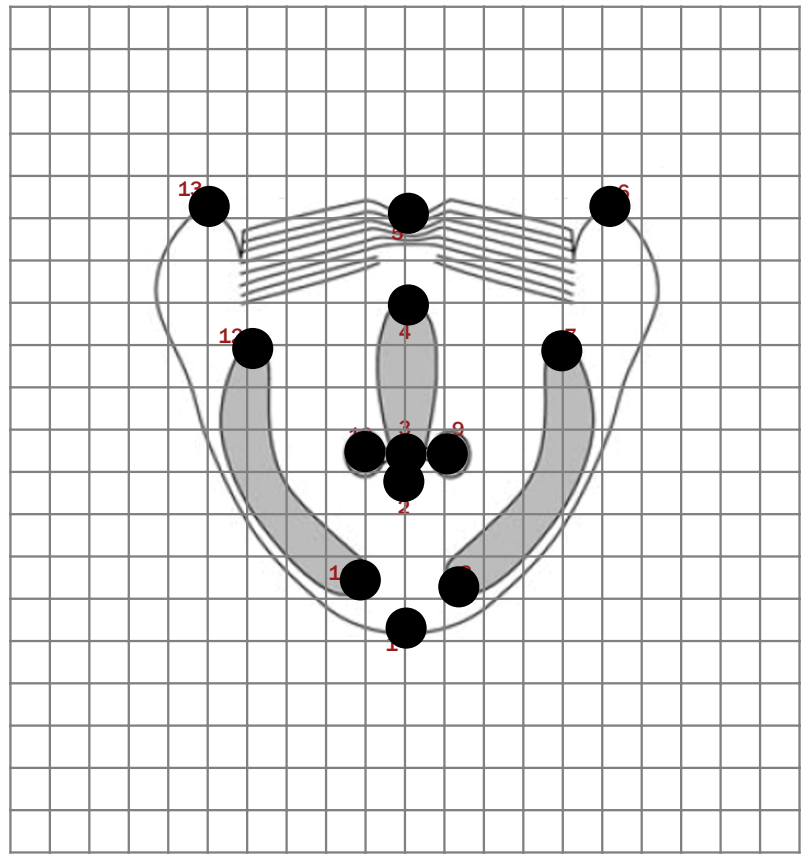
Introduction to geometric morphometrics

Day 1



P. David Polly
Department of Earth and Atmospheric Sciences
Adjunct in Biology and Anthropology
Indiana University
Bloomington, Indiana 47405 USA
pdpolly@indiana.edu

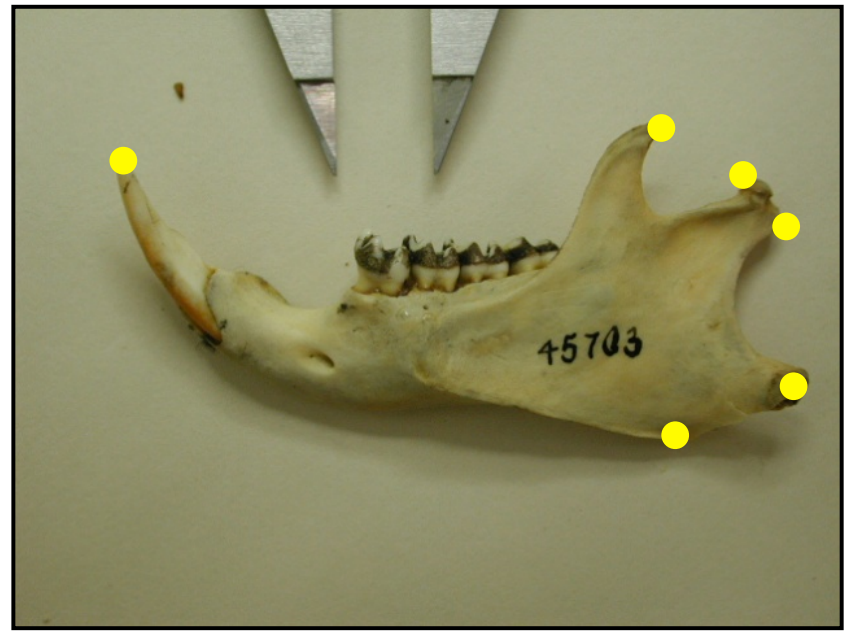
Katrina E. Jones
Museum of Comparative Zoology
Harvard University
26 Oxford St
Cambridge MA 02138 USA
katrinajones@fas.harvard.edu

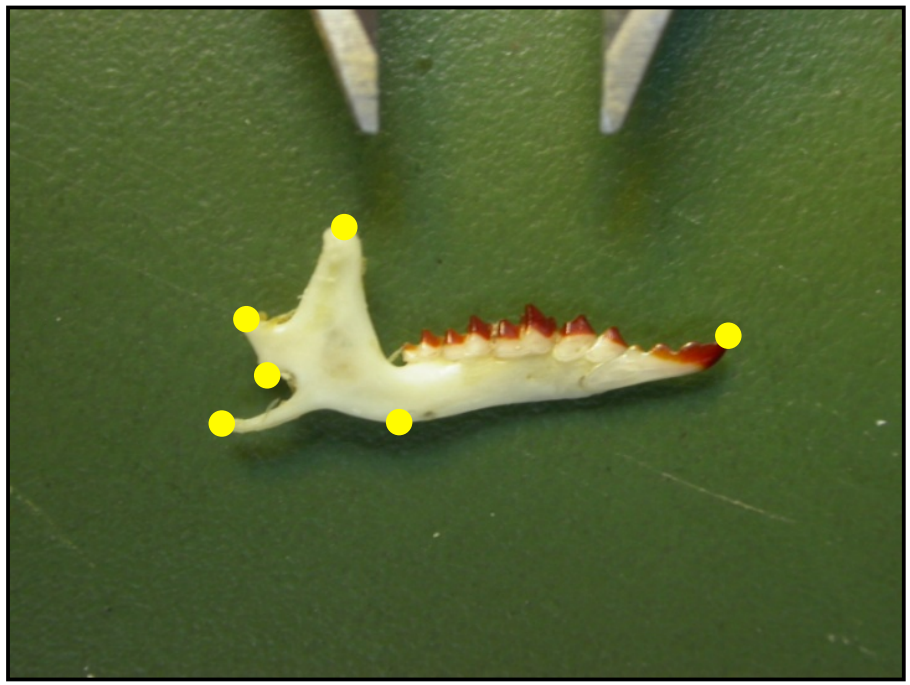
Introduction to geometric morphometrics

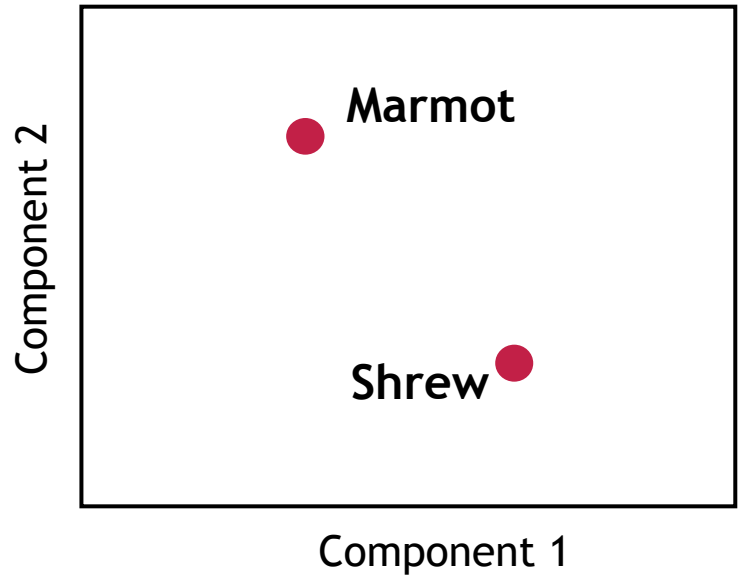
- What is GMM?
- How does GMM differ from other forms of morphometrics?
- Basic steps in GMM
  - Landmarking
  - Procrustes superimposition
  - Shape variables (PCA)
  - Subsequent analyses
- GMM and size
- Landmarks, semilandmarks, and sliding semilandmarks
- Brief history of GMM
- Strengths and weaknesses compared to alternative methods
  - Size issues
  - Procrustes and variation at individual landmarks
  - Analyzing individual principal components
  - Visualization: pictures versus numbers
- Software for GMM

Geometric morphometrics (GMM)

The quantitative representation and analysis of morphological shape using geometric coordinates instead of measurements







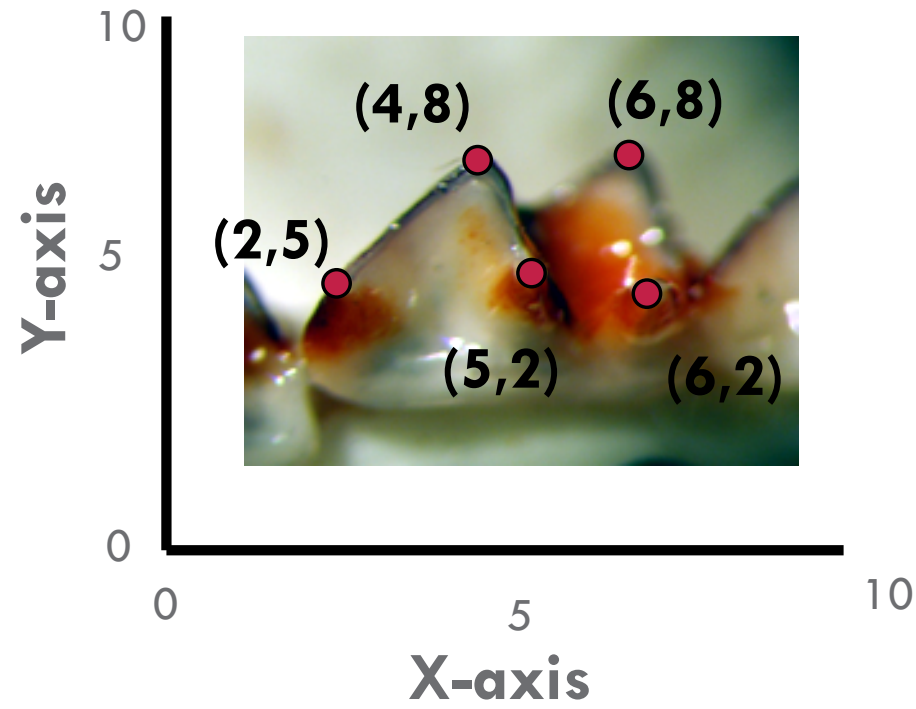
Landmarks

Landmarks are coordinate points used to represent a shape

Landmarks can be two-dimensional or three-dimensional

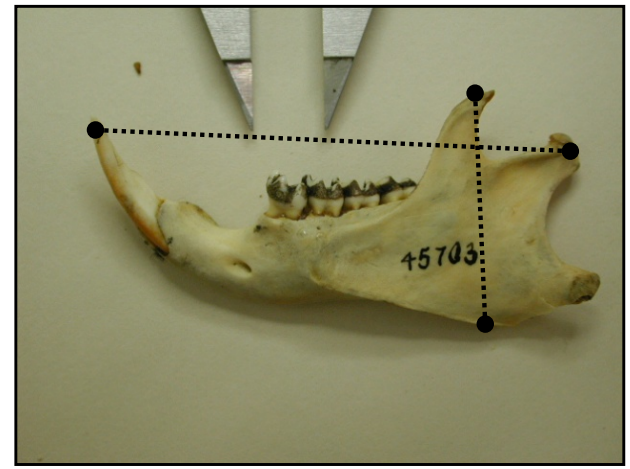
They are quantified as Cartesian coordinates (x,y[,z])

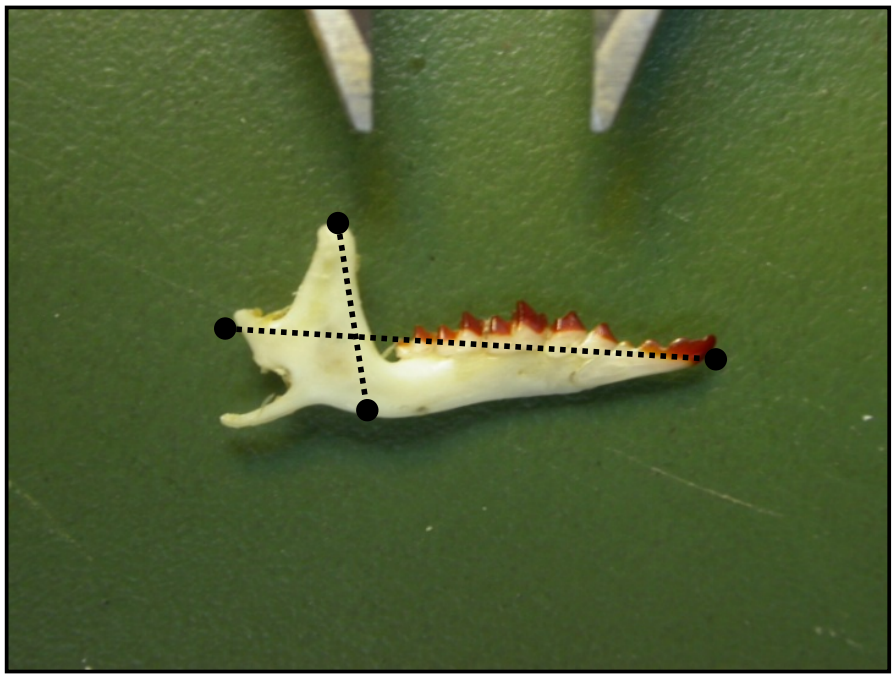
At least 3 landmarks are required (two points make a line and all lines have the same shape)

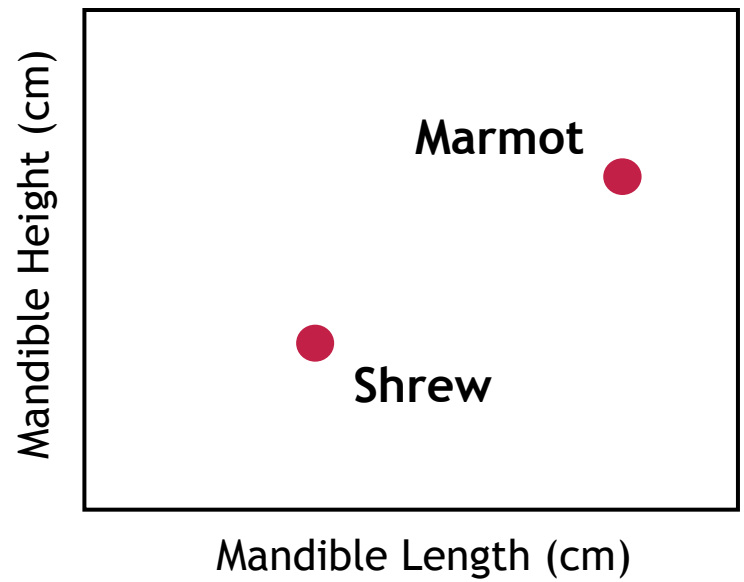


Measurement-based morphometrics

Any quantitative measurement and analysis of morphological traits







The true goal of all morphometric analyses...

