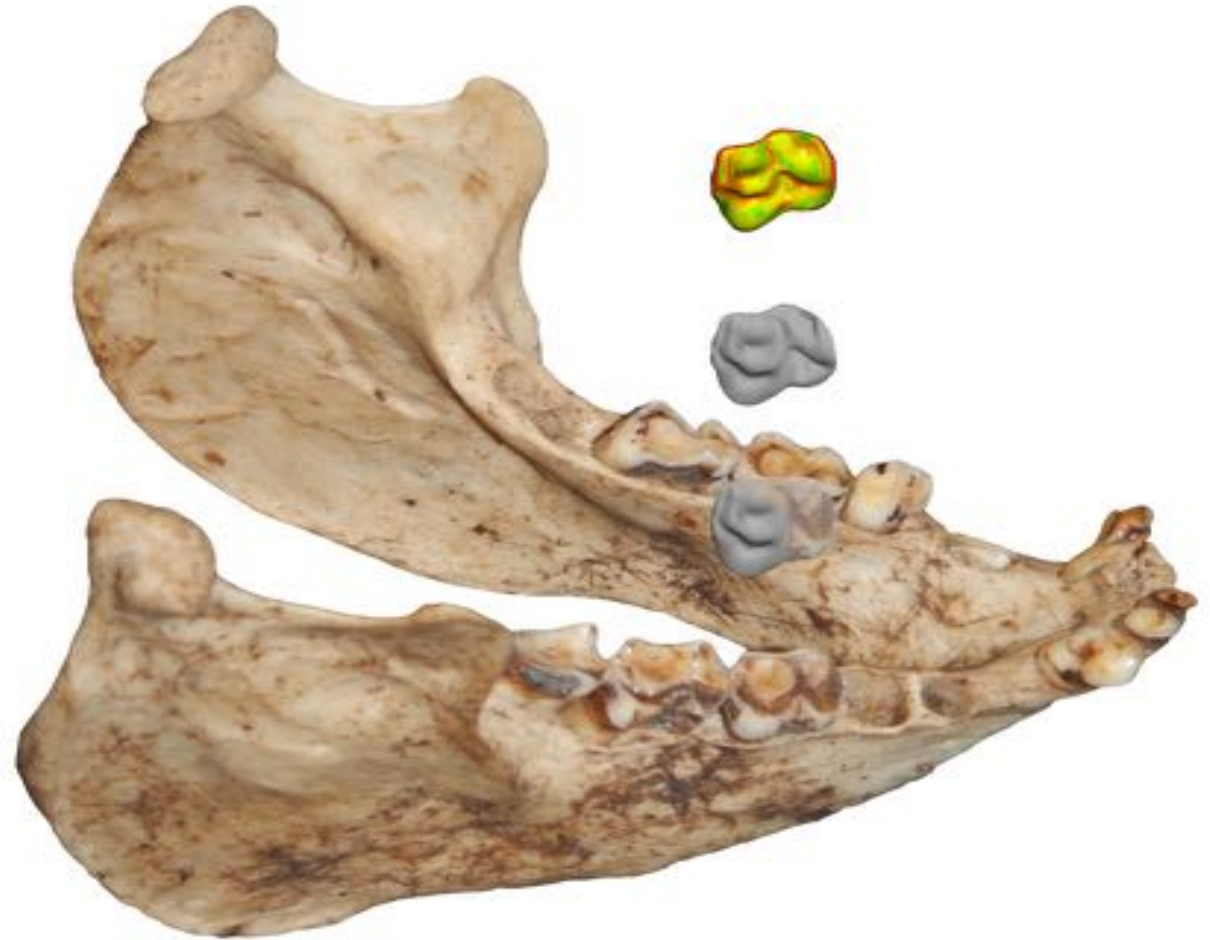


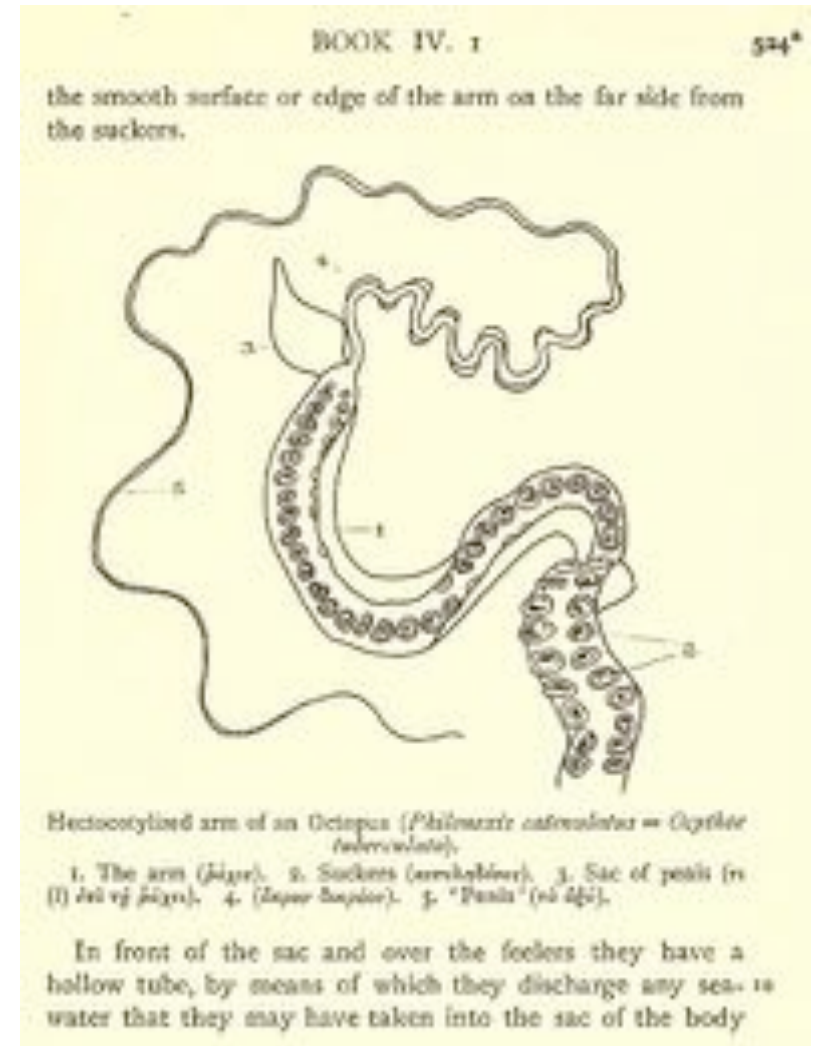
Automated shape specification using digital data

Ethan Fulwood
Julia Winchester
Tingran Gao
Doug Boyer
Ingrid Daubechies



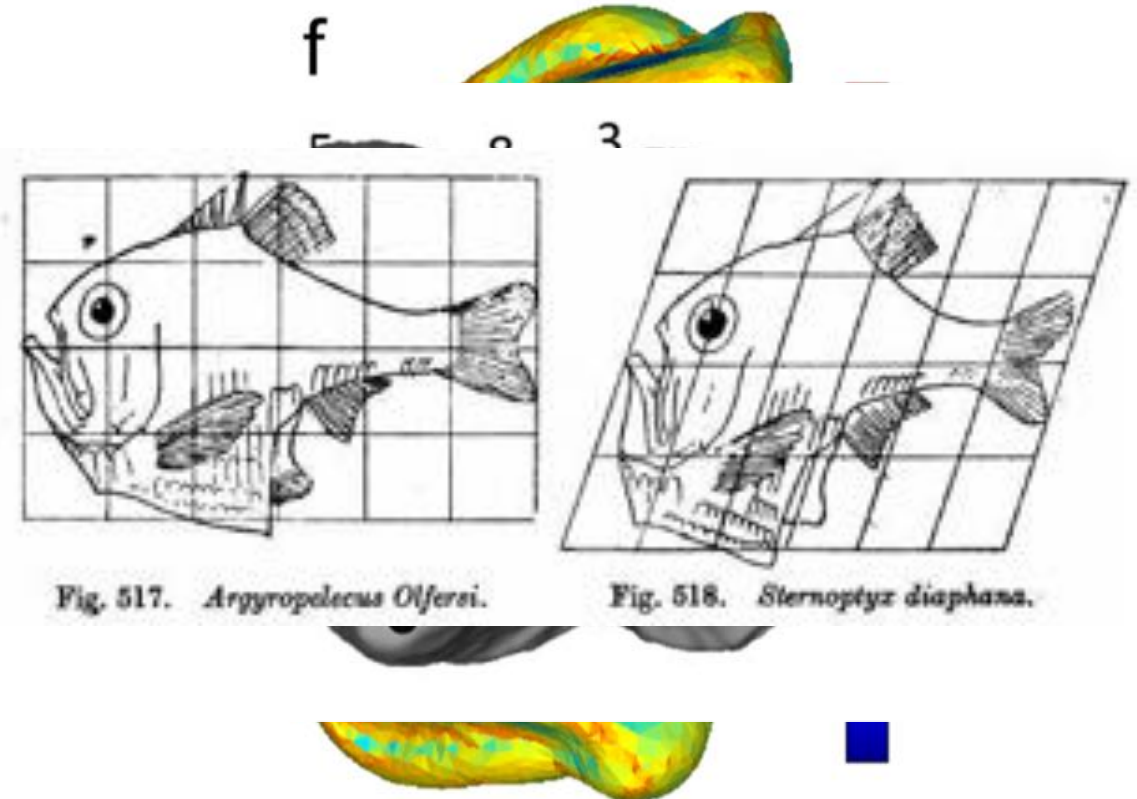
Thinking about Form

- Everything begins with Aristotle (*De Partibus Animalium*, *De Historia Animalium*, 350 BC)
 - How does *form* relate to the *function* of a structure or an organism?
 - What does form tells us about the natural groups organisms compose?
- Versions of these questions continue to animate the study of functional morphology, systematics, and macroevolution.



Thinking about form

- “Shape description”
 - Shape metrics of hypothesized functional significance
 - Measurements of limb ratios
 - Dental topography metrics
- “Shape specification”
 - Encoding the whole shape of an object
 - Qualitative observation
 - Combinations of linear measurements
 - Shear mapping (D’Arcy Thompson)
 - Geometric morphometrics

