Advanced topics

(the leftovers)

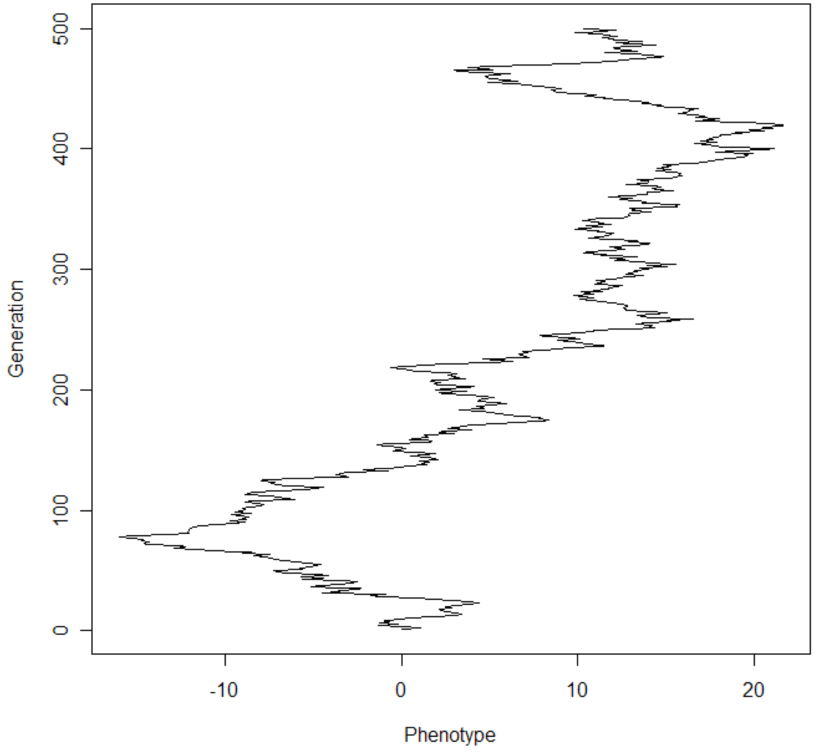
- Models of evolution as applied to shape
    - o Brownian motion
    - o Directional selection
    - o Stabilizing selection (OU)
- Curved spaces and tangent spaces
    - o Why is morphospace curