

Mathematicians

Ingrid Daubechies*

Tingran Gao

Rob Ravier

Shahar Kovalsky

Nadav Dym

Barak Sober

Statisticians

Sayan Mukherjee*

Henry Kirveslahti

Biologists

Doug Boyer*

Julie Winchester

Ethan Fulwood

Automated Shape Correspondence: *Why and How*

Shan Shan

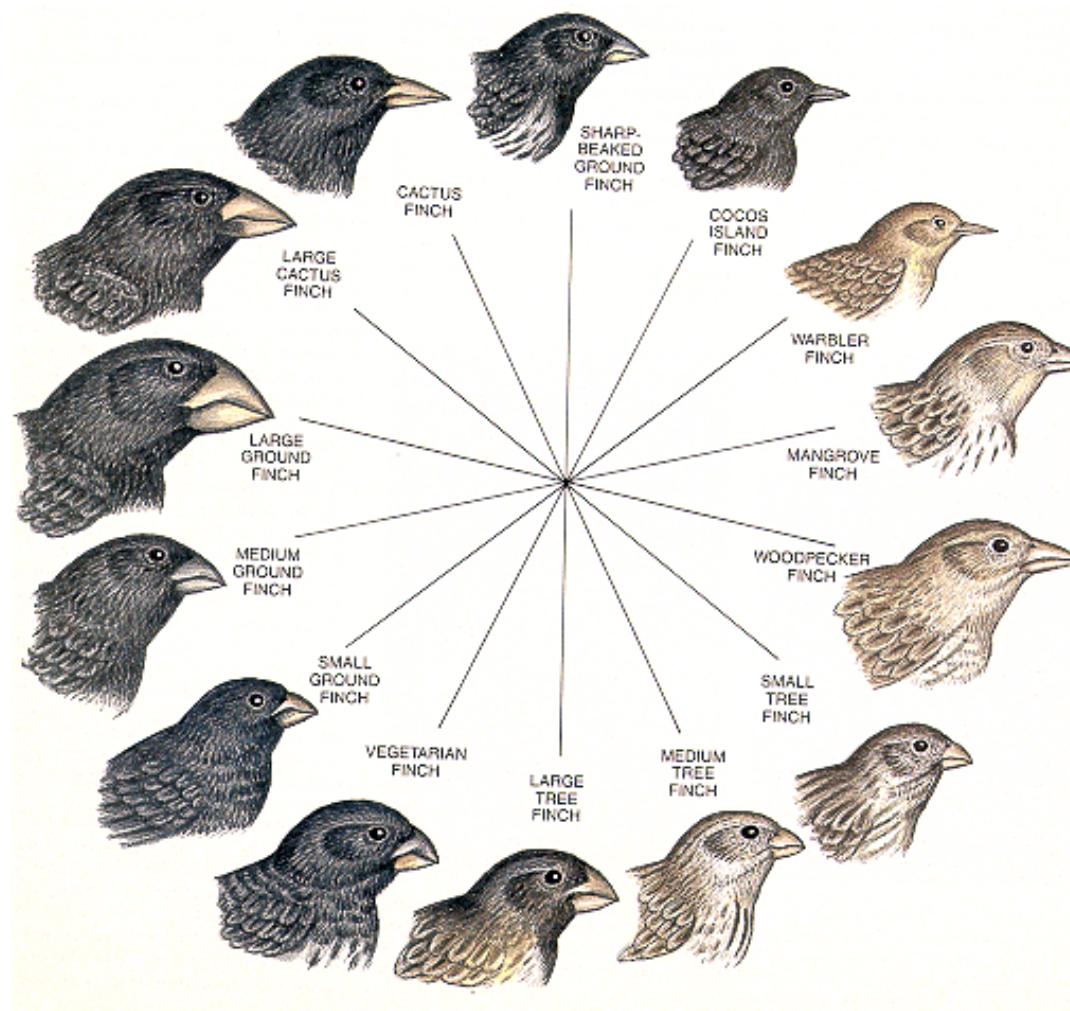
Department of Mathematics

Duke University



SLICERMORPH

Modeling variations in shapes



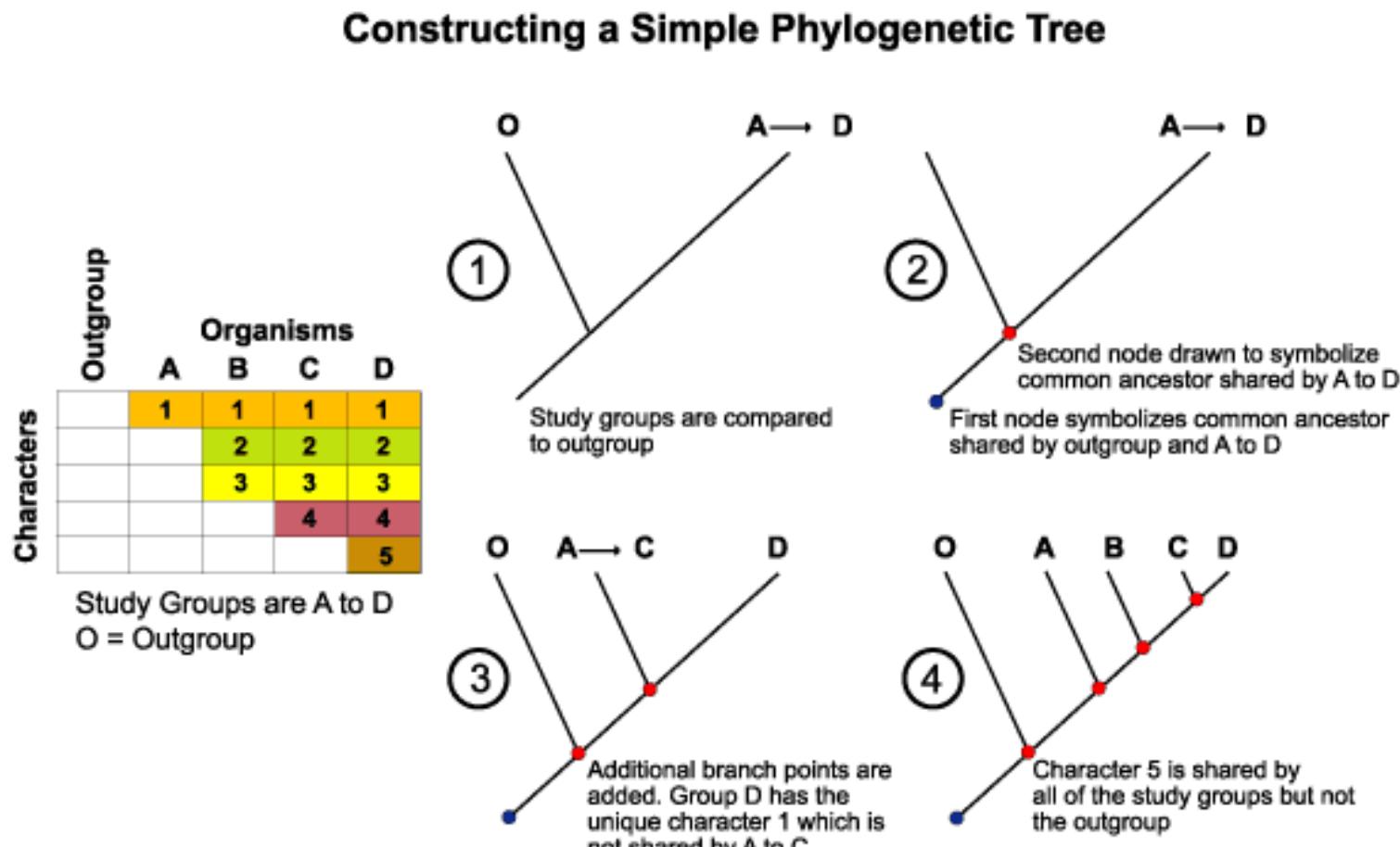
S. J. Gould

Variations in calcanei



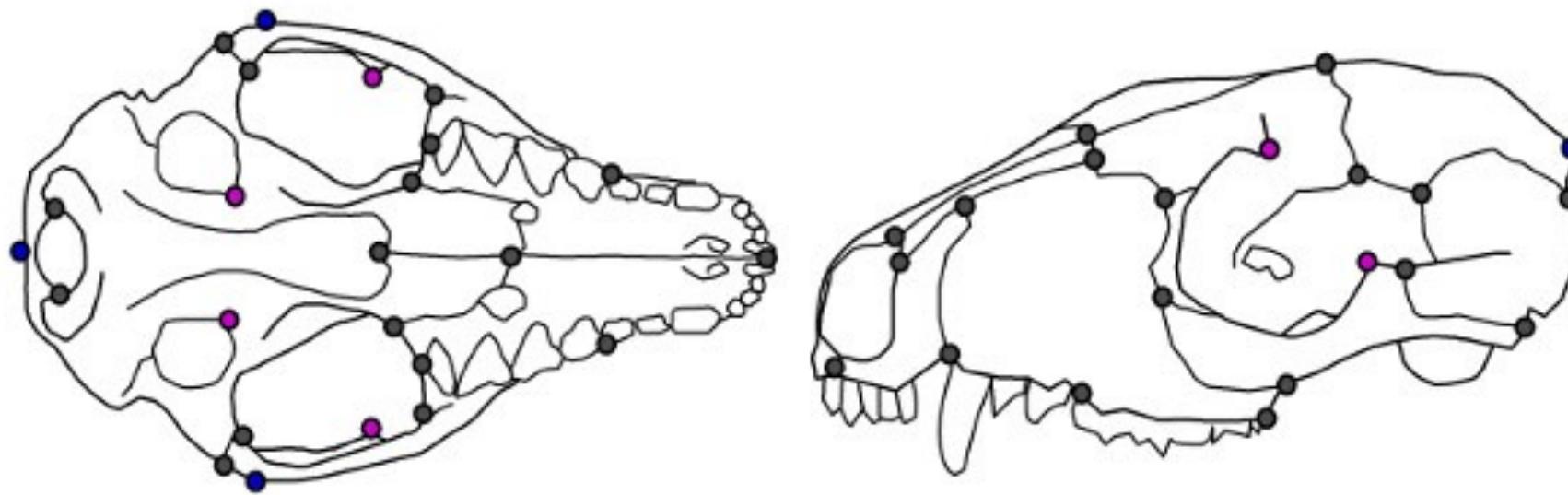
D. Boyer.

From distances to tree



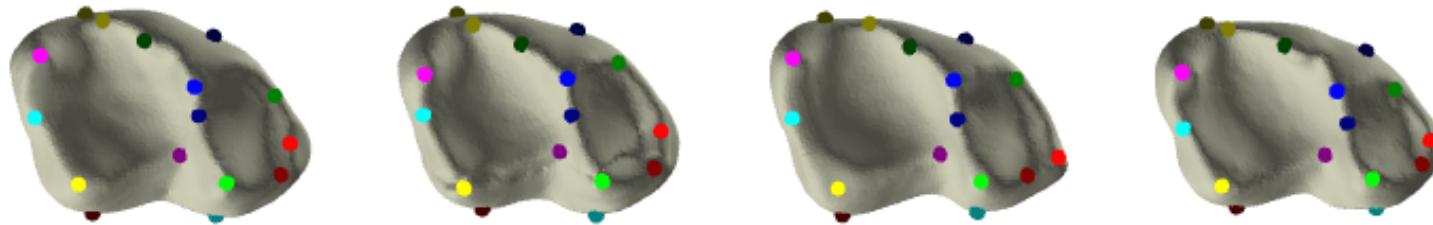
Adapted from Campbell "Biology" 4th Edition

Landmarks



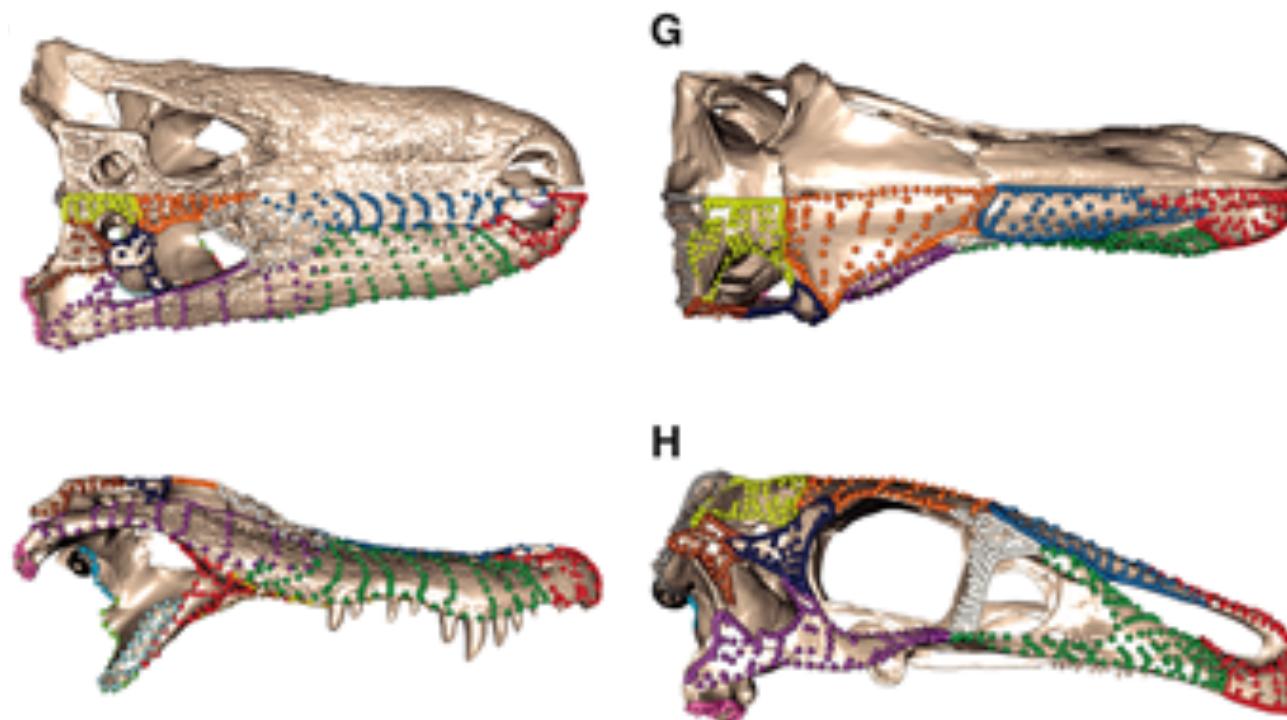
Landmarks

Landmarks are points of correspondence on each object that matches between and within populations.

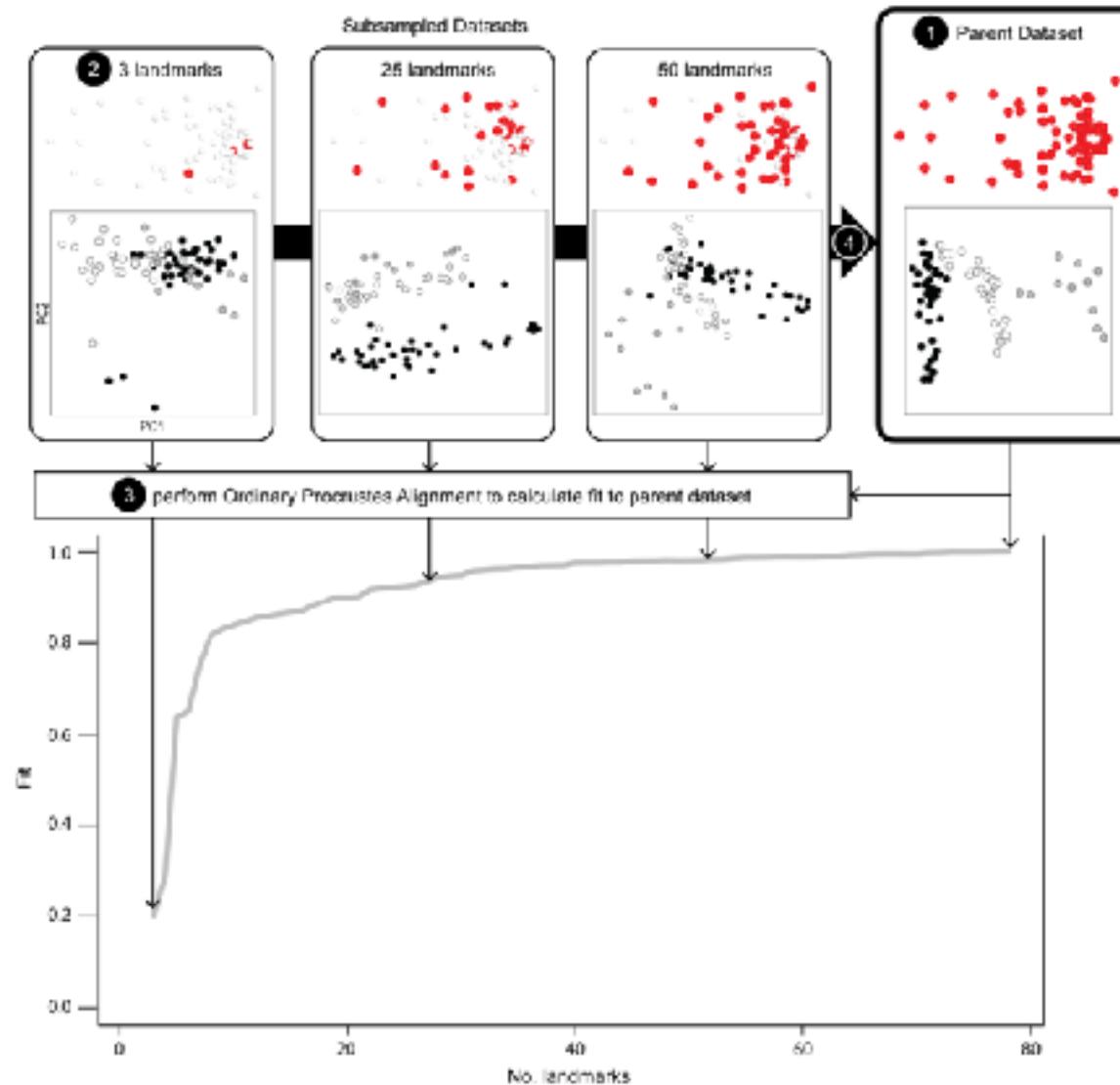


Limitations of traditional landmarks

- 1) Time consuming for richly corresponding shapes
 - Inadequate coverage → unstable shape characterization

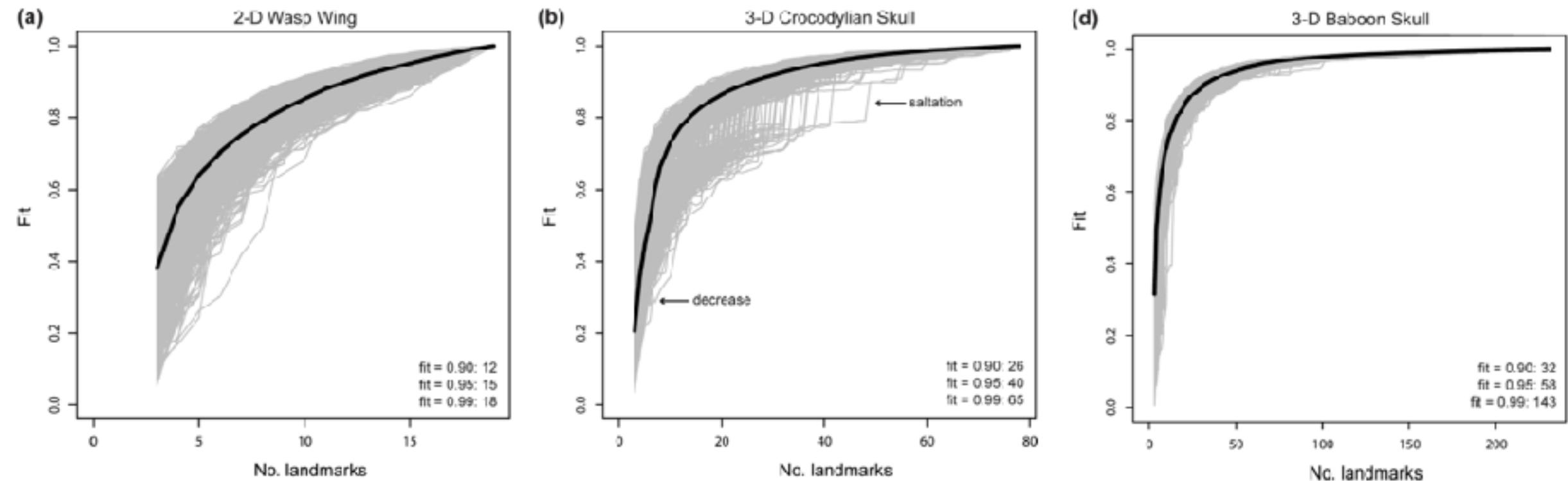


Limitations of traditional landmarks



Watanabe, 2018 (PLoS)