...to measure morphological similarity and difference

Distinguishing features of GMM

- Shape is represented with landmarks (or semilandmarks)
- Variables consist of Cartesian coordinates instead of measurements
- Coordinates must be registered with Procrustes analysis
- Coordinate system has no objective scale so size is not present in the analysis
- Results can be visualized as pictures using the Cartesian coordinates
- Variants analyze pairwise distances between landmarks (e.g. EDMA) or angles between successive semilandmarks on an outline or curve (e.g. Fourier, Eigenshape)

Steps in a Geometric Morphometric Methods (GMM) Analysis

1. Collect landmark coordinates

2. Do a Procrustes superimposition

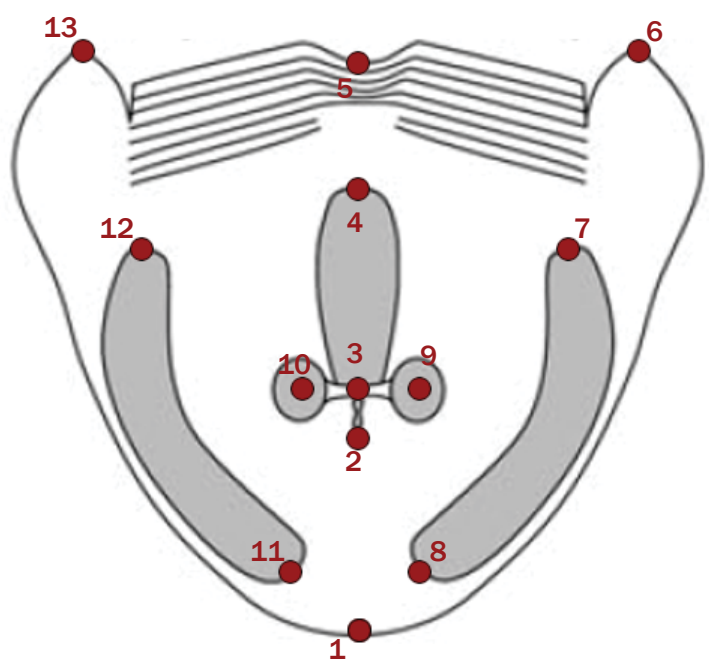
   Standardizes landmarks by rescaling them and rotating them to a common orientation using least-squares fitting

3. Analyze similarity and difference of shape

   Analysis usually starts with a Principal Components Analysis, which (A) shows similarity and differences as simple scatter plots, and (B) provides new variables for further statistical analysis

Step 1: Collecting landmarks

1. Each shape must have the same number of landmarks
2. The landmarks on all shapes must be in the same order
3. Landmarks are ordinarily placed on homologous points, points that can be replicated from object to object based on common morphology, common function, or common geometry



Osteostracan head shield from Sansom, 2009

Step 2: Procrustes superimposition

Procrustes superimposition is the “standardization” step in GMM.

Procrustes removes

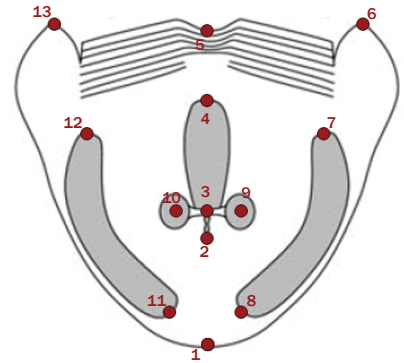
1. size
2. translation
3. rotation

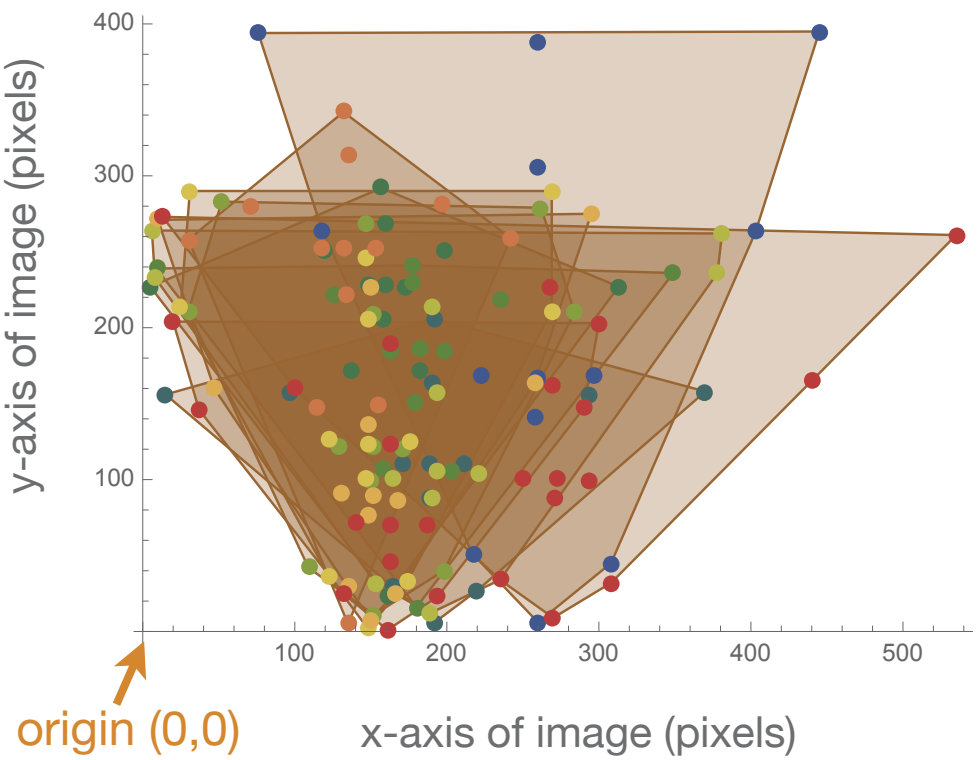
In other words, it centers the shapes on the same point, scales them to the same size, and rotates them into the same orientation.

These manipulations remove statistical degrees of freedom, which has implications for later statistical analyses.

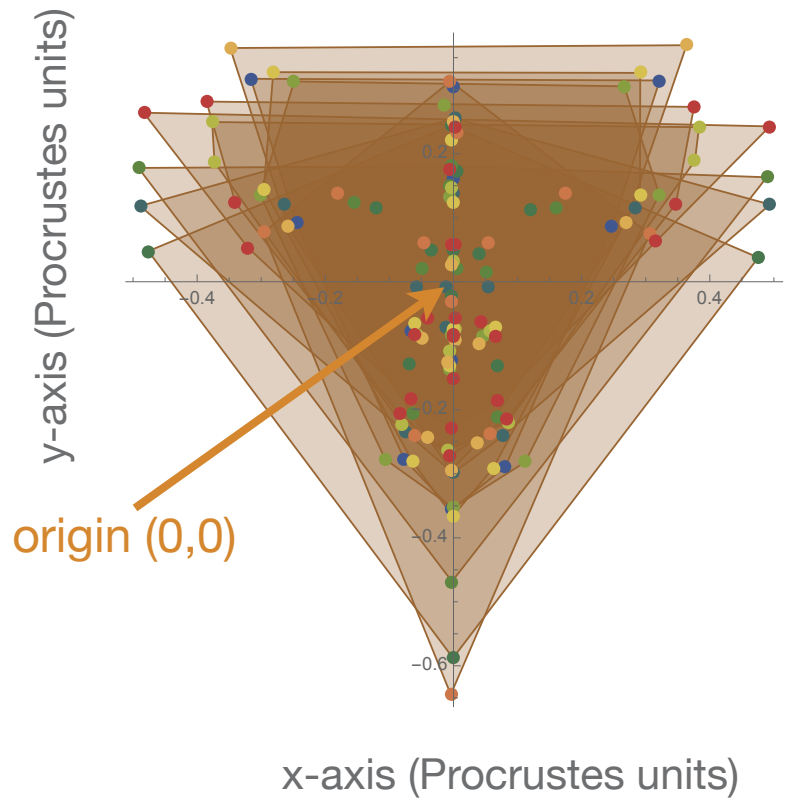
After landmarks have been superimposed, the similarities and differences in their shape can be analyzed.

Procrustes superimposition





Original coordinates



Procrustes coordinates

Procrustes superimposition

also known as…

- Procrustes analysis
- Procrustes fitting
- Generalized Procrustes Analysis (GPA)
- Generalized least squares (GLS)
- Least squares fitting

Centers all shapes at the origin (0,0,0)

Usually scales all shapes to the same size (usually “unit size” or size = 1.0)

Rotates each shape around the origin until the sum of squared distances among them is minimized (similar to least-squares fit of a regression line)

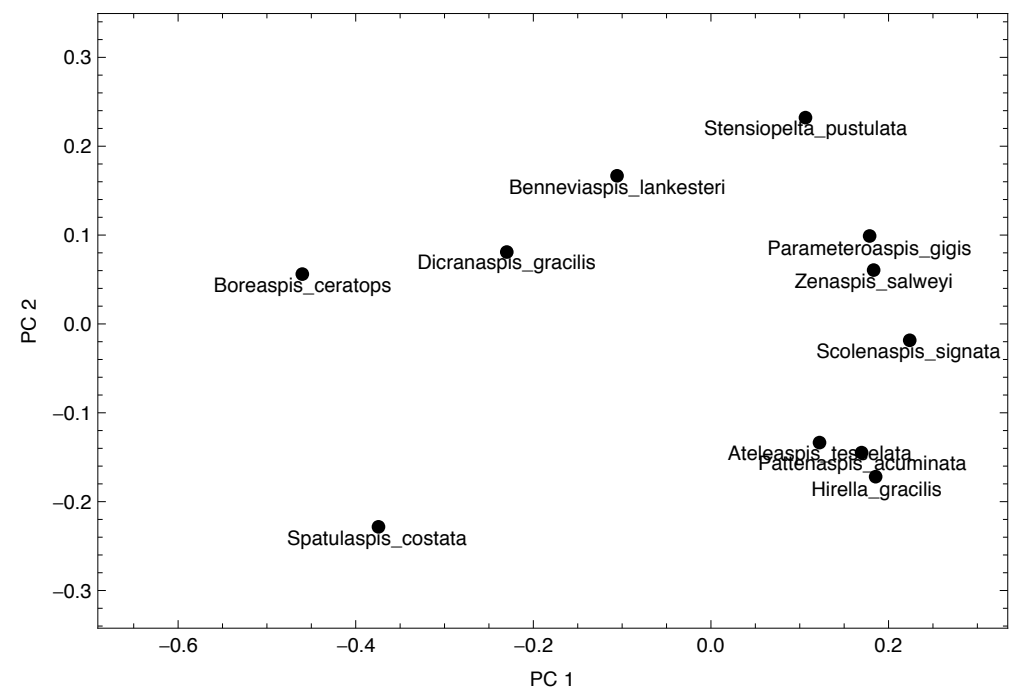
Ensures that the differences in shape are minimized

Step 3: Principal Components Analysis

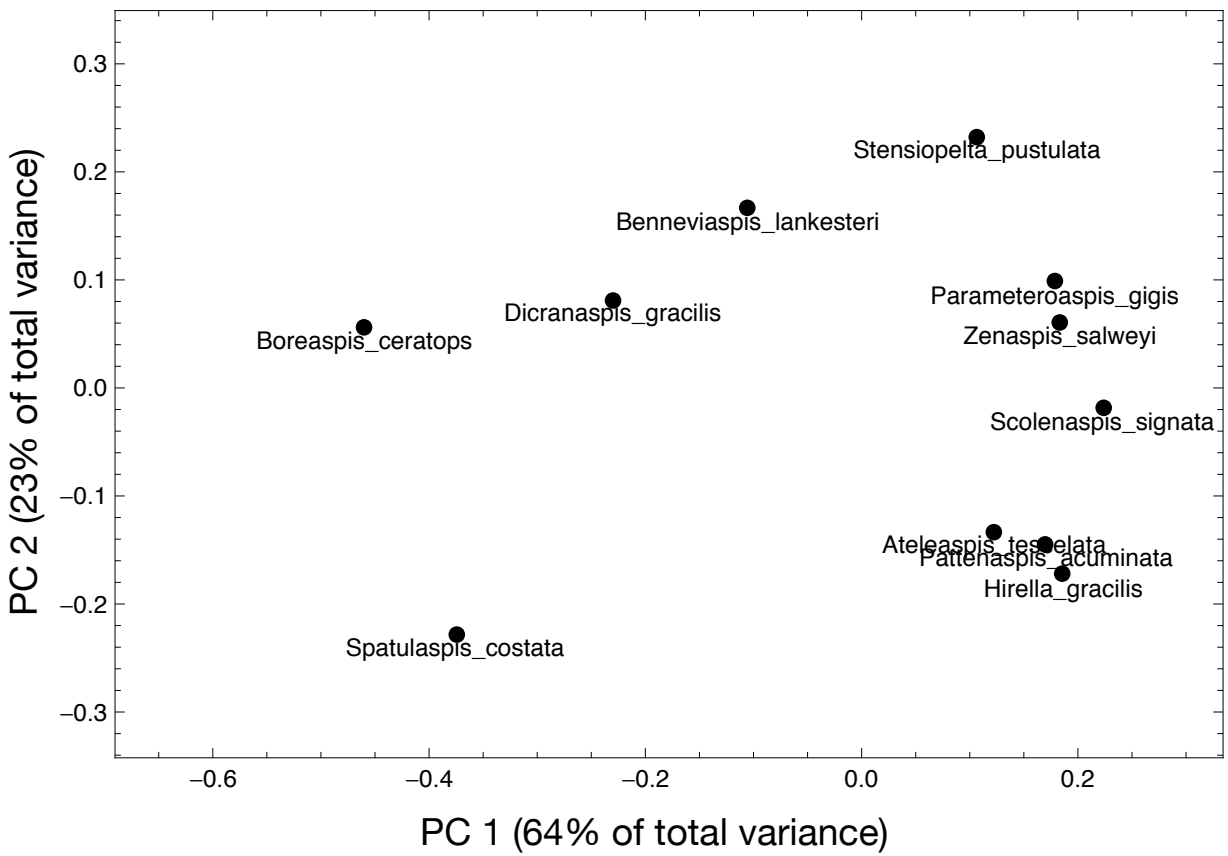
Principal Components Analysis (PCA) ordinates objects by arranging them in a shape space. Similarities and differences can easily be seen in a PCA plot.