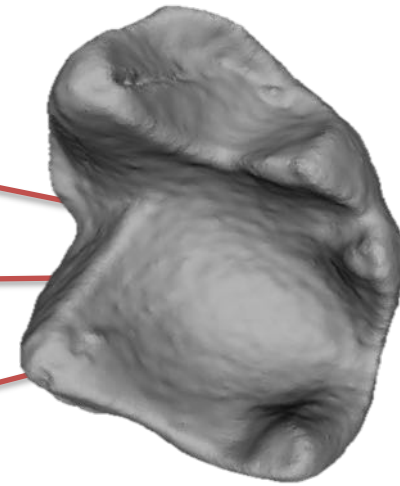


Case Study – Specifying the Paraconid Qualitatively



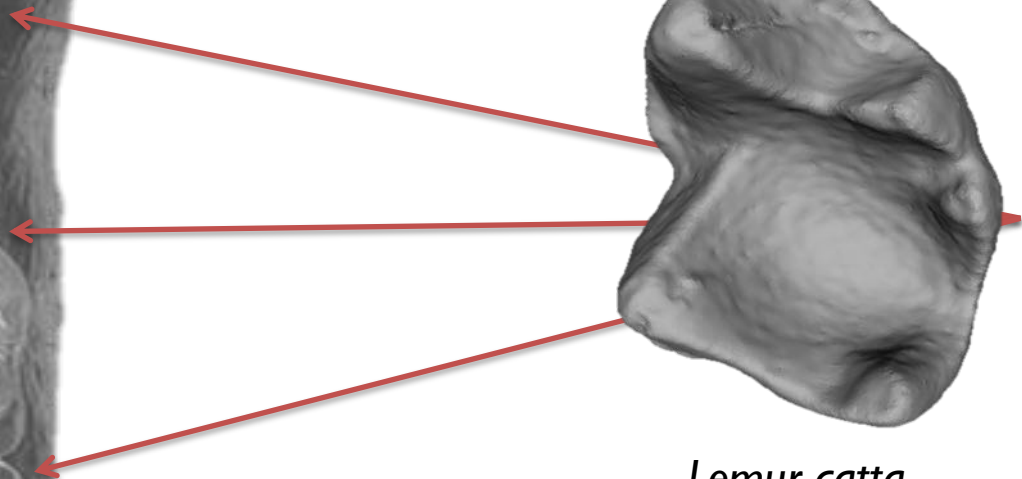
Teilhardina asiatica

Problem

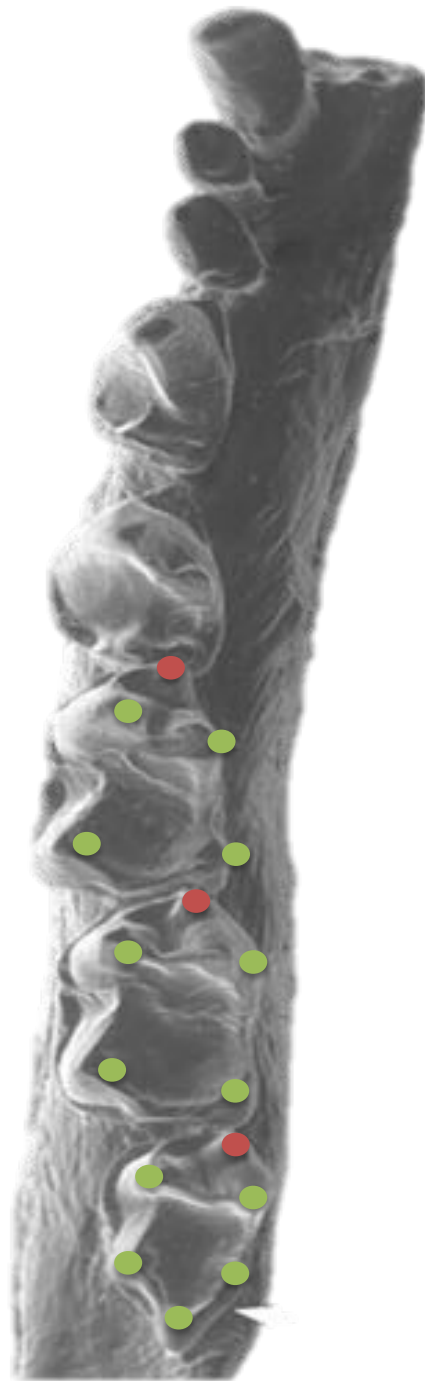


Lemur catta

Paraconids of M_{1-3}



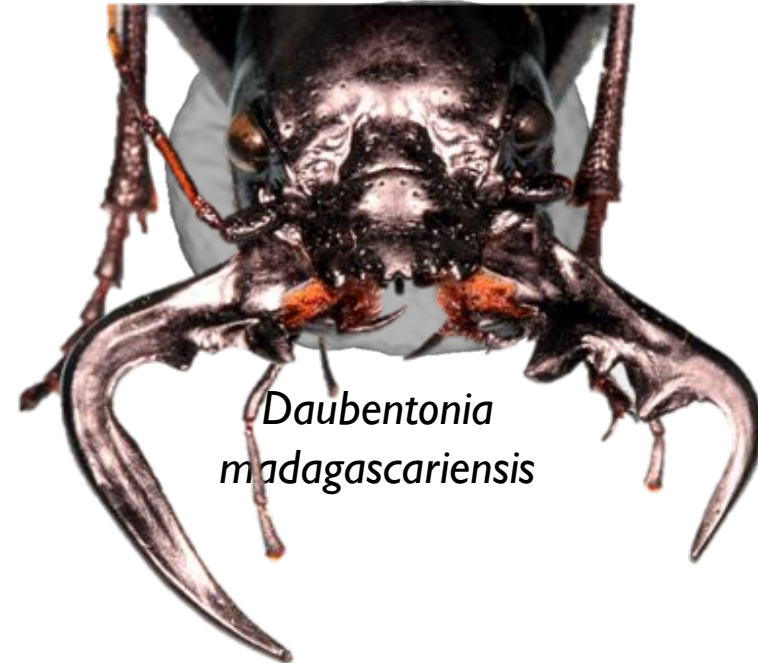
Case Study – Specifying the paraconid using traditional geometric morphometrics



Teilhardina asiatica

Problem

Paraconids of M_{1-3}



*Daubentonia
madagascariensis*

Automated methods

Advantages

1. Can be applied to large samples efficiently
2. No need to accurately identify “homologous” structures
3. Gives a better accounting of overall shape.

Implementations

- Automated 3D geometric morphometrics (Auto3dgm)
 - Uses “pseudolandmarks” distributed across a surface to align shapes and facilitate the calculation of distances among them (Boyer 2015).
- Diffusion-based segmentation (hecate)
 - Uses point to point similarities between surfaces to probabilistically identify regions of corresponding shape on a sequence of structures.

Steps to auto3dgm

1. Generating pseudolandmarks
2. Initial pairwise alignment
3. Construction of a minimum spanning tree
4. Permutation and rotation matrices are propagated along the minimum spanning tree to produce transitive alignments

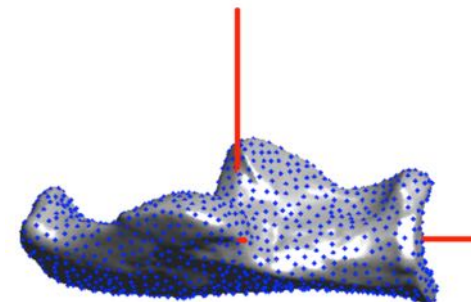
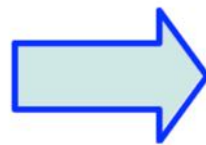
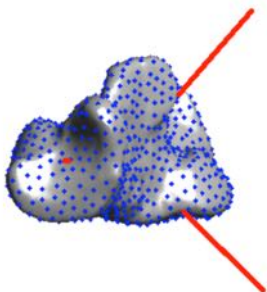
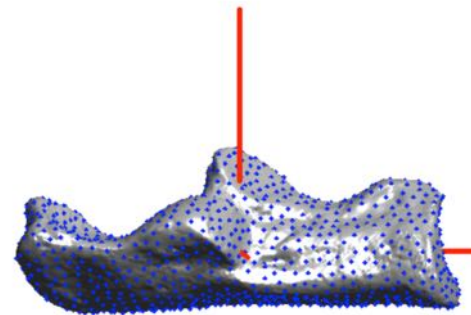
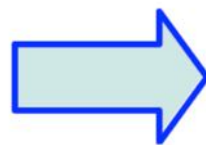
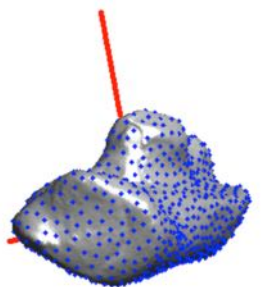
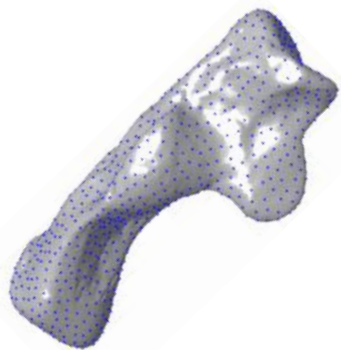
Traditional landmarks vs pseudolandmarks



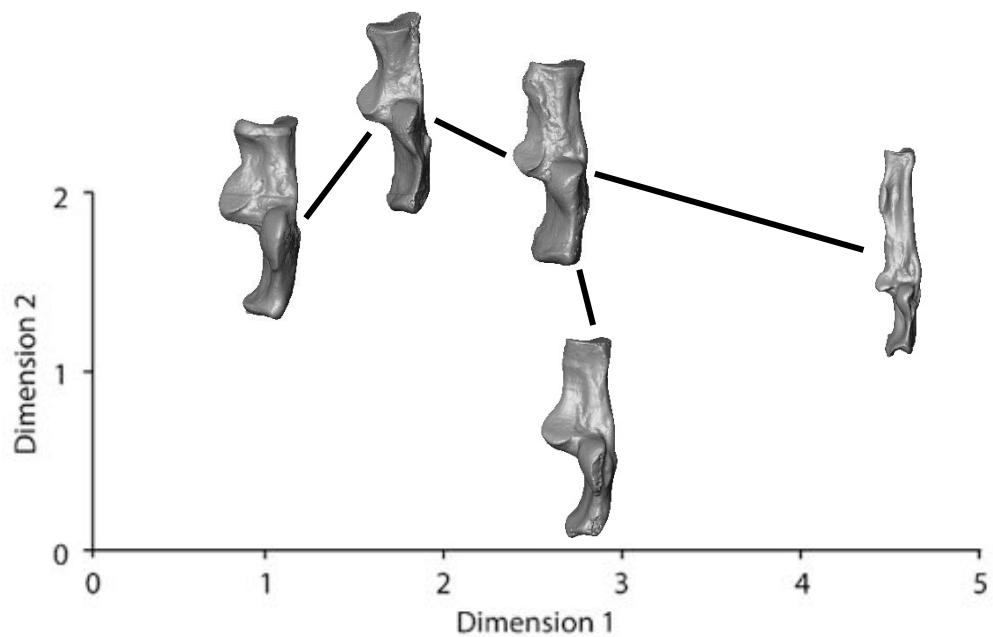
Traditional landmarks vs pseudolandmarks