Limitations of traditional landmarks

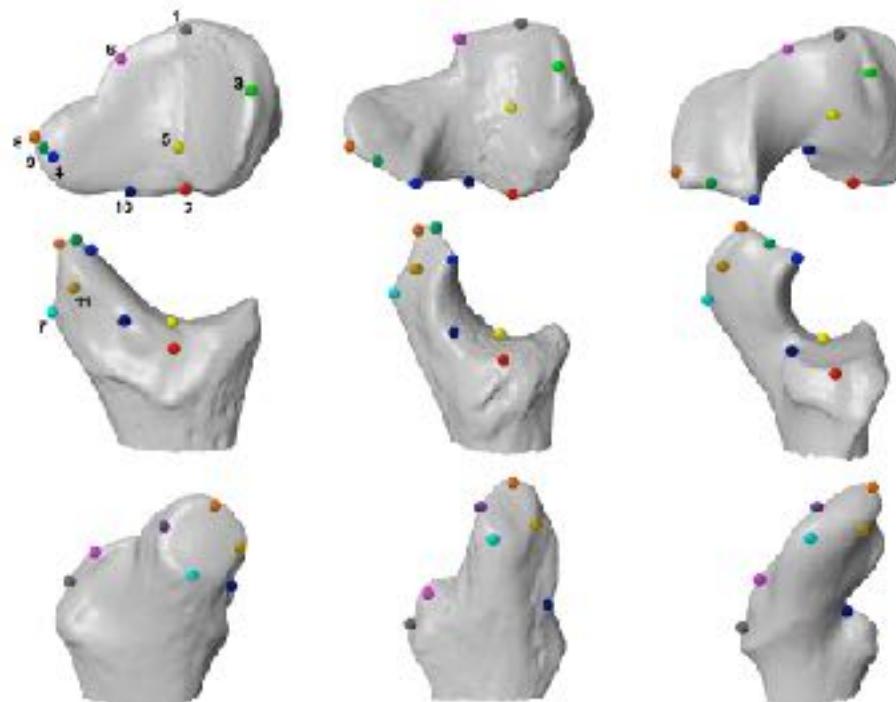


Watanabe, 2018 (PLoS)

Limitations of traditional landmarks

2) Data to represent shape must be chosen *a priori*, maybe subjectively

“As a general rule, landmarks should be chosen so you can quantify any differences that you can see” (Zelditch et al., 2004)

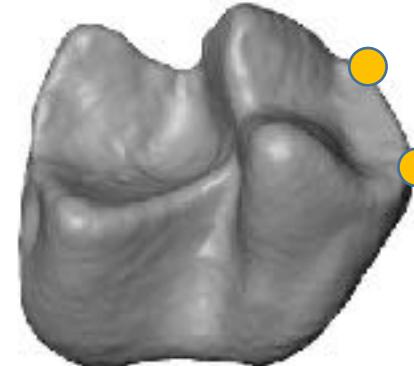


Boyer et al., 2011

Limitations of traditional landmarks

3) Criterion of *homology* can be difficult to substantiate in practice

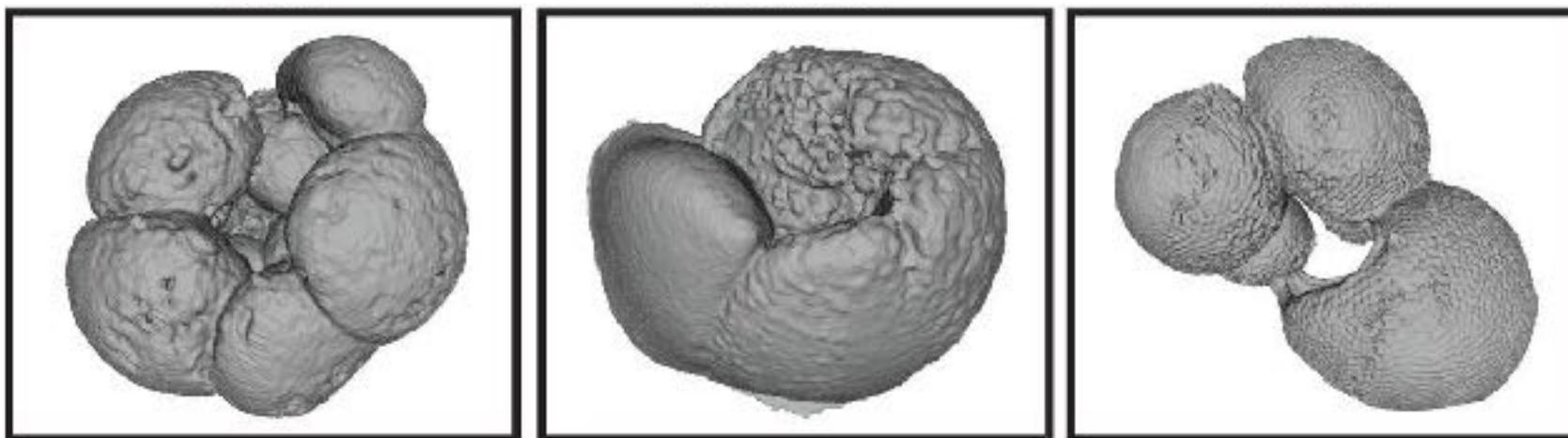
- Leads to differing interpretations of homologous points
- Also expressed as inflated inter- or intraobserver error



Limitations of traditional landmarks

4) Cannot be used for “topology” changes

- i.e., differences reflecting missing features will not be quantified (Zelditch et al., 2004)
- *Coverage* will not be adequate
 - too few landmarks
 - Landmarks not repeatable



Goals & Rationale

Goal

- To achieve data driven understanding of the pattern of micro- and macroevolution that describes the anatomical diversity of life.

Requirement

- Quantitative approaches capable of meaningfully aligning and/or comparing “the same organ in different animals under every variety of form and function”

Rationale

- Without such a framework, comparative morphology will remain arbitrarily limited in its ability to quantitatively describe the morphological (and potentially evolutionary) continuity uniting different individuals and species.

Definition of ‘automated correspondence’

“biologically meaningful (possibly homologous) points are identified between pairs of digitized structures without human intervention except through manipulation of the algorithm parameters”

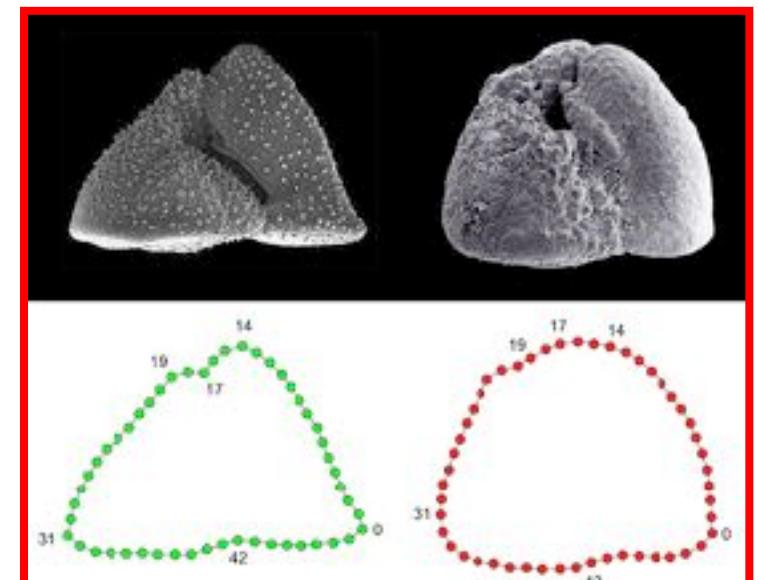
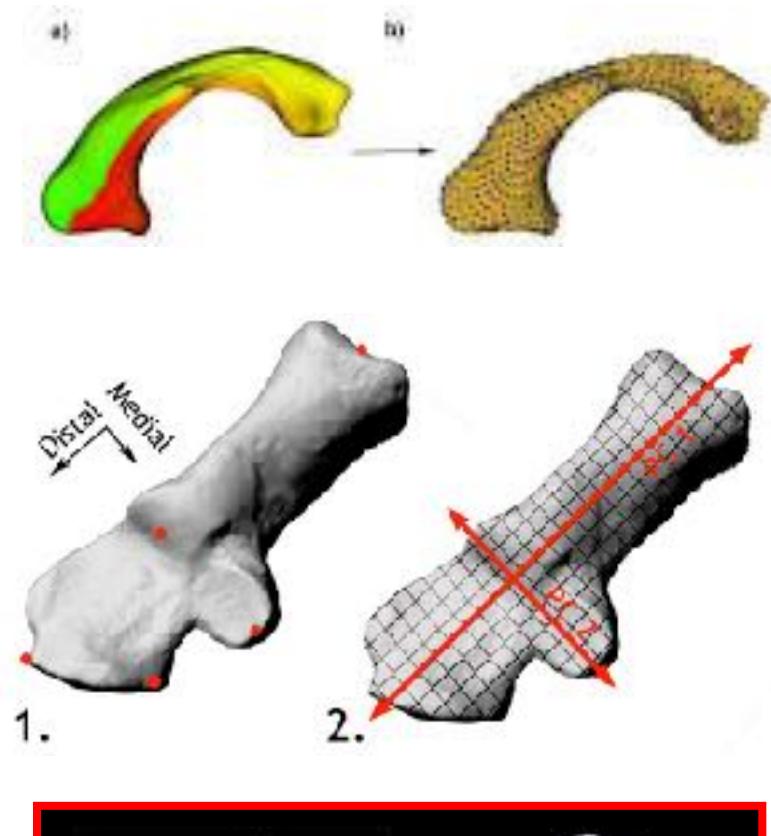
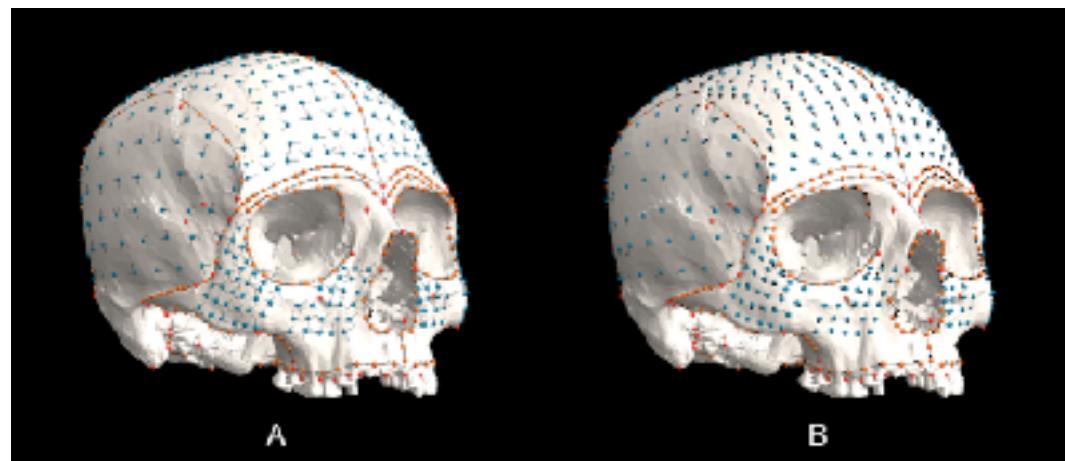
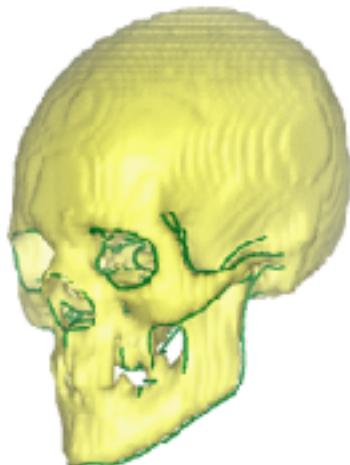
Initial Progress

Semi-automated methods

- Lohmann 1983, Weeks 1999, various papers in MacLeod 2007, Polly & MacLeod 2008, Siewwright et al. 2012, Mitteroecker & Gunz 2009

Automated methods restricted to conservative samples

- Thirion & Guordon 1996, Subsol et al. 2002, Styner et al. 2006, Paniagua et al. 2012



Automated approaches and their limitations thus far

- Require an observer to supply an orientation or initial guess
 - For photography (to work from 2D images)
 - For choosing a orientation (Eigensurface of Polly & MacLeod, 2008)
 - Aligning “crest lines” (Thirion & Guordon, 1995; Subsol et al., 2002)
- Expect minimal variation
 - Clinical work on human-only samples (Styner et al. 2003)
 - Work with model animals (intraspecific variation)
- Computationally prohibitive
 - Biggest hurdle is finding correct initial alignment for diverse samples (e.g., Iterative Closest Points algorithm)

Outline

- *Auto3dgm*: Concepts & Logistics
- Diffeomorphsim approach
- Topological data analysis

Part I: Auto3dgm

Auto3dgm Overview

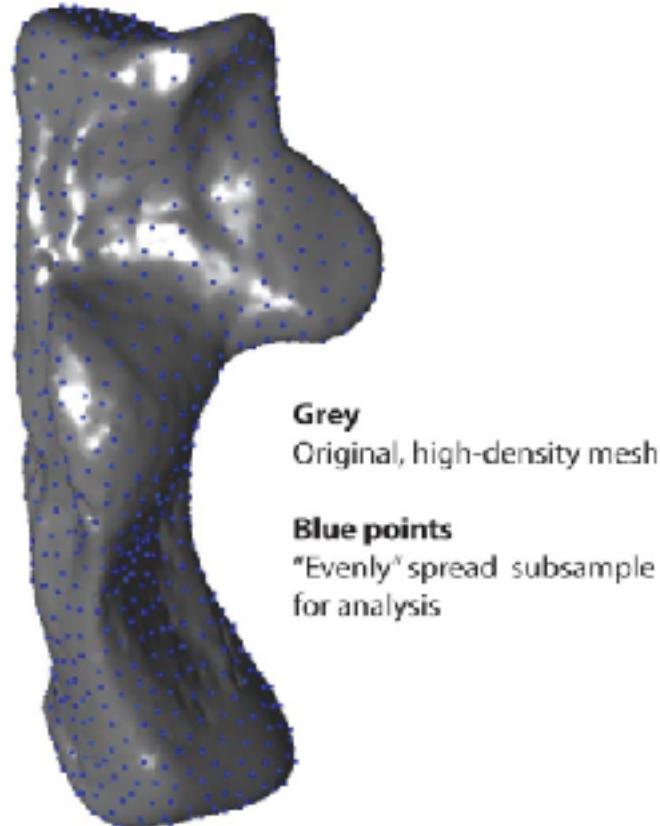
- Resampling
- Pairwise alignment
- Global alignment

Boyer, Doug M., et al. "A new fully automated approach for aligning and comparing shapes." *The Anatomical Record* 298.1 (2015): 249-276.

Auto3dgm Overview

- **Resampling**
- Pairwise alignment
- Global alignment

FPS: Farthest Point Sampling



Grey
Original, high-density mesh

Blue points
"Evenly" spread subsample
for analysis

Boyer, Doug M., et al. "A new fully automated approach for aligning and comparing shapes." *The Anatomical Record* 298.1 (2015): 249-276.

Auto3dgm Overview

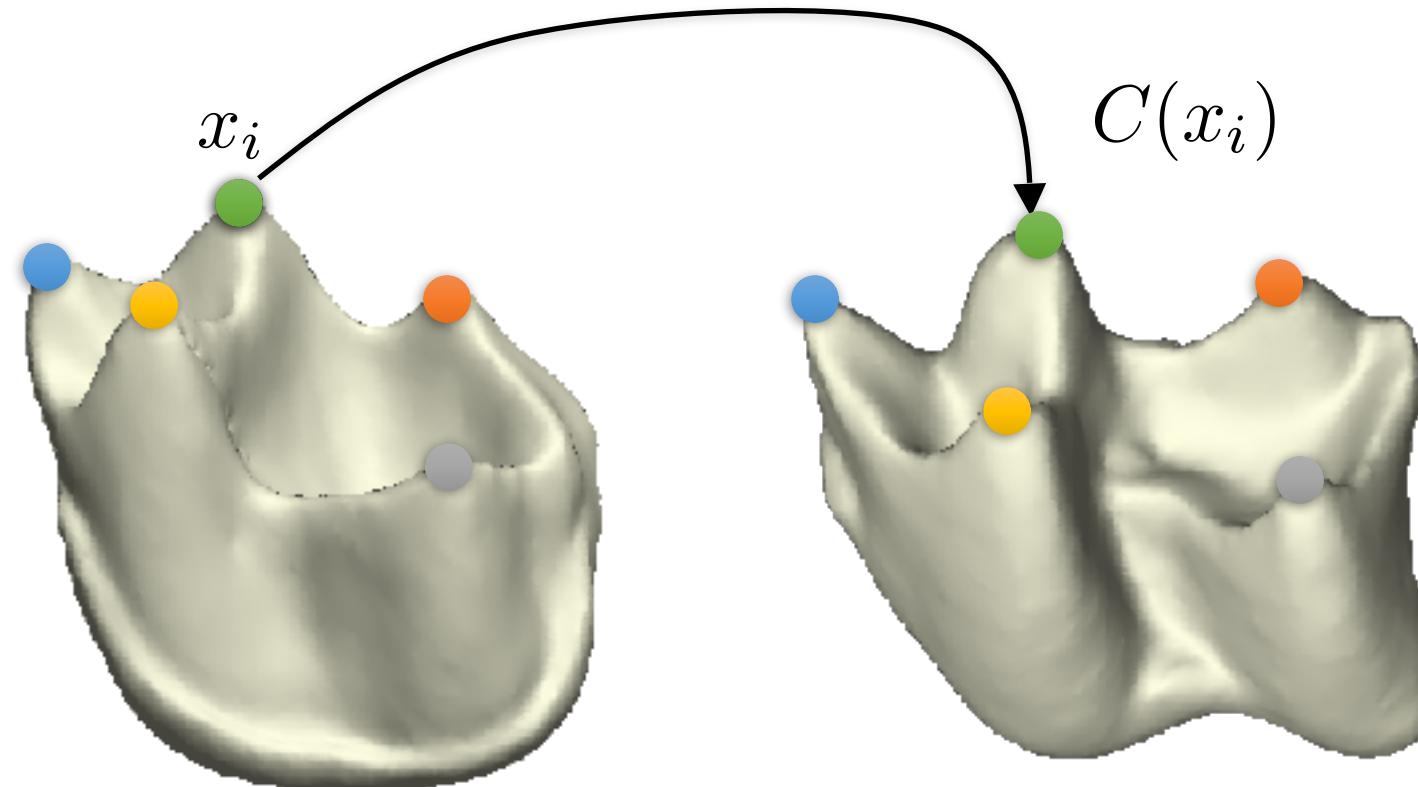
- Resampling
- **Pairwise alignment**
- Global alignment

Iteratively search:
Best possible rotation via Kabsch
Best possible permutation via Hungarian



Boyer, Doug M., et al. "A new fully automated approach for aligning and comparing shapes." *The Anatomical Record* 298.1 (2015): 249-276.