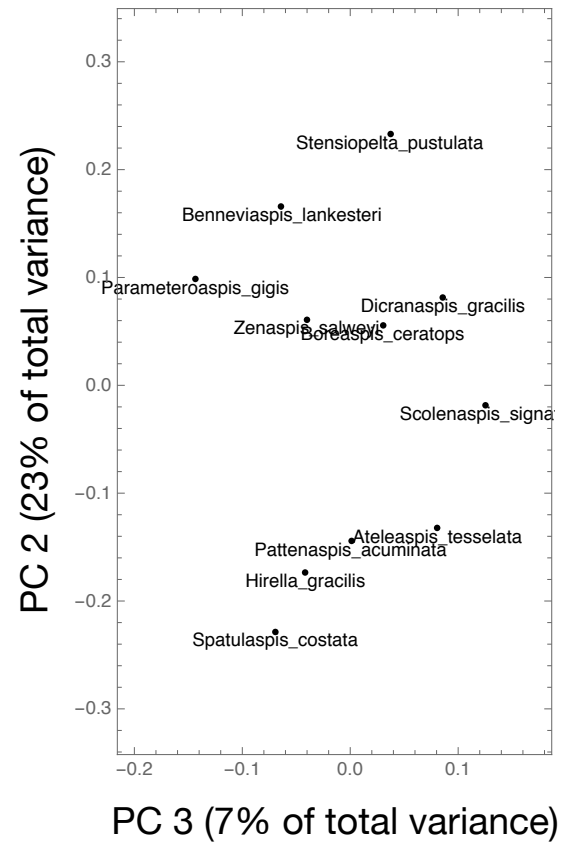
The axes of a PCA plot are Principal Components (PCs). The first PC is, by definition, the line that spans the largest axis of variation in shape. The second PC spans the next largest axis of variation at right angles to the first, the third PC spans the third largest axis of variation, and so on.

Each point on a PCA plot represents the shape of a single object from your analysis. The closer two objects are, the more similar they are in shape.



Basic PCA plots of Paleozoic fish heads



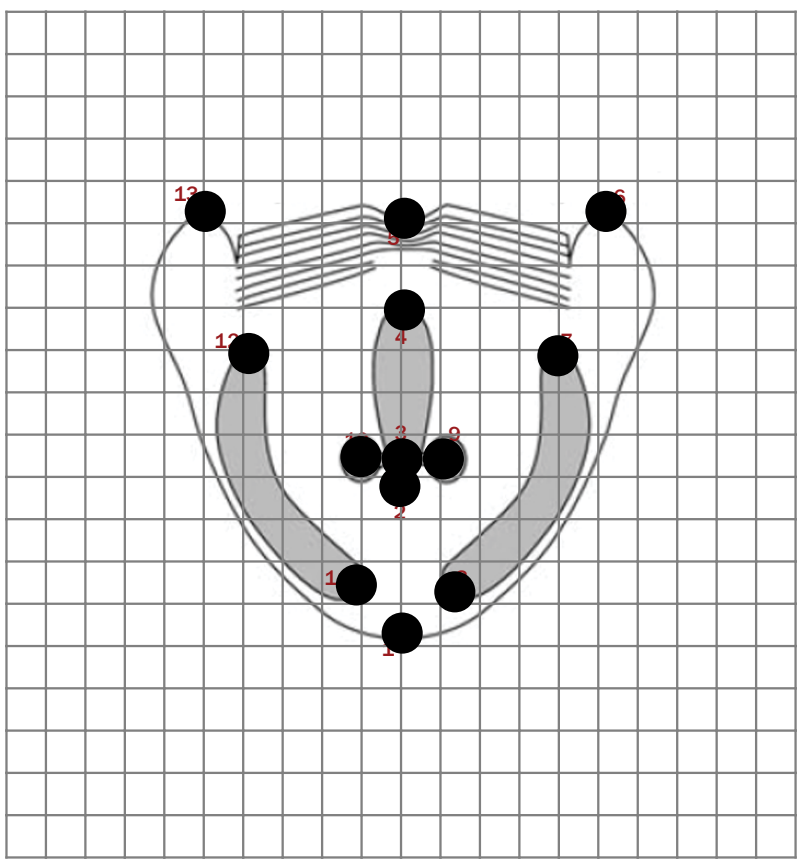


PCA plot is a *morphospace*

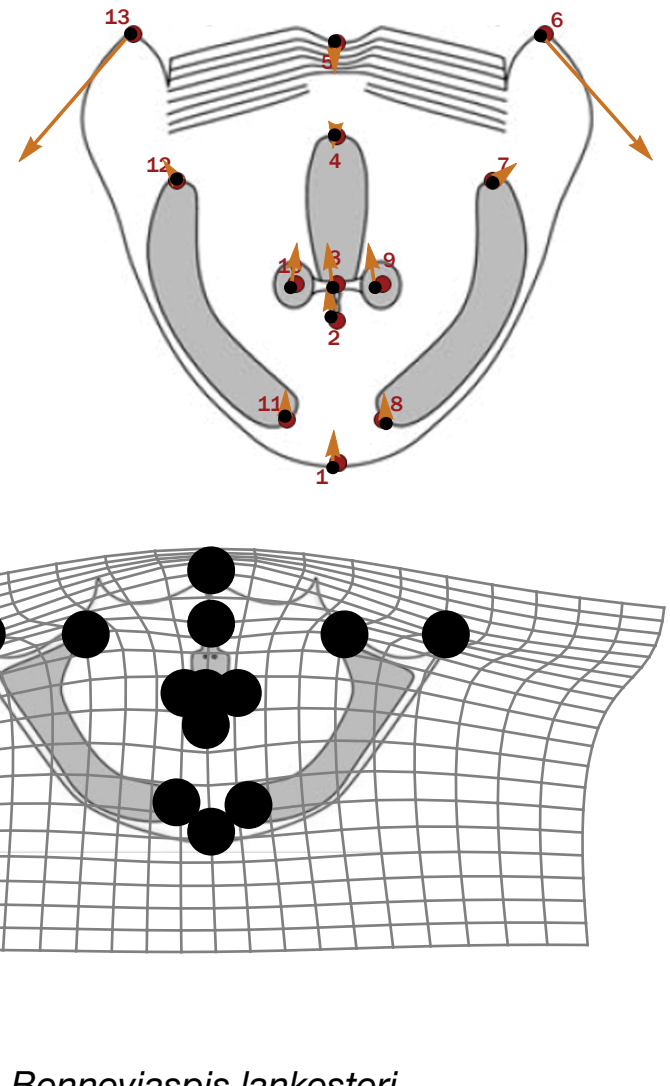
Exploring the gradients of shape along PC axes is an important part of GMM analysis.

- Each point represents a unique shape (configuration of landmarks)
- A "shape model" can be constructed for every point
- Shape forms a continuous gradient through the PC space
- Each PC axis "describes" a different component of shape
- Real shapes are found where their landmarks correspond to the shape space
- Coordinates of the shapes are "scores"
- Each axis is "orthogonal" or uncorrelated
- Each axis accounts for a descending proportion of the total shape variance
- Axis directions (positive vs. negative) are arbitrary and can be flipped
- Axis are scaled in Procrustes units (nearly meaningless except as a within-study metric)

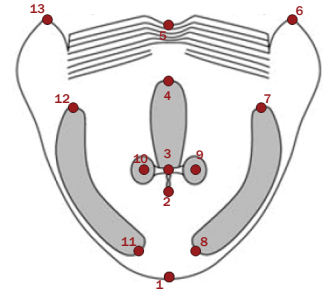
Thin-plate spline deformations as visual tools



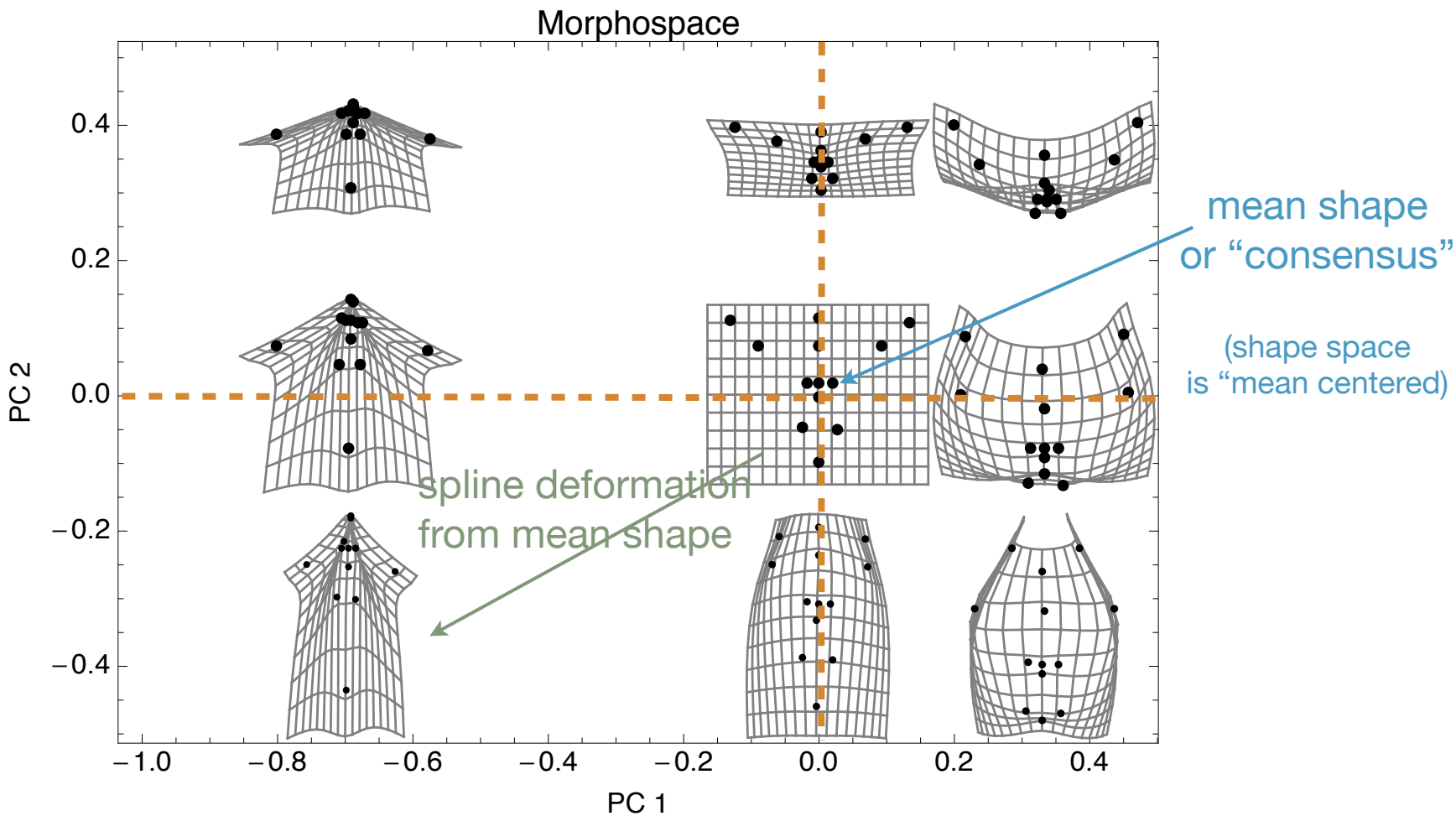
Ateleaspis tesselata



Benneviaspis lankesteri

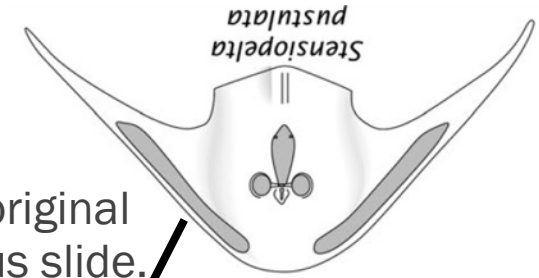


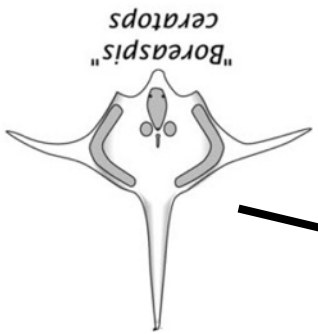
Shape models to explore morphospace

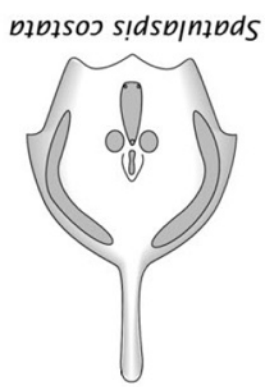


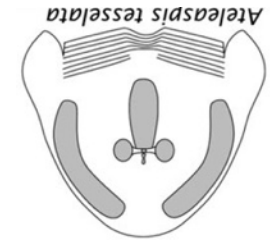
Comparing original shapes

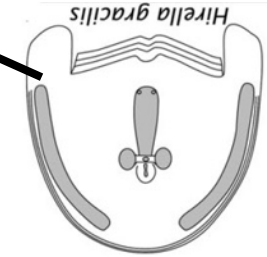
Also explore the distribution of shape by referring back to your original photographs. Compare these shapes to the grids on the previous slide.