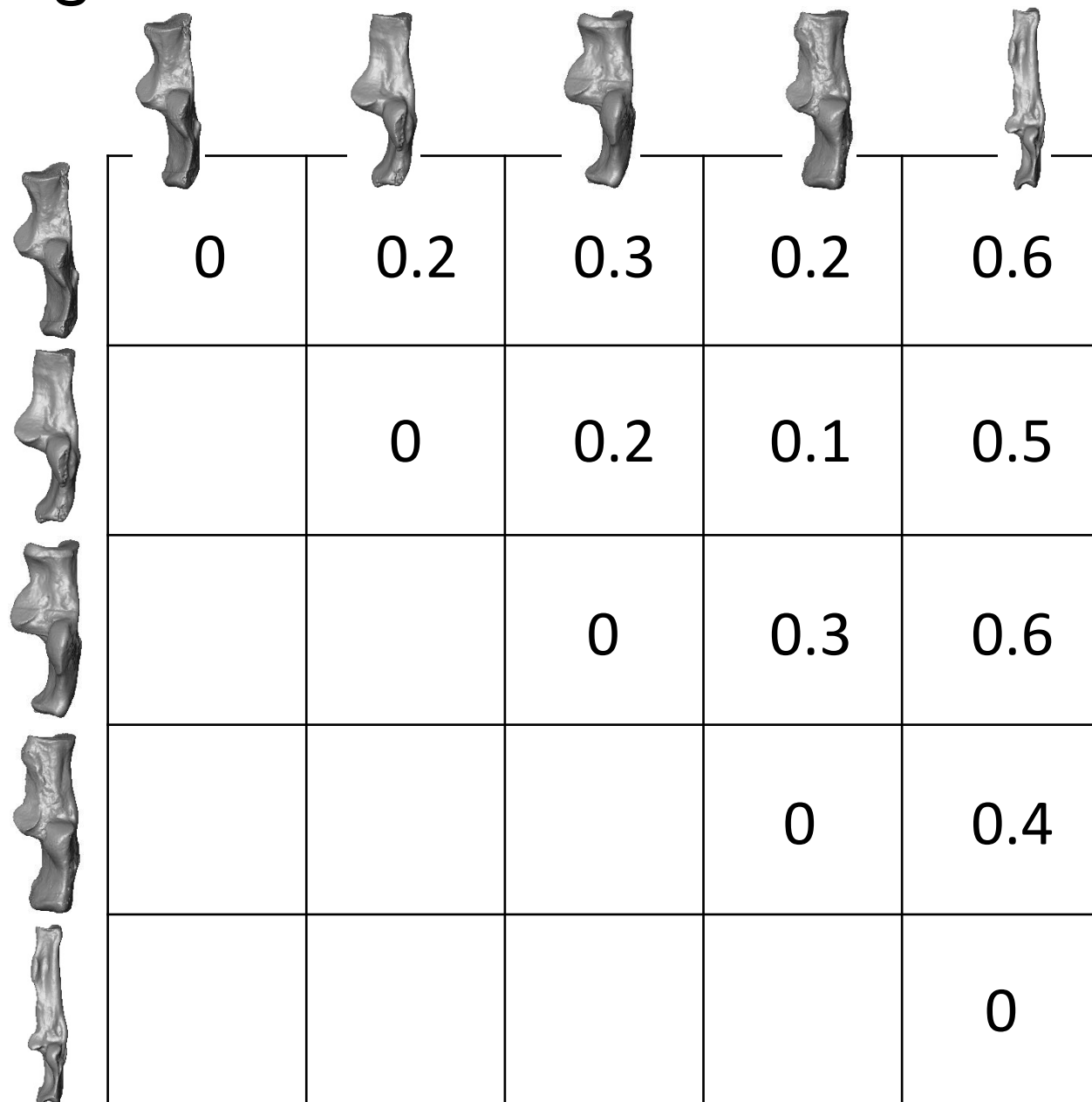
Initial pairwise alignment



Construction of a minimum spanning tree

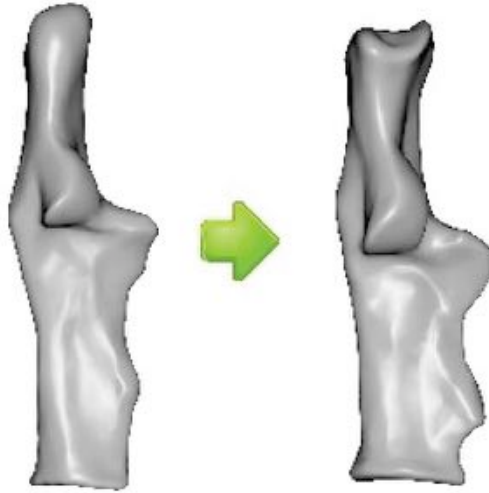


PCD matrix



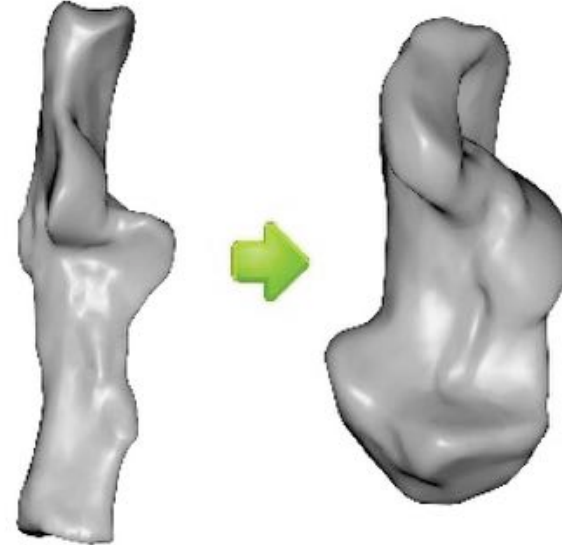
Purpose of MST & Propagation

Similar shapes align well



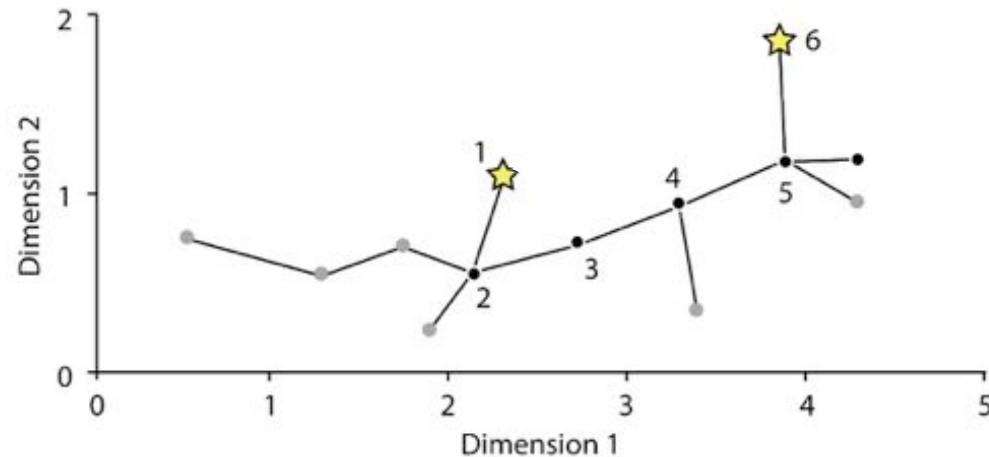
PCD = 0.1

Dissimilar shapes may align incorrectly



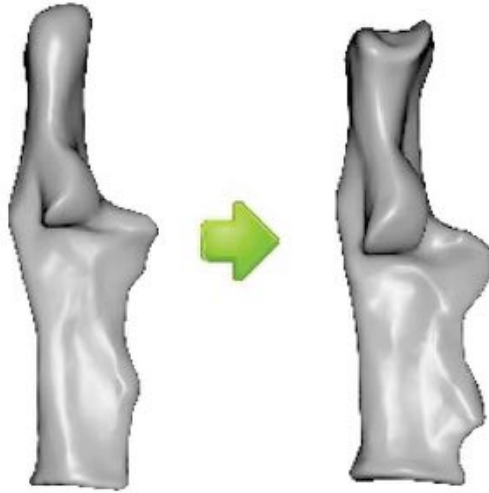
PCD = 0.7

All Procrustes distances used to construct initial minimum spanning tree



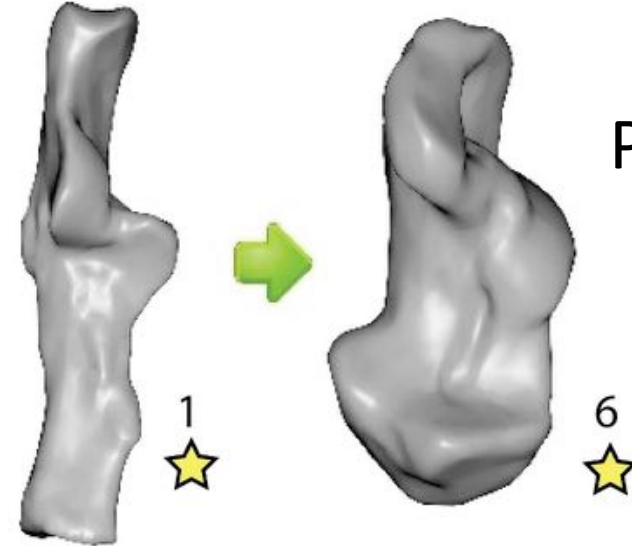
Purpose of MST & Propagation

Similar shapes align well



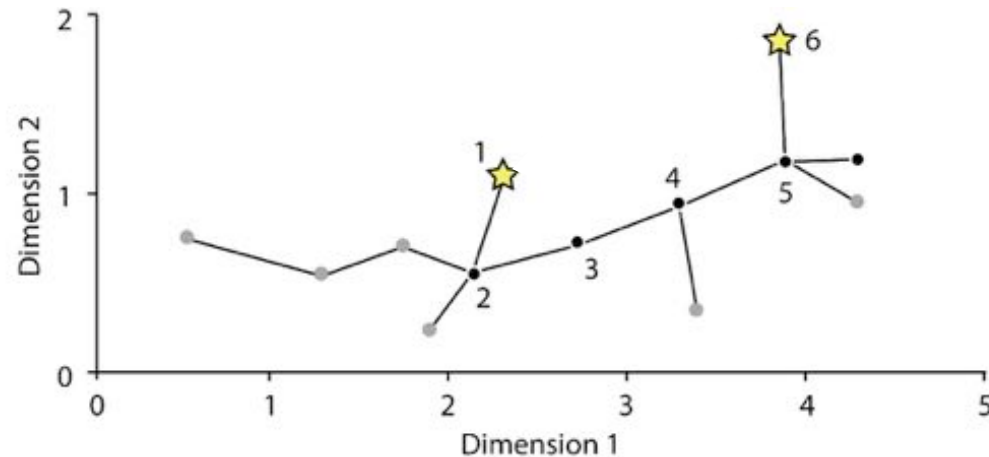
PCD = 0.1

Dissimilar shapes may align incorrectly



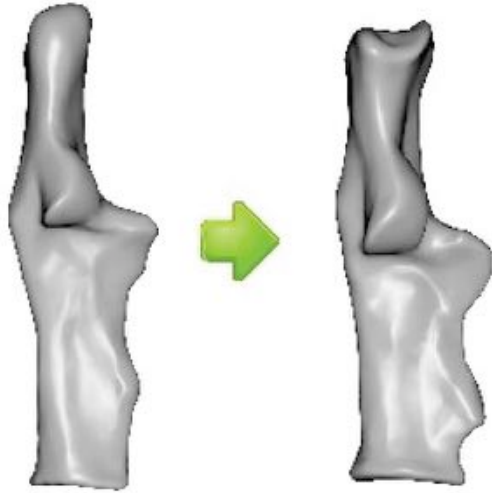
PCD = 0.7

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



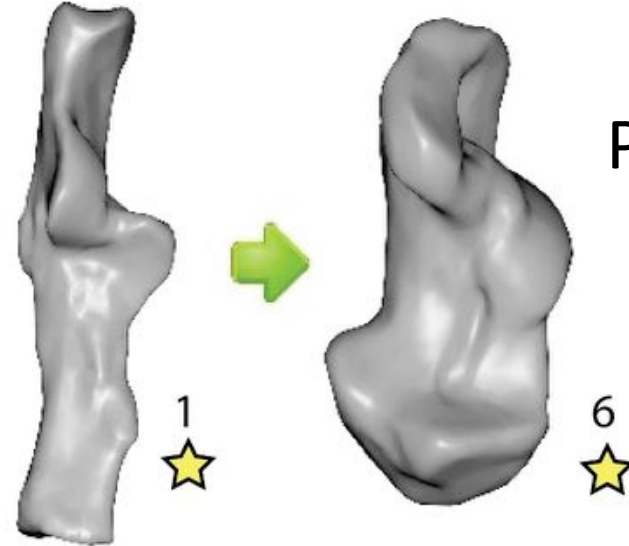
Purpose of MST & Propagation

Similar shapes align well



PCD = 0.1

Dissimilar shapes may align incorrectly

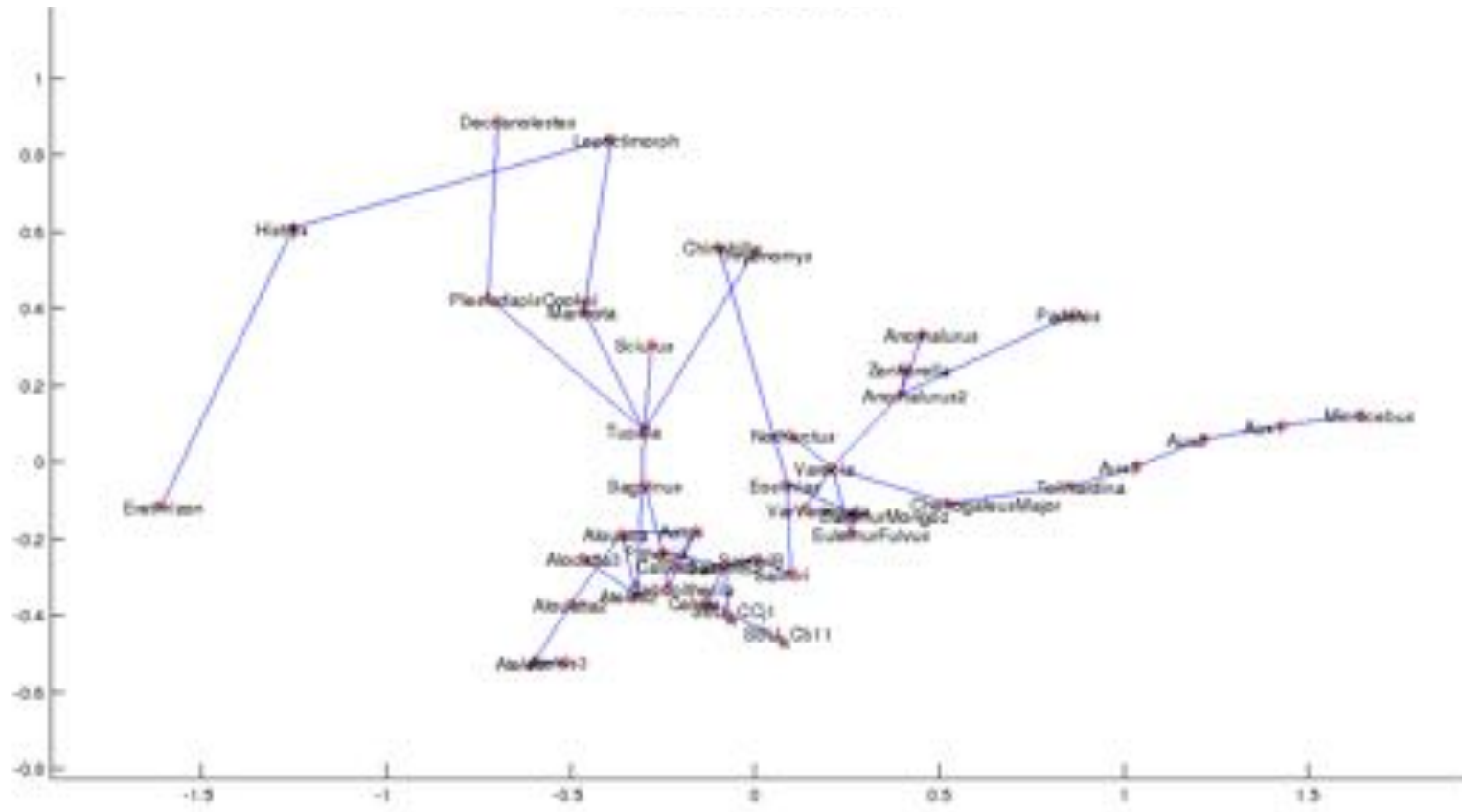


PCD = 0.7

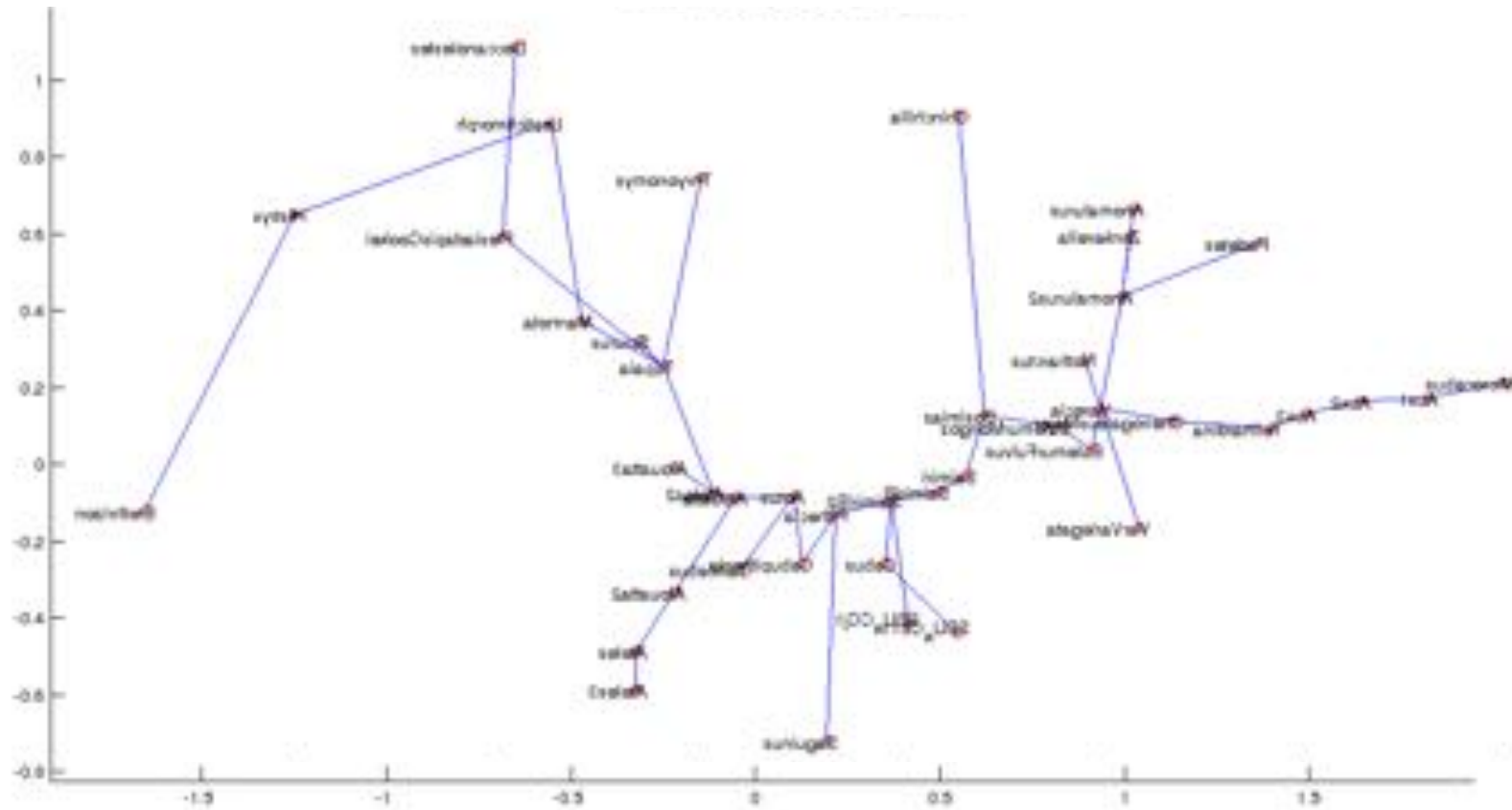
The MST path can be used to find correct alignments, and recompute distances



MST w initial correspondences

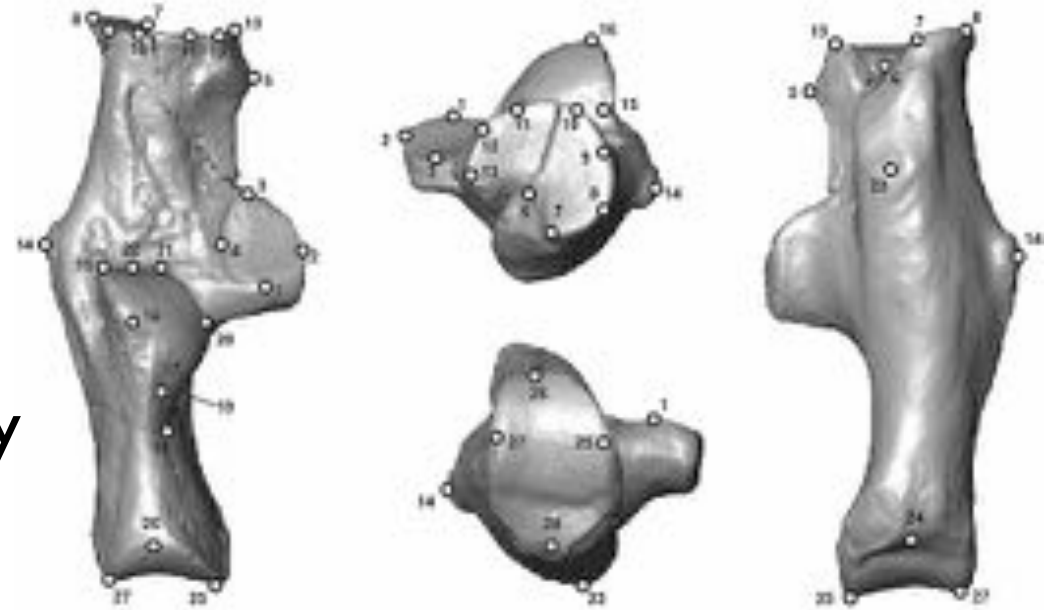


MST w transitive correspondences



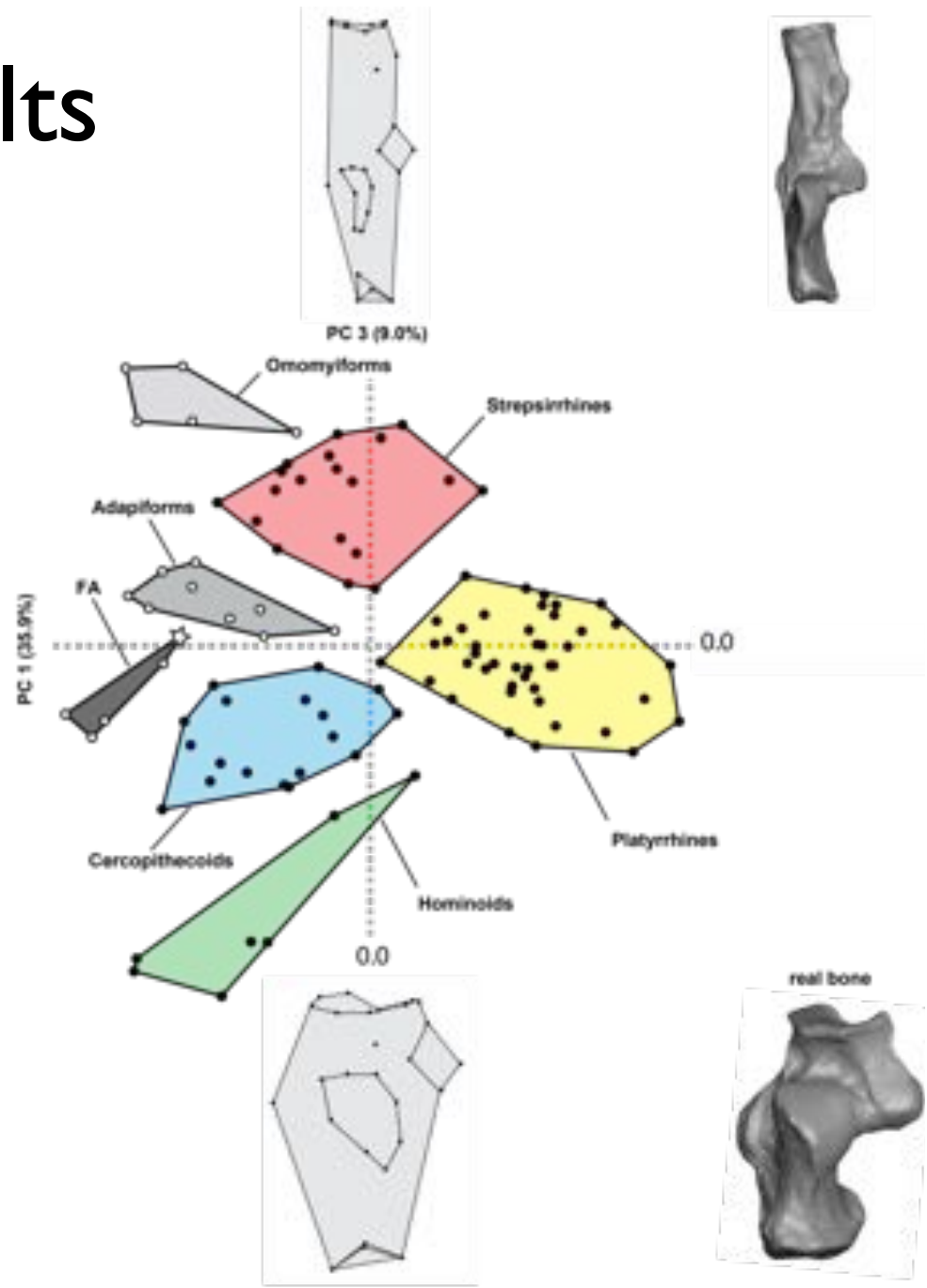
Comparison to researcher-based data set

- 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

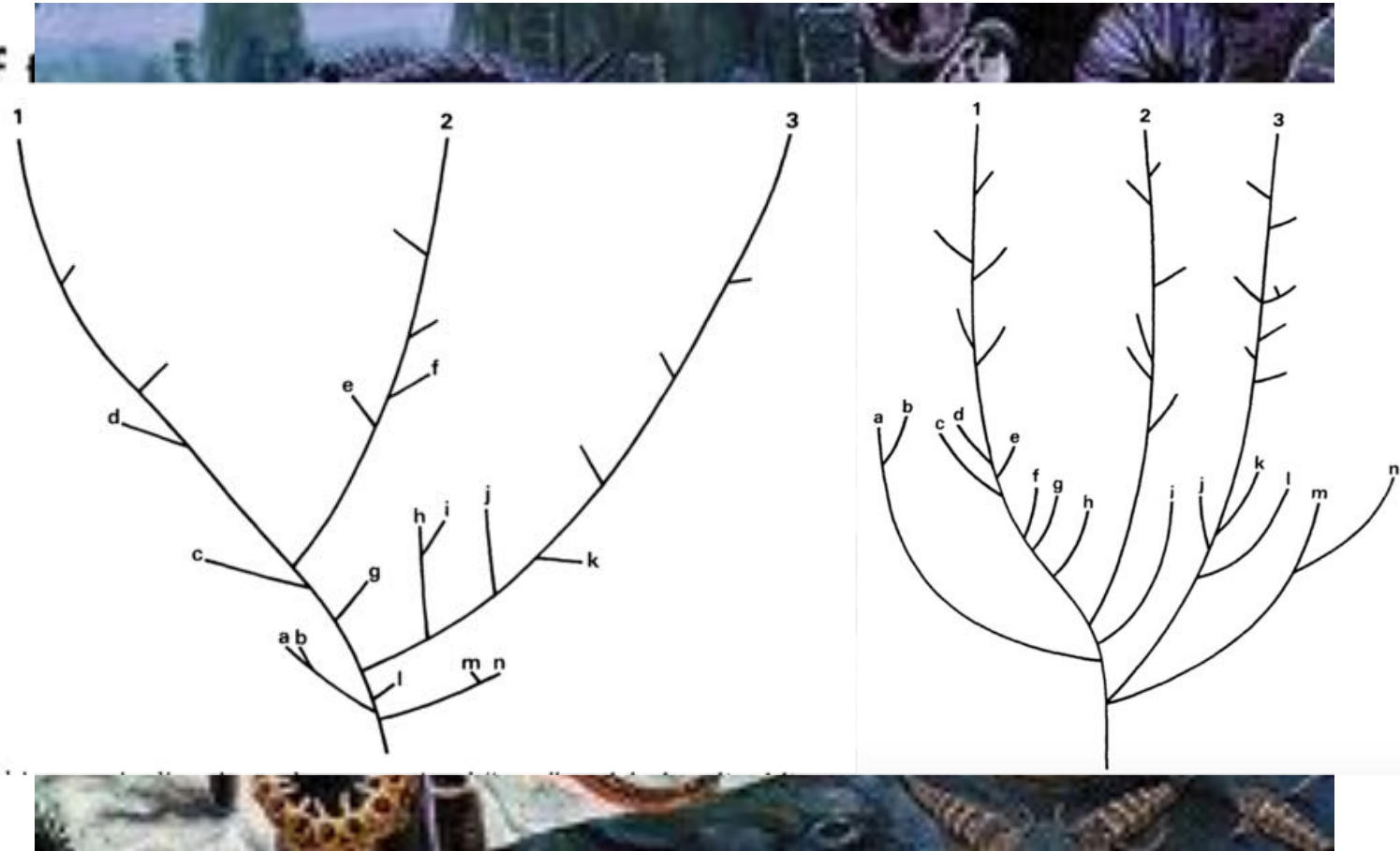


Potential application – Quantifying disparity

Paleobiology, 17(4), 1991, pp. 411–423

**The disparity of
fauna and the
why we must**

Stephen Jay Gould

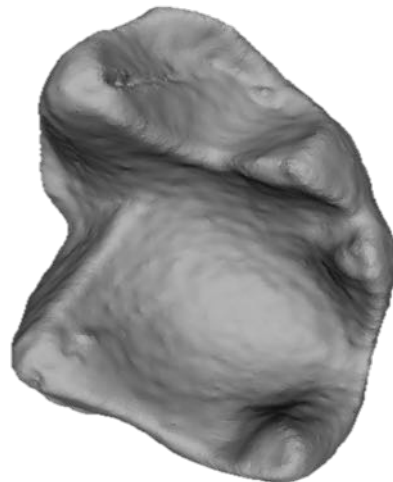


Conclusions

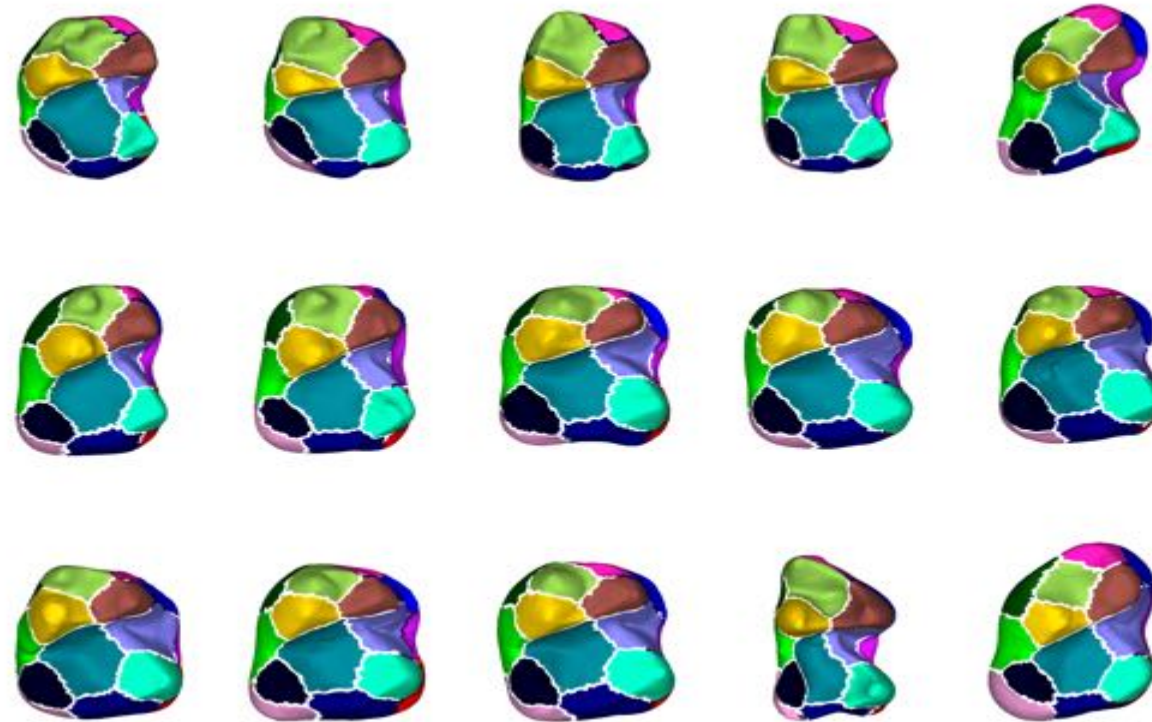


Auto3dgm is available for application to your data through an R package (“auto3dgm”) and an implementation in Matlab distributed through github.

What about the parts of larger shapes? Can they be identified using automated methods?