Standard Procrustes Distance



$C(x_i)$ “qualitatively determined”

Standard Procrustes Distance



Find rigid motions (translation & rotation) that minimize sum distance

Standard Procrustes Distance

Write as $3 \times n$ matrix



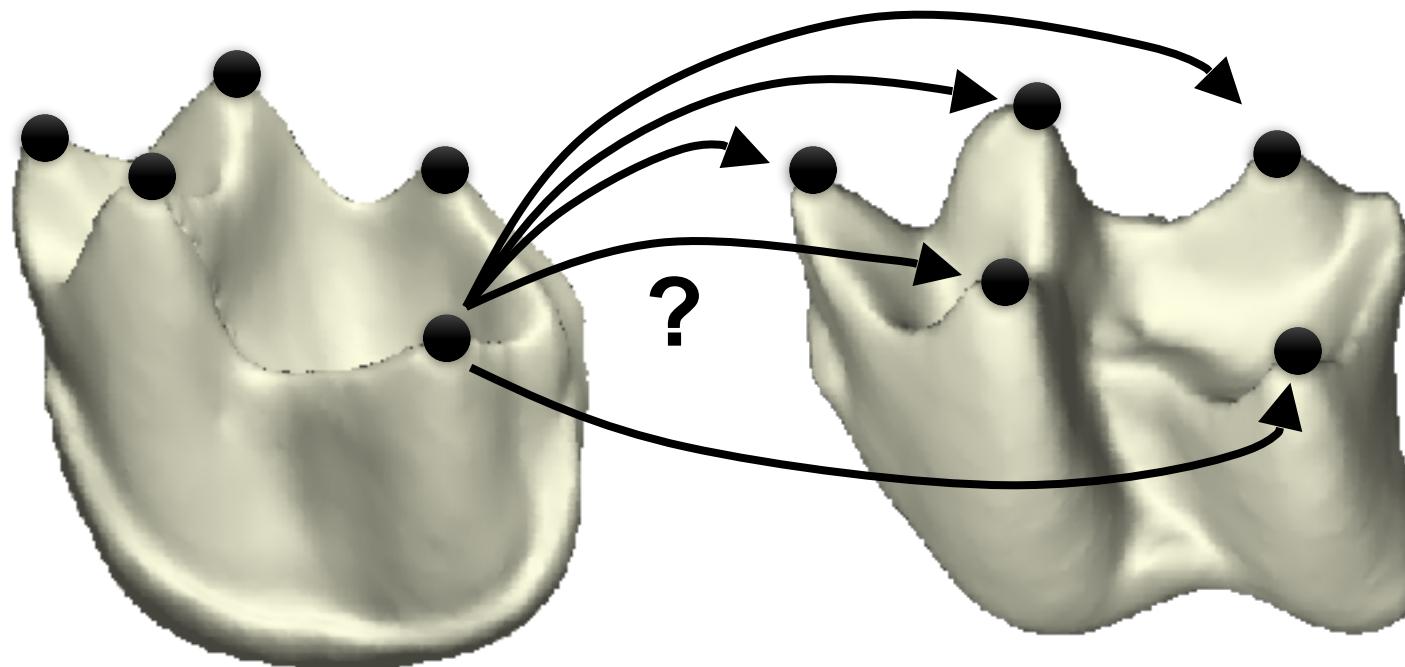
Find rotation \mathbf{R} (3 by 3) minimizing

$$d(\mathbf{R}A, B)$$

Modified Procrustes Distance

5 points = 120 possibilities....

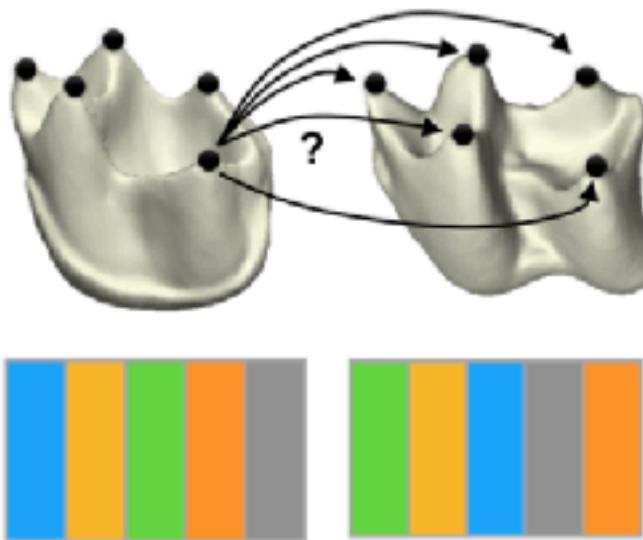
10 points = 3.6 million possibilities



Modified Procrustes Distance

Key math observation:

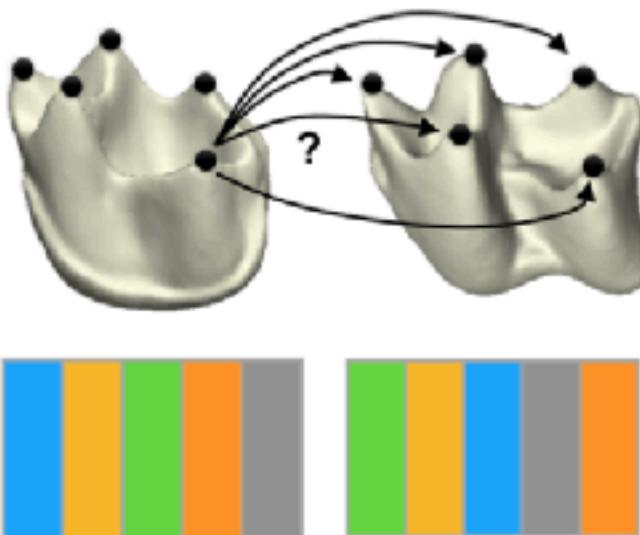
Landmarking is ordering points. Points are given as columns in a matrix. Reordering the columns is right-multiplying a permutation matrix.



Modified Procrustes Distance

Key math observation:

Landmarking is ordering points. Points are given as columns in a matrix. Reordering the columns is right-multiplying a permutation matrix.



Find permutation \mathbf{P} (n by n) and \mathbf{R} minimizing
 $d(\mathbf{RA}, \mathbf{BP})$

Modified Procrustes Distance

1. subsample: spread n points evenly on each surface
2. iterative search:
find \mathbf{R} minimizing $d(\mathbf{RA}, \mathbf{BP})$
until converges

Modified Procrustes Distance

1. subsample: spread n points evenly on each surface
2. iterative search:
find \mathbf{P} minimizing $d(RA, B\mathbf{P})$
until converges

Modified Procrustes Distance

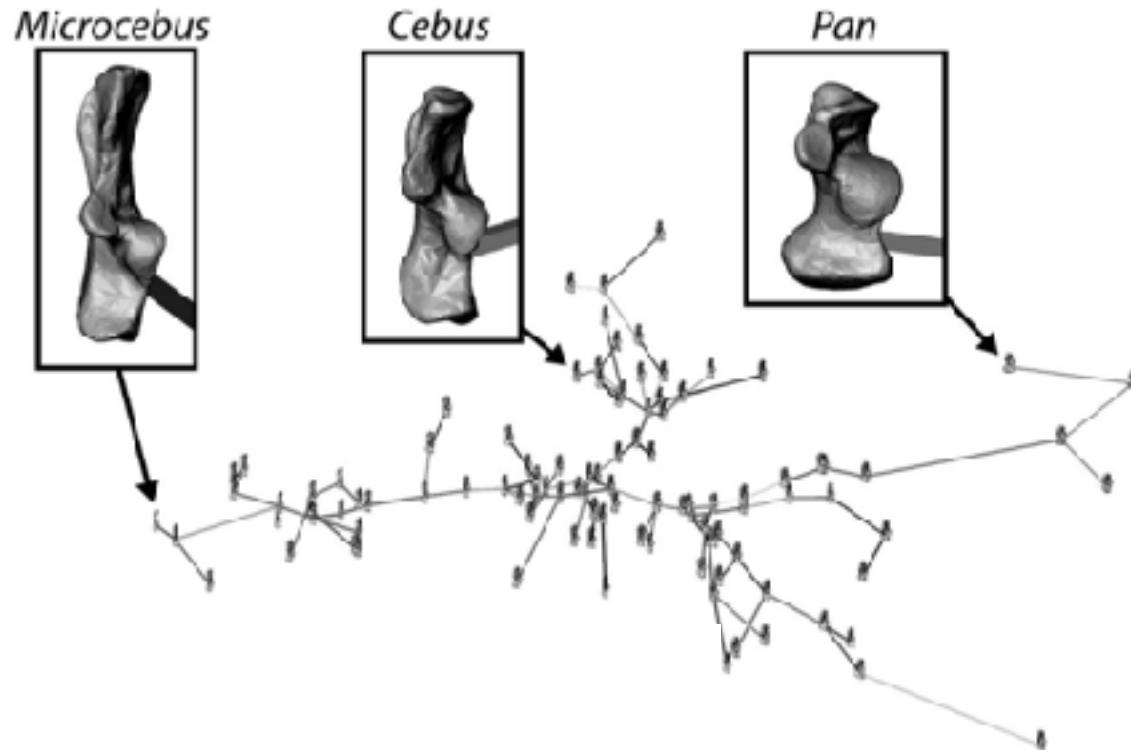
1. subsample: spread n points evenly on each surface
2. iterative search:
find \mathbf{P} minimizing $d(RA, B\mathbf{P})$
until converges

larger $n \rightarrow$ more accurate
but computationally more expensive

Auto3dgm Overview

- Resampling
- Pairwise alignment
- **Global alignment**

MST: Minimum Spanning Tree

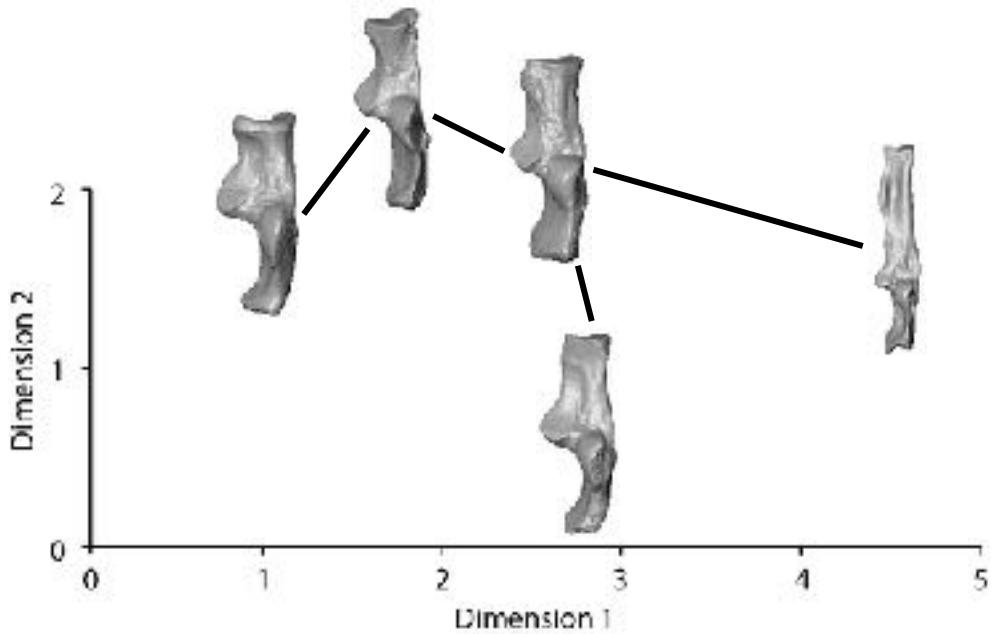


Boyer, Doug M., et al. "A new fully automated approach for aligning and comparing shapes." *The Anatomical Record* 298.1 (2015): 249-276.

Steps to the method

Alignment refinement

- Compute Procrustes for all pairs
- Determine minimum spanning tree (MST)
- Propagate correspondences through MST for transitivity



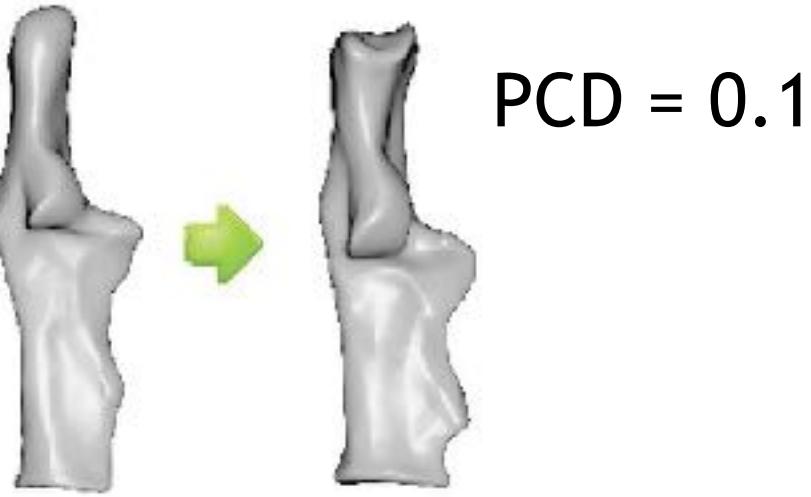
PCD matrix

The PCD matrix is a 5x5 grid where each cell contains a grayscale 3D point cloud visualization and a numerical value representing the Procrustes distance or similarity between the corresponding point clouds. The values are:

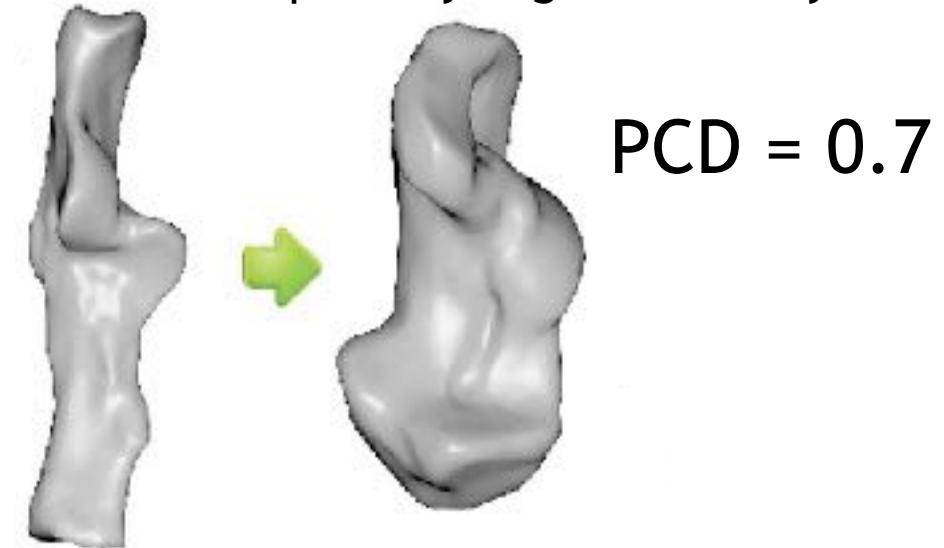
	0	0.2	0.3	0.2	0.6
0	0	0.2	0.3	0.2	0.6
0.2		0	0.2	0.1	0.5
0.3			0	0.3	0.6
0.2				0	0.4
0.6					0

Purpose of MST & Propagation

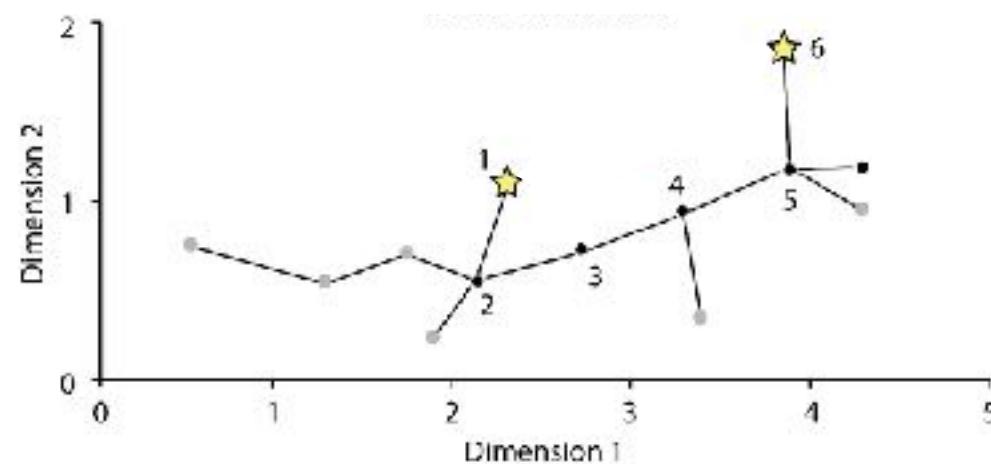
Similar shapes align well



Dissimilar shapes may align incorrectly

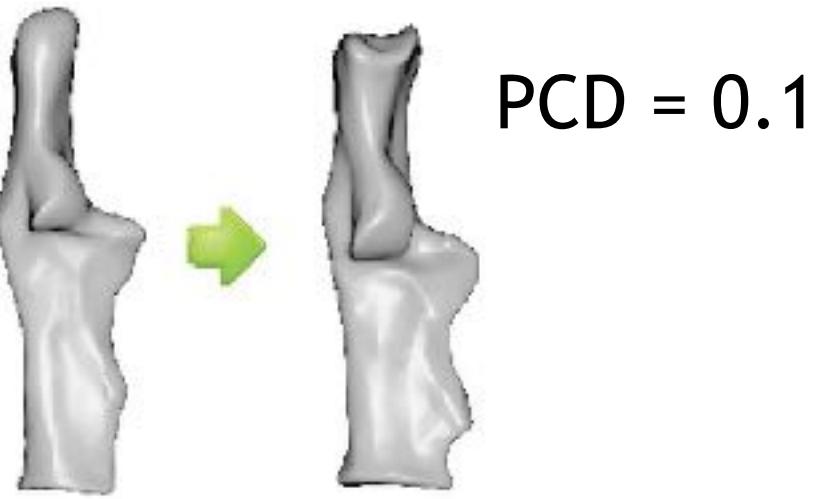


All Procrustes distances used to construct initial minimum spanning tree



Purpose of MST & Propagation

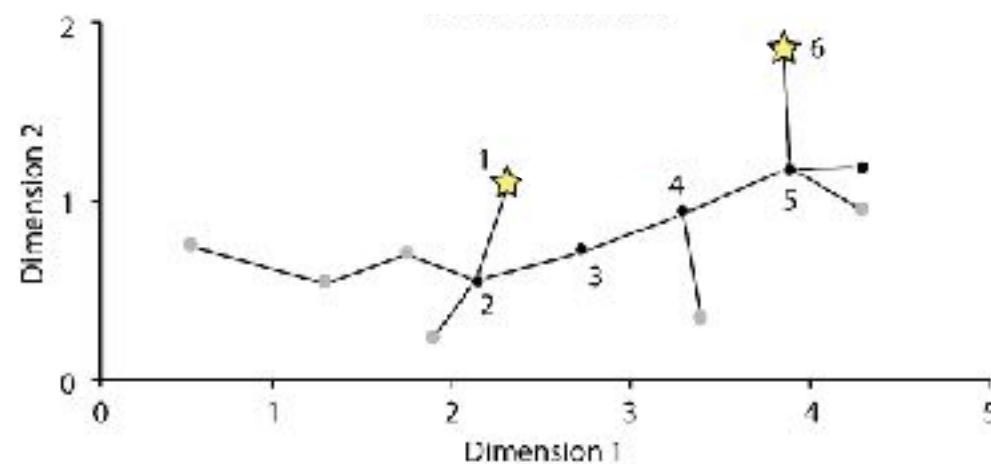
Similar shapes align well



Dissimilar shapes may align incorrectly

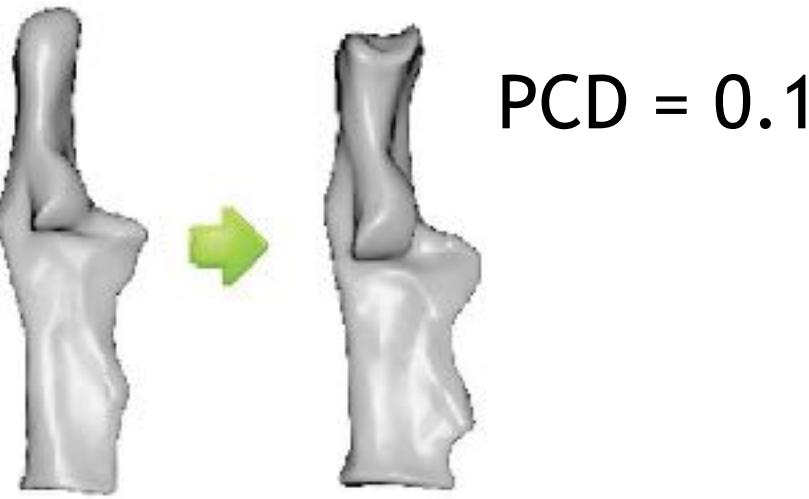


Incorrect alignments typically have high PCD and won't be in the MST



Purpose of MST & Propagation

Similar shapes align well



Dissimilar shapes may align incorrectly



The MST path can be used to find correct alignments, and recompute distances



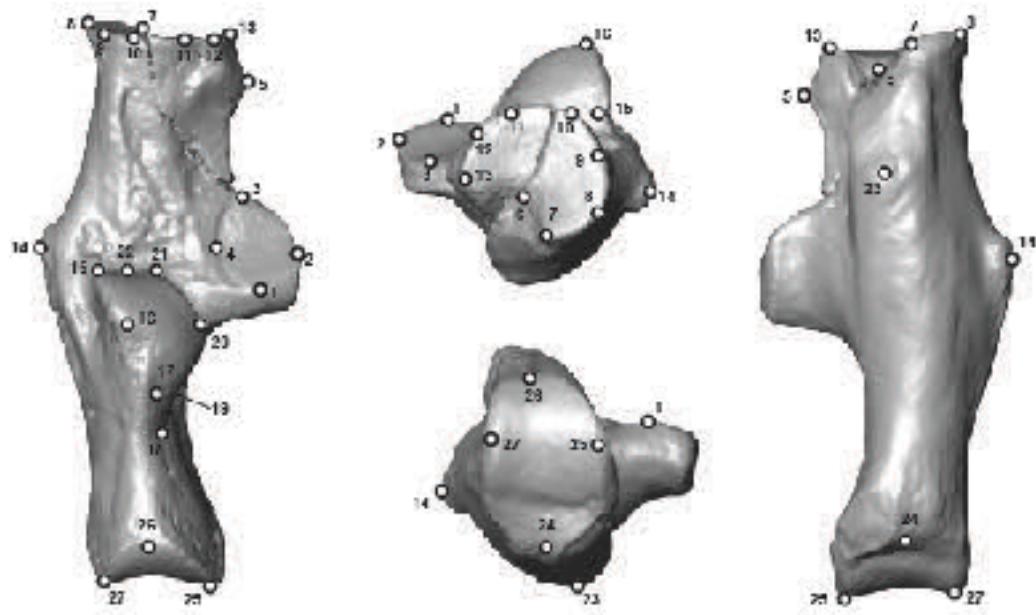
Auto3dgm Overview

- Resampling
- Pairwise alignment
- Global alignment

Boyer, Doug M., et al. "A new fully automated approach for aligning and comparing shapes." *The Anatomical Record* 298.1 (2015): 249-276.

