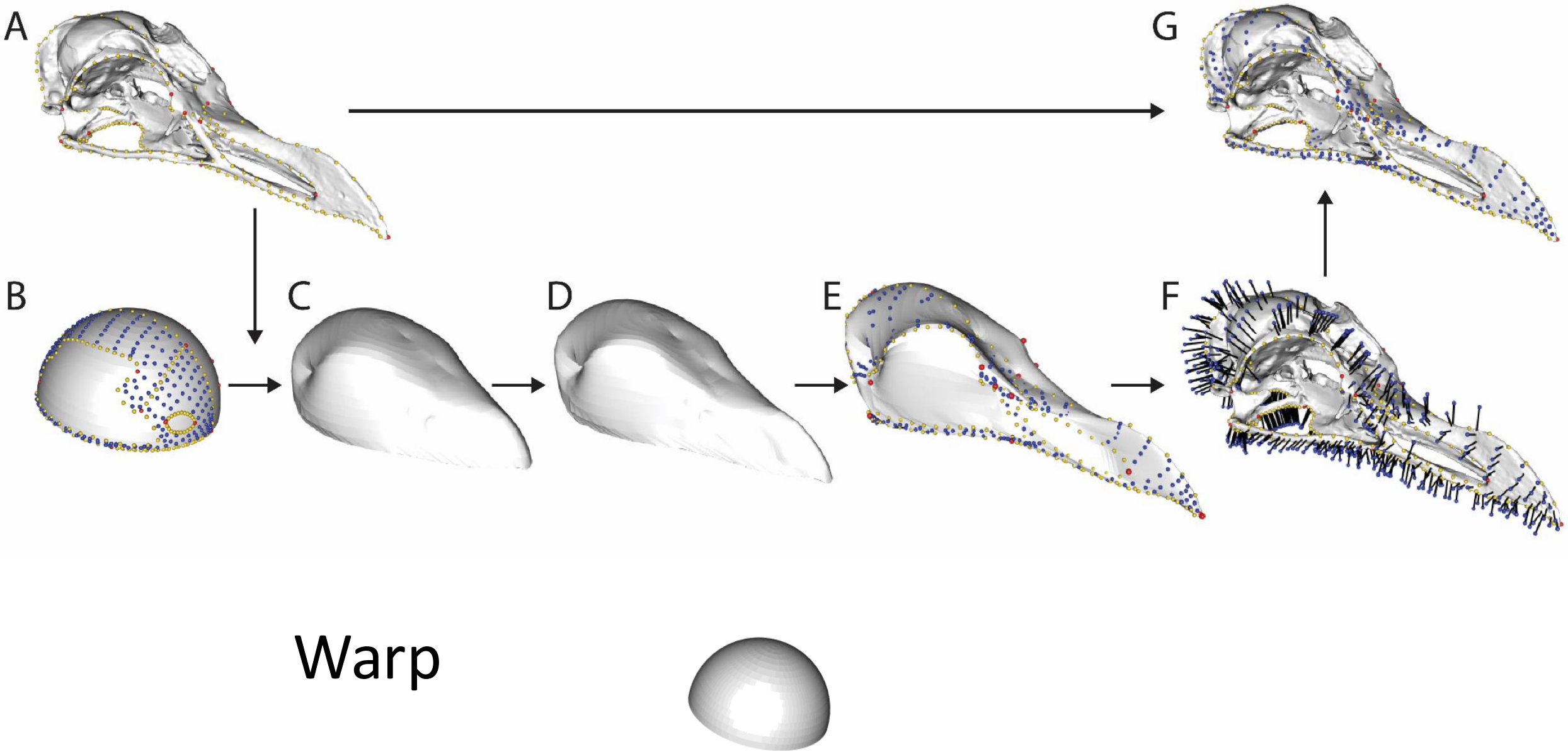


Morpho R package

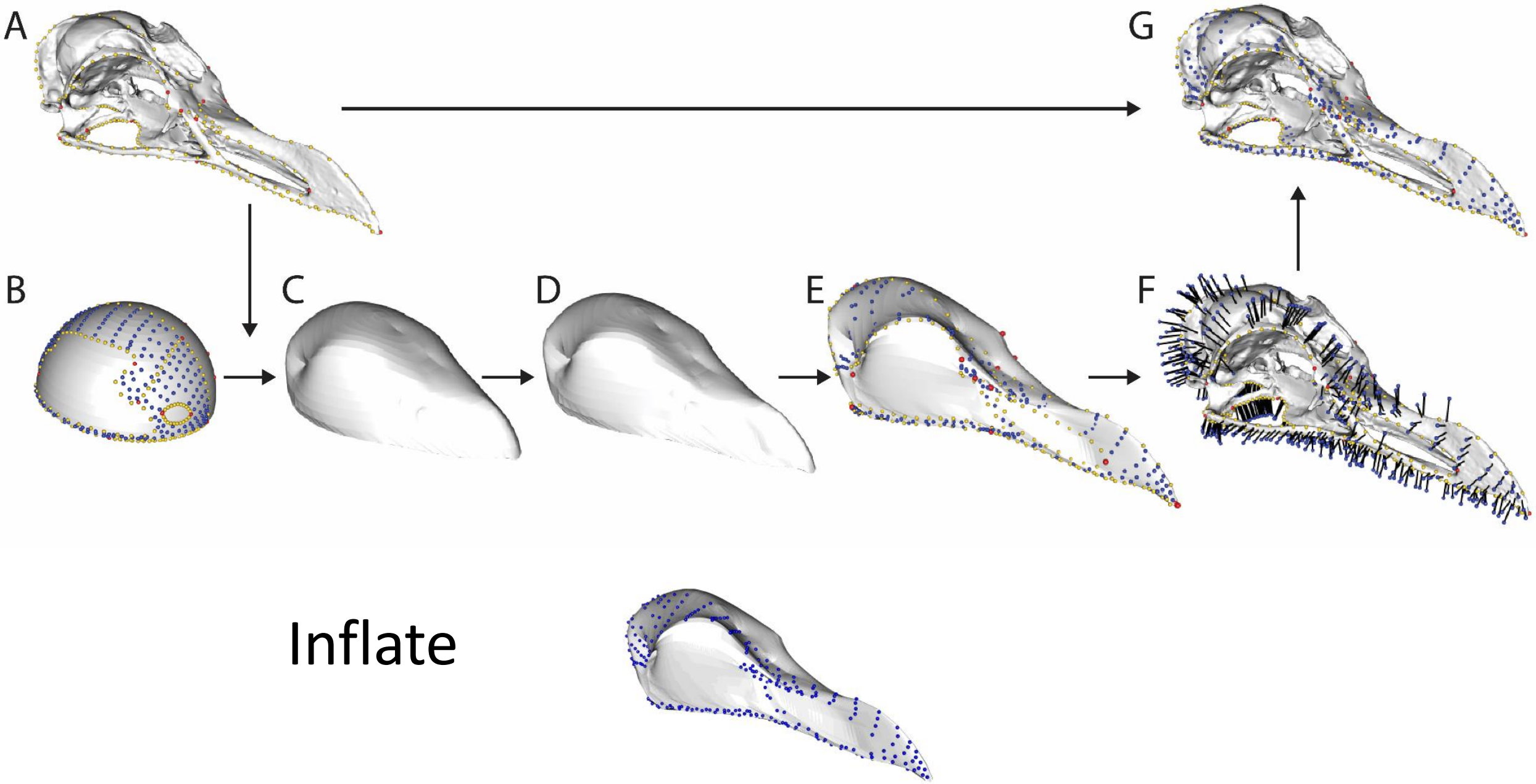
Schlager 2017



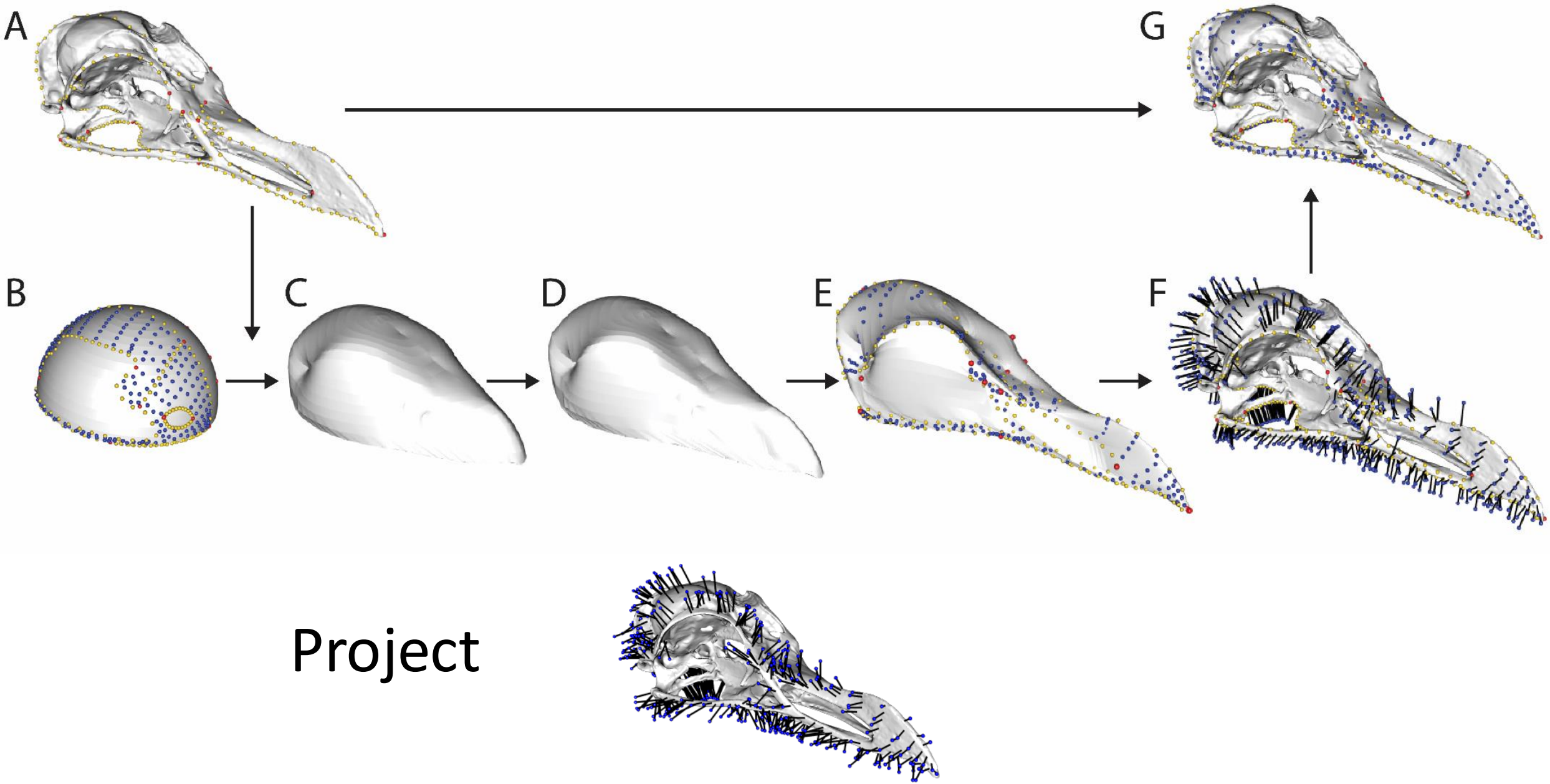
Bardua et al 2019



Bardua et al 2019

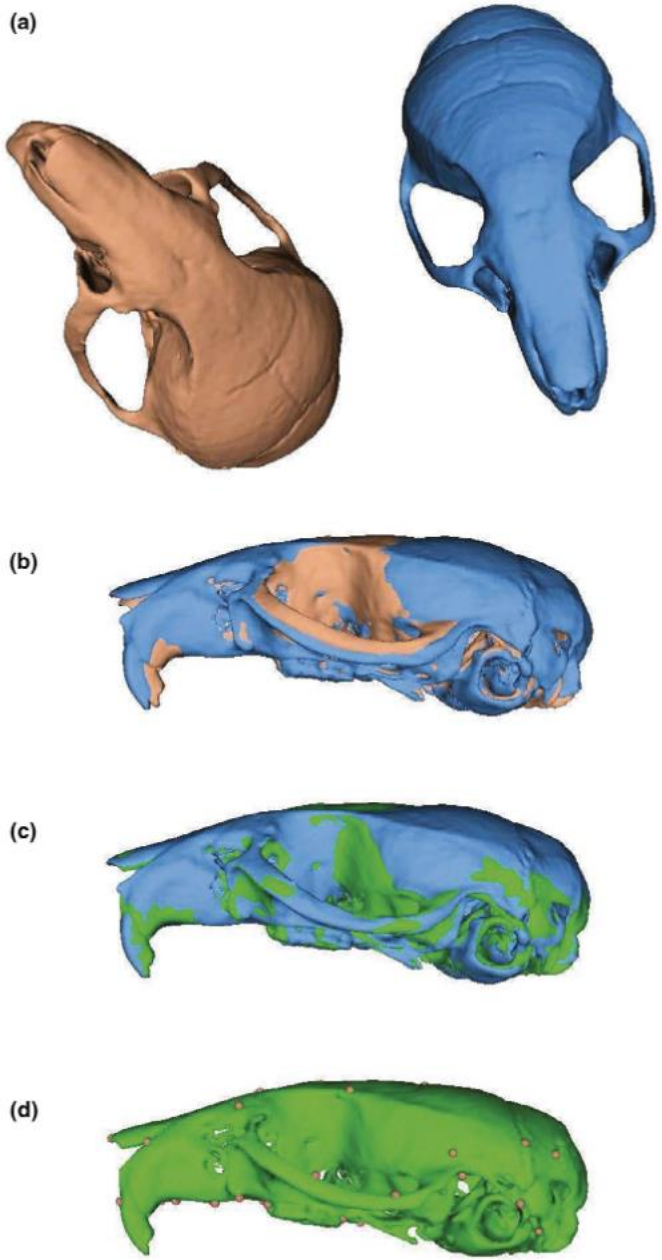


Bardua et al 2019



Project

Bardua et al 2019



Accepted: 11 July 2021

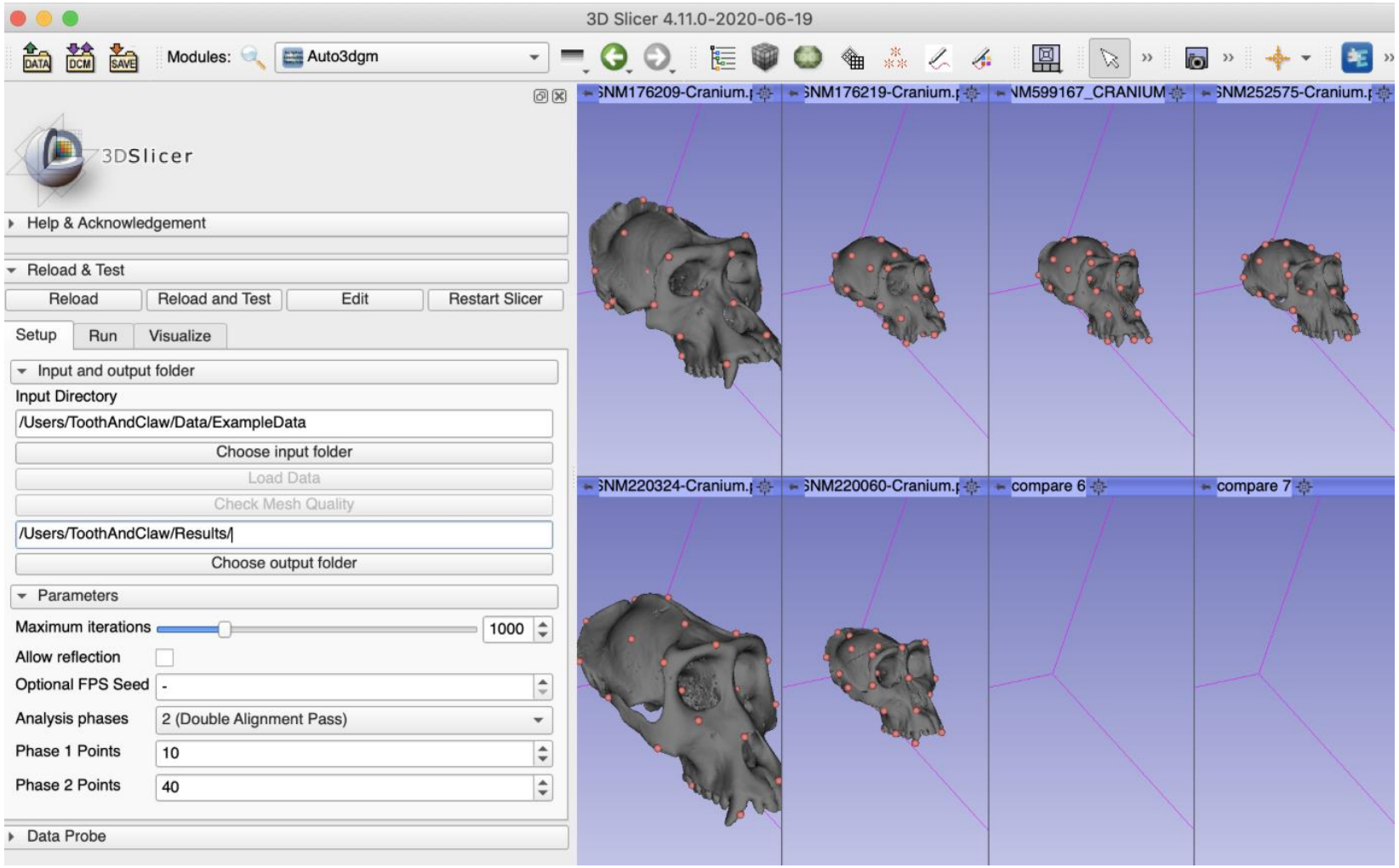
DOI: 10.1111/2041-210X.13689