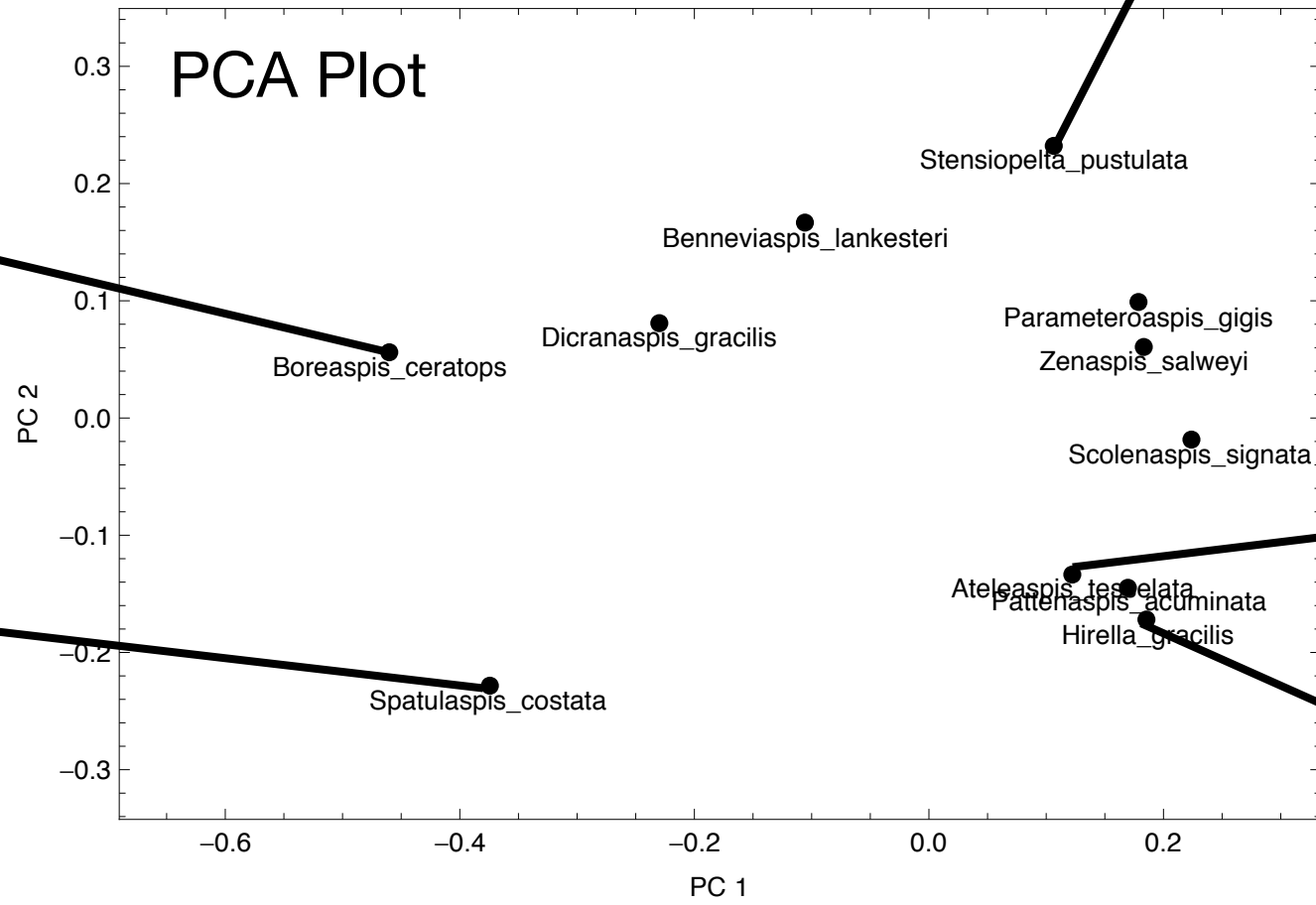












Why is PCA a standard step in GMM?

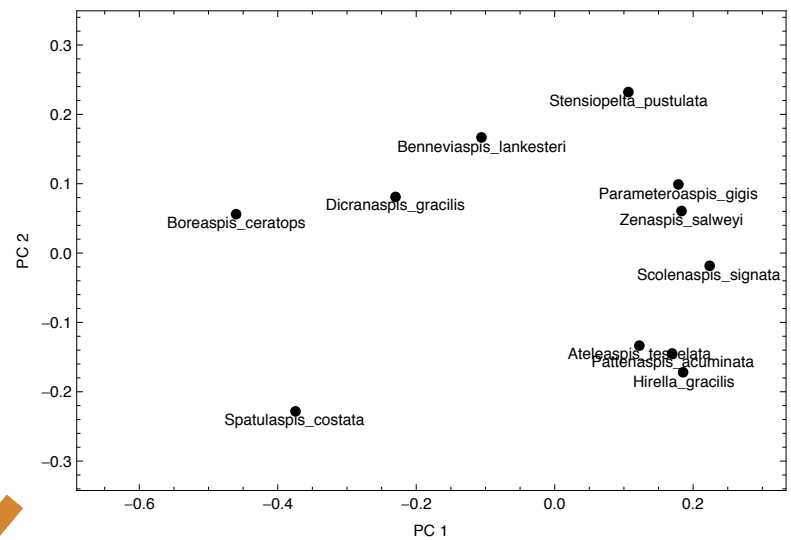
1. Rotates data to its major axes for better visualization

2. Preserves original distances between data points
(in other words, PCA does not distort the variation data, but only if the covariance method is used, which is standard in geometric morphometrics)

3. Removes correlations between landmark coordinates and adjusts to proper degrees of freedom to simplify statistical analysis

Step 4: further analyses

Almost all other analyses are performed on PC scores, which are known as "shape variables"



Regression to test for size allometry

MANOVA to test for group differences

Evolutionary shape modeling

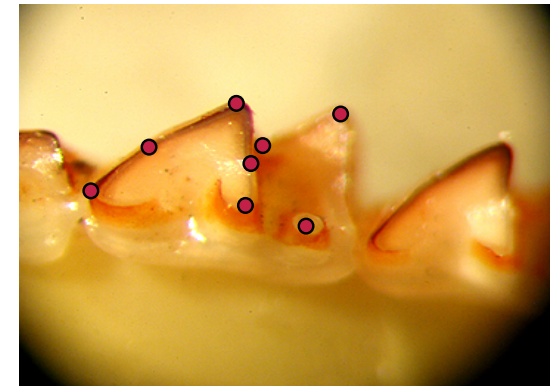
Definitions

Landmark.-any point described with cartesian coordinates (x, y, z) used to represent the shape of a structure.

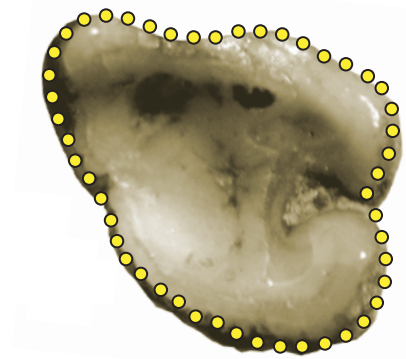
Landmark (2).– any point that can be placed on a biologically or geometrically homologous point on the structure.

Semi-landmark.– a point that is placed arbitrarily using an algorithm, often by defining endpoints at biologically homologous points and placing a specified number of semilandmarks between them.

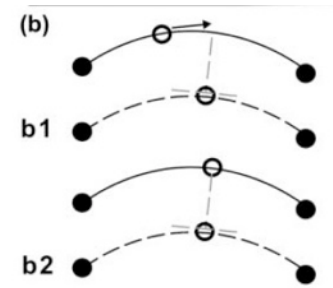
Sliding semi-landmark.-semilandmark points whose positions are algorithmically adjusted to minimize either the Procrustes distance or "bending energy". (use with caution: placement is sample dependent)



Polly, P. D. 2001. *Genetica*, 111-112, 339-357.



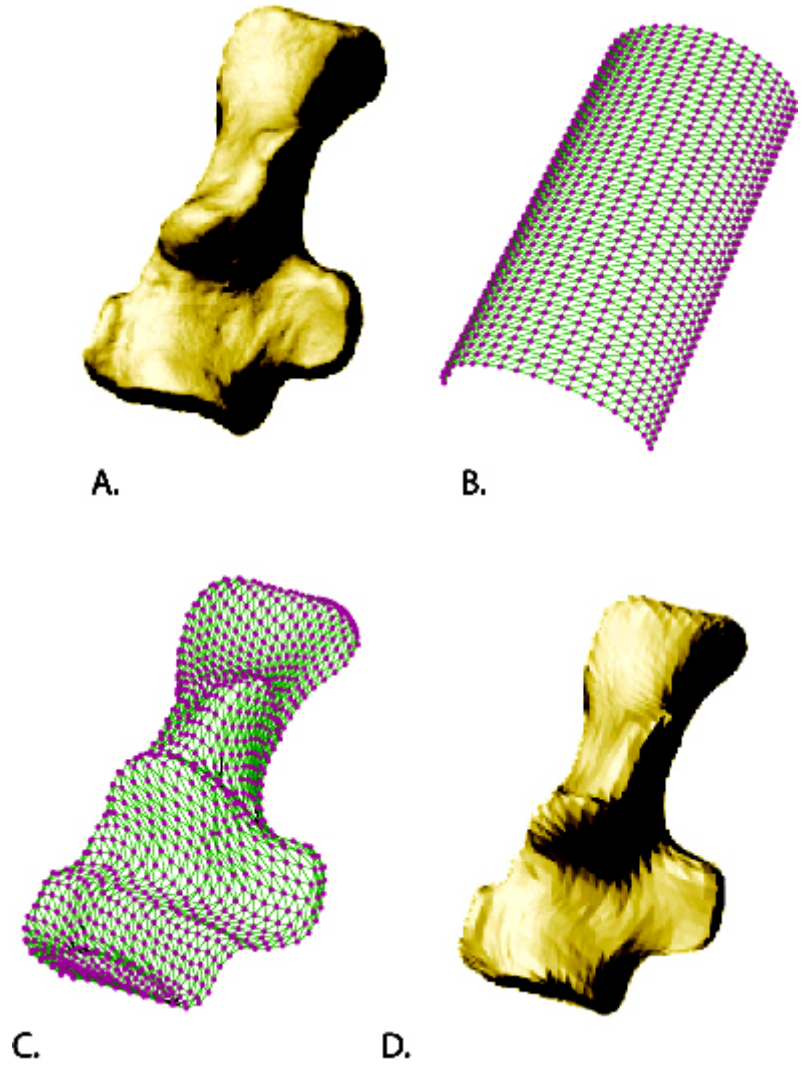
Polly, P. D. 2008. *Evolutionary Biology*,35, 83-96.



Perez, et al. 2006. Journal of Anatomy, 208: 769-784.

Surfaces

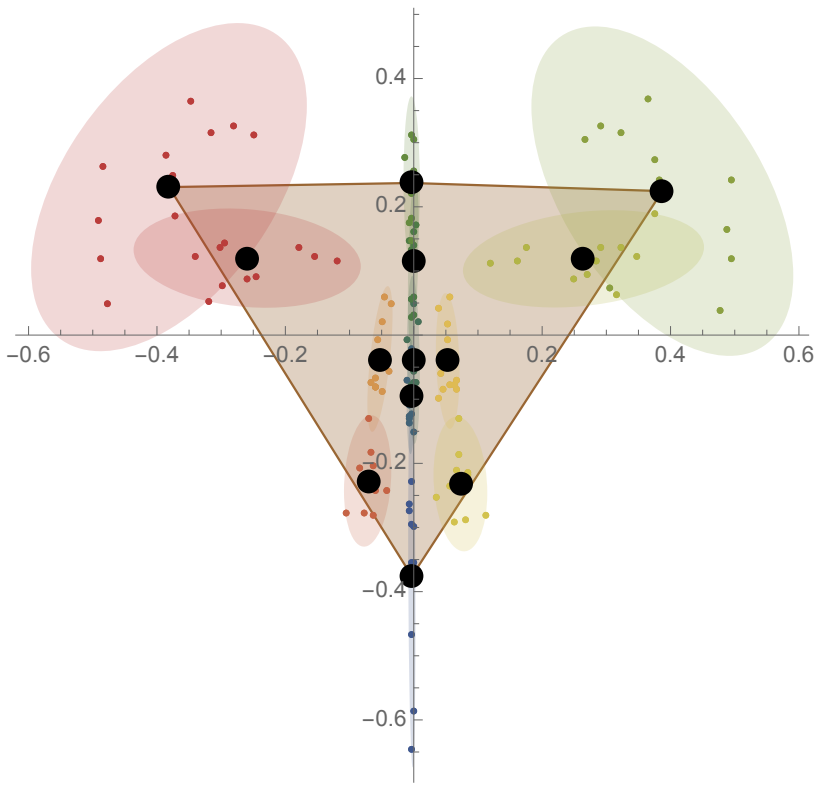
Surfaces.-semilandmark representation of the 3D surface of an object. Semilandmarks are quantified as Cartesian coordinates (x,y,z). Either ordinary (object-dependent) or sliding (sample dependent) semilandmarks can be used.

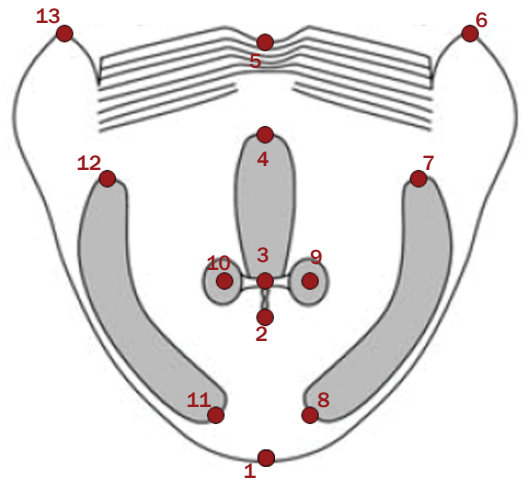


Polly, P. D. 2008. Adaptive Zones and the Pinniped Ankle: A 3D Quantitative Analysis of Carnivoran Tarsal Evolution. Pp. 165-194 in (E. Sargis and M. Dagosto, Eds.) Mammalian Evolutionary Morphology: A Tribute to Frederick S. Szalay. Springer: Dordrecht, The Netherlands.

Definitions

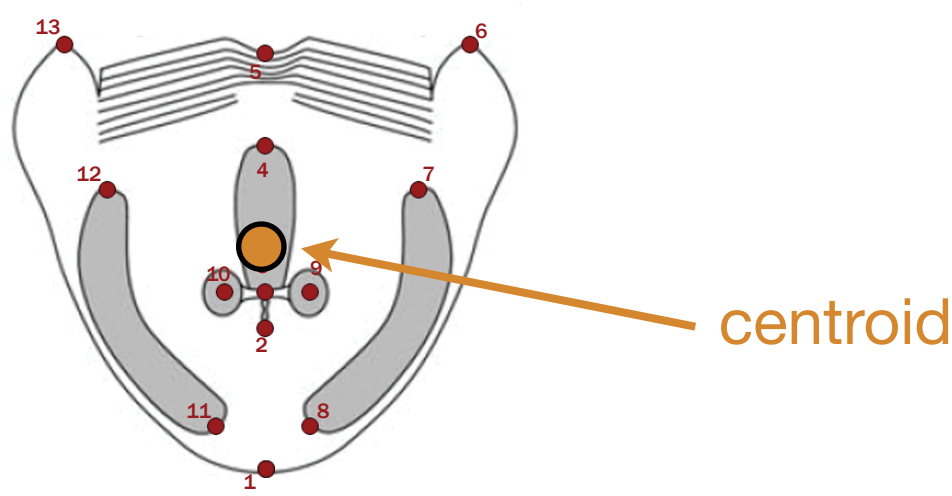
The consensus shape is the mean of the Procrustes coordinates, landmark by landmark.