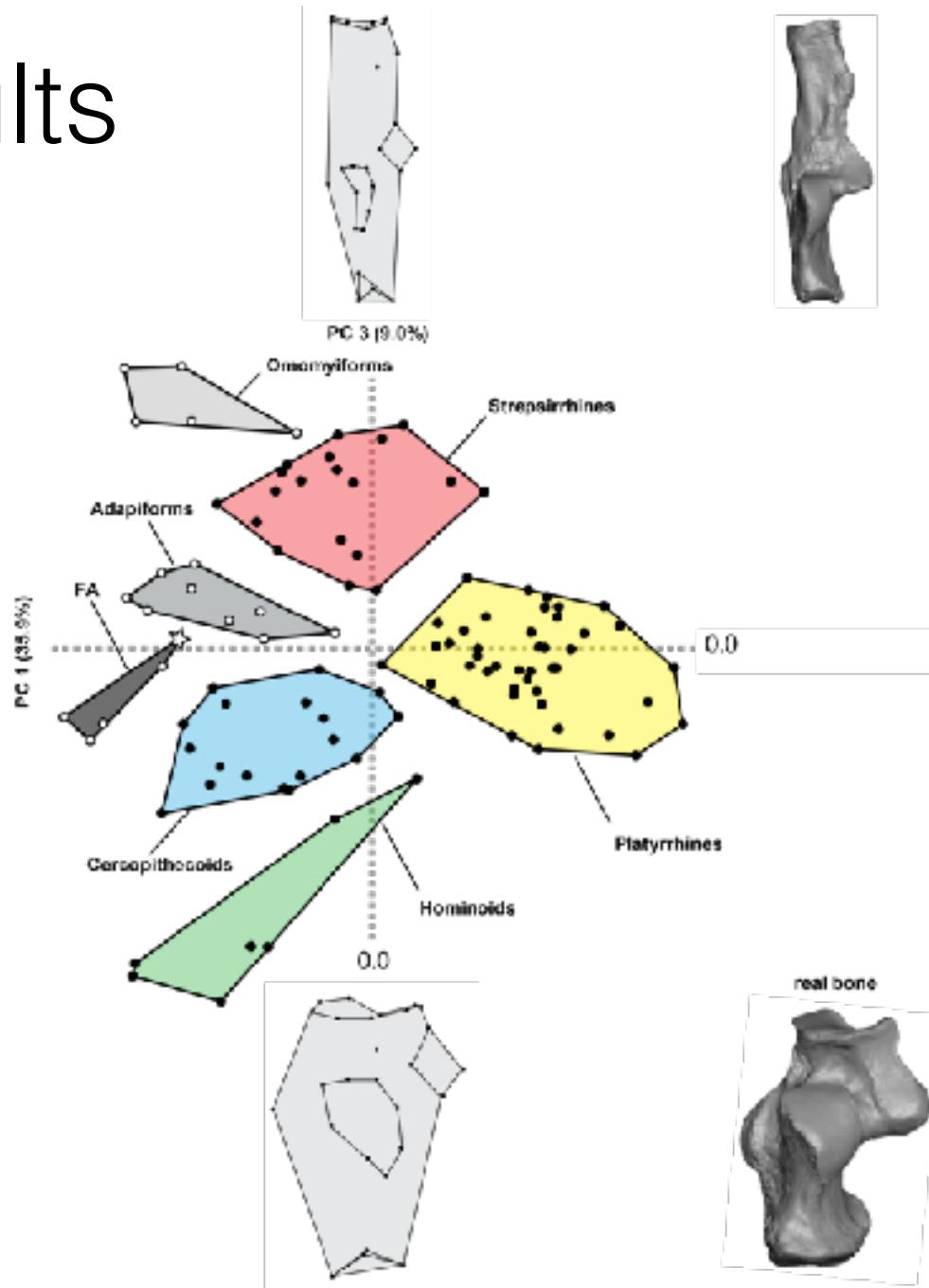
Comparison to researcher-based data set

- Sample - 106 calcanei
- Data points
 - Observer (27 landmarks)
 - Algorithm (1,024 points)
- Software
 - *Morphologika2.5*
- Both datasets analyzed identically

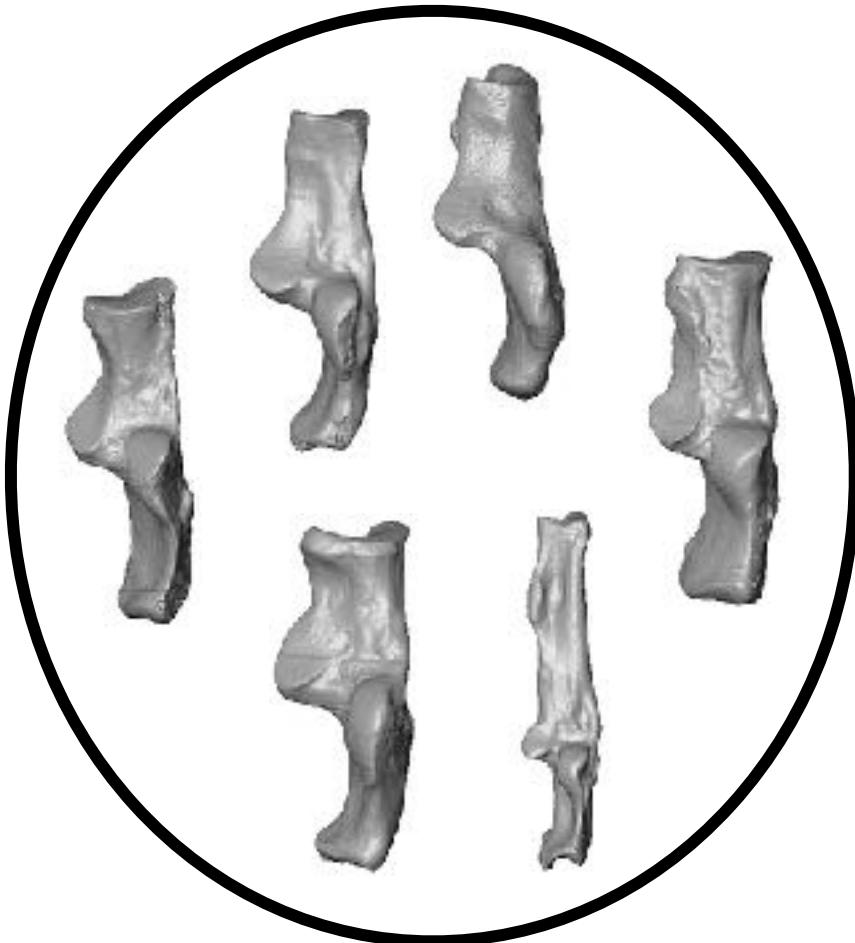


From Gladman et al. (2013)

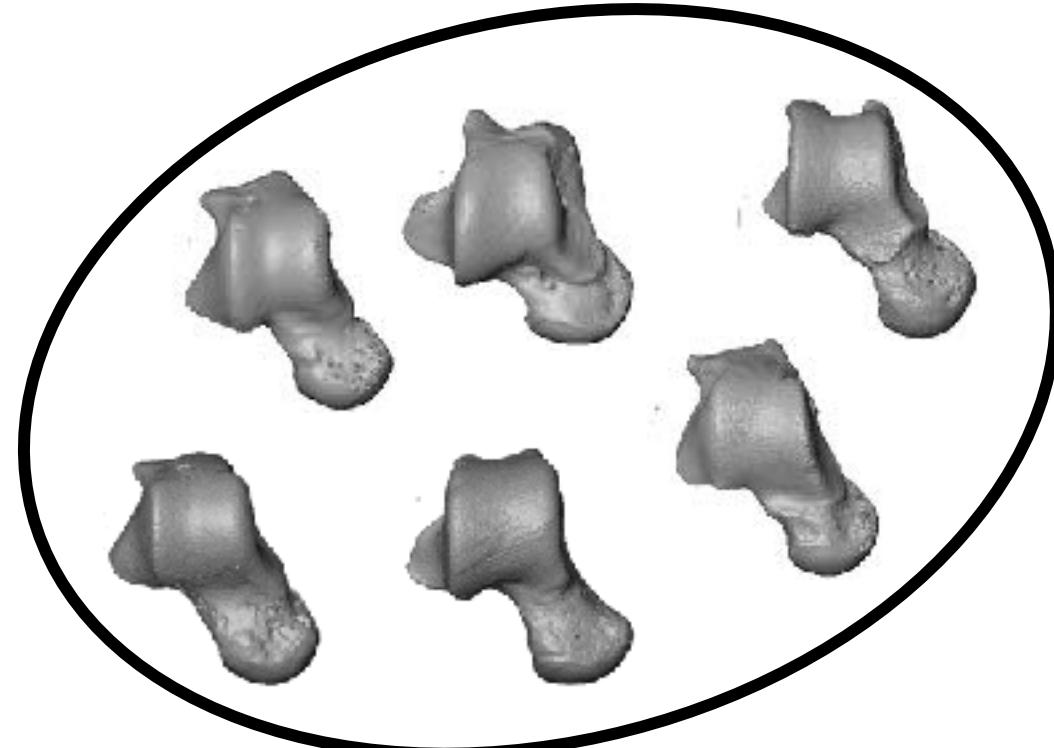
Results



Comparison between bones

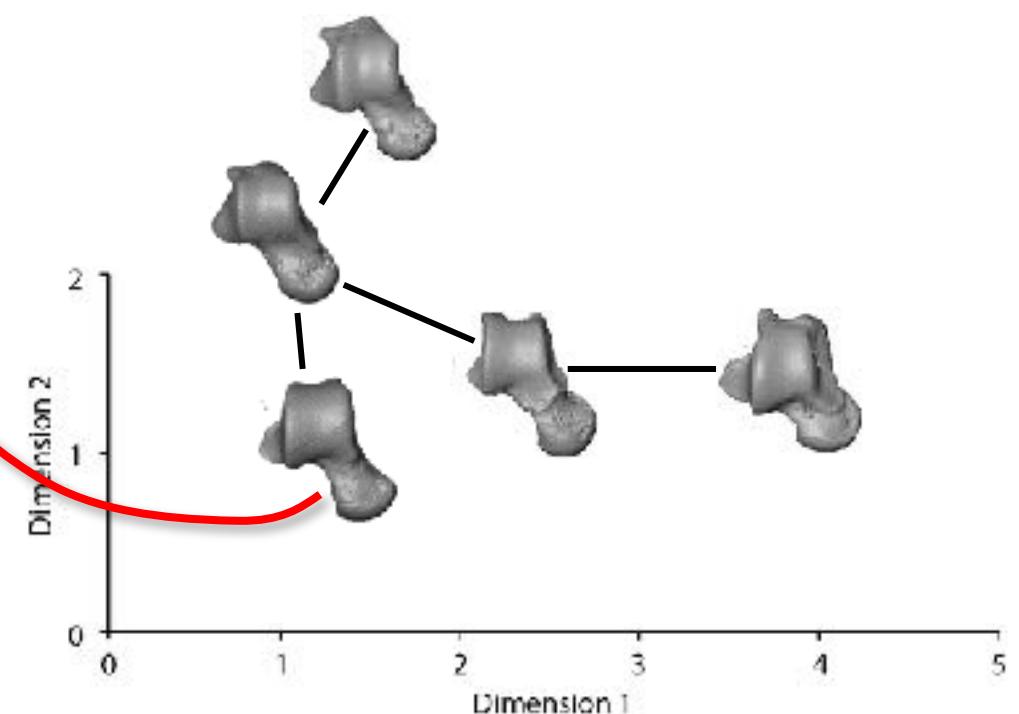
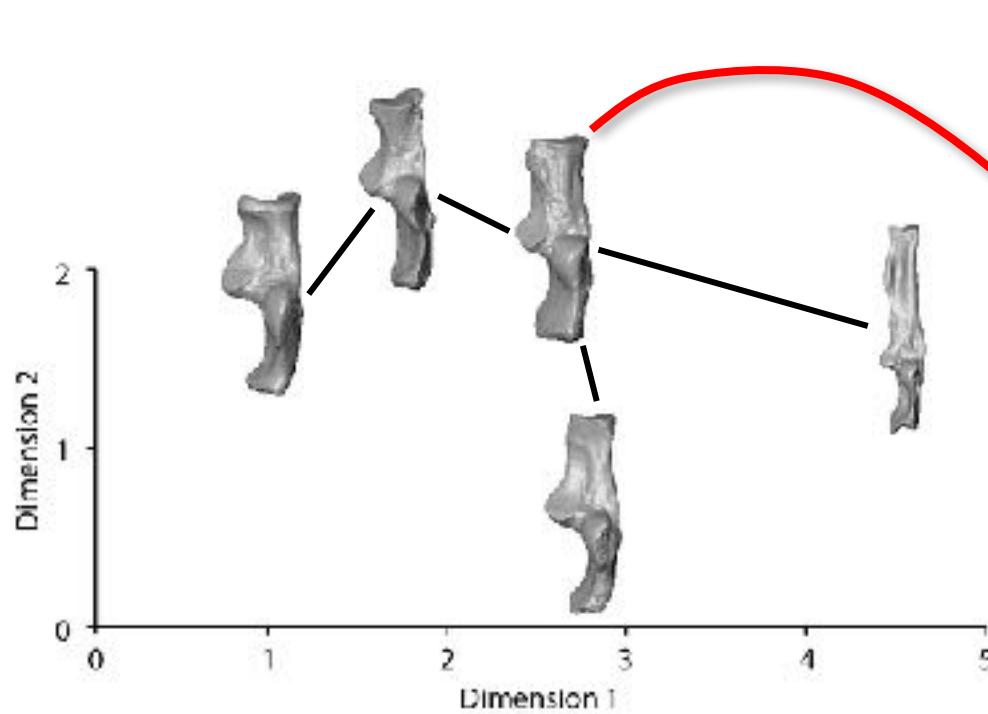


versus



Steps to the method

- Use standard number of points in each sample
- Analyze each bone type separately
- Compute Procrustes minimum distance for all pairs astragali+calcaneaei
- Link minimum spanning trees using pair with smallest PCD



Steps to the method

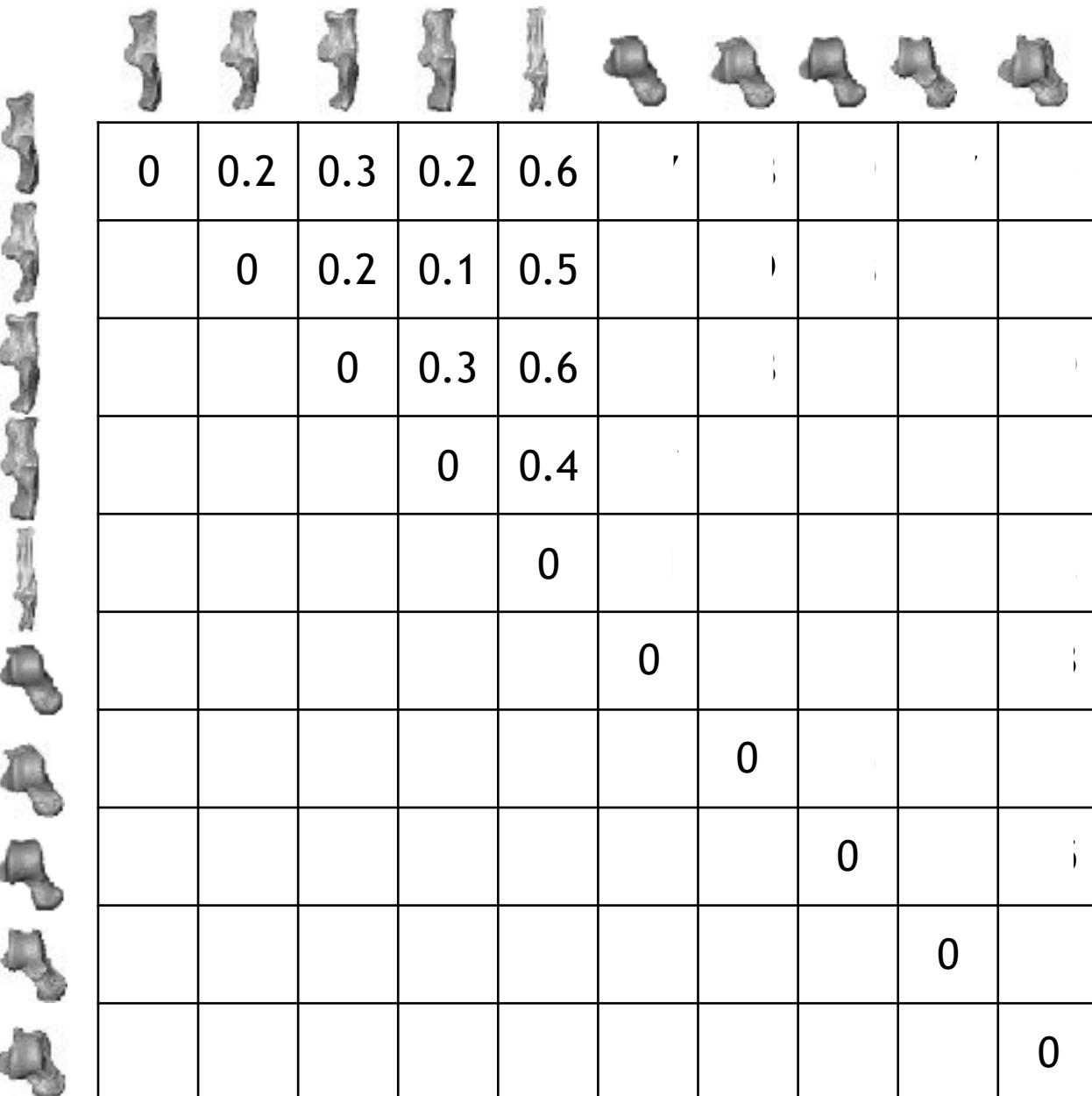
Yields global landmark dataset that applies to both bone types

- Combined distance matrix
 - Procrustes alignment and PCA can be run on mixed bone sample