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

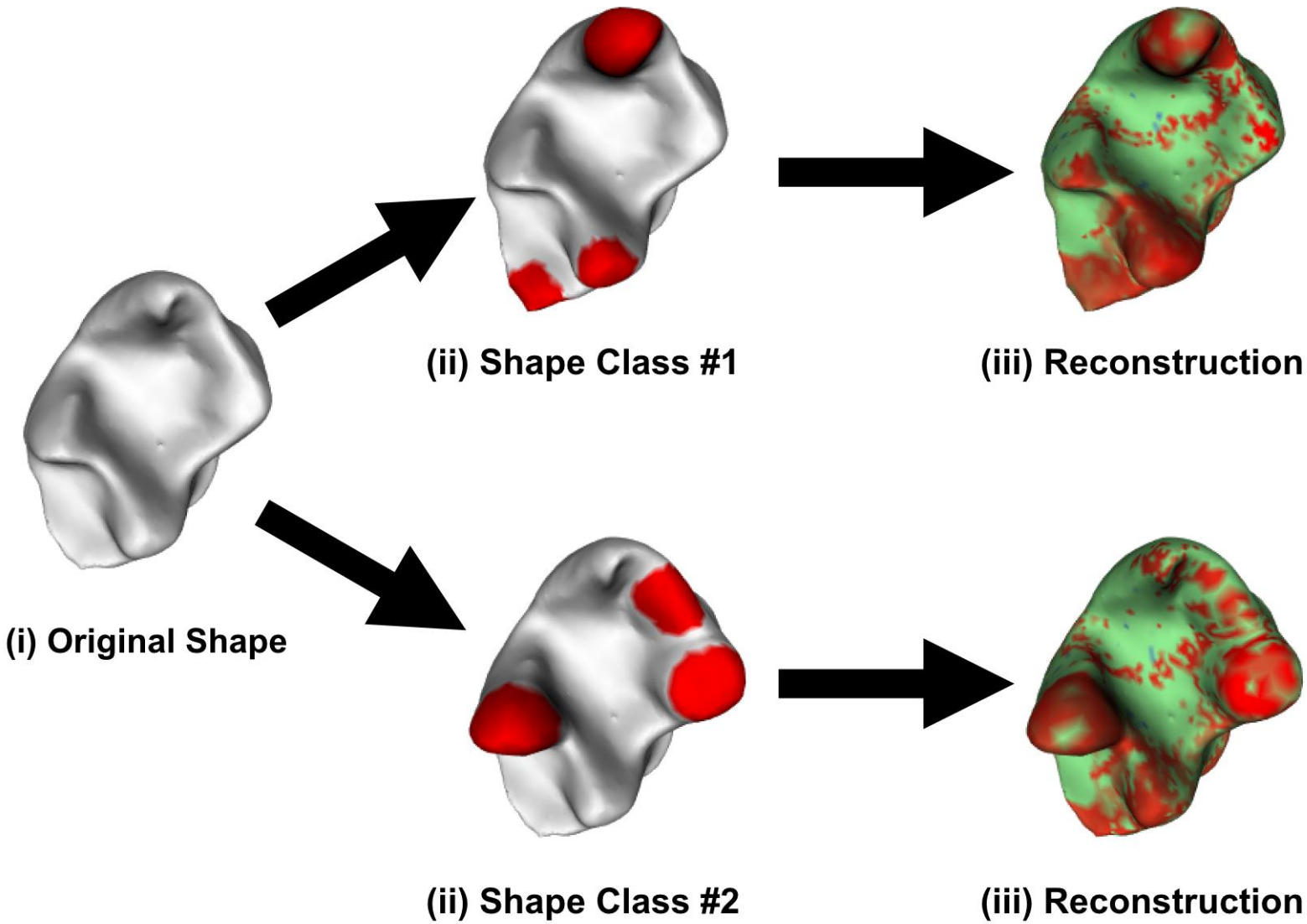
ALPACA: A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures

Arthur Porto[1,2] | Sara Rolfe[3,4] | A. Murat Maga[4,5]



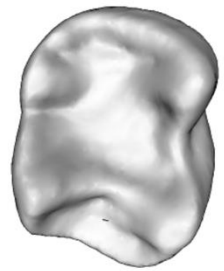
Pseudolandmarks-Auto3dGm

https://toothandclaw.github.io


(i) Original Shape
(ii) Shape Class #1
(iii) Reconstruction
(ii) Shape Class #2
(iii) Reconstruction

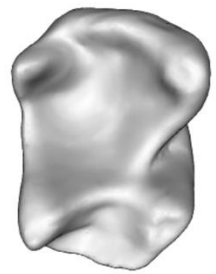
(a) Input 3D Shapes

Data from Species #1



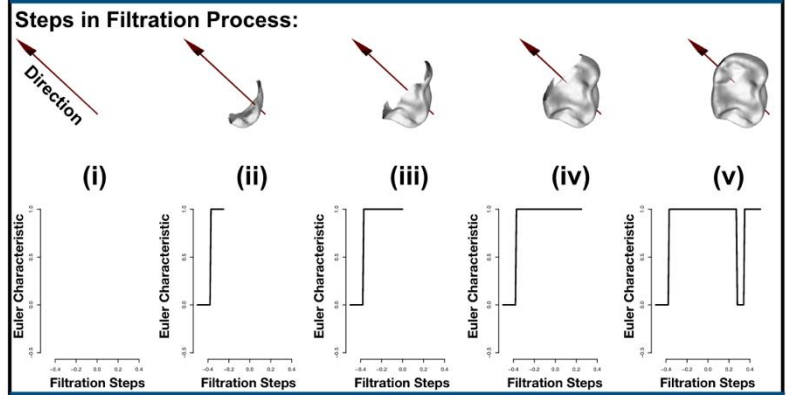
Outcome: $\mathbf{y}_i = 0$

Data from Species #2

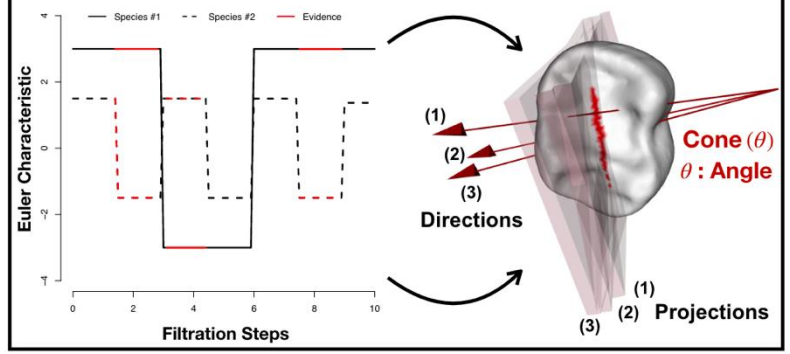


Outcome: $\mathbf{y}_i = 1$

(b) Derive Topological Summary Statistics

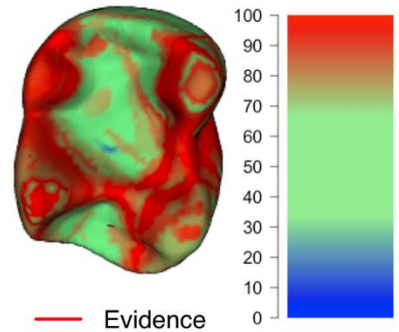


(c) Variable Selection and Reconstruction

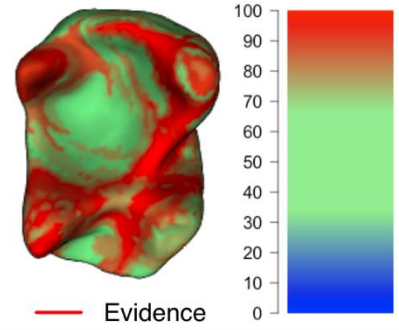


(d) Visualize Enrichment

Enrichment in Species #1



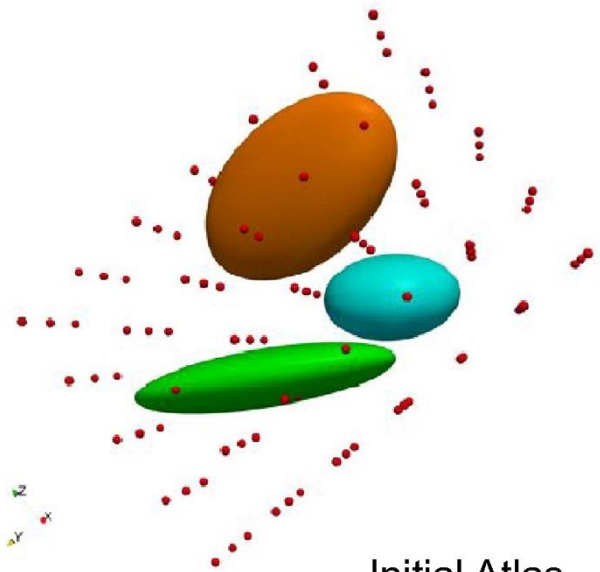
Enrichment in Species #2



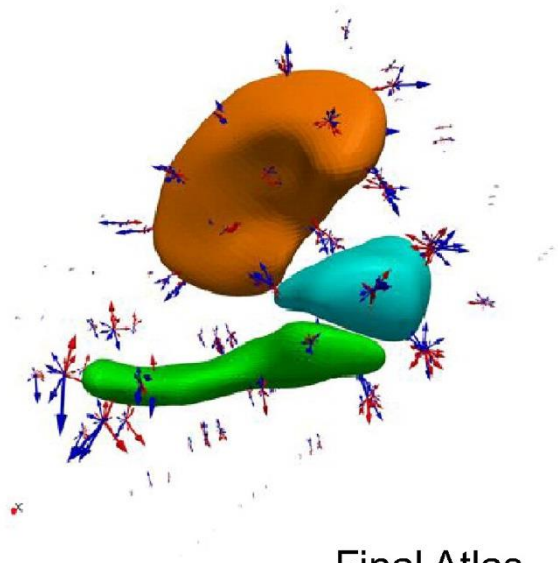
A Statistical Pipeline for Identifying Physical Features that Differentiate Classes of 3D Shapes

Bruce Wang[1,2,*], Timothy Sudijono[3,*], Henry Kirveslahti[4,*], Tingran Gao[5], Douglas M. Boyer[6], Sayan Mukherjee[4,7-9,†], and Lorin Crawford[10,†]

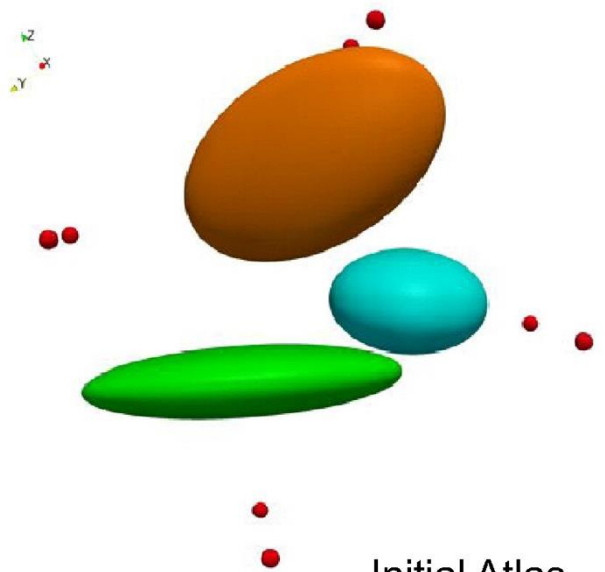
Deterministic Atlas Analysis



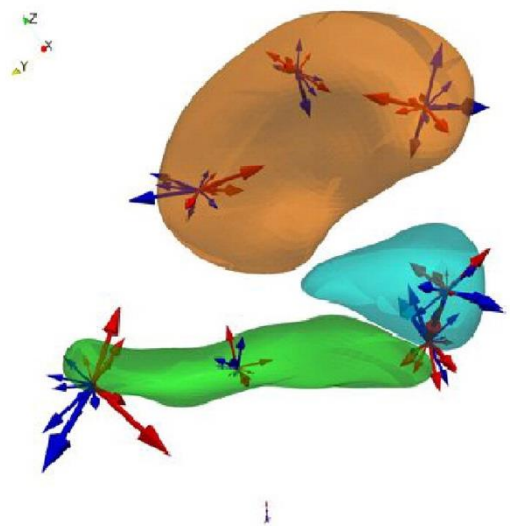
Initial Atlas



Final Atlas

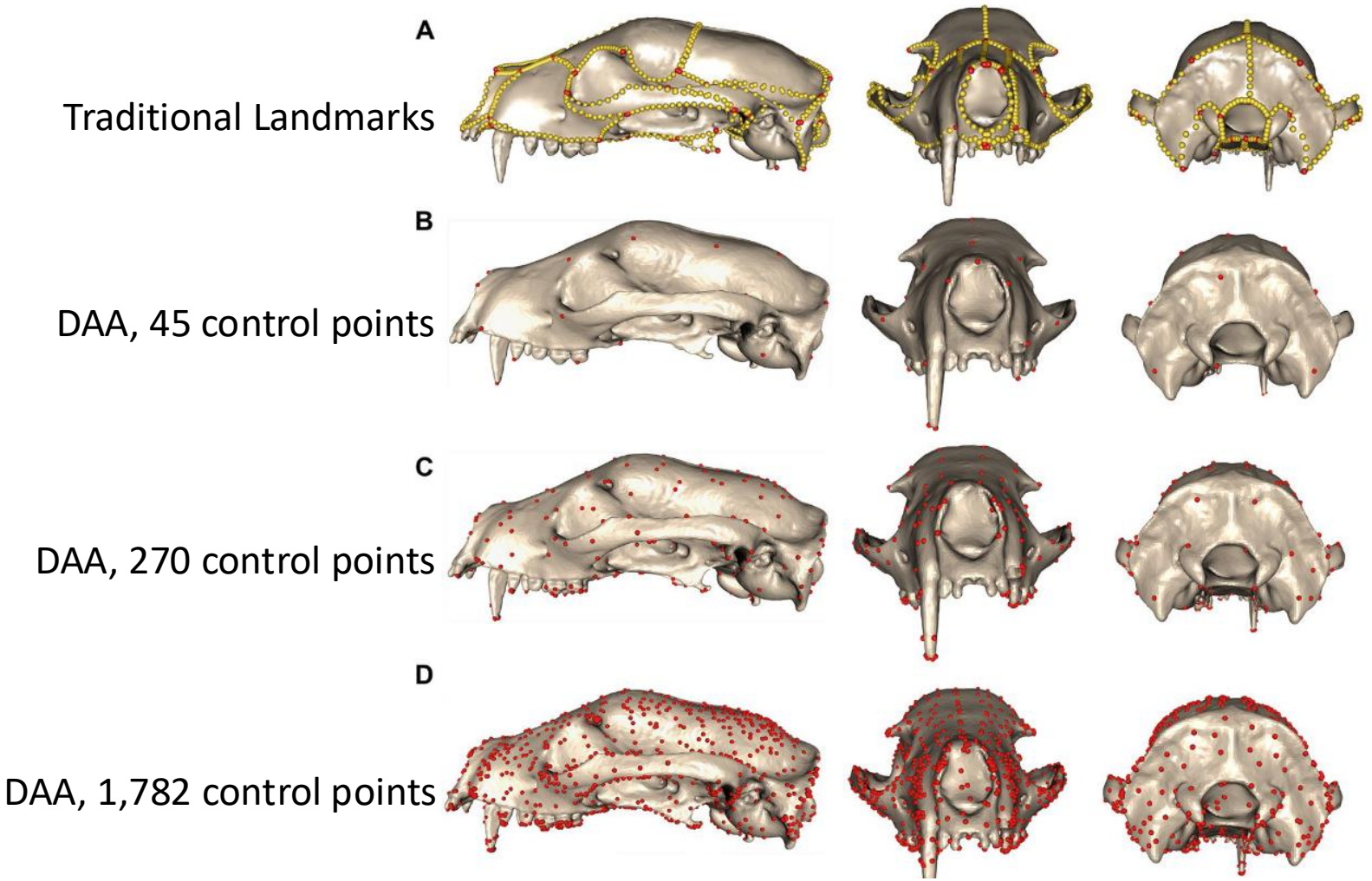


Initial Atlas



Final Atlas

Deterministic Atlas Analysis

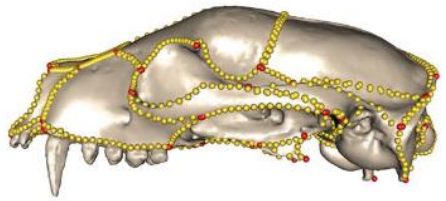


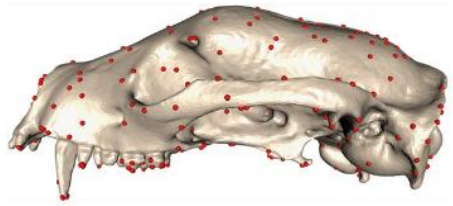
RESEARCH Open Access

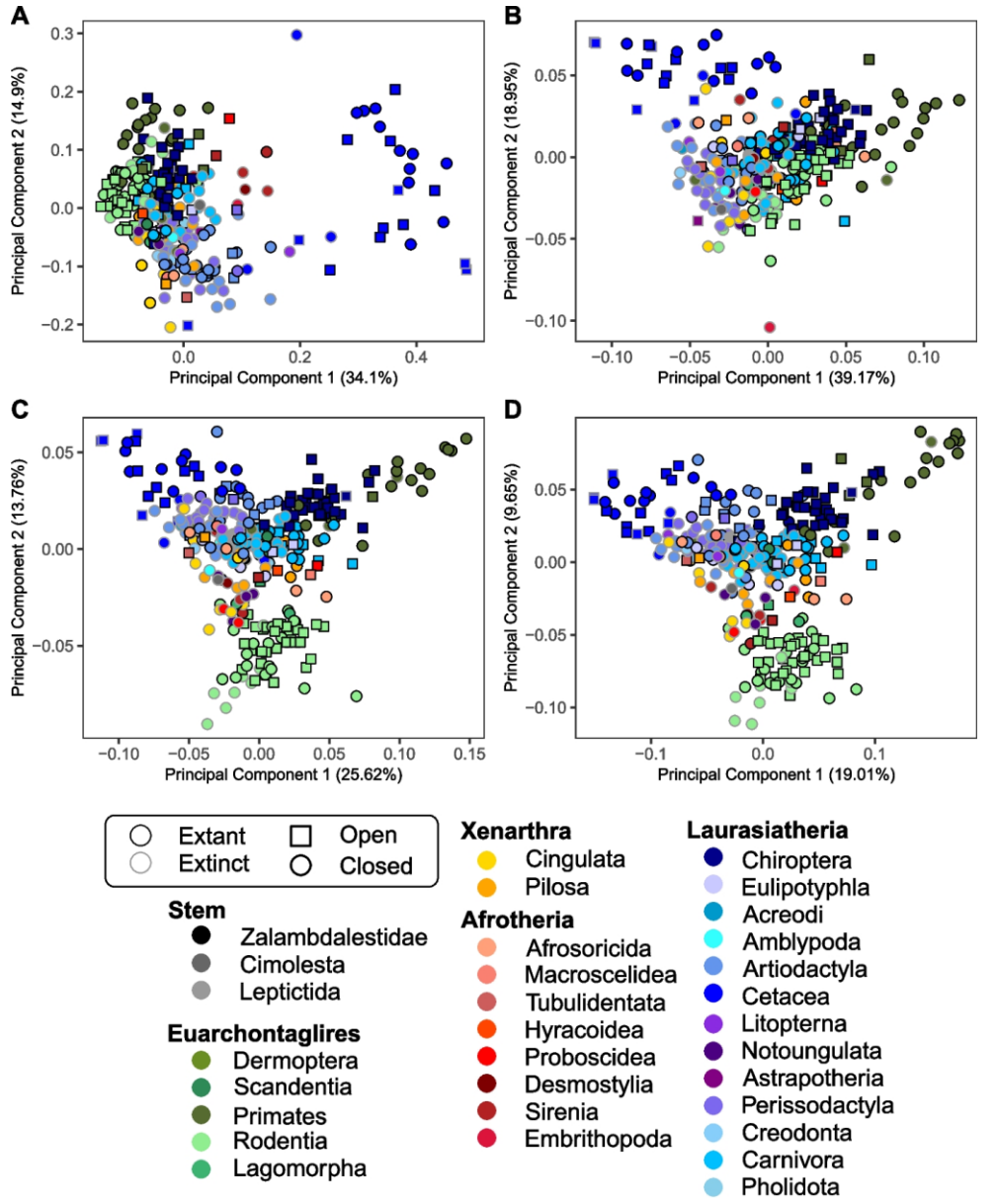
Assessing the application of landmark-free morphometrics to macroevolutionary analyses

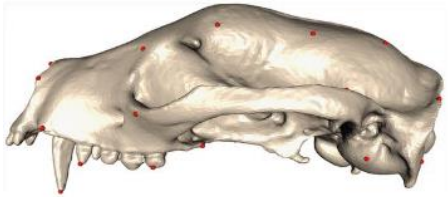
James M. Mulqueeney[1,2*], Thomas H. G. Ezard[1] and Anjali Goswami[2]

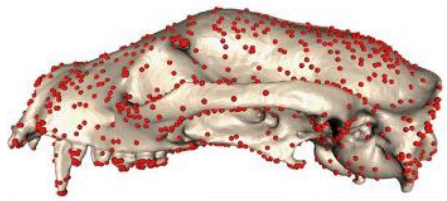
Deterministic Atlas Analysis











How much firepower do I need?



VS.