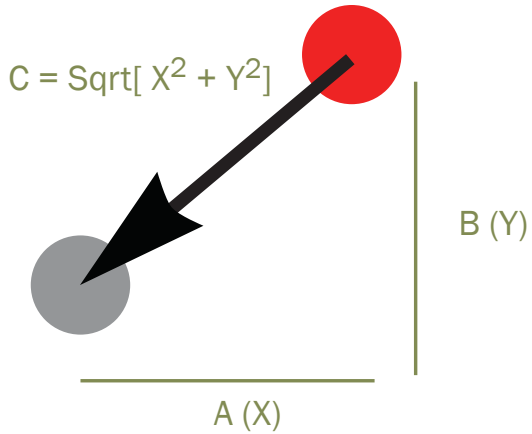
Definitions

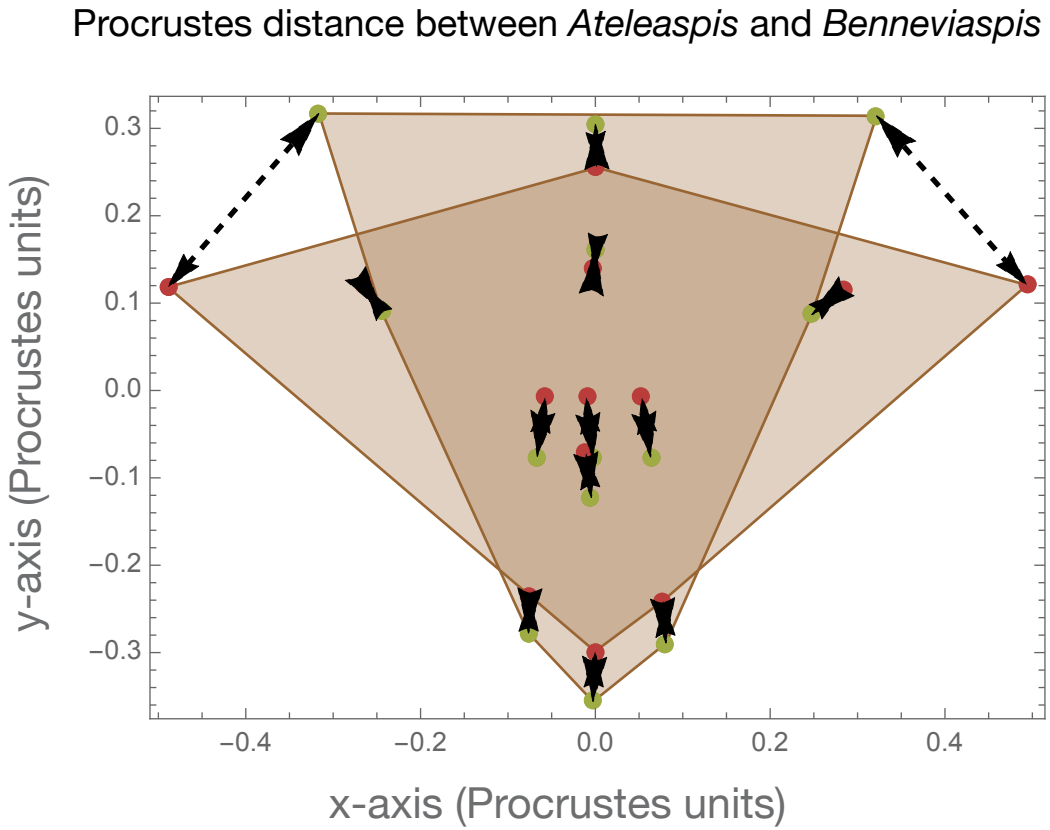
The centroid is the geometric center (the average of all the x- and all the y- coordinates).



Definitions

A Procrustes distance is the sum of all the distances between the corresponding landmarks of two shapes.





The true goal of all morphometric analyses...

...to measure morphological similarity and difference

Morphometric distances are the main measure of difference

Measured as the difference between objects (which might be specimens or means of species, or whatever) on all the variables being used

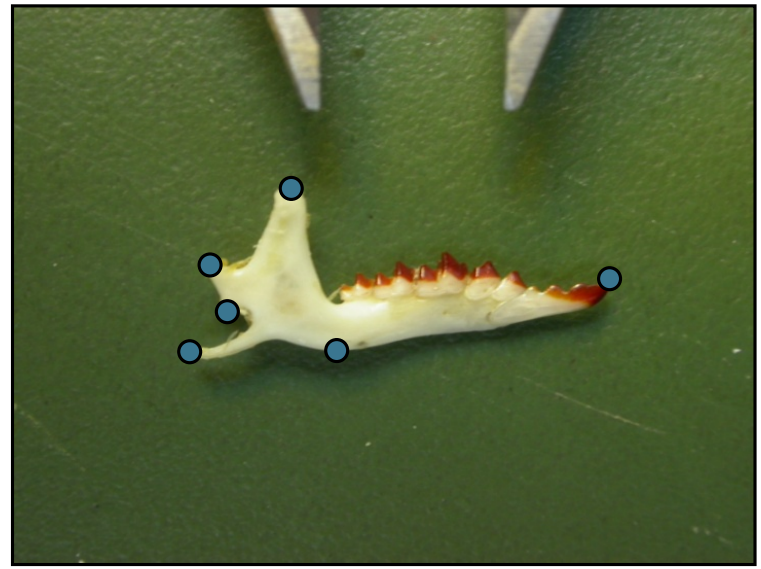
In GMM, the main measure of difference is the Procrustes distance, the distance between shapes after they have been superimposed

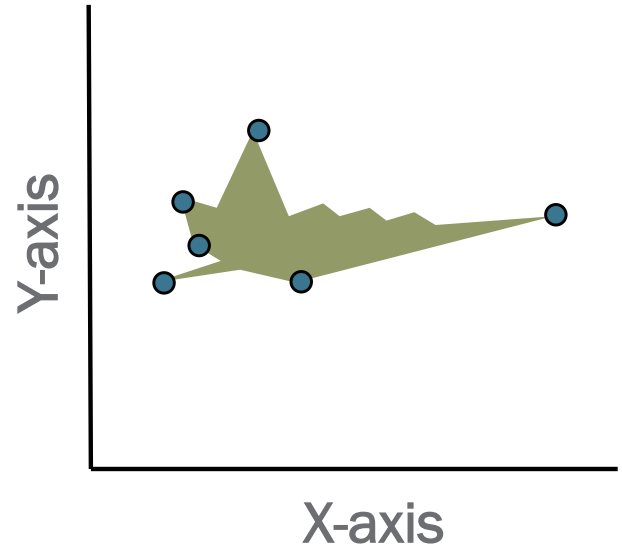
Advantages of geometric morphometrics

Results can be presented visually as a “shape” than tables of numbers

Data are easily collected from digital photographs

Size is mathematically removed from the analysis to focus on pure shape

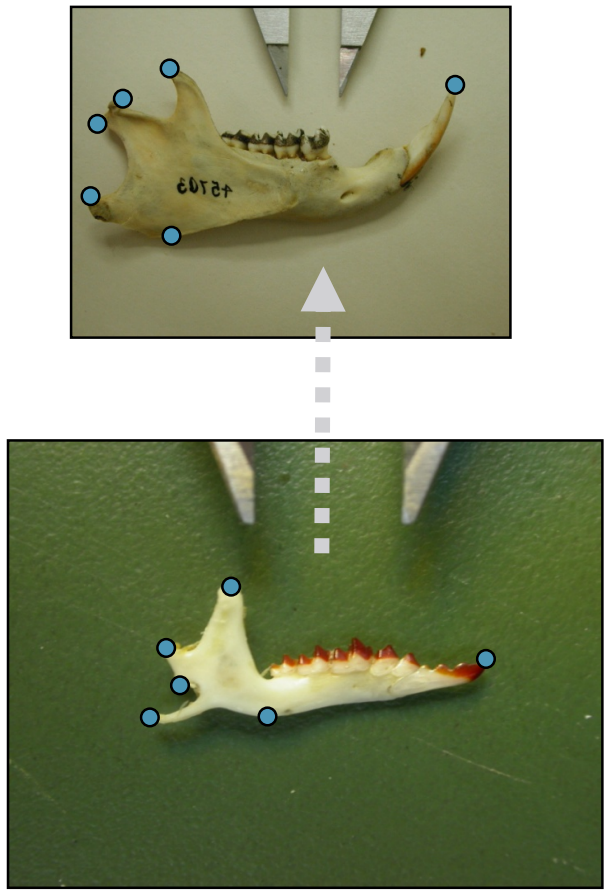




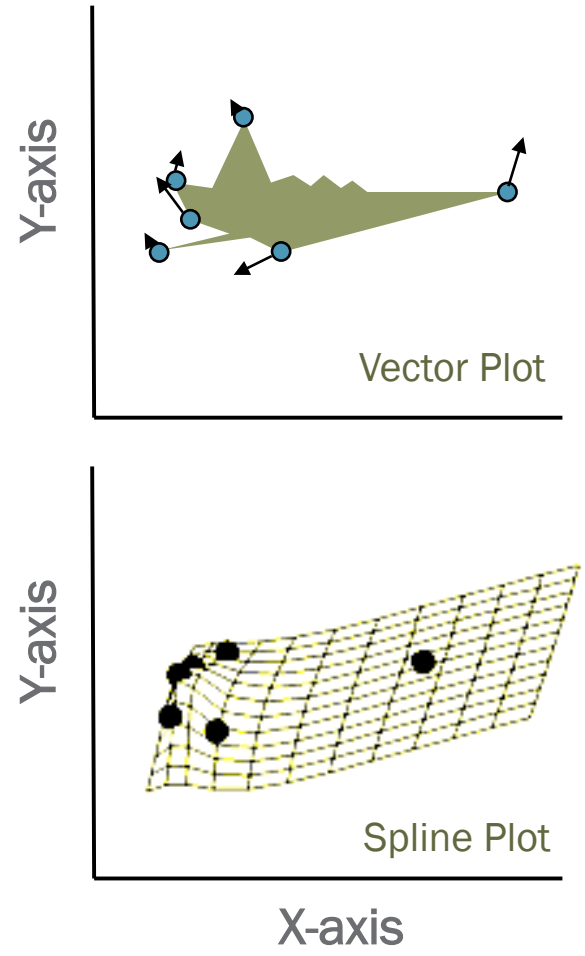
GMM results can be presented graphically

Difference in shape of mandibles of shrew and marmot

Difference in shape of mandibles of shrew and marmot

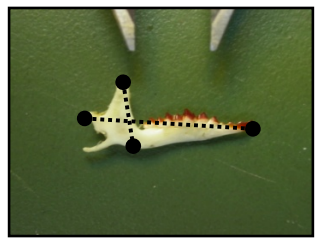


Two examples of how the differences are shown graphically

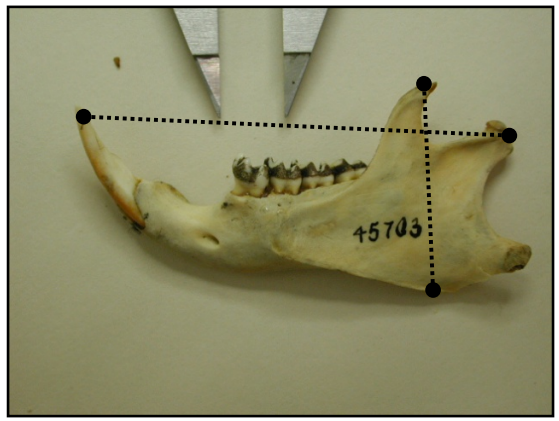


Traditional morphometrics mixes size and shape

The size of the animal affects all measurements so that primary morphometric difference between two taxa is size rather than shape



Shrew

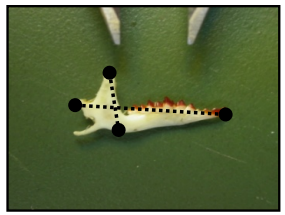


Marmot

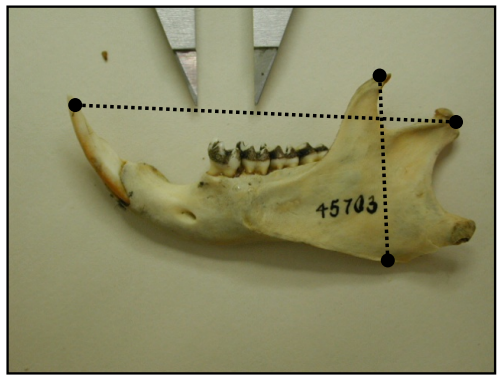
Geometric morphometrics removes size by rescaling

Shapes are enlarged or reduced to achieve a standard, equal size

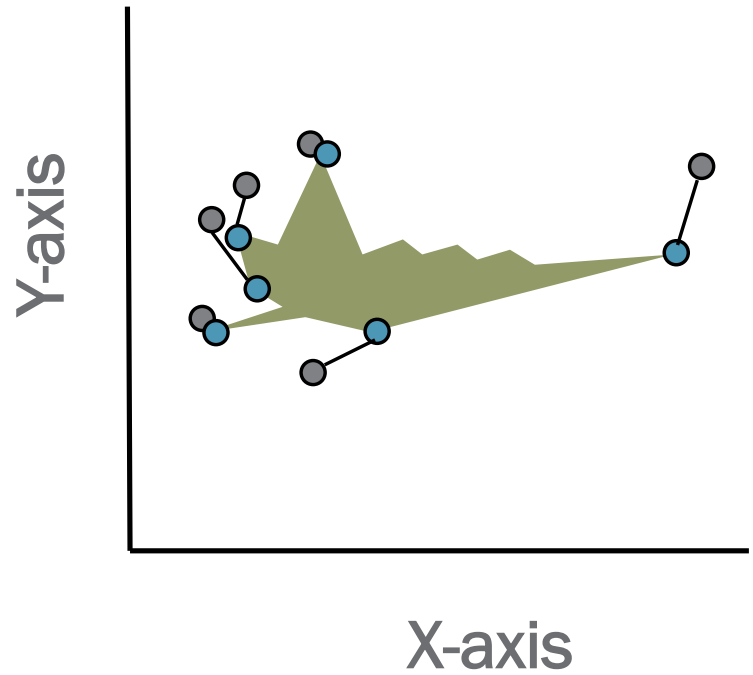
Coordinates of rescaled landmarks show differences only in their relative positions



Shrew



Marmot

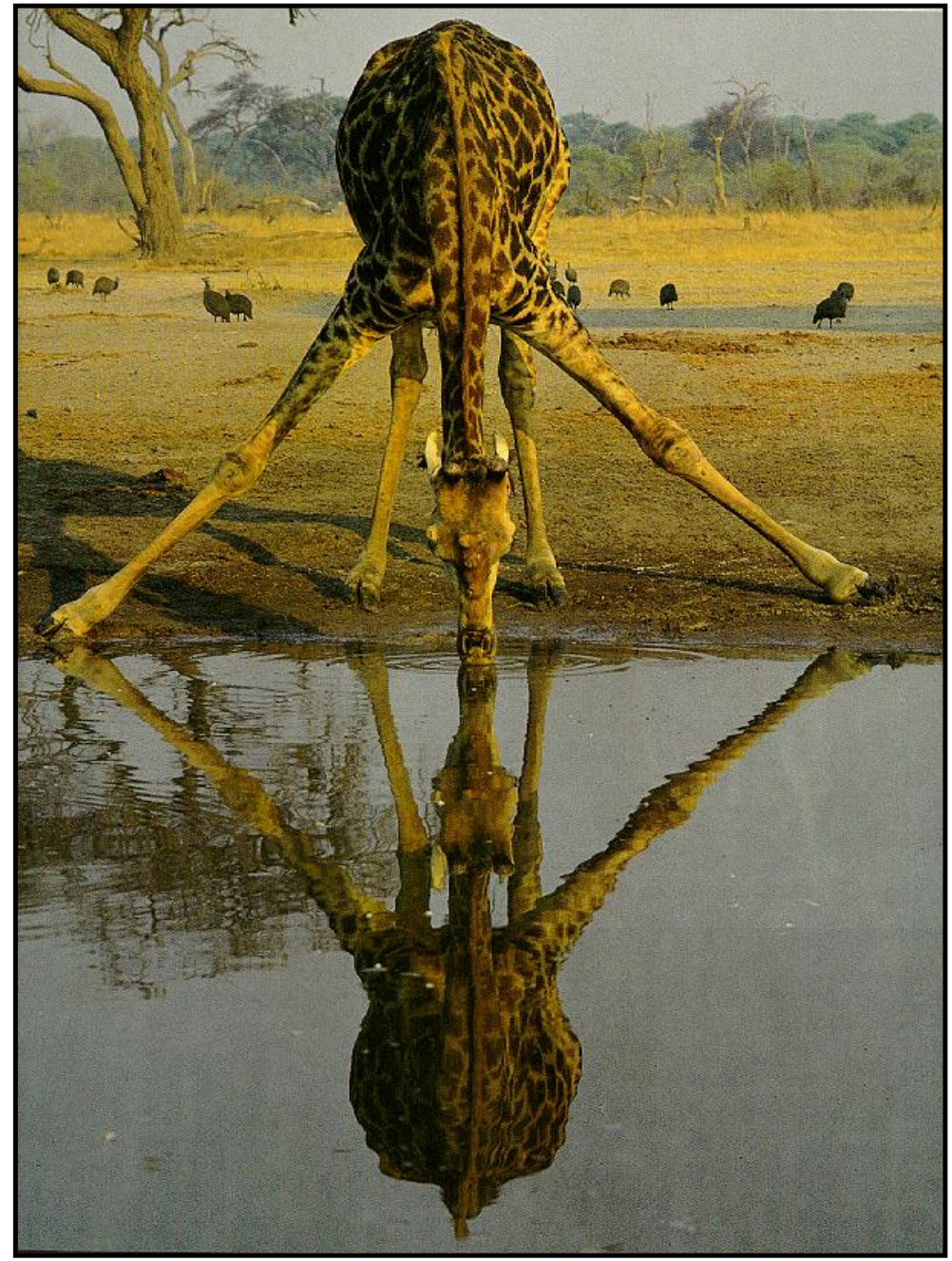


Disadvantages of geometric morphometrics

Size is completely absent from the analysis, and size may be biologically relevant

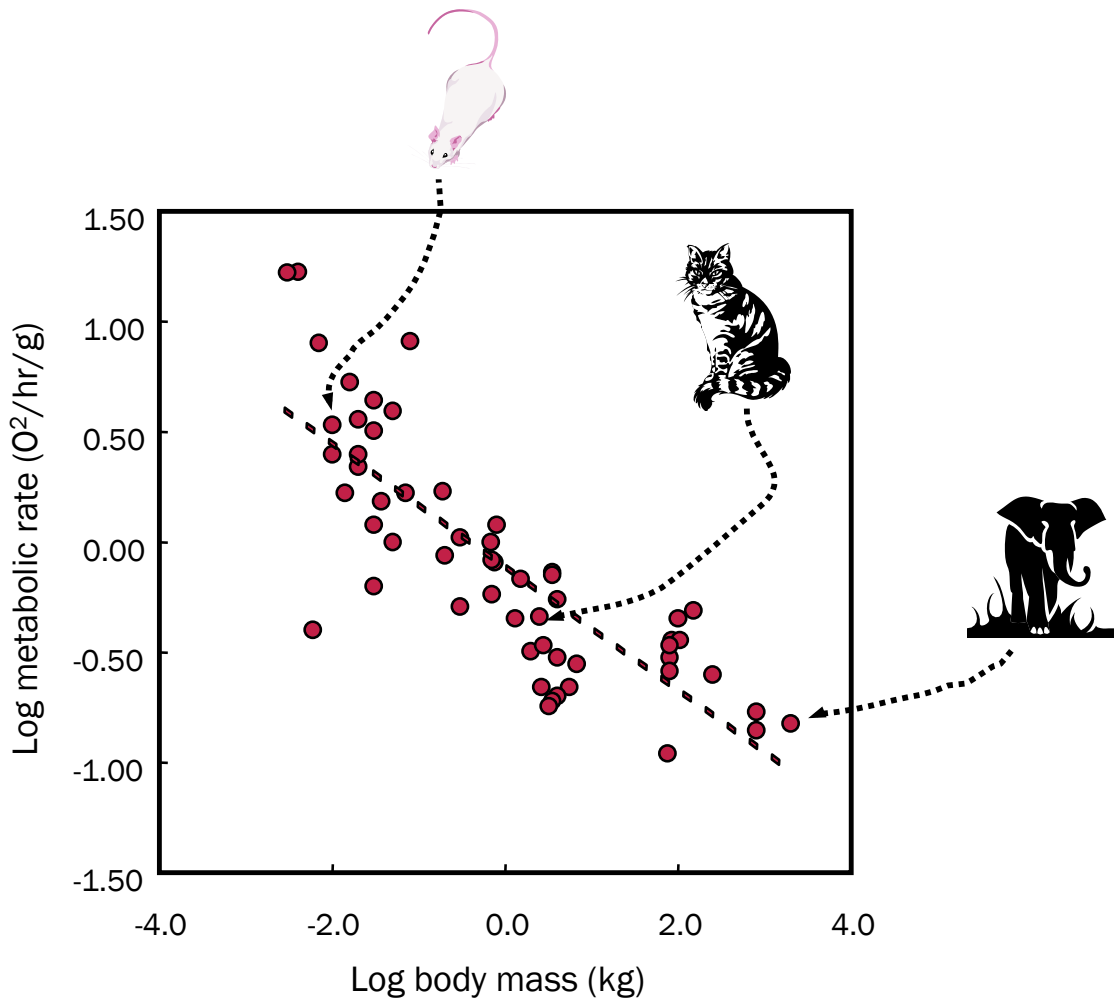
Only single rigid structures can be easily analyzed

GMM analyzes variation in entire shape, not in individual landmarks



Size is biologically important

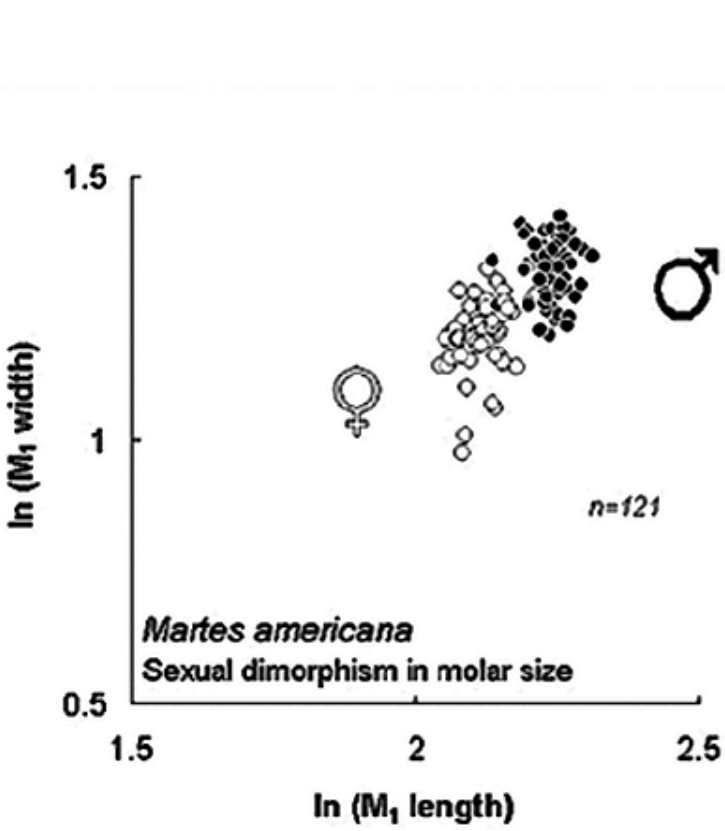
and it may be of interest in a morphometric analysis

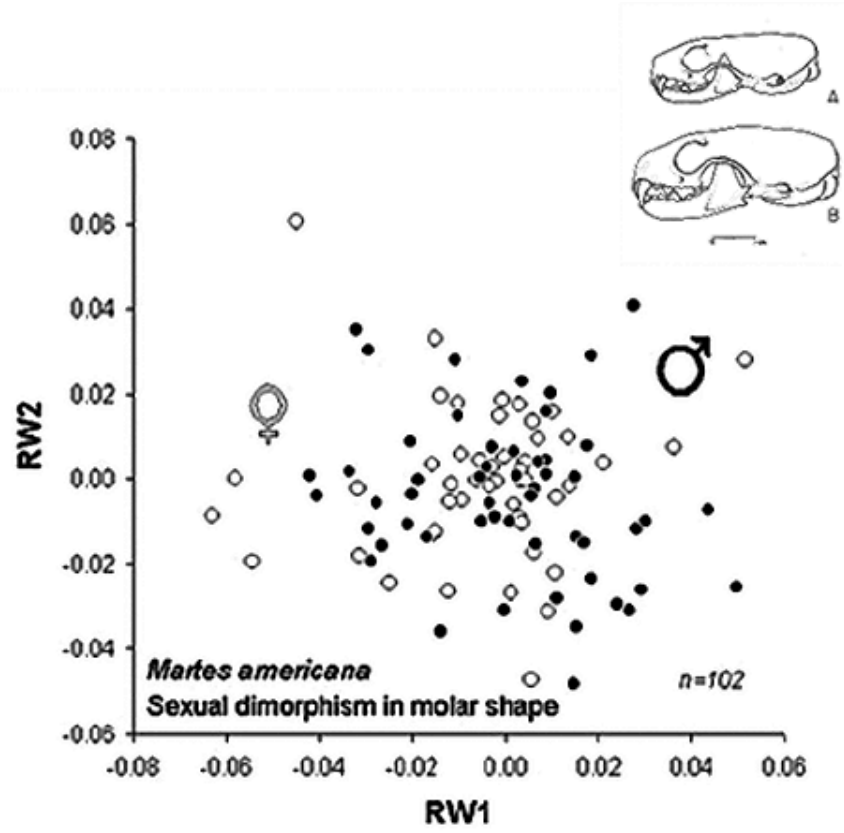


data from Eisenberg, 1981

Size and shape may behave differently

Size or shape may be desired in different analyses

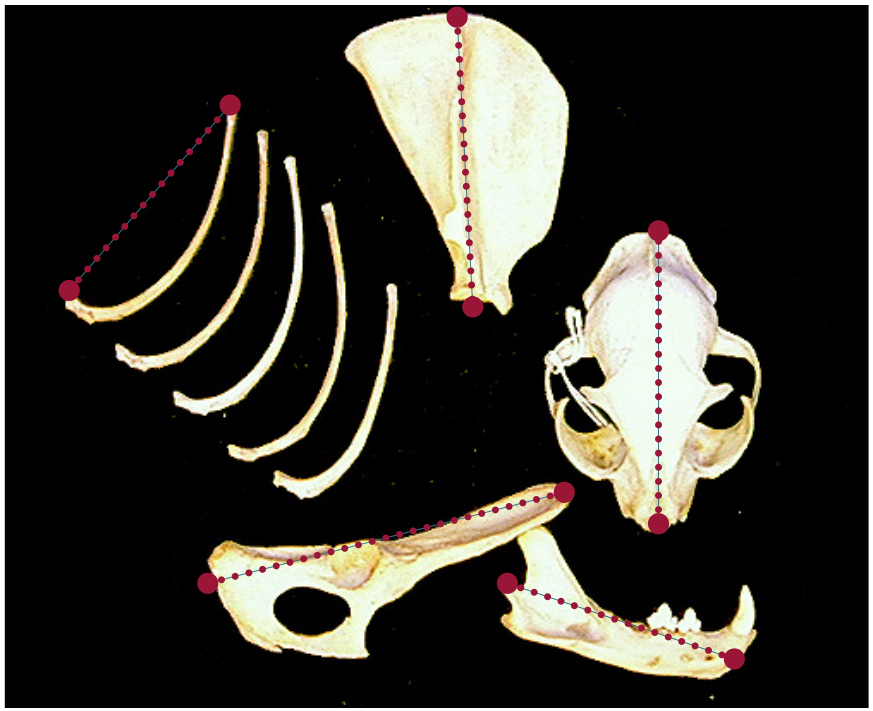




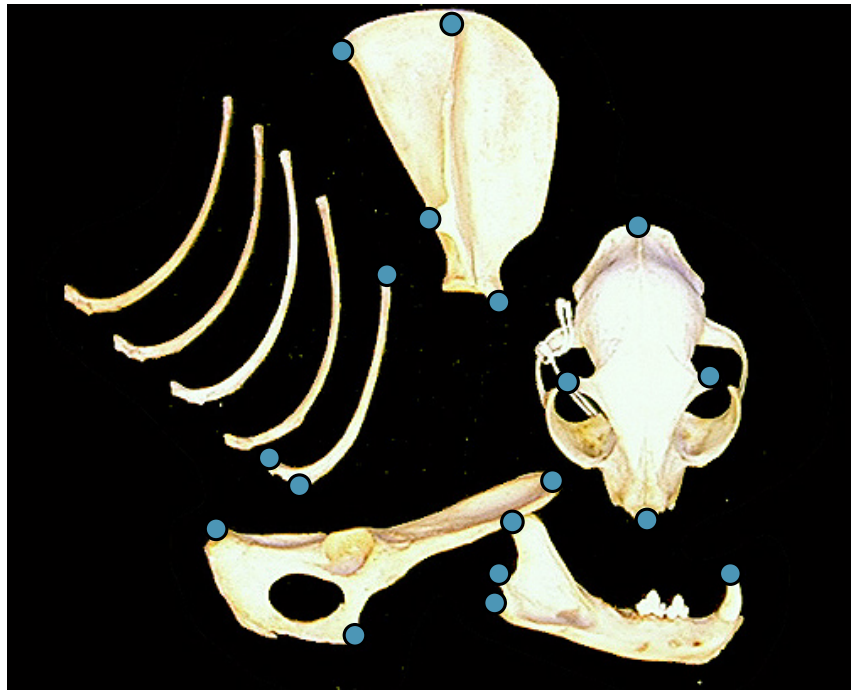
Polly, P. D. 1998. Variability in mammalian dentitions: size-related bias in the coefficient of variation. Biological Journal of the Linnean Society, 64: 83-99.

Only single rigid structures can represented with geometric morphometrics

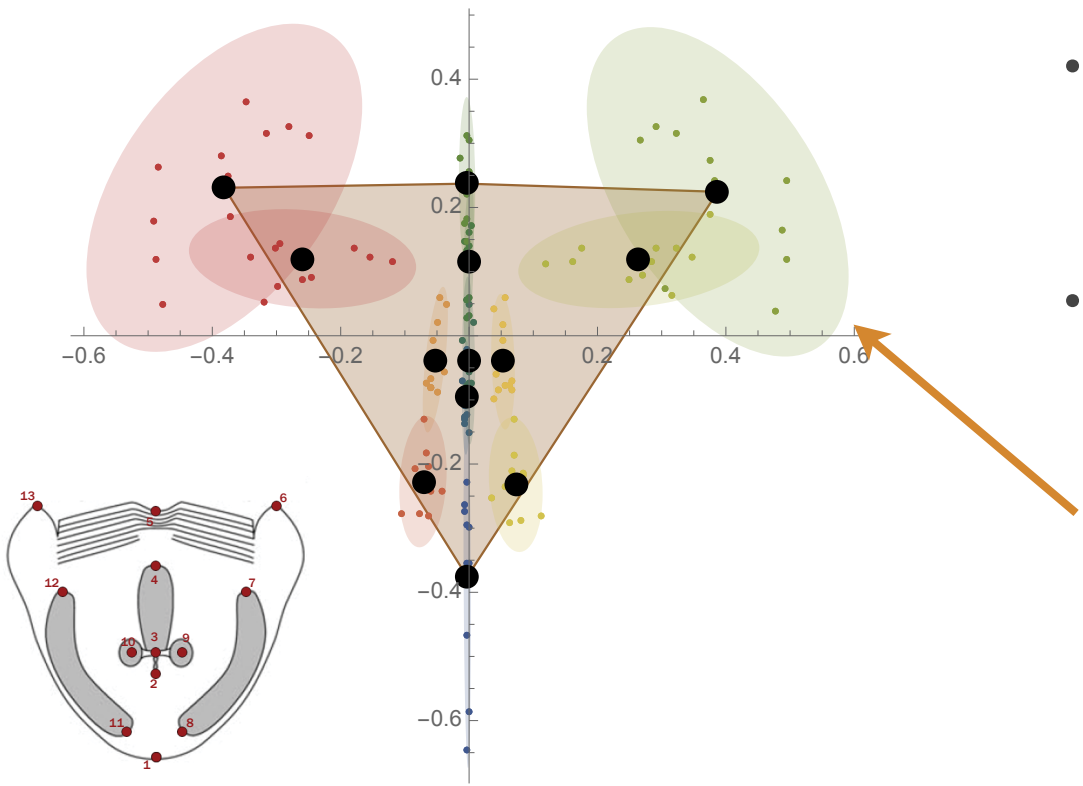
Okay



Not okay



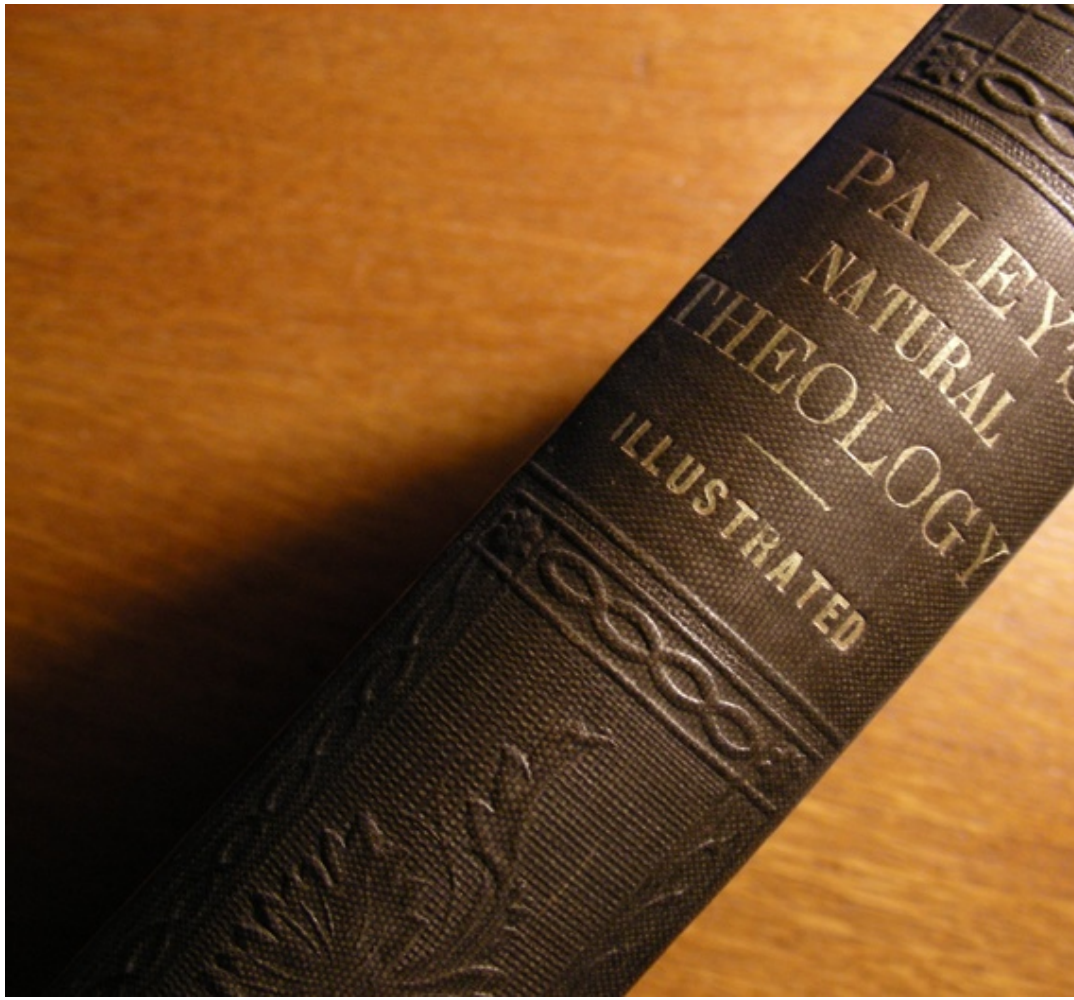
GMM cannot measure variation at individual landmarks



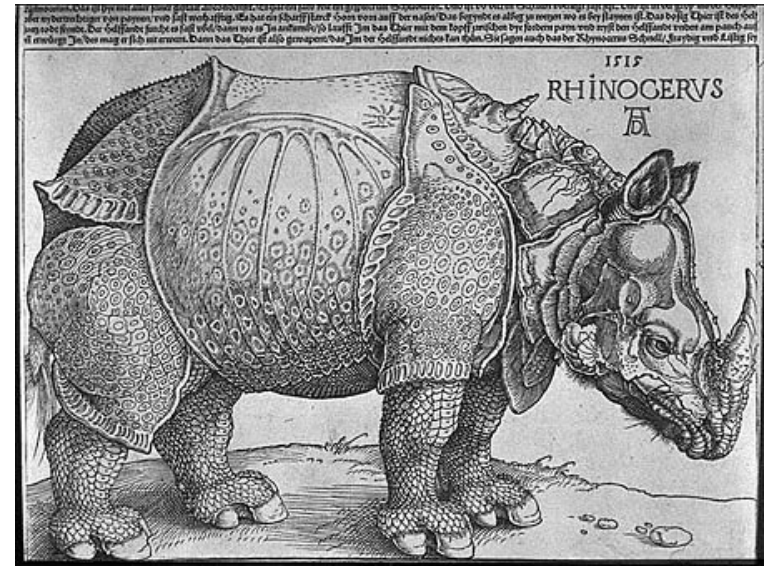
- Procrustes distributes variation by least-squares to minimize differences between whole shapes.
- Variation at an individual landmark cannot be interpreted as biological variation.
- Use EDMA (Euclidean distance matrix analysis) instead.

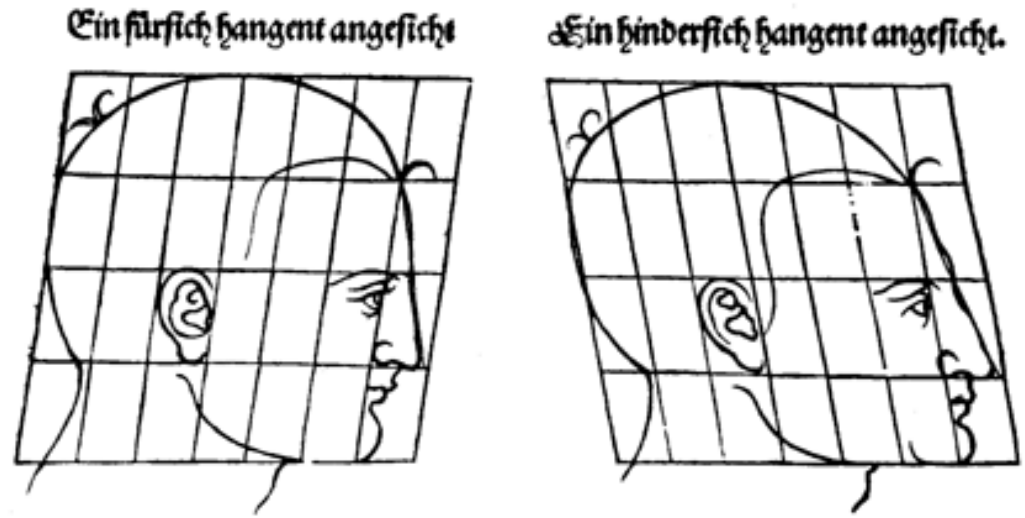
Is this landmark more variable than others?

A short history of geometric morphometrics....

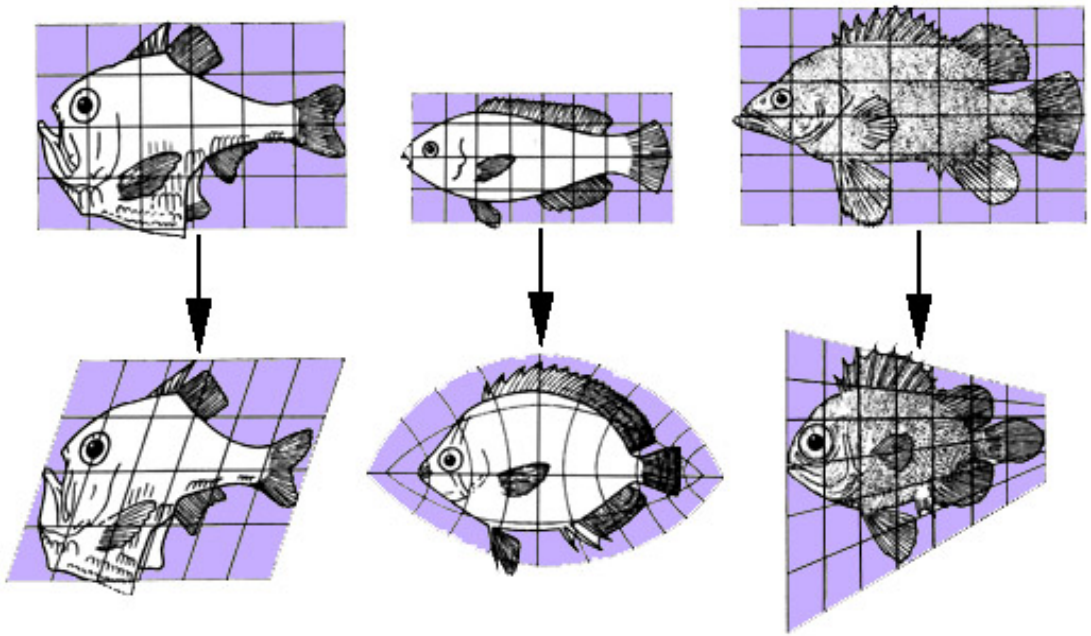


Albrecht Dürer (1471 -1528)





D'Arcy Thompson (1860-1948)



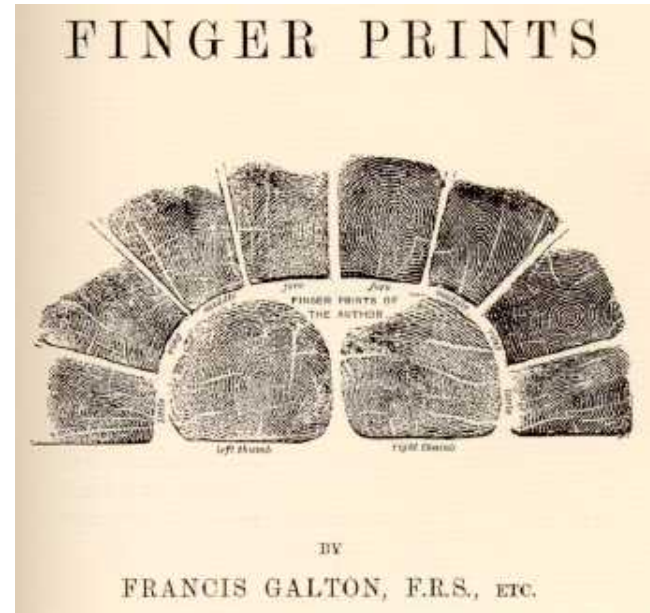
Francis Galton (1822-1911)

1891: starts biometric laboratory at University College London

Biometric approach to genetics: regression & correlation

Composite portraiture: photographs of different subjects combined (through repeated limited exposure) to produce a single blended image