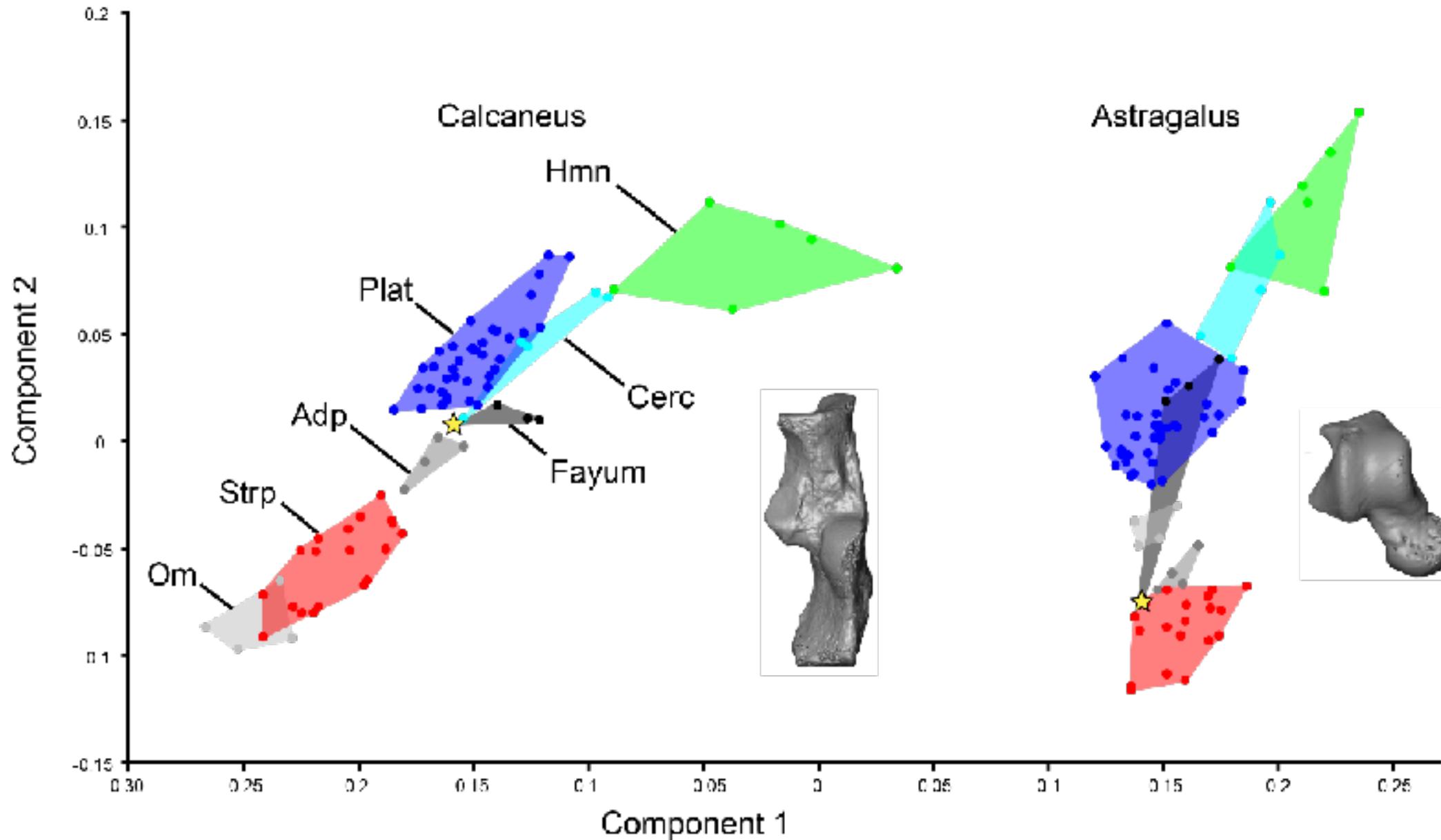
Example

- 80 calcanei and 80 astragali
 - Taxon matched
 - Stats
 - Compare morphospace area
 - Proxy for disparity
 - Compare phylogenetic signal
 - Tree from 10k trees website
 - Check for correlations in PCA

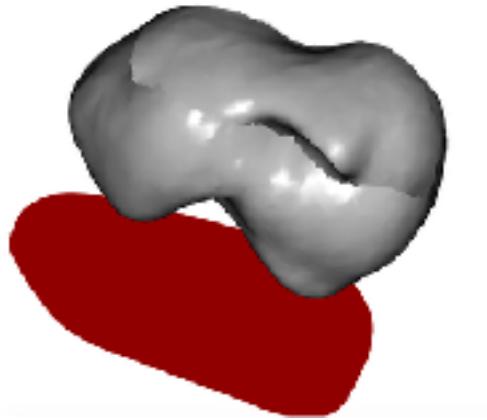
PCD matrix



Results

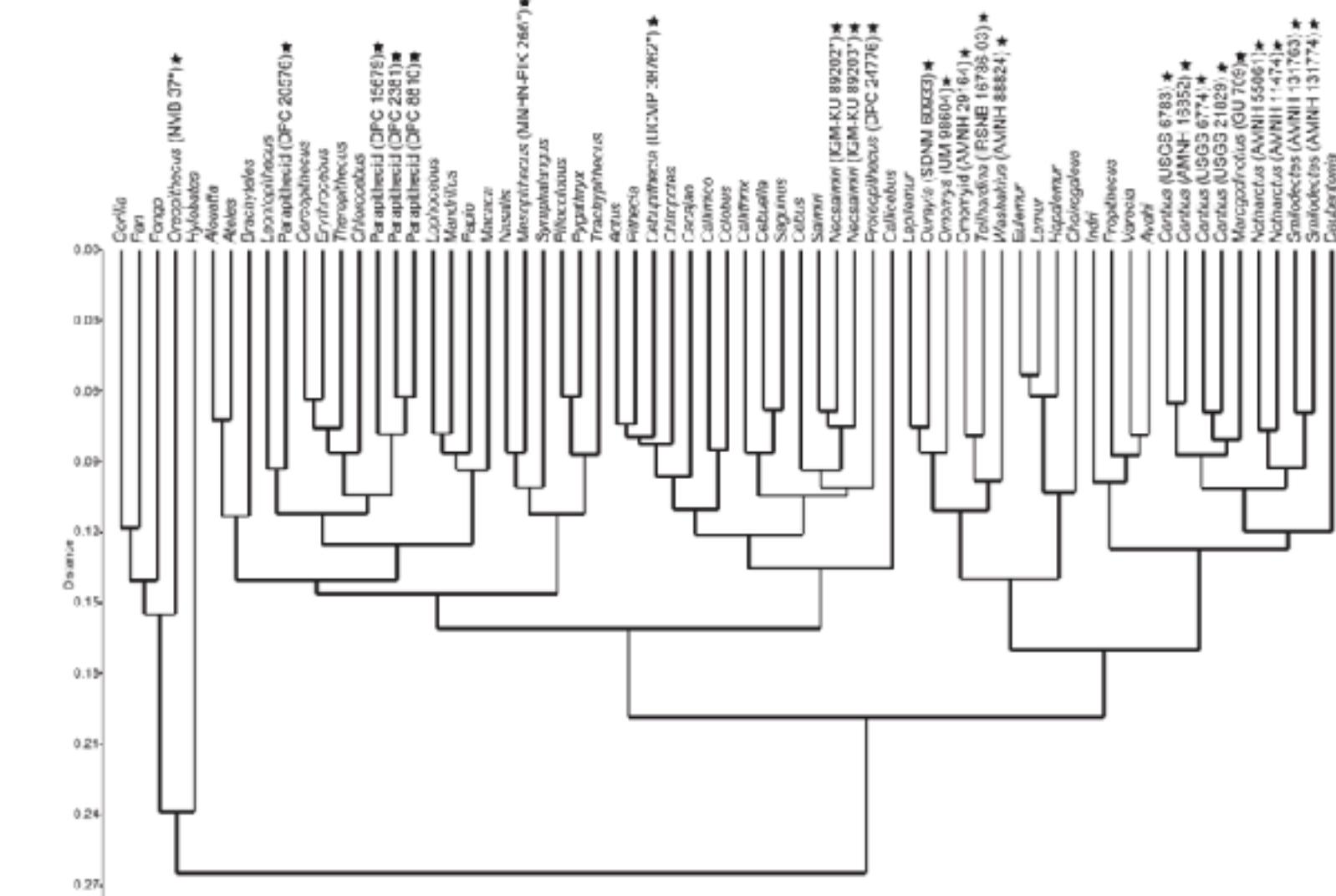


Other applications: Data Pre-processing



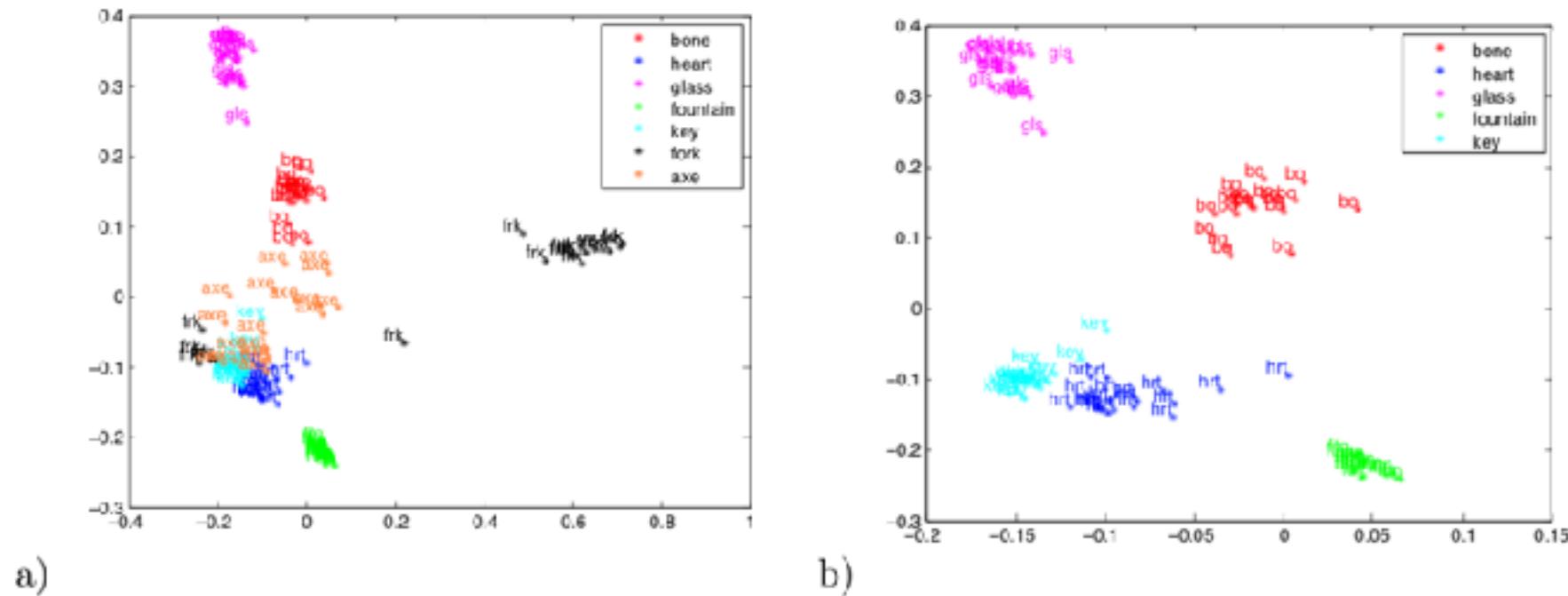
Boyer, Doug M. "Relief index of second mandibular molars is a correlate of diet among prosimian primates and other euarchontan mammals." *Journal of Human Evolution* 55.6 (2008): 1118-1137.

Other applications: Phenotypic affinities



Boyer, Doug M., et al. "A new fully automated approach for aligning and comparing shapes." *The Anatomical Record* 298.1 (2015): 249-276.

Other applications: Align shape spaces



Turner, Katharine, Sayan Mukherjee, and Doug M. Boyer. "Persistent homology transform for modeling shapes and surfaces." *Information and Inference: A Journal of the IMA* 3.4 (2014): 310-344.

Auto3dgm: Publications

Methodological/validation

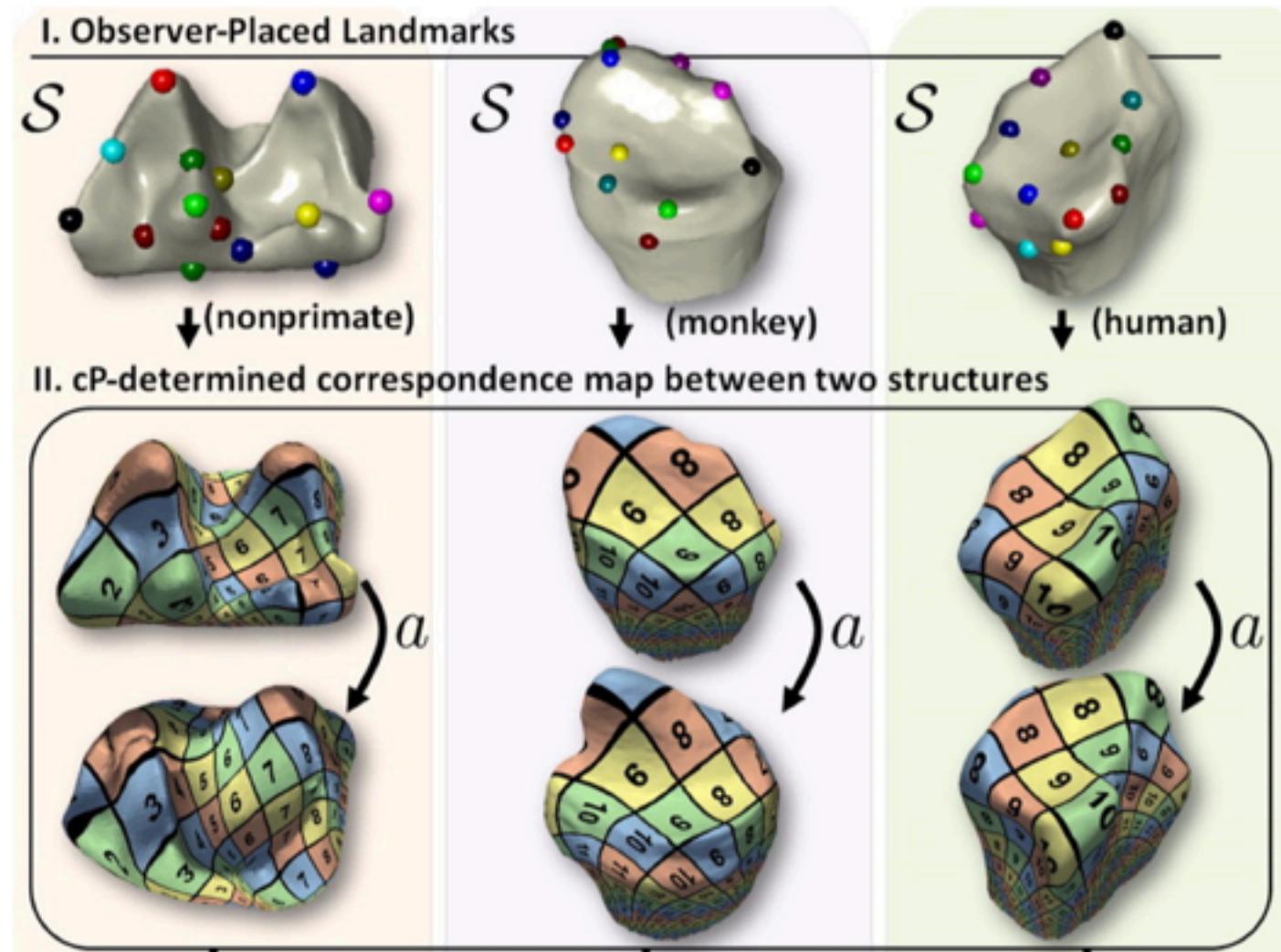
- Boyer et al. (2015): <https://onlinelibrary.wiley.com/doi/full/10.1002/ar.23084>
- Boyer et al. (2016): <https://onlinelibrary.wiley.com/doi/abs/10.1002/ar.23202>
- Gonzalez et al. (2016): <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.22934>
- Gao et al. (2017): <https://onlinelibrary.wiley.com/doi/abs/10.1002/ar.23700>
- Gao et al. (2018): <https://arxiv.org/pdf/1807.11887.pdf>

Applications (e.g.):

- Turner et al. (2014): <https://doi.org/10.1093/imaiai/iau011>
- Sieffert et al. (2015): <https://peerj.com/articles/1036/>
- Hsiang et al. (2016): <http://rstb.royalsocietypublishing.org/content/371/1691/20150227>
- Gingerich (2016): <https://link.springer.com/article/10.1007/s12542-017-0362-8>
- Boyer et al. (2017): <https://www.sciencedirect.com/science/article/pii/S0047248417301021>
- Vitek et al. (2017): <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.3058>
- Renaud et al. (2017): <https://www.sciencedirect.com/science/article/pii/S0003996917301383?via%3Dihub>
- Renaud et al. (2018): <https://link.springer.com/article/10.1007/s11692-018-9459-6>
- Boyer et al. (2018): <https://www.sciencedirect.com/science/article/pii/S0047248417305134>
- Gunnell et al. (2018): <https://www.nature.com/articles/s41467-018-05648-w>
- Wang et al. (2019): <https://www.biorxiv.org/content/10.1101/701391v1>

Part II: Continuous Procrustes

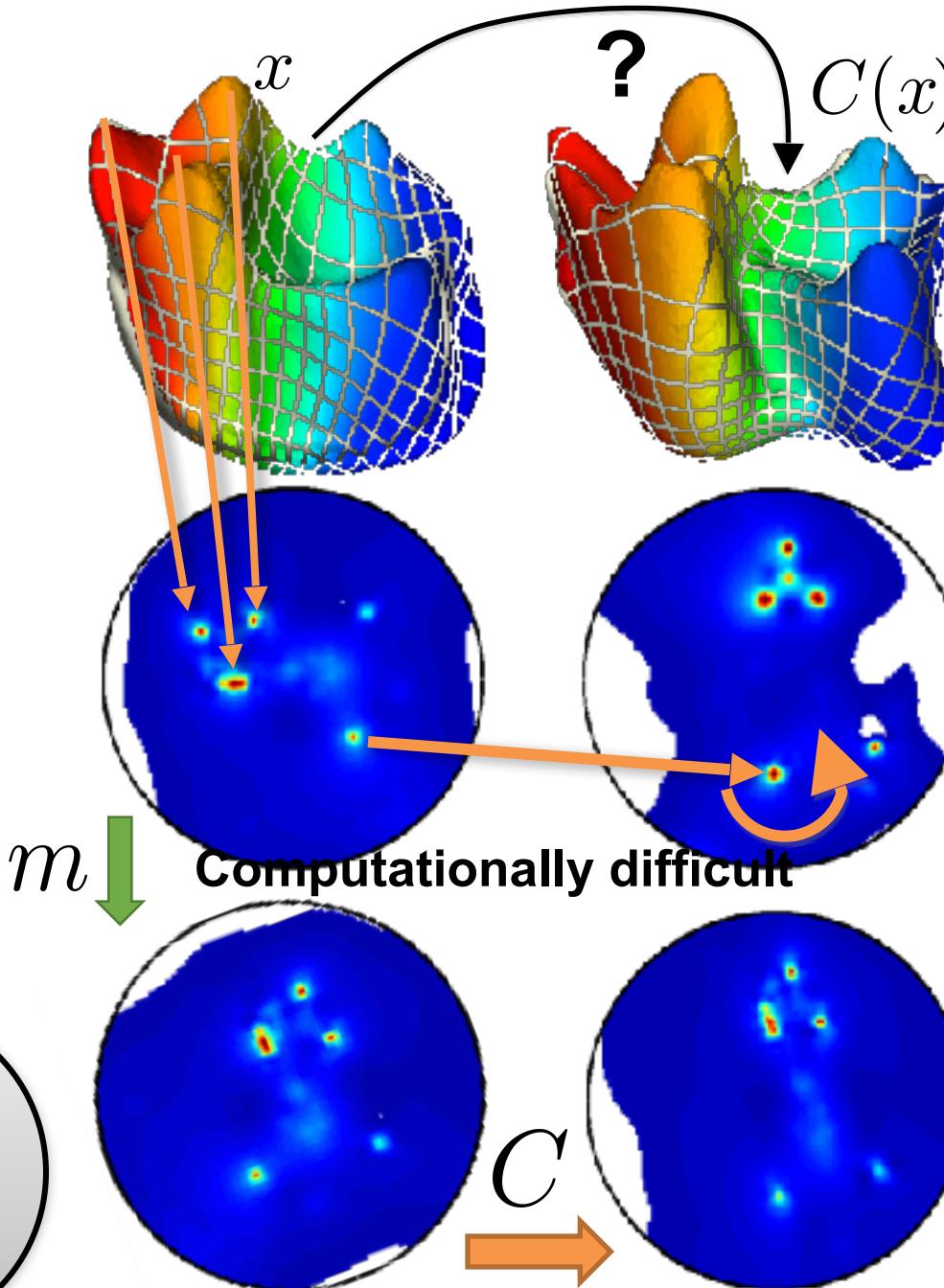
Diffeomorphism based approach



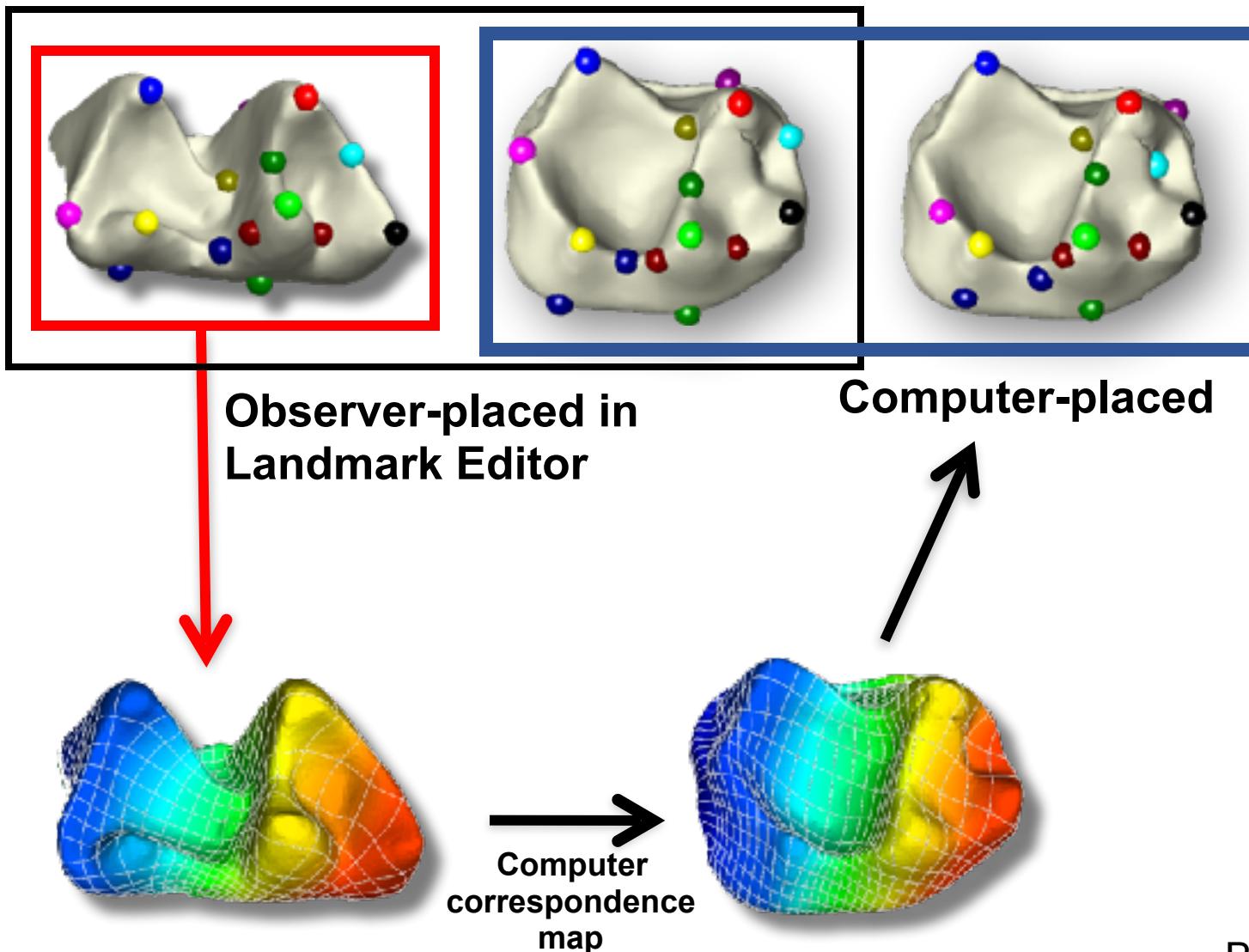
Search area of
Correspondence
Maps...

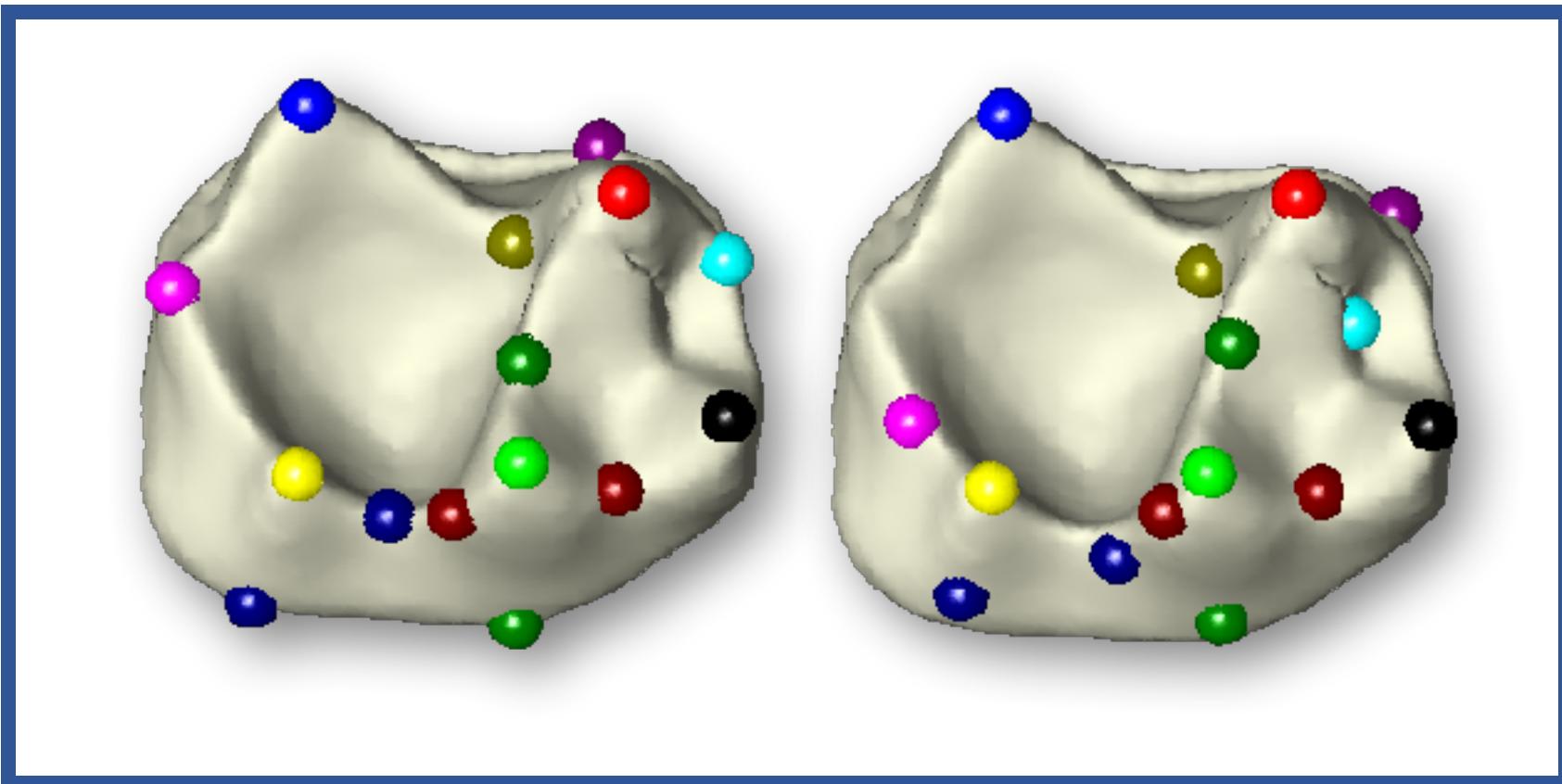
∞ Degrees of freedom

C
TPS
 m
Möbius
Maps
3 Degrees



Feature-aware landmarks





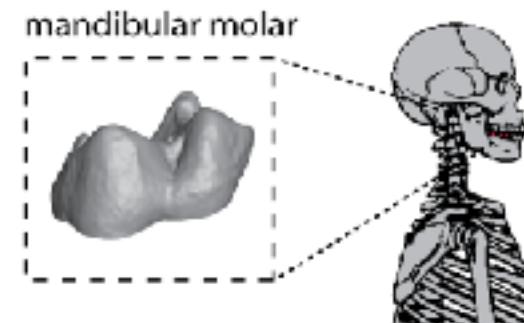
**Observer-placed in
Landmark Editor**

Computer-placed

CLASSIFICATION ANALYSIS

3 data sets (mainly primates)

116 mandibular 2nd molars
5 orders, 19 families, 35 genera



45 distal radii
1 order, 2 families, 4 genera

61 proximal first metatarsals (Mt1)
1 order, 9 families, 13 genera

HUMAN correspondences

Same 3 data sets: Geometric landmarks applied

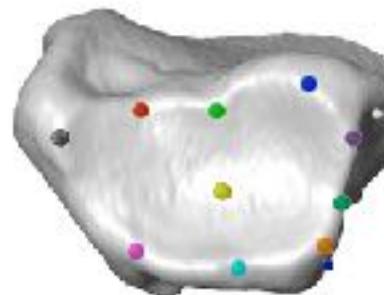
116 mandibular 2nd molars
5 orders, 19 families, 35 genera

16 Landmarks



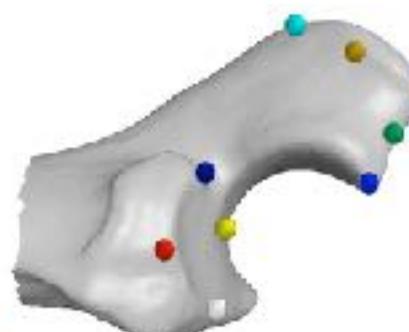
45 distal radii
1 order, 2 families, 4 genera

13 Landmarks



61 proximal Mt1
1 order, 9 families, 13 genera

11 Landmarks

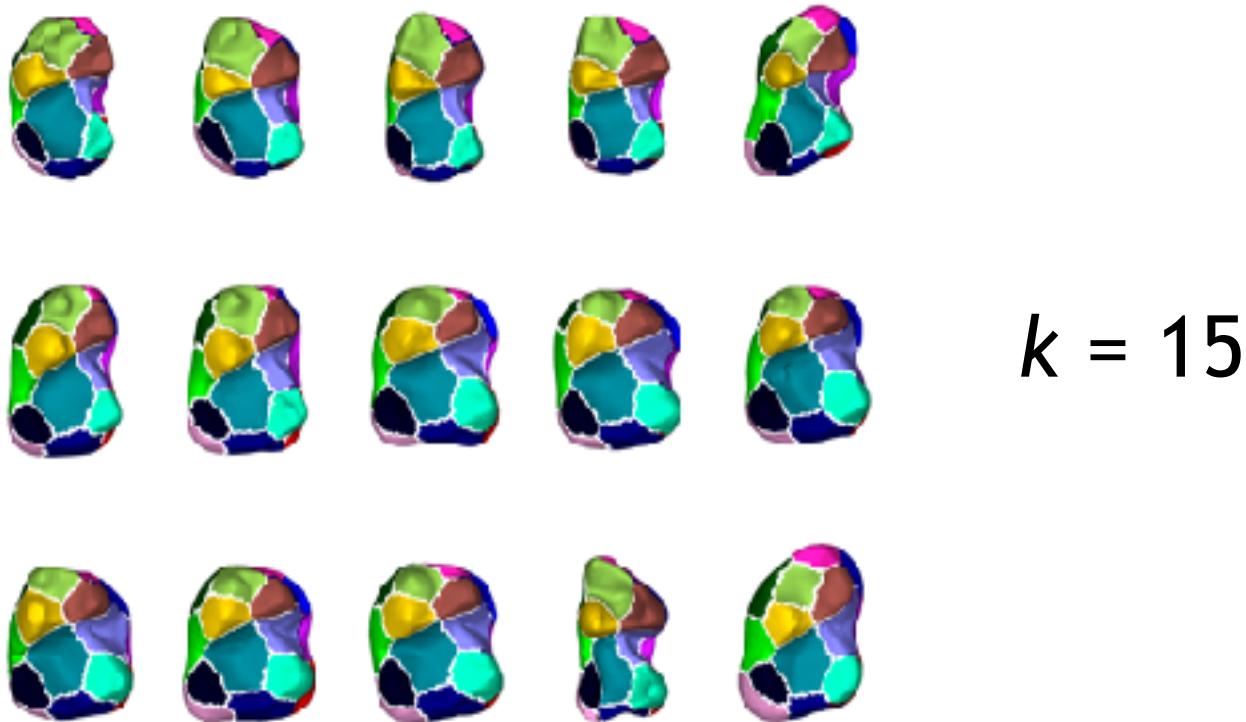


Results

Classification success (mean % correct per genus)

Dataset	HUMAN correspondences	COMPUTER correspondences
Teeth 24 grps	90%	90%
Radii 5 grps	83%	88%
Mt1 13 grps	76%	82%

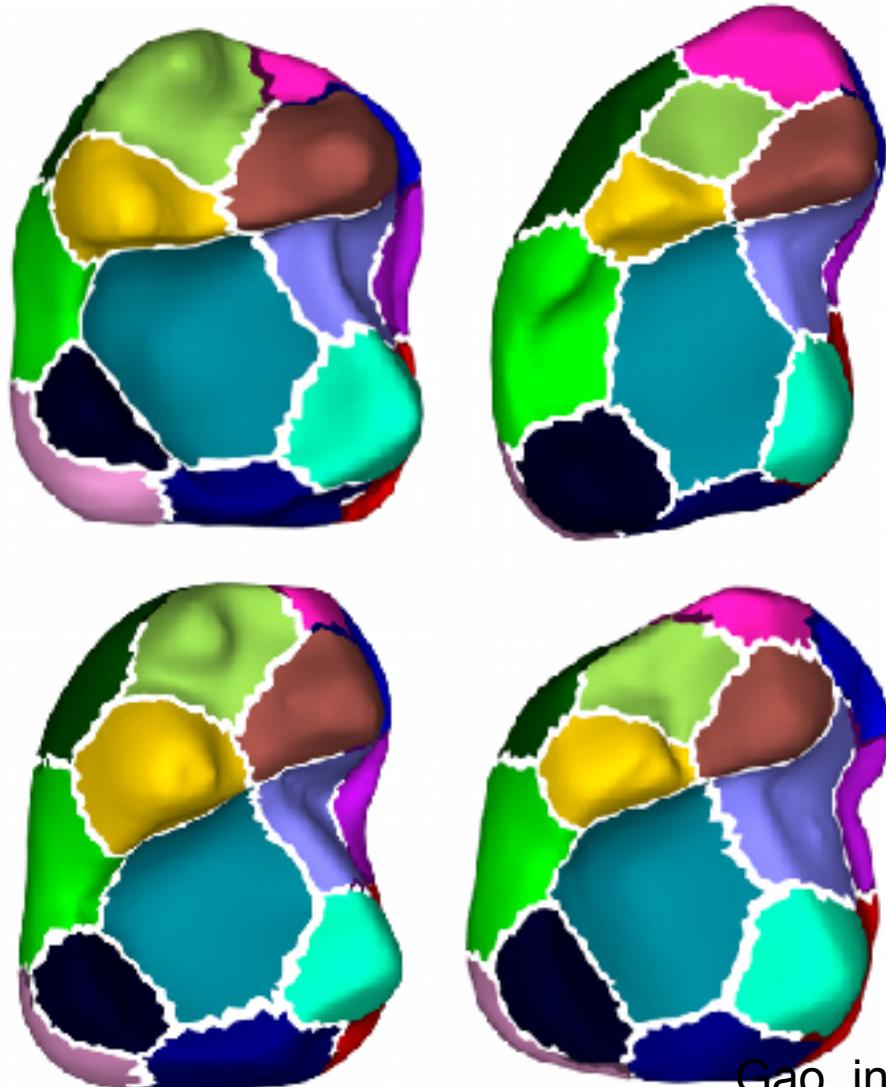
Diffusing surfaces to model evolution of form



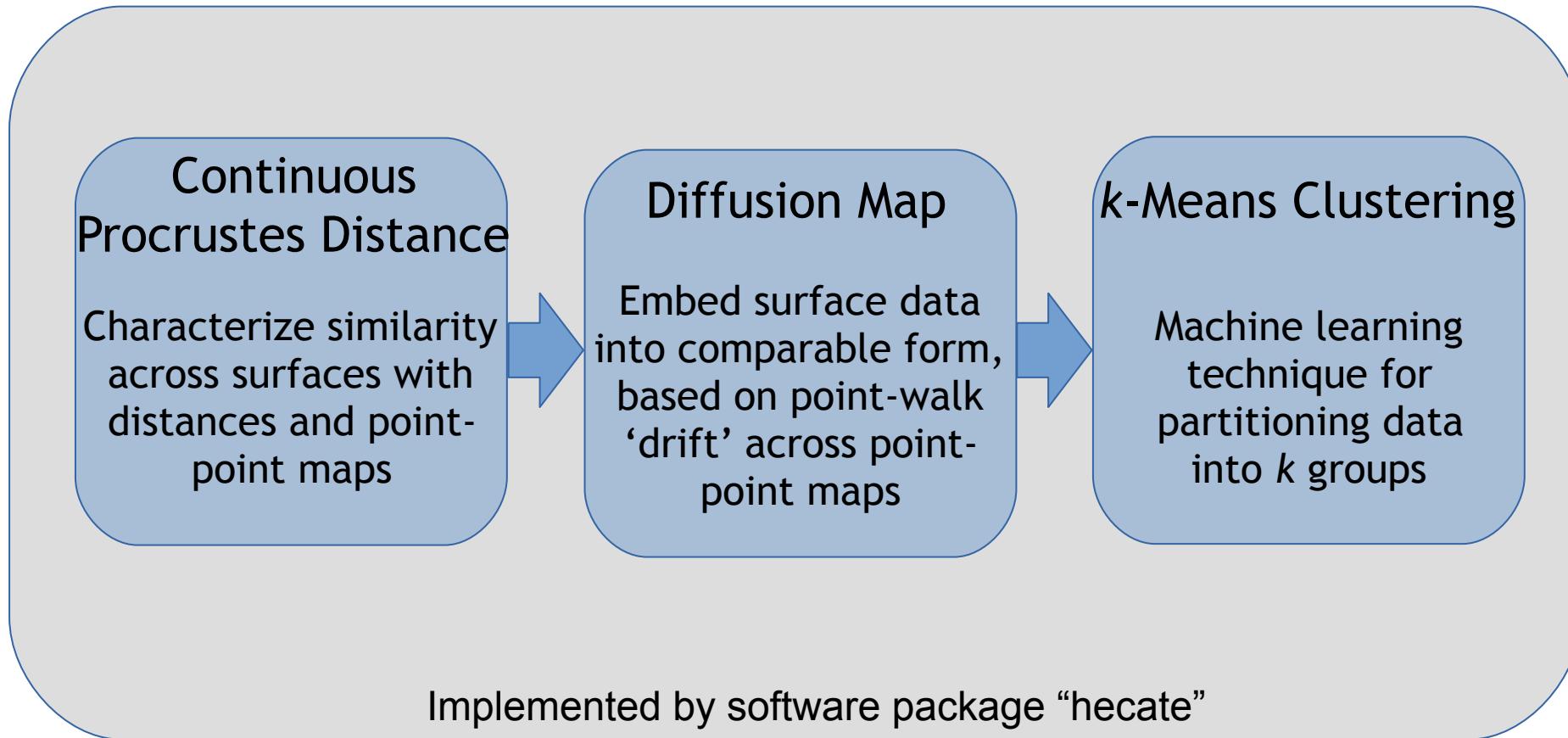
- Attempt to objectively recognize discrete surface region ‘characters’