Modelling High Dimensional Data

- What do we mean by high dimensional data?
- N = number of observations
- P = number of variables

	Trait 1	Trait 2	Trait 3
Species 1			
Species 2			
Species 3			
Species 4			
Species 5			
Species 6			
Species 7			

	Trait 1	Trait 2	Trait 3	Trait 4	Trait 5	Trait 6
Species 1						
Species 2						
Species 3						
Species 4						

Modelling High Dimensional Data

- Who cares?
- Everything is linear models
- Estimating F-statistic and likelihood of linear models with p>N is impossible (you can't invert a matrix that has p>N)
- "Curse of dimensionality"

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma$$

$$F = \frac{\widehat{\boldsymbol{\beta}}^{\top}(\boldsymbol{X}^{\top}\boldsymbol{X})\widehat{\boldsymbol{\beta}}/p}{\widehat{\sigma}^2}$$

The Solution:

- RRPP: Residual randomization in permutation procedure
- F-statistic (and significance) is computed by randomizing the residuals of the model and comparing to a null

$$F = \frac{MS_{Model}}{MS_{residuals}}$$

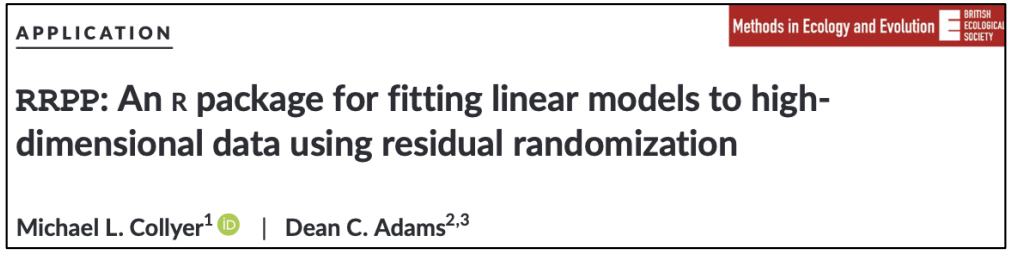
MS = Mean squares

The Solution:

- RRPP: Residual randomization in permutation procedure
- F-statistic (and significance) is computed by randomizing the residuals of the model and comparing to a null

Recommendations:

1) Always think about what methods are appropriate for your data
2) Use procD.lm or procD.pgls functions in geomorph or lm.rrpp in the RRPP package



https://doi.org/10.1111/2041-210X.13029