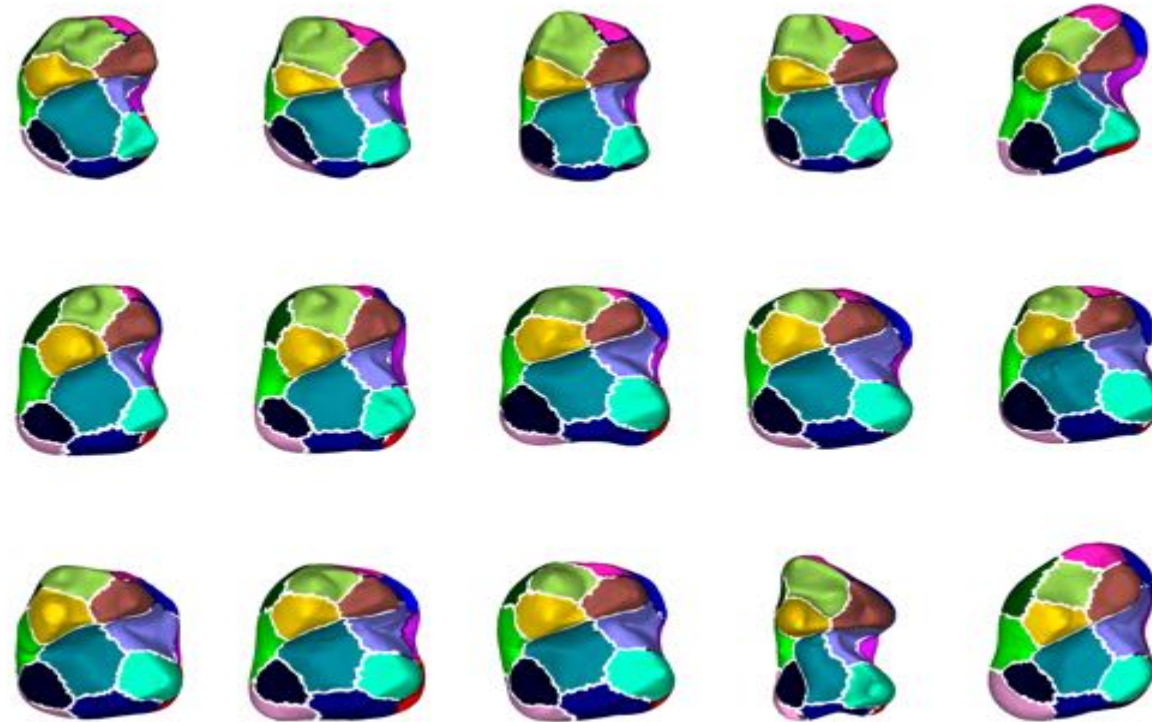
Consistent segmentation of biological surface regions



$k = 15$

- Identify k surface regions across variable shape sample (k is user specified)

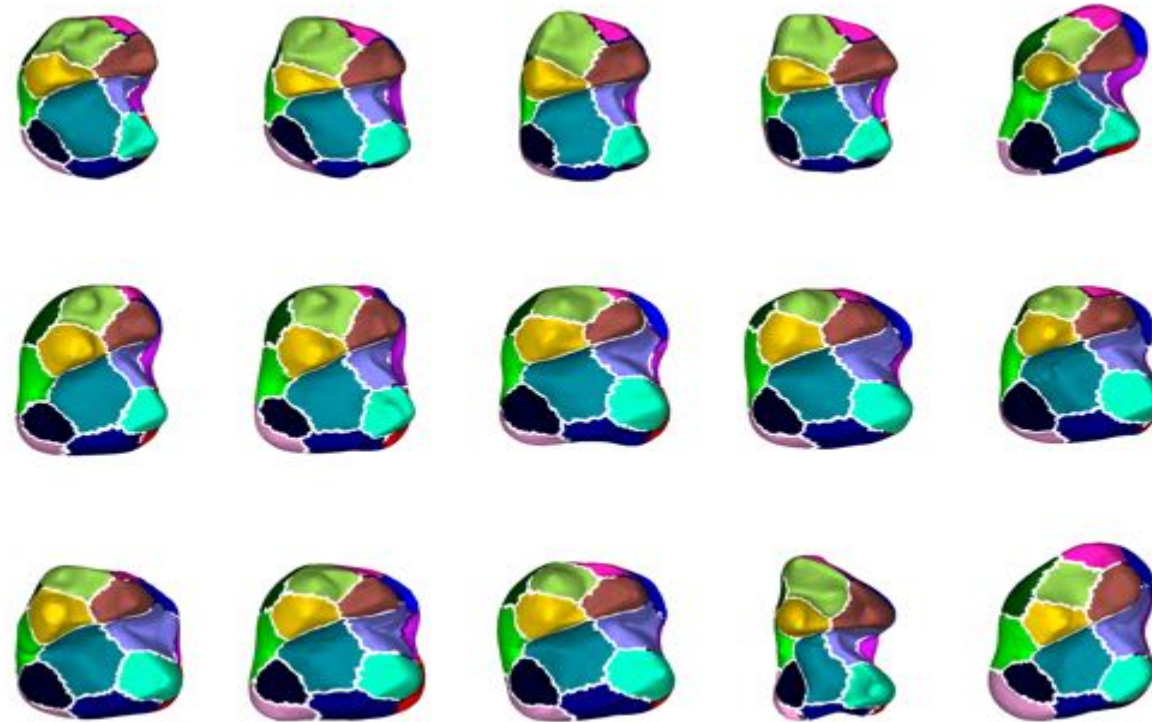
Consistent segmentation of biological surface regions



$k = 15$

- Surface regions reflect local shape similarity

Consistent segmentation of biological surface regions



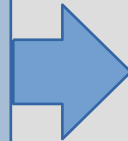
$k = 15$

- Attempt to objectively recognize discrete surface region 'characters'

Segmentation method

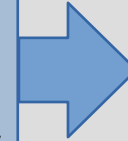
Continuous
Procrustes Distance

Characterize similarity
across surfaces



Diffusion Map

Embed surface data into
comparable form based on
probabilistic notions of similarity

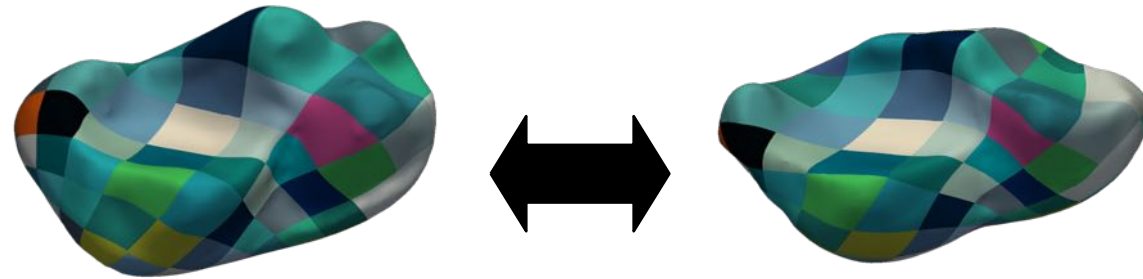


k-Means Clustering

Machine learning
technique for partitioning
data into *k* groups

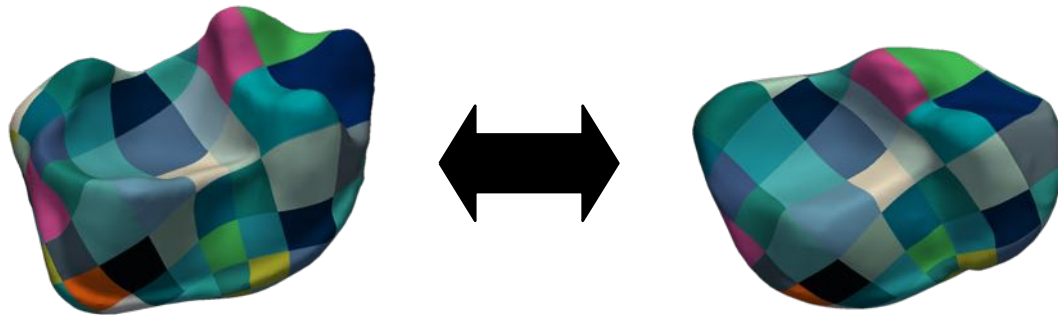
Implemented by software package "hecate"