Segmentation results

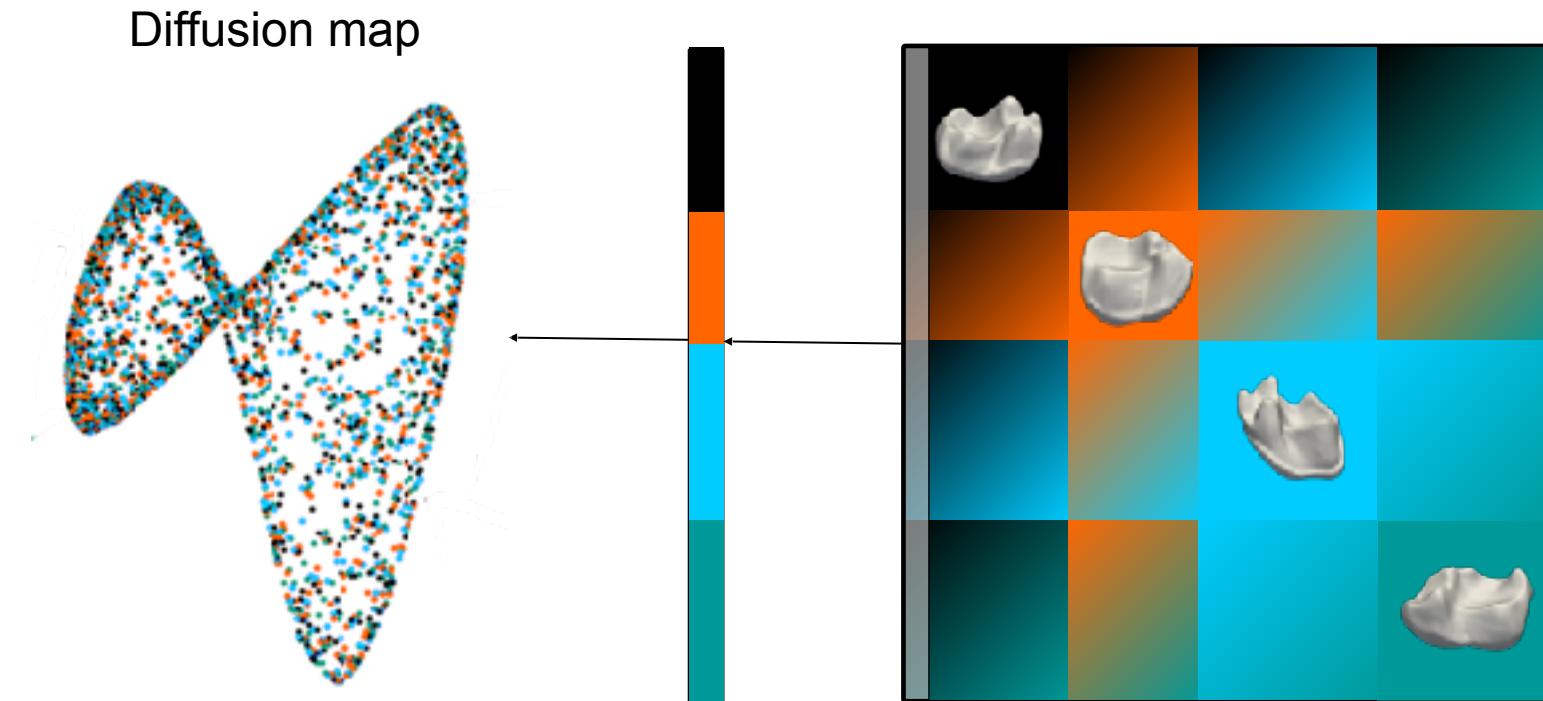
- Surface regions ($k = 15$) capture major features
 - Cusps
 - Basins
 - Side walls
- Surface feature presence/absence important for questions of taxonomy, phylogeny, function, etc.
- Surface regions may be able to provide an objective quantitative criterion for feature presence



Segmentation method

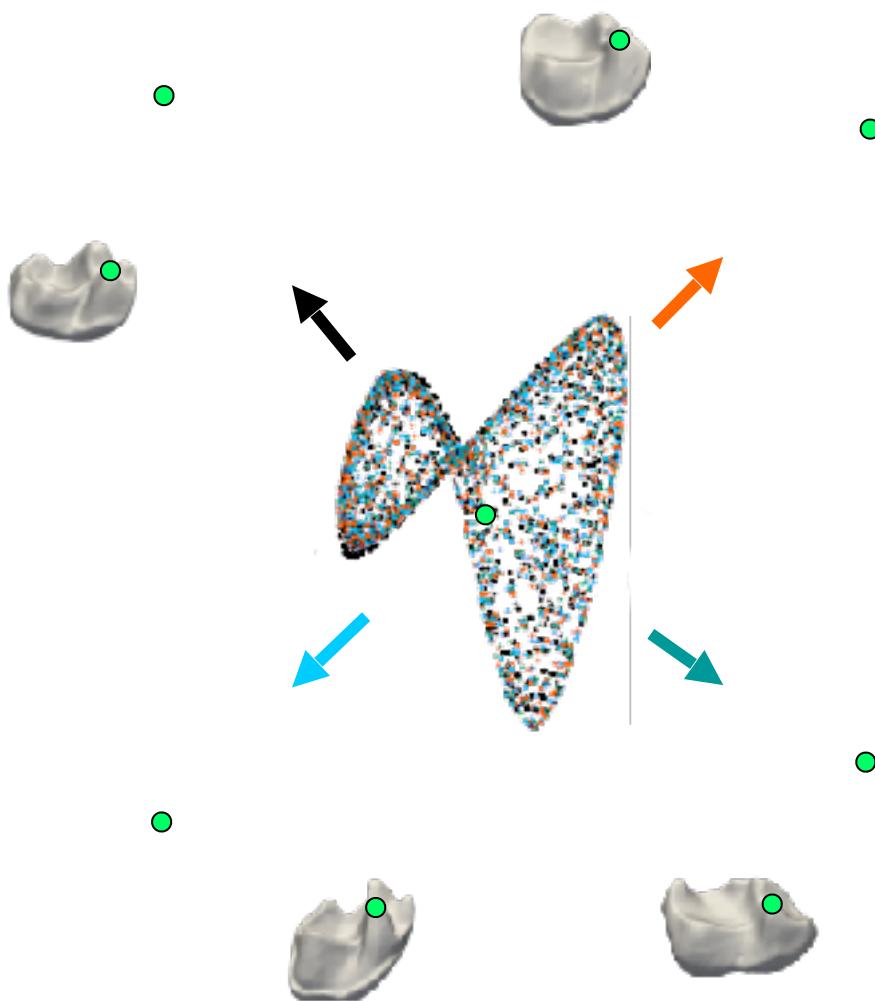


Diffusion operator



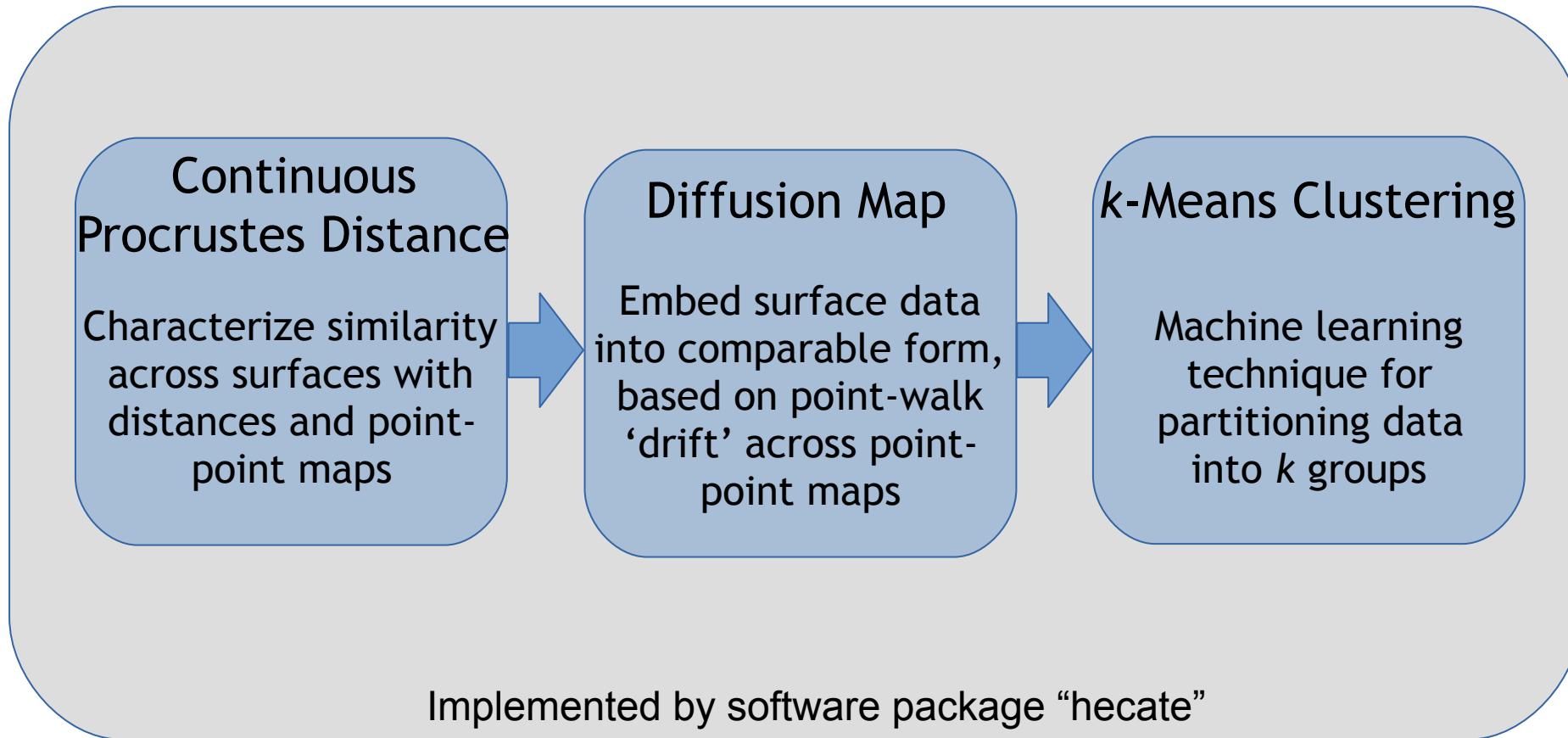
- Matrix column: Single point compared to **all other points from all meshes (including its own)**

Diffusion map

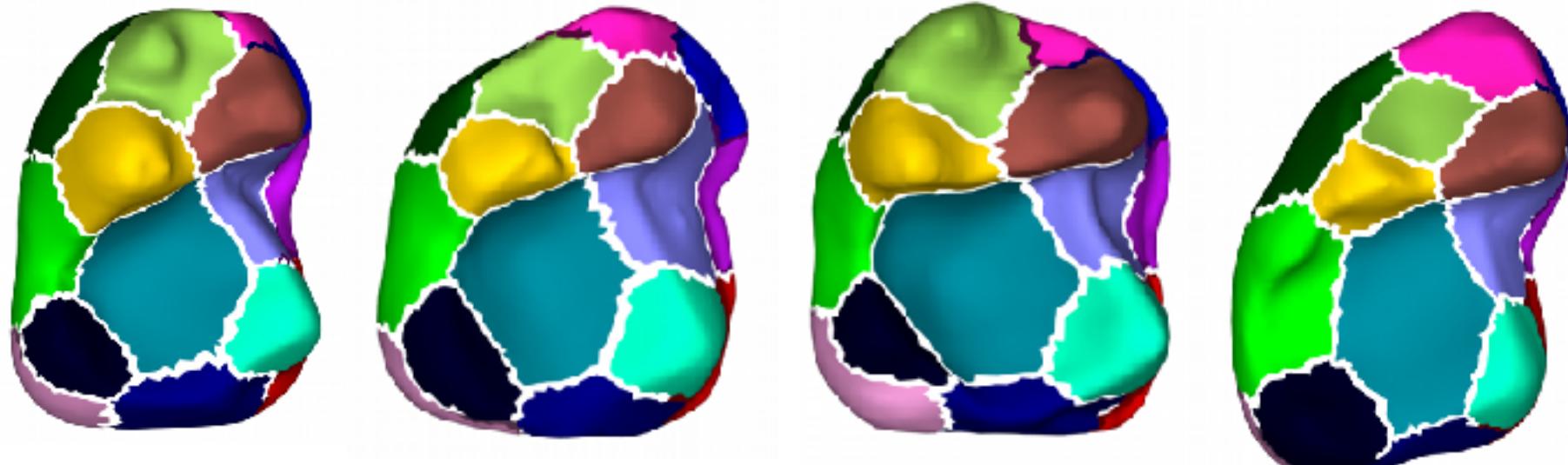


- Diffusion map coordinates embed mesh data in new multidimensional space
- Mesh data now similar in form
- Nearness in diffusion map = similarity across original surfaces

Segmentation method



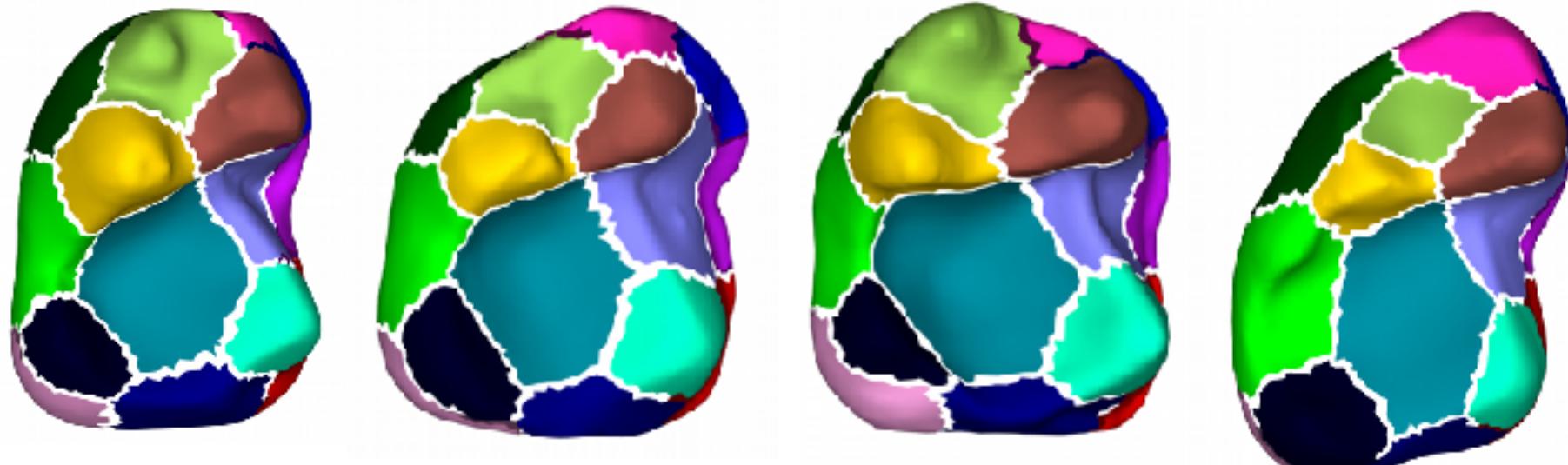
k -Means Clustering



$k = 15$

- Partition diffusion map coordinates into k groups, translate to original surface coordinates

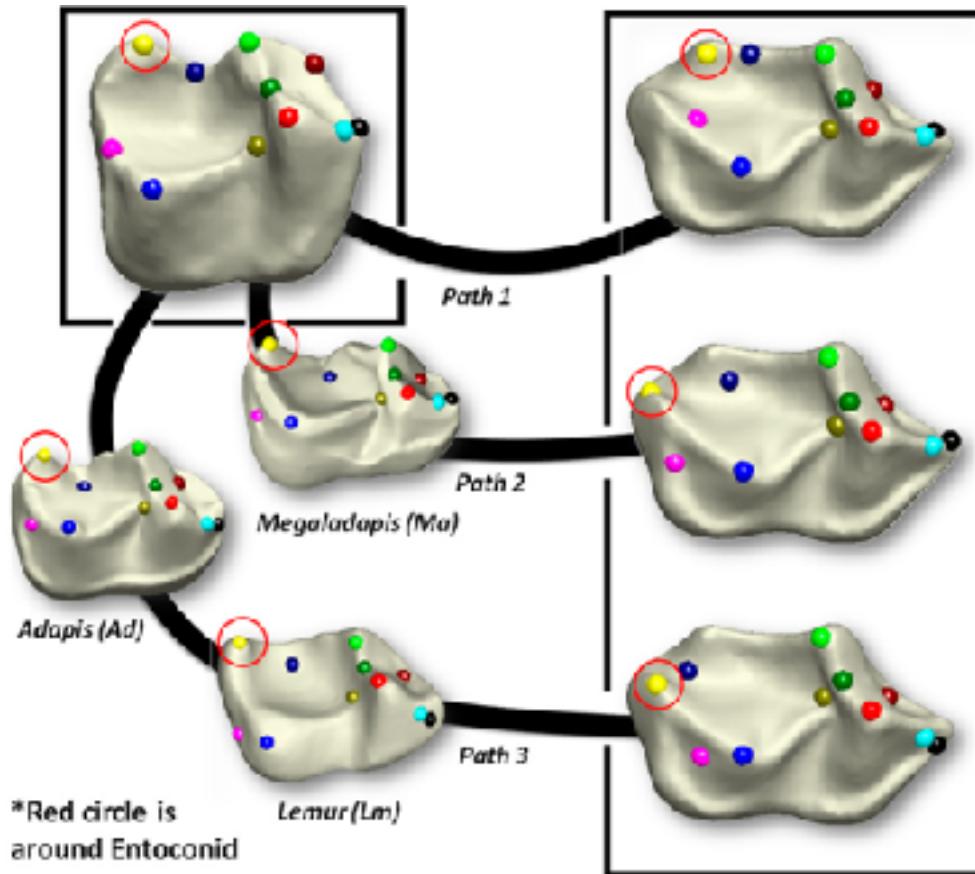
k -Means Clustering



$k = 15$

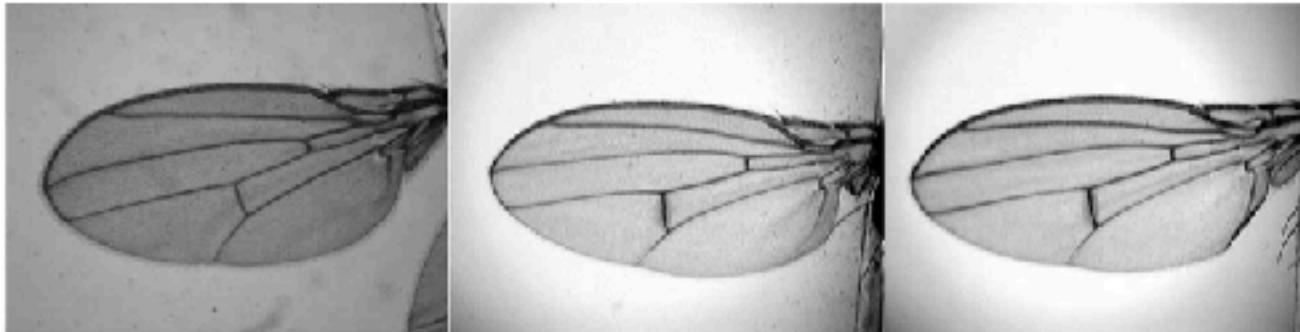
- Regions reflect local similarity based on probability of walking point ‘drift’ on surface maps

Limitations of Continuous Procrustes Approach

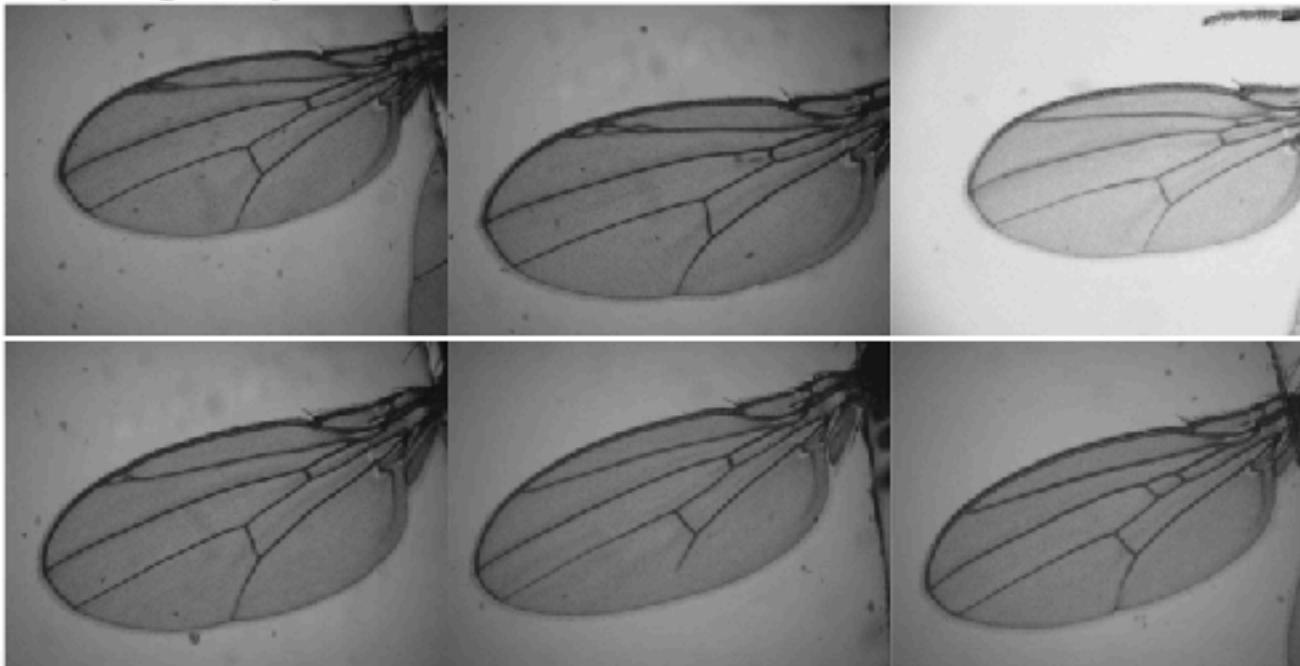


Limitations of Continuous Procrustes Approach

Normal fly wings [photos from David Houle's lab]:

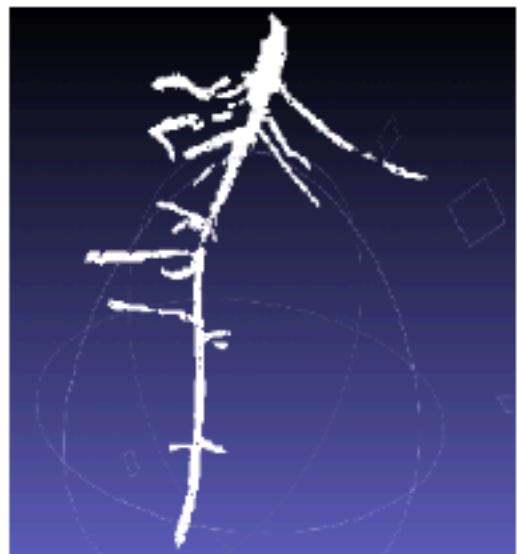


Topologically abnormal veins:



- Fly wings are not homologous

Limitations of Continuous Procrustes Approach



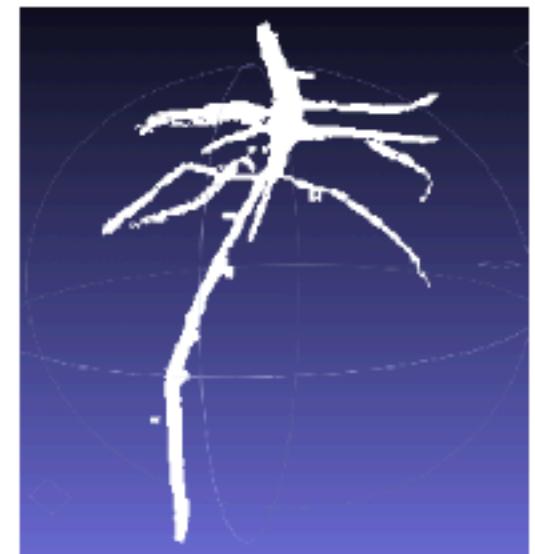
(a)



(b)



(c)



(d)

Part III: Topological data analysis

Goals and rationales

Model shapes without requiring landmarks or diffeomorphisms.