Anthropometry & differential psychology: quantitative analysis of fingerprints



Modern Geometric Morphometric Methods

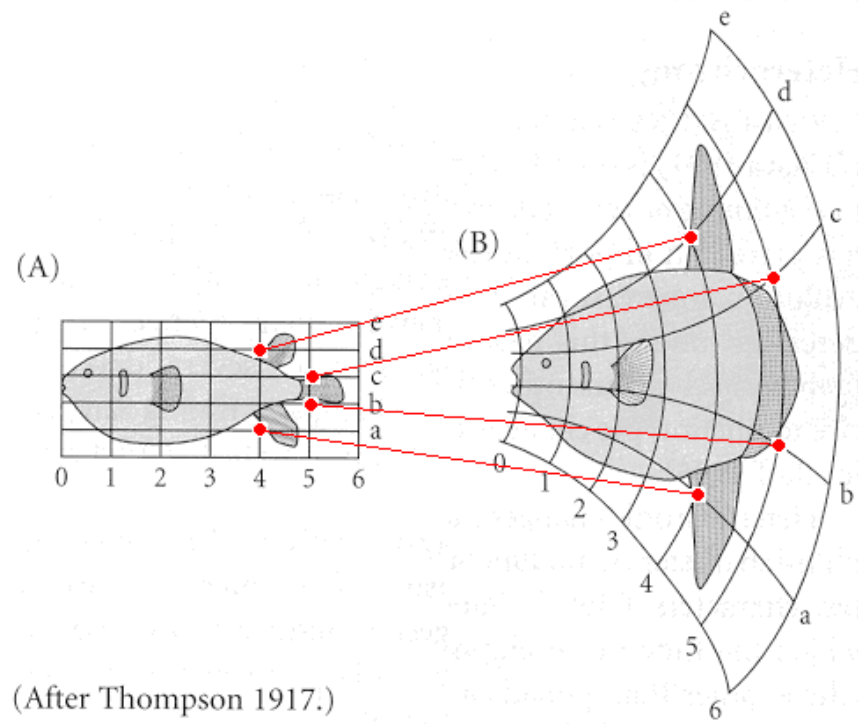
Bookstein

Development of landmark geometrics was driven by Fred Bookstein (long of University of Michigan, now Washington and Vienna)

Joined very productively by F. James Rohlf (Stony Brook)

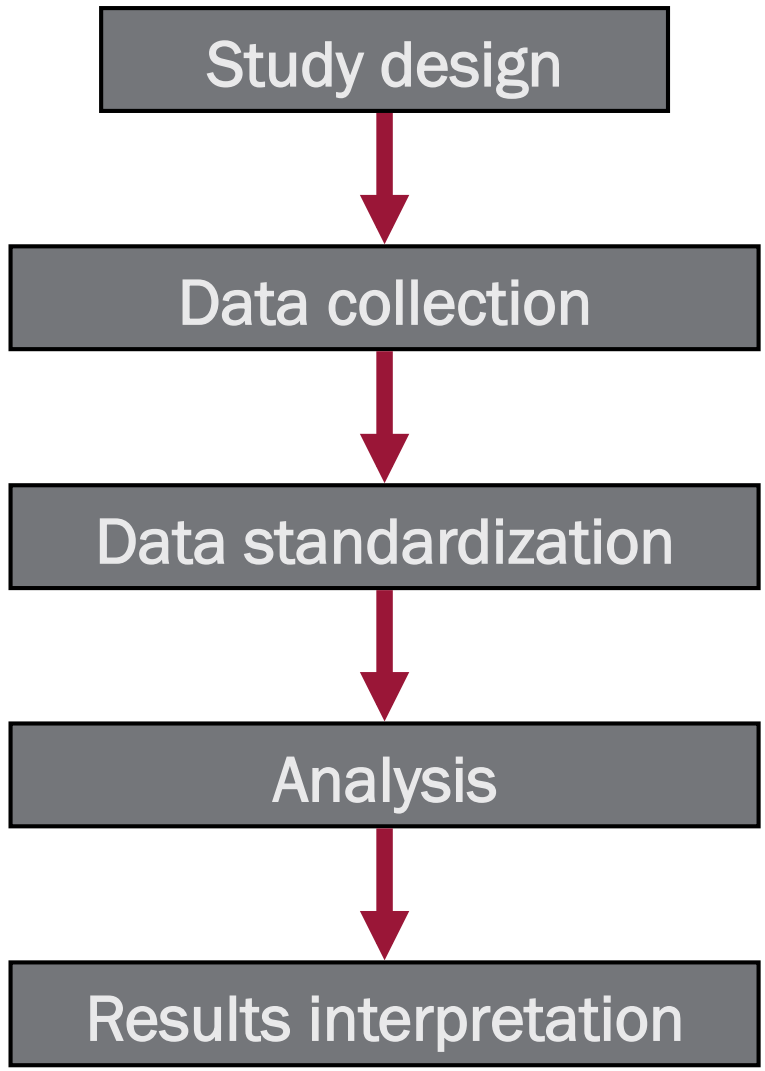
Ian Dryden, Kanti Mardia, Les Marcus, and Dennis Slice have been important names in developing techniques and theory.

Bookstein was originally intent on creating a truly quantitative way of producing d'Arcy Thompson's transformation grids.



(After Thompson 1917.)

Steps in a geometric morphometric study



Careful thought

Landmarking

Procrustes, PCA

Other statistical tests

Careful thought

How do you choose landmarks (or outlines, or surfaces)?

1. The data must reflect a hypothesis
2. The data must represent the shape adequately
3. Landmarks must be present on all specimens

Measurement Error and Sample size

1. Measurement error (ME) always exists in any collection of data, but ME doesn't matter if it is substantially less than the differences you want to measure.
2. Sample size required for a particular study depends on the within-group variation relative to differences between groups.

How many specimens do I need?

- Depends on the question being addressed
- Depends on the error in your data
- You need more specimens when the differences you want to measure are small compared to the variation within your group (natural or due to error)
- For sexual dimorphism in skulls of humans or other primates, 10 individuals of each sex might be enough
- For differences in genetic strains of mice where the mutation doesn't obviously affect the skeleton, 50 individuals of each strain is more realistic
- For species that belong to different families or orders, 1 specimen per species is almost always sufficient

What morphometrics can't answer for you..

- Morphometrics does not tell you what 'large' or 'difference' or 'shape' mean
  (These are definitions you must supply and your results depend upon them)
- Morphometrics does not tell you whether you unwittingly have two unrecognized groups in a single sample
  (Although comparison with known groups may help such an endeavour)
- How to identify cladistic characters
  (For the first two reasons combined)

Examples of available software

Digitizing landmarks and outlines: tpsDIG, ImageJ

Superimposition: Morpheus (plus integrated in some below)

Outline analysis: Eigenshape, PAST

MANOVA: Statistica, PAST

Discriminant functions, CVA: Statistica, PAST

Principal components analysis of landmarks: tpsRELW, PST

Construction of trees: PHYLIP, PAUP, NTSYSpc, PAST

All of the above plus simulations: Mathematica, *R, MorphoJ*

Links and downloads at SUNY Stony Brook morphometrics site: http://life.bio.sunysb.edu/morph/

Equipment: 2D outlines and coordinates

High-quality digital cameras

(resolution doesn't matter as much as the possibility of lens distortion: test your camera first by photographing a piece of graph paper and looking for "fish eye" distortion)

Calipers or scale bar

Equipment: 3D landmarks and outlines

Reflex Microscope for collecting three-dimensional landmarks, outlines and measurements (good for objects the size of a cat skull down to things about 2-5 mm long)

Microscribe G2

Microscribe robotic arm for collecting 3D landmarks and measurements (good for objects the size of a human skull down to a rat skull)

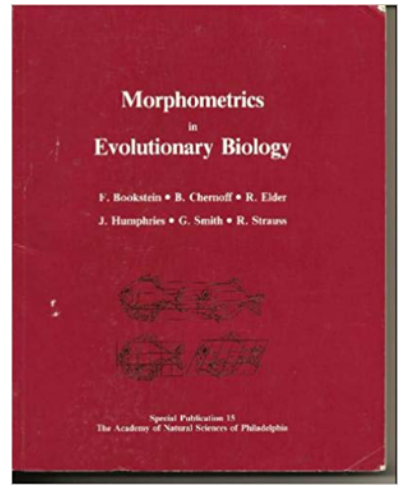
3D surfaces

Microscribe G2LX with the MicroScan Laser System modification

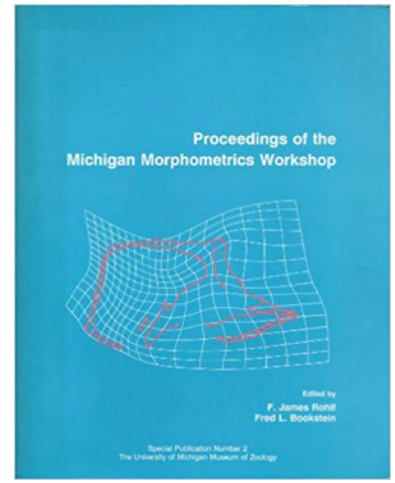
Microscan Laser scanner for scanning surfaces (good for objects the size of a cat skull down to about 2-3 cm long)

NextEngine laser scanner (good for objects the size of a horse skull down to a single tarsal bone)

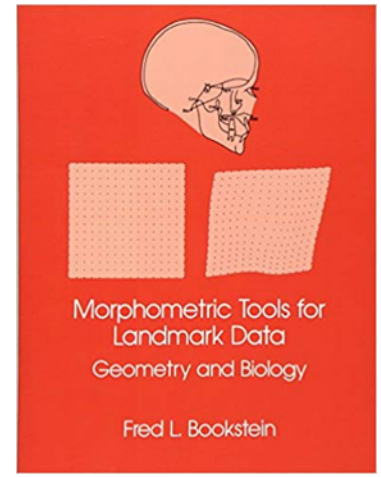
Methods (The Color Books)



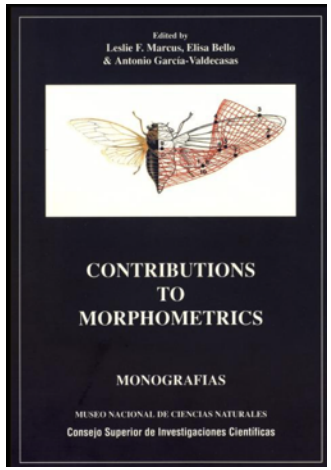
The Red Book (Bookstein et al. 1985)
Prelude to GMM, edited volume



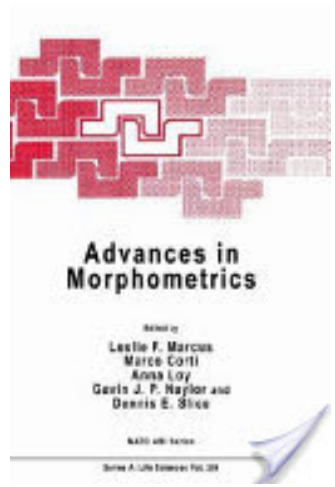
The Light Blue Book (Rohlf and Bookstein 1990)
edited volume



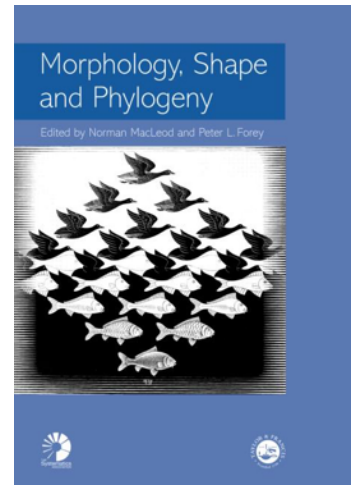
The Orange Book (Bookstein 1991)
THE primary source for GMM



The Black Book (Marcus et al. 1993)
Edited volume



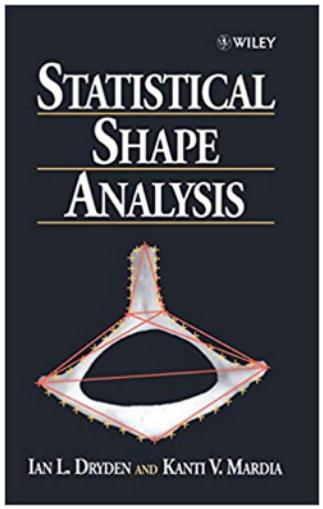
The White Book (Marcus et al. 1996)
Edited volume



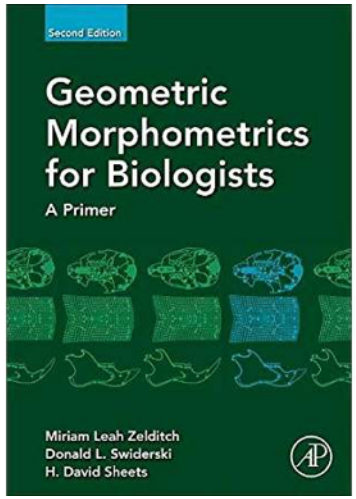
The Dark Blue Book (MacLeod and Forey 2002)
edited volume

The Yellow Book (MacLeod and Forey 2002)
edited volume

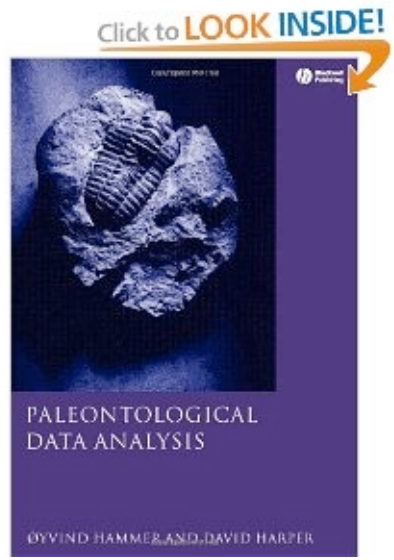
Methods (other key references)



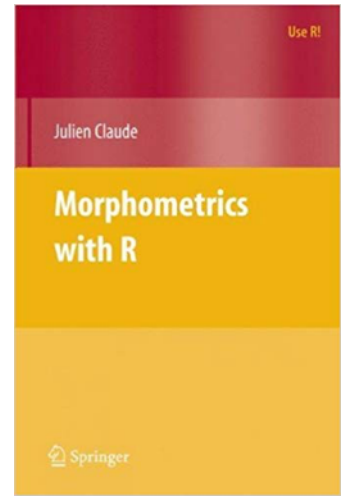
(Dryden and Mardia 1998)
Serious mathematical treatment of GMM



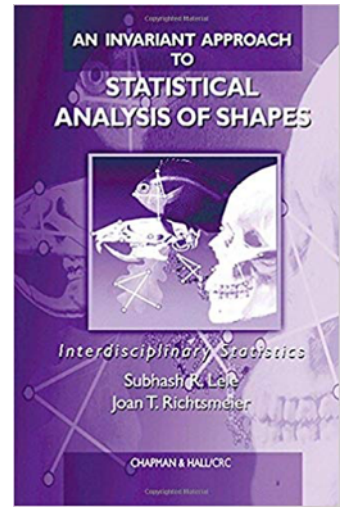
(Zelditch et al 2004)
Primer on GMM with full explanations of principles and methods



(Hammer and Harper 2006)
Encyclopedic reference with explanations and equations



(Claude 2008)
Hands on guide to theory and coding



(Lele and Richtsmeier 2001)
Shape analysis with EDMA