Continuous Procrustes distance



Donrusselia gallica

Lemur catta



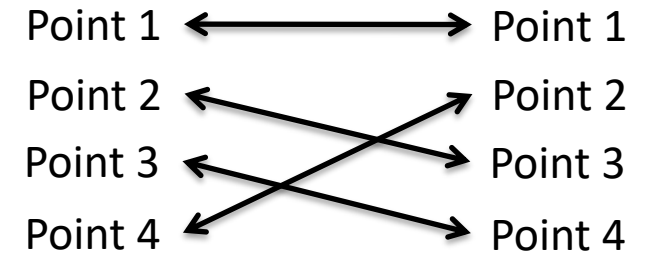
Tarsius spectrum

Mirza coquereli

Point-to-point
correspondence

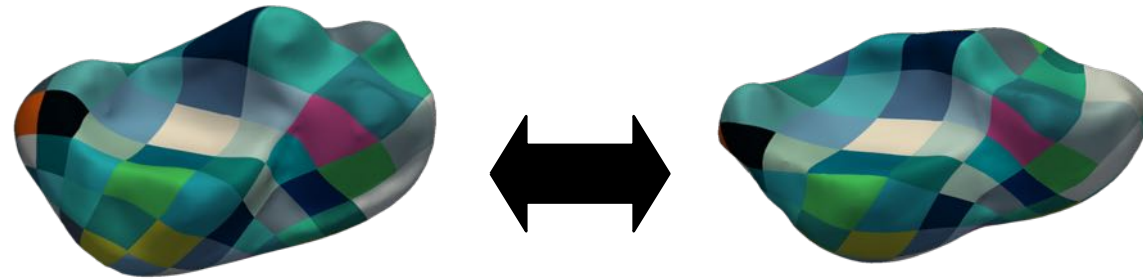
Mesh A

Mesh B



- Produces surface point-to-point correspondence maps and mesh-to-mesh distances

Continuous Procrustes distance



Donrusselia gallica

Lemur catta



Tarsius spectrum

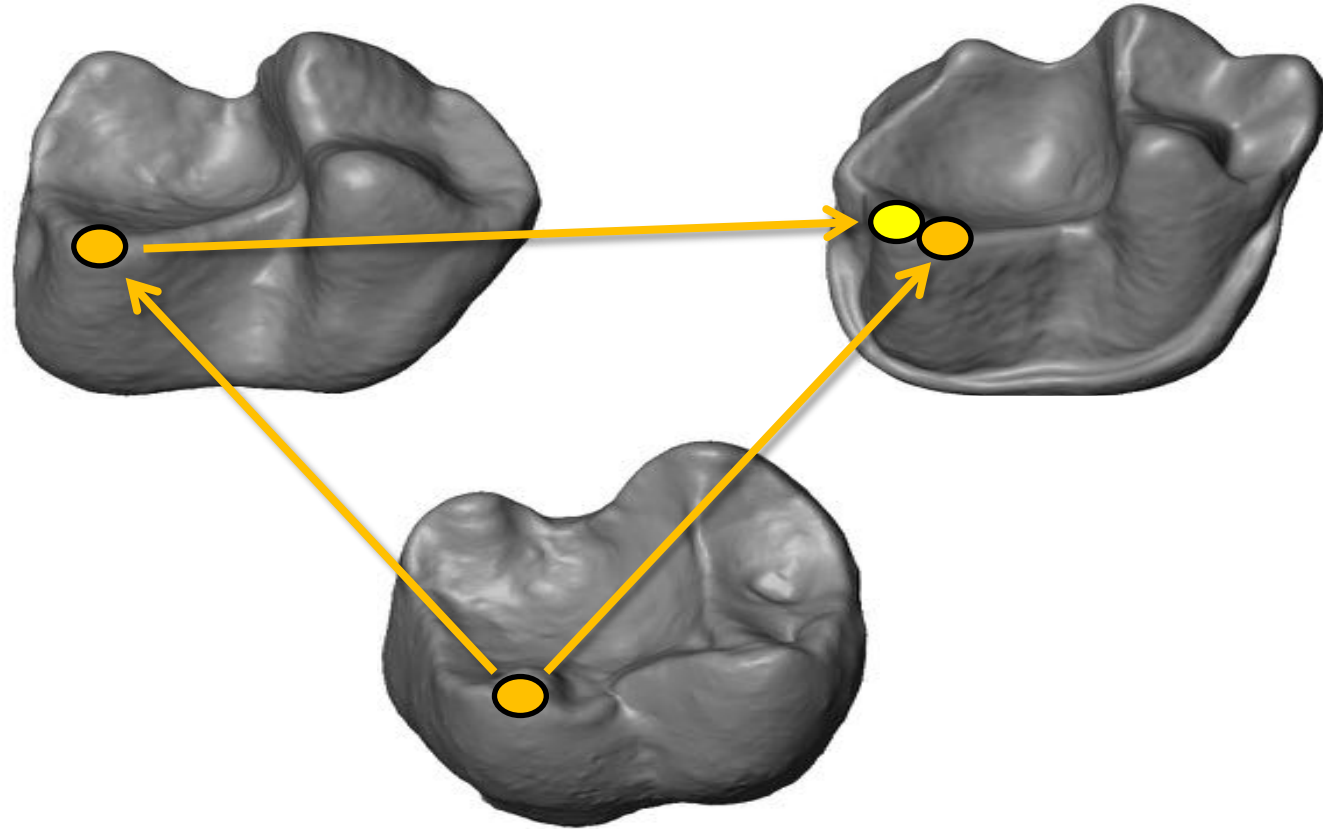
Mirza coquereli

Mesh Continuous
Procrustes Distances

	A	B	C	D
A		.2	.7	.1
B	.2	.3	.4	
C	.7	.3	.5	
D	.1	.4	.5	

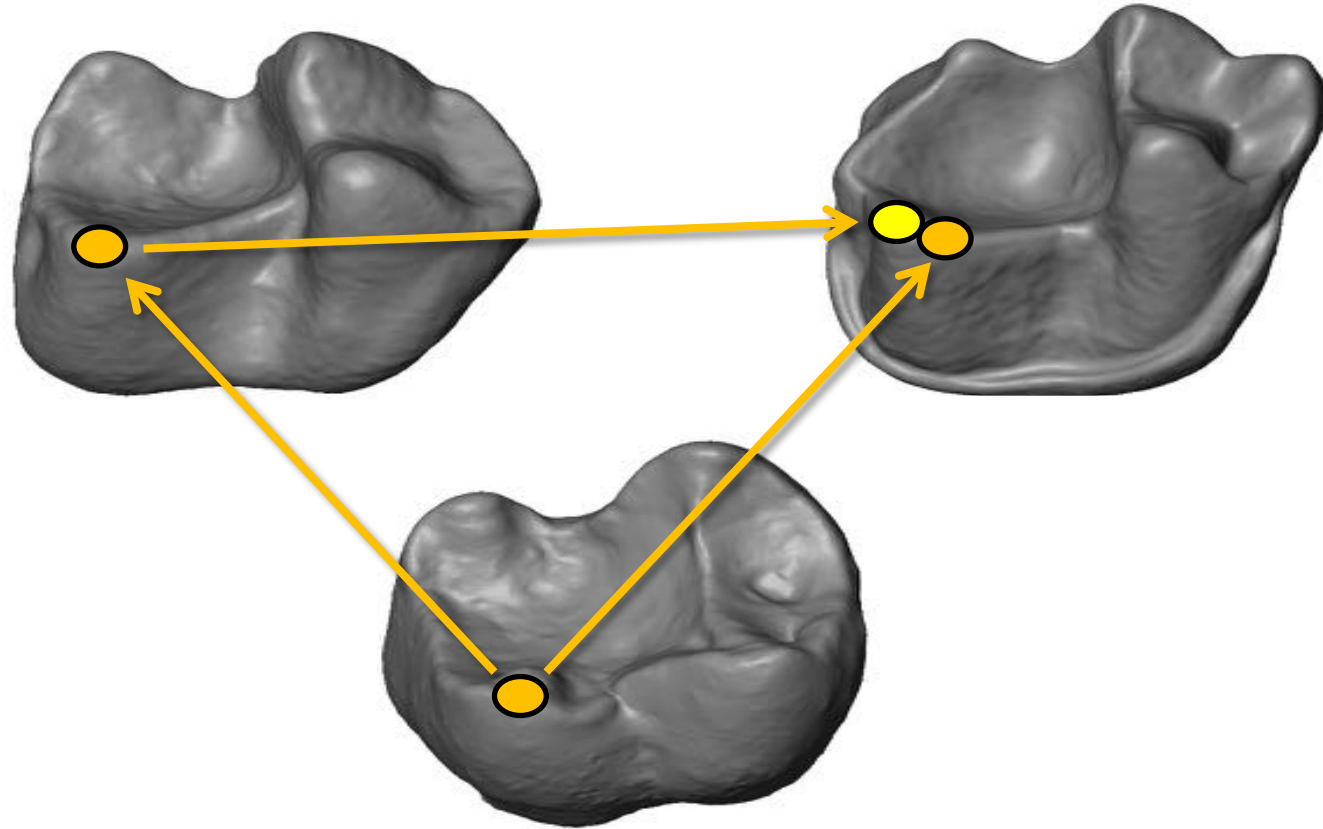
- Produces surface point-to-point correspondence maps and mesh-to-mesh distances

Continuous Procrustes distance



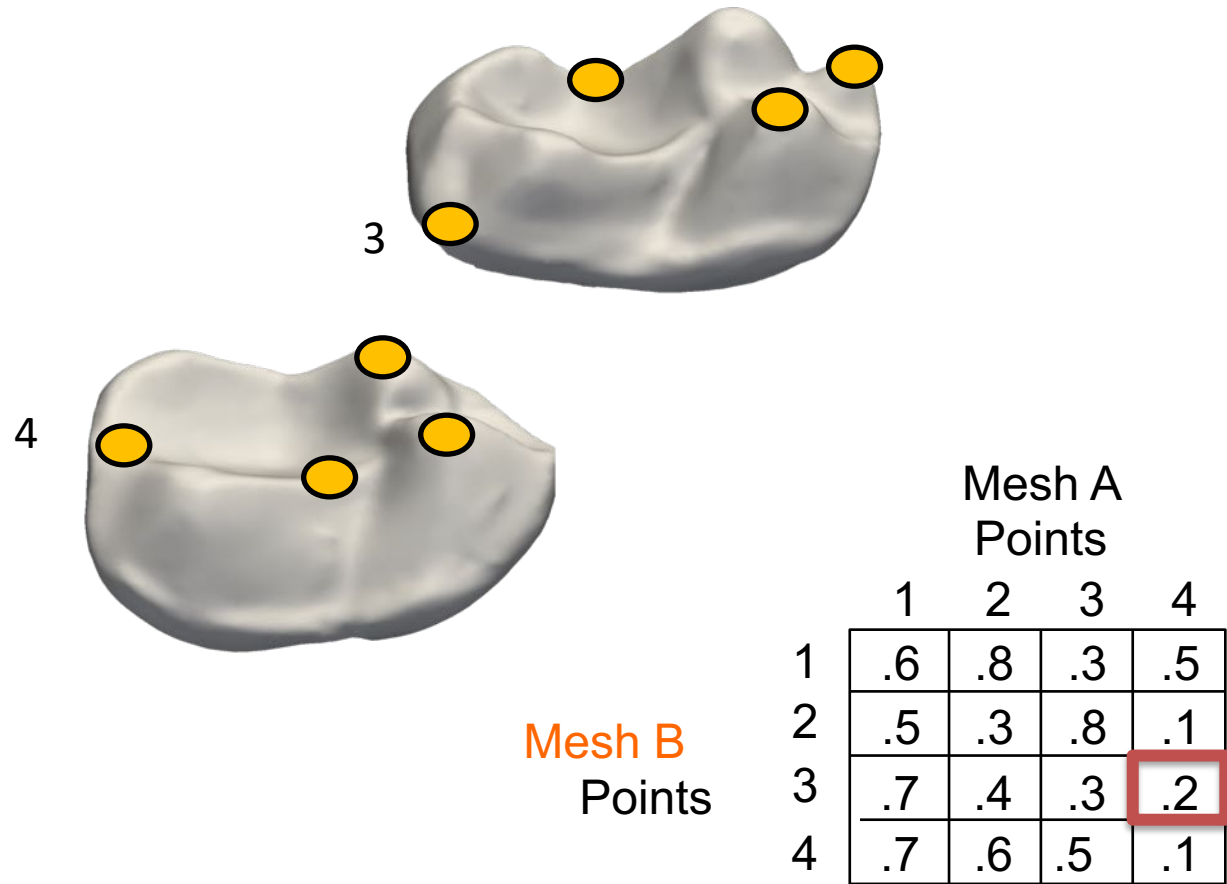
- Possible concern: 'walking' point drifts (diffuses) across combinations of surface maps

Continuous Procrustes distance



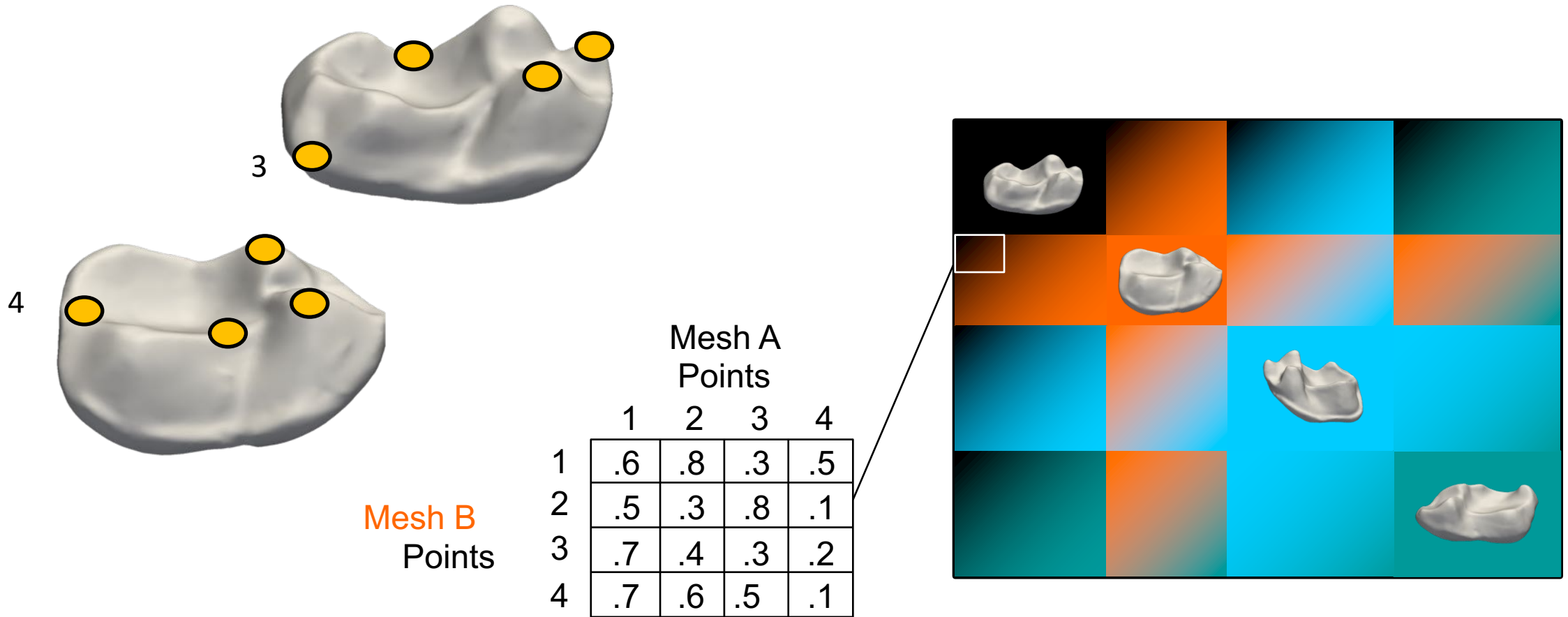
- Upside: For regions of local similarity, points walk ('diffuse') in neighborhoods