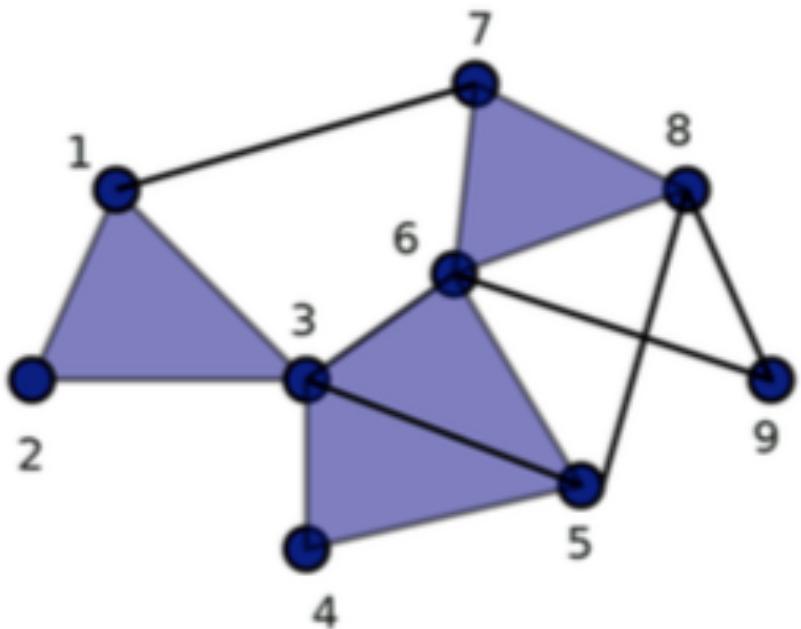
Transform the data/object into a representation that can be modeled using standard methods.

Topological summaries:

- (1) Euler characteristics
- (2) Persistent homology

Why summaries? We are not interested in capturing the exact geometry, certain perturbations or invariants should not matter.

Mesh data



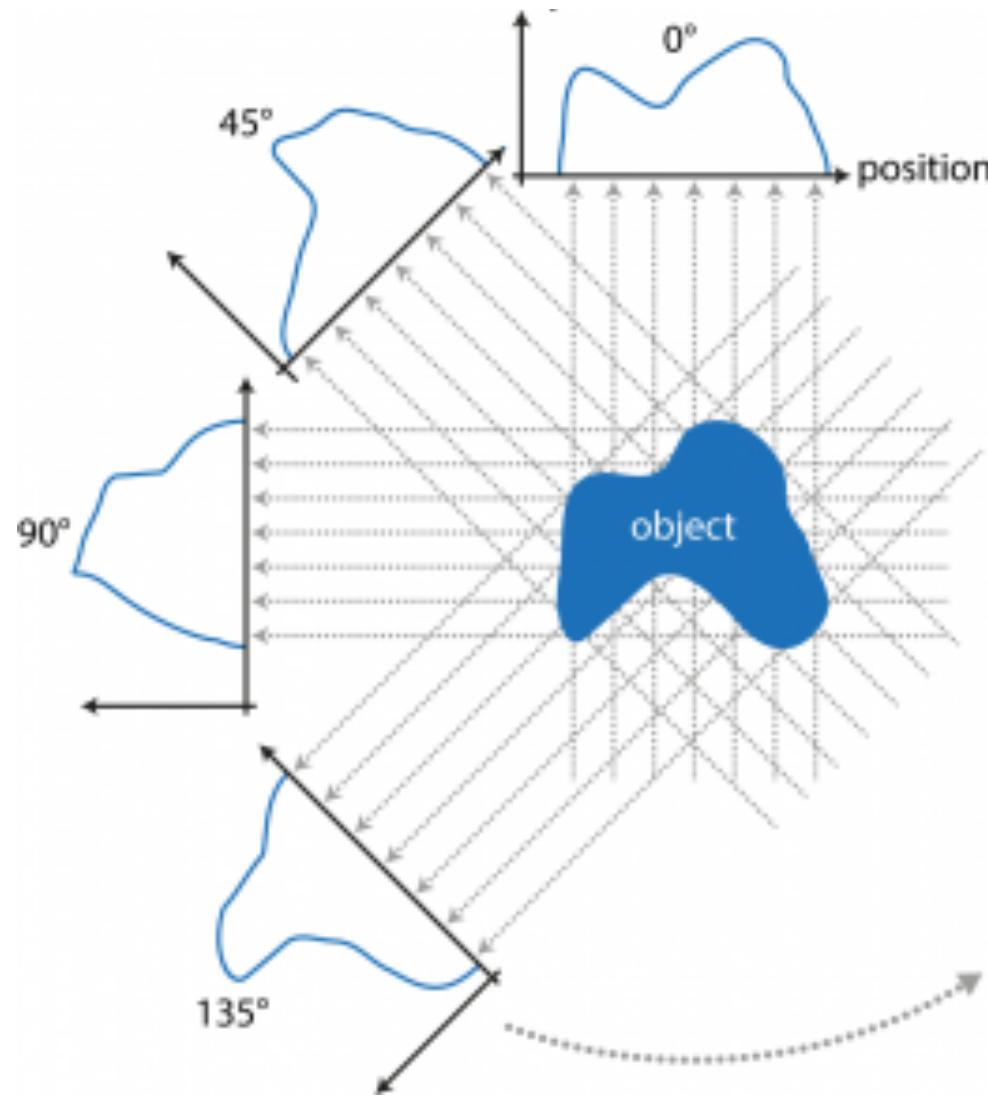
OFF
9 4 4 #V #T #E
0.85 2.37 0
0.32 1.22 0
1.89 1.22 0
1.89 0.32 0
3.45 0.63 0
2.72 1.83 0
2.84 2.91 0
4.07 2.31 0
4.54 1.21 1
3 1 2 3
3 3 4 5
3 3 5 6
3 6 7 8
2 1 7
2 5 8
2 6 9
2 8 9

Coordinates

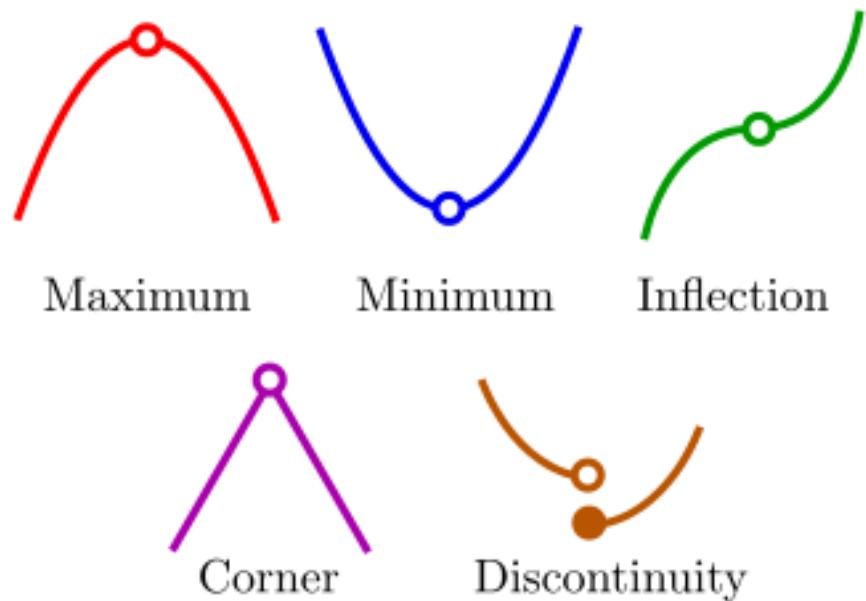
Triangles

Edges

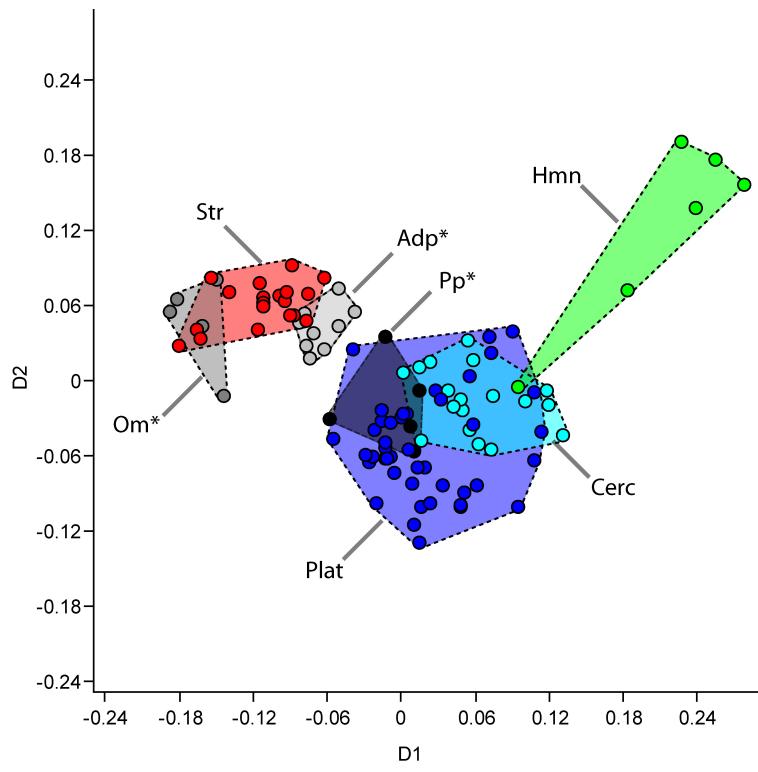
Euler characteristic representation



Critical points



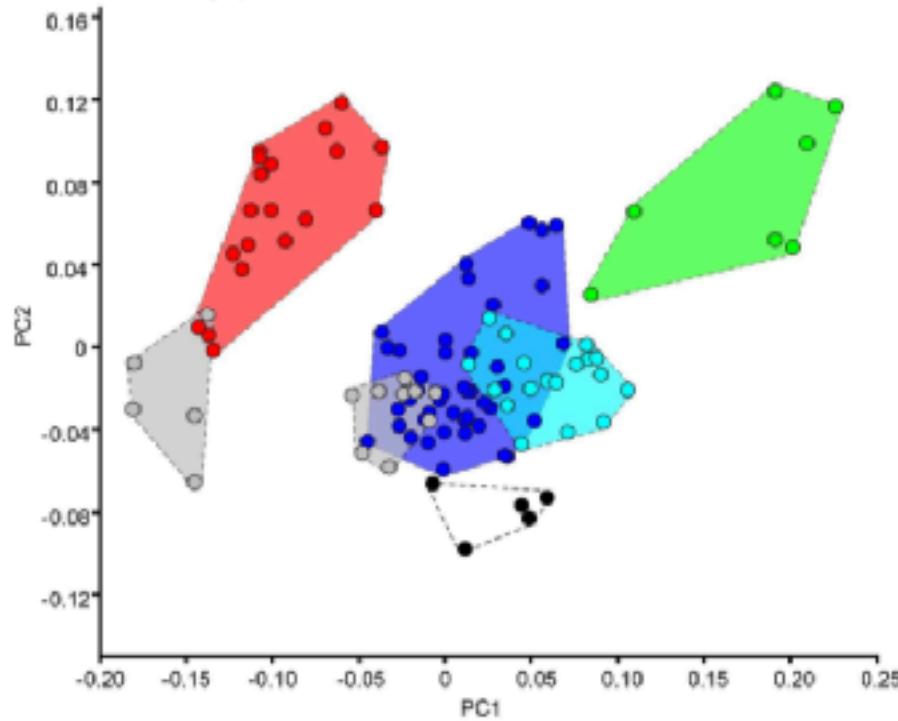
Primate calcanei



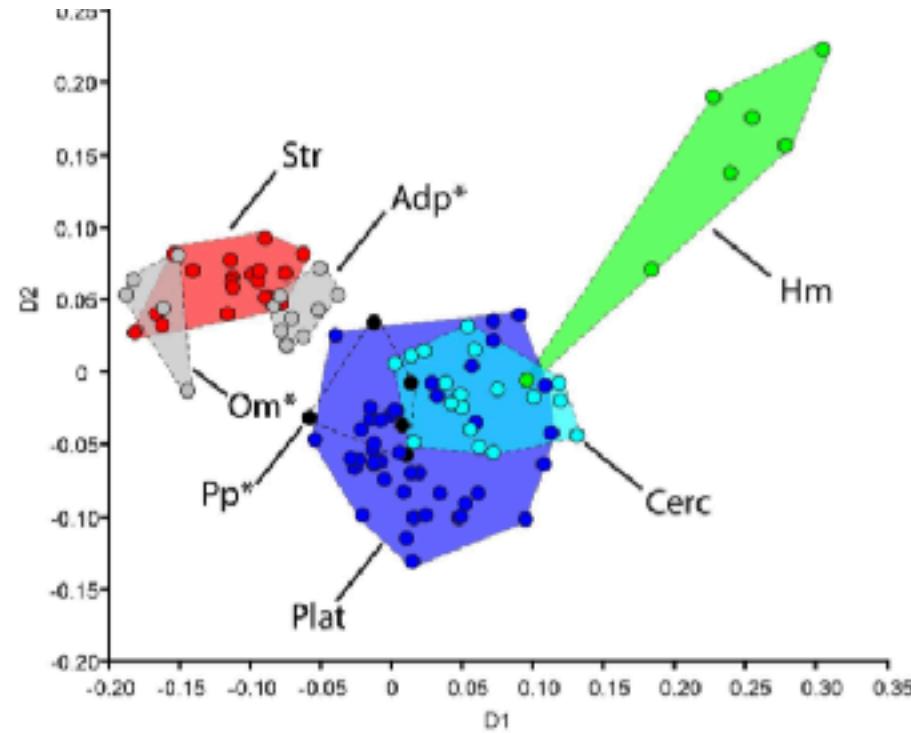
Phylogenetic groups of primate calcanei with 67 genera. Asterisks indicate groups of extinct taxa. Abbreviations: Str, Strepsirrhines; Plat, platyrhines; Cerc, Cercopithecoids; Om, Omomyiforms; Adp, Adapiforms; Pp, parapithecids; Hmn, Hominoids.

Comparing methods

A. Manually placed landmark data



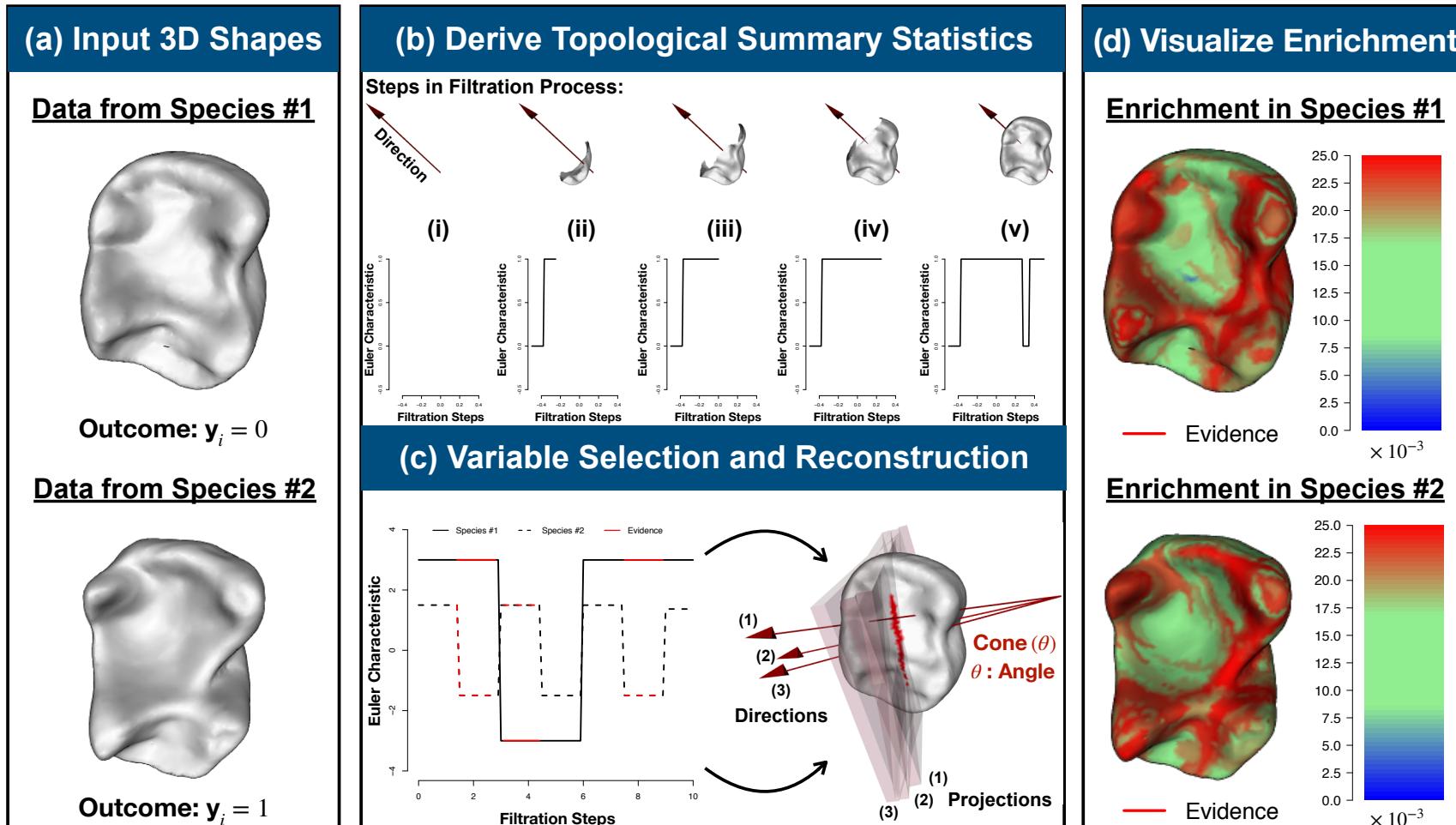
B. Topological data analysis



Comment from Doug

"In at least one way the method matched shapes with family groups better than any of the other previous methods... it linked a *Hylobates* specimen with the the other ape specimens (*pan*, *gorilla*, *pongo*, and *oreopithecus*). Previous both hylobatids (which ARE apes) always ended up closest to some *Alouatta* specimens."

Subimage Selection



Wang, Bruce, et al. "SINATRA: A Sub-Image Analysis Pipeline for Selecting Features that Differentiate Classes of 3D Shapes." *bioRxiv* (2019): 701391.

On the horizon



On the horizon

- Statistical model without landmarks
- Fiber bundle for unifying shape theories
- Building a cauliflower model for shapes