Diffusion probability matrix

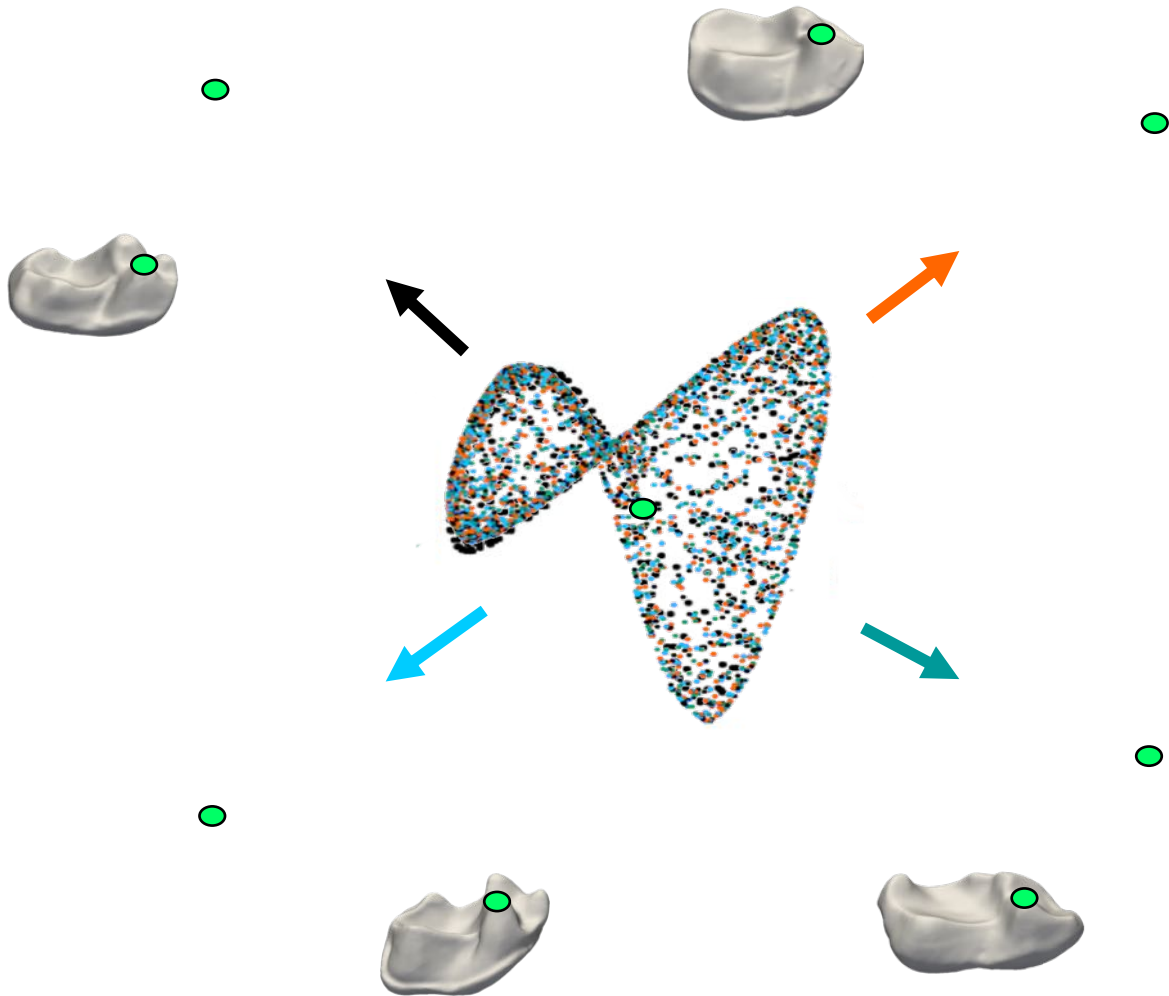


- Matrix showing probability of point-to-point diffusion

Diffusion probability matrix

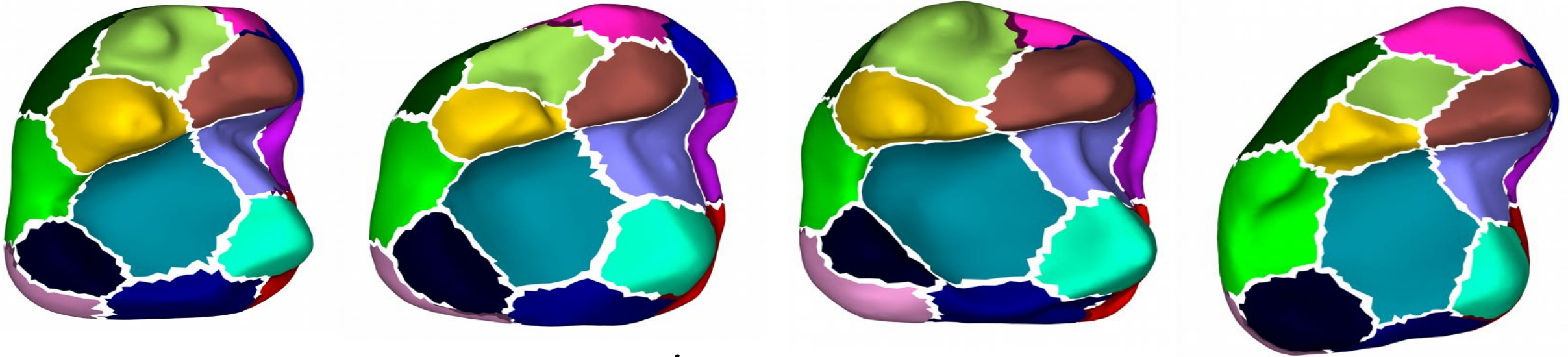


Diffusion map



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

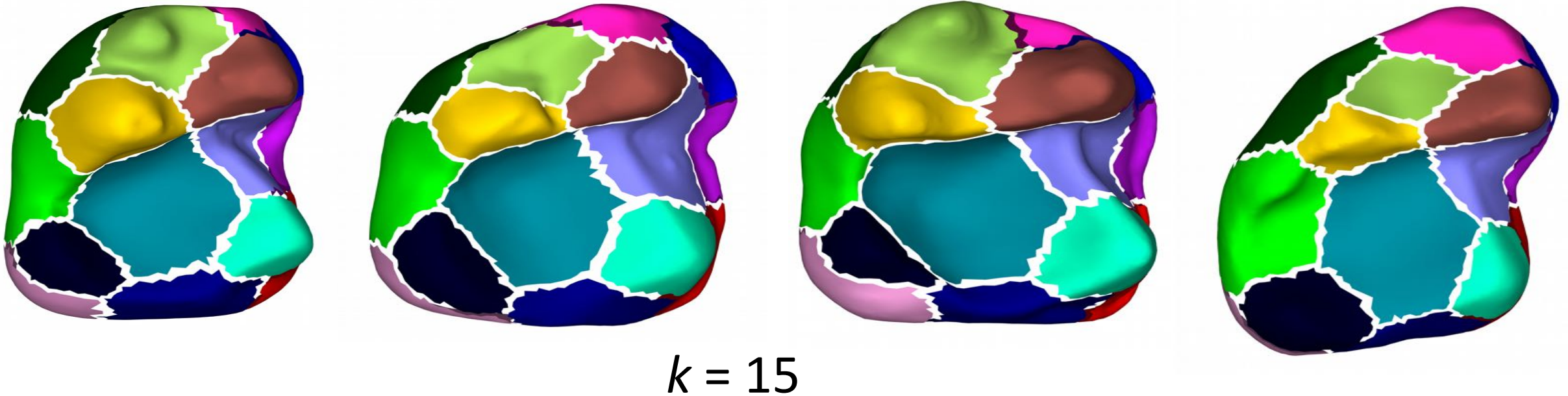
k -Means Clustering



$k = 15$

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

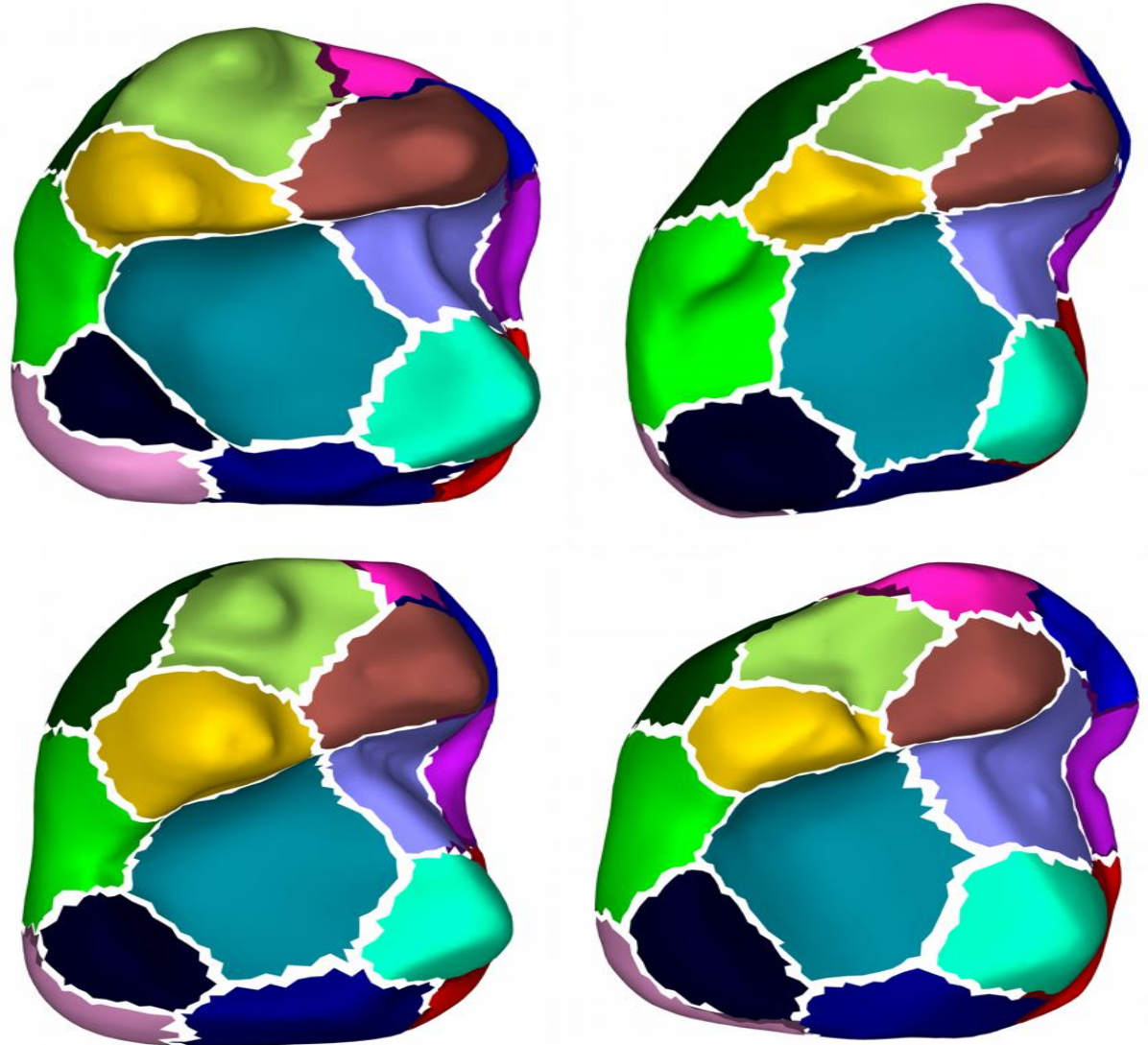
k -Means Clustering



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

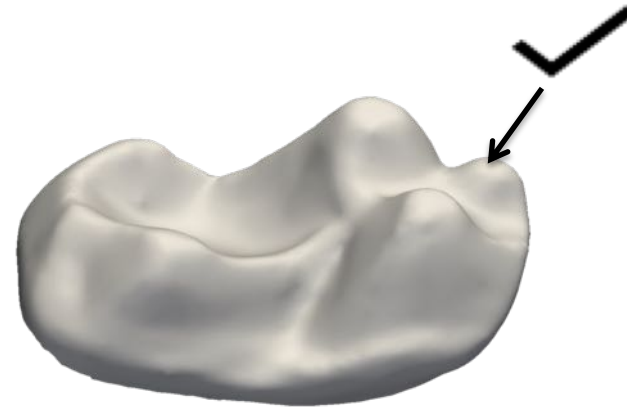
Segmentation results

- 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



Test: paraconid cusp

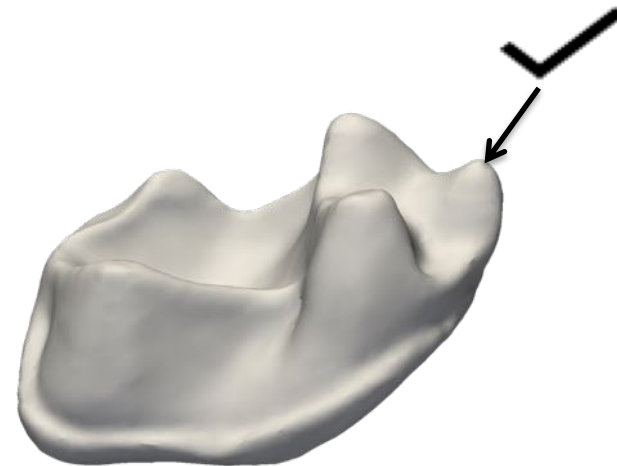
- Variable among prosimians
- Qualitative characterizations differ, e.g. *Lemur catta*
 - Ni et al.: weakly present
 - Herrera and Dávalos: absent



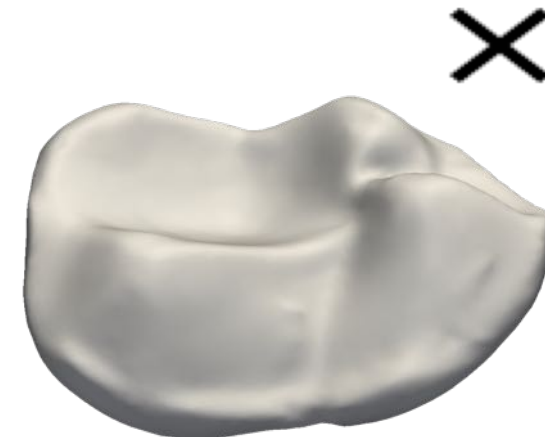
Donrusselia gallica



Lemur catta



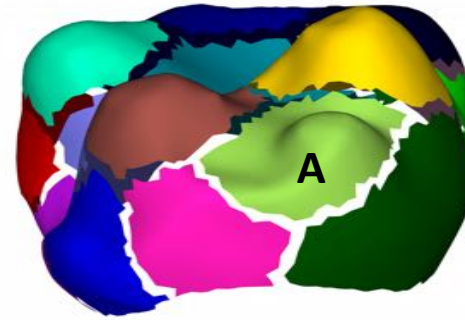
Tarsius spectrum



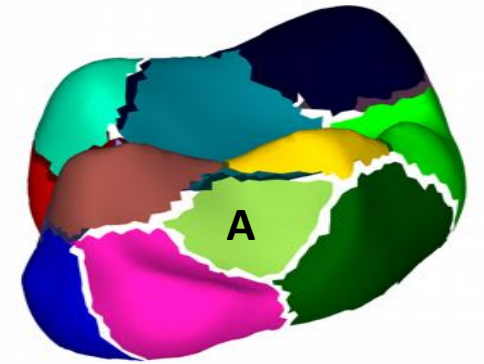
Mirza coquereli

Test: paraconid cusp

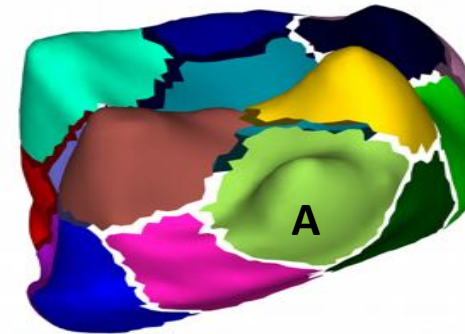
- Sample: 116 prosimian M_2 s, 15 segments
- Species-level paraconid presence/absence
- Quantify surface curvature of segment A
 - Dirichlet normal energy
 - Species means
- Compare groups
 - Welch's T test



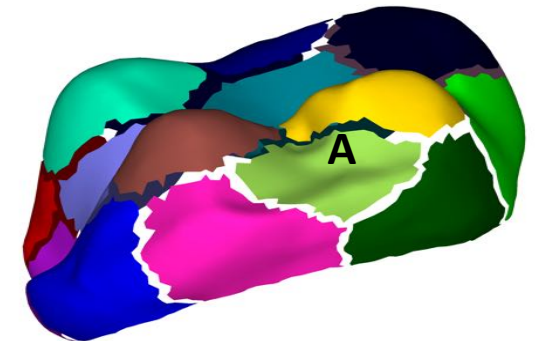
Donrusselia gallica



Lemur catta

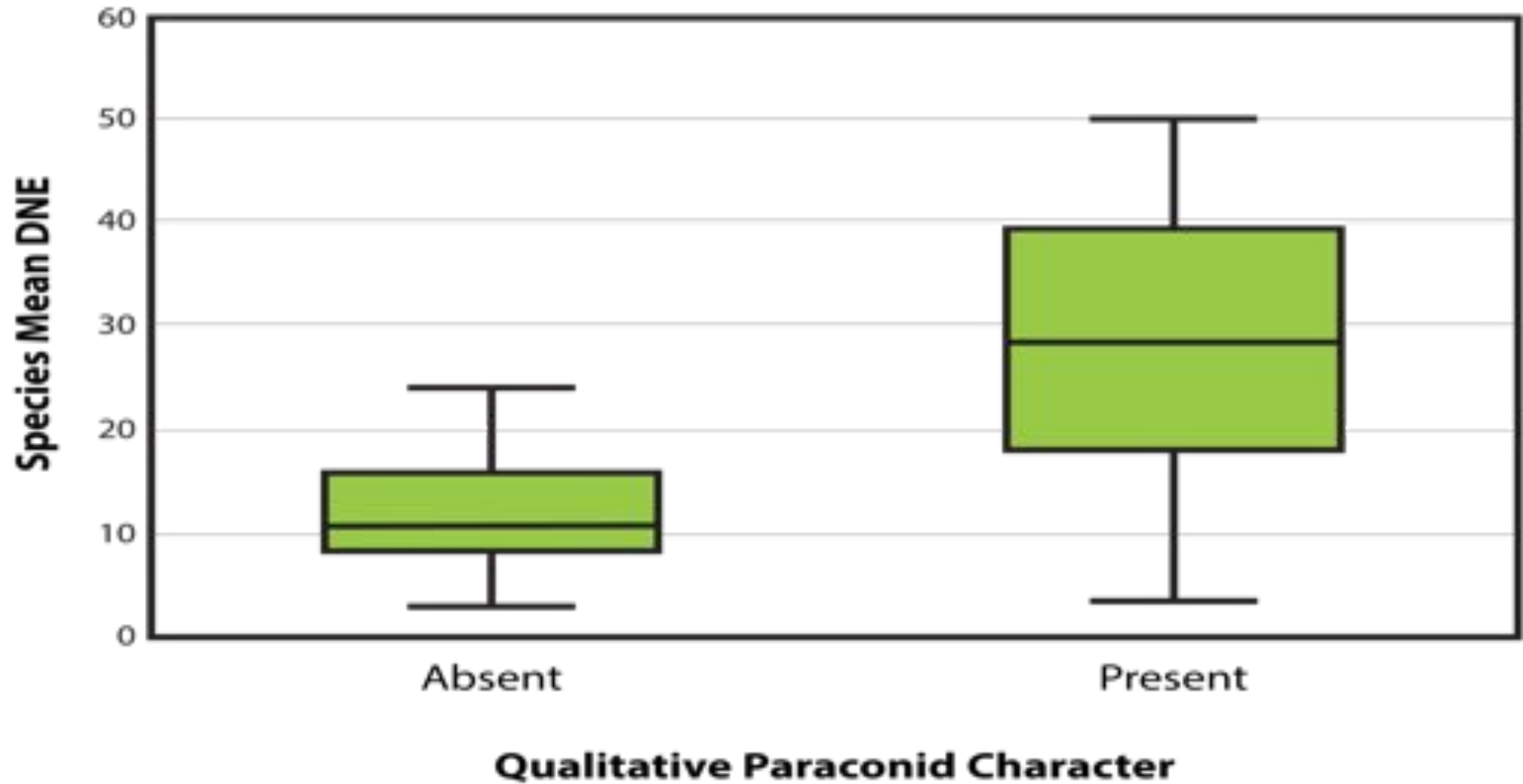


Tarsius spectrum

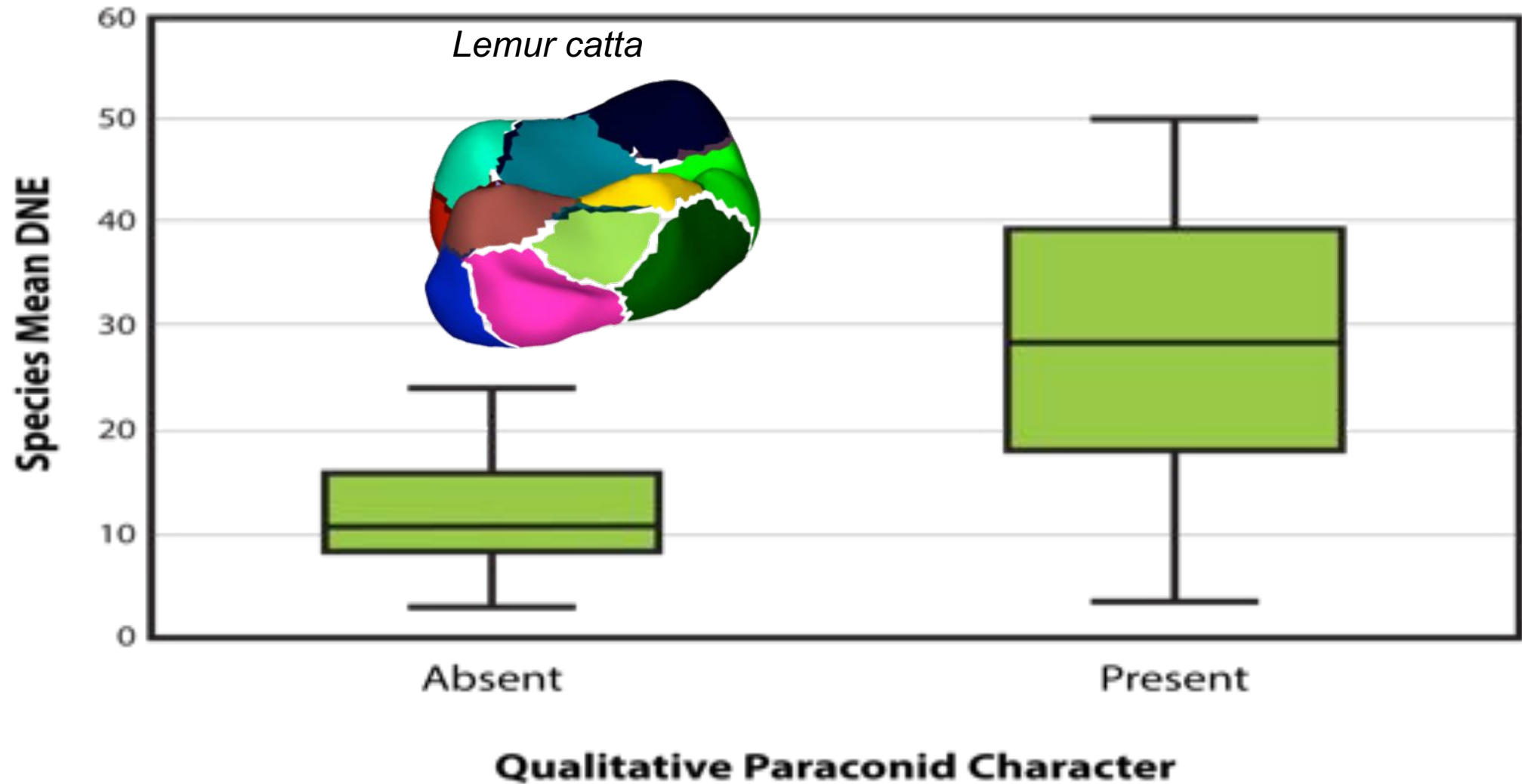


Mirza coquereli

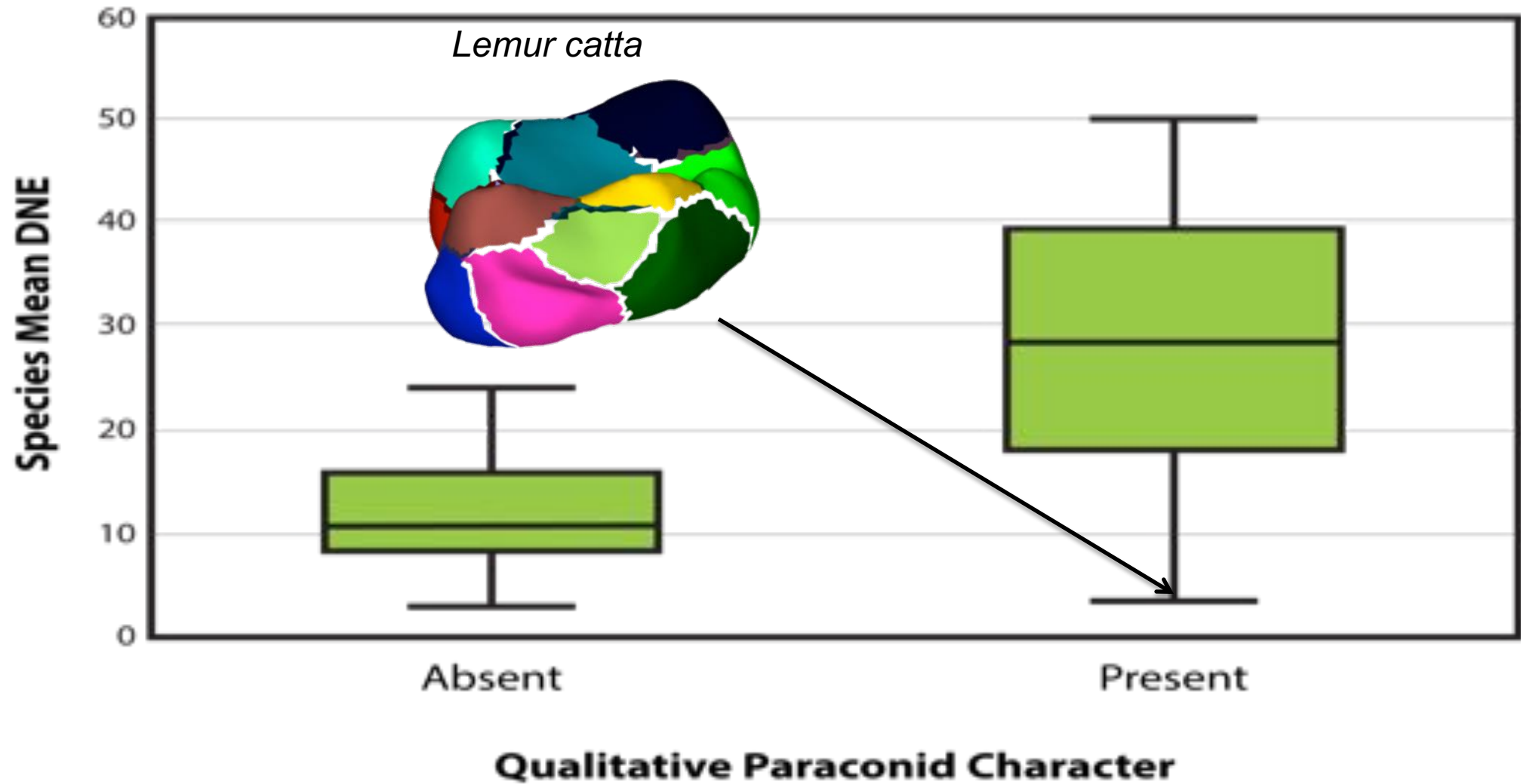
Test: paraconid cusp



Test: paraconid cusp



Test: paraconid cusp



Potential application – Testing key innovations

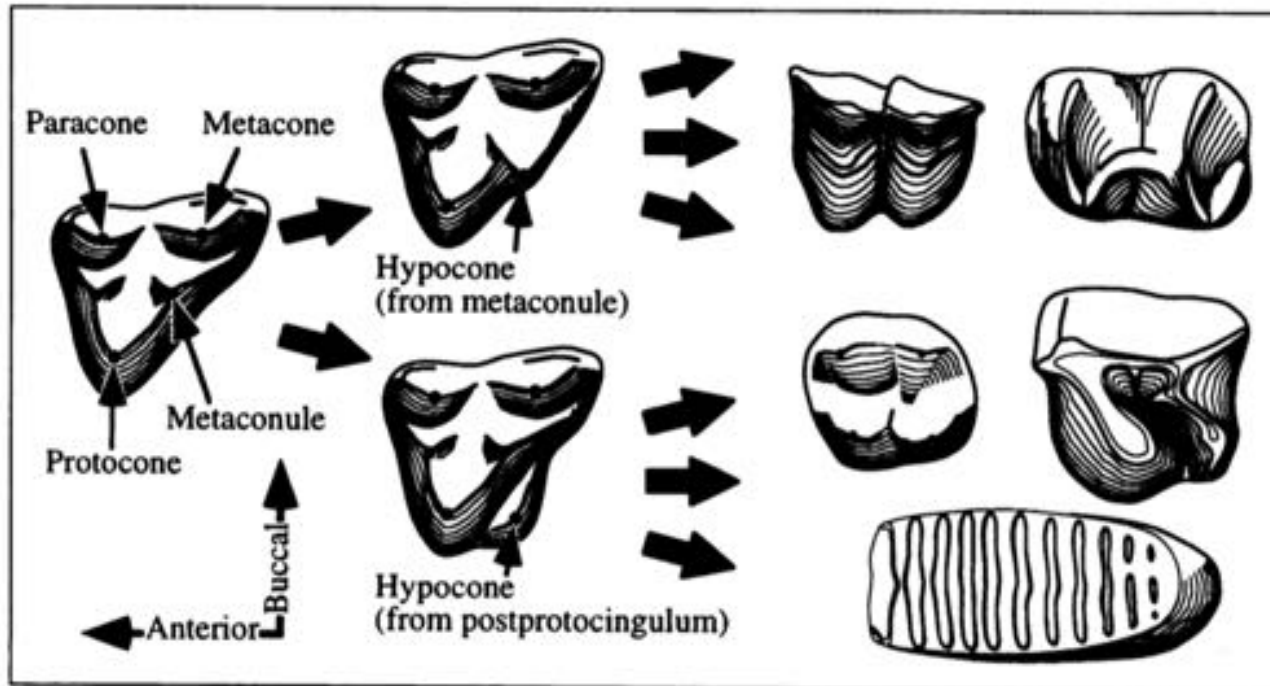
Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 10718–10722, November 1995
Evolution

The hypocone as a key innovation in mammalian

(adaptive zone/convergent evolution/dentition/diversity/Theria)

JOHN D. HUNTER* AND JÜRGEN IEDVALL†

*Depart
Biotech
Comm



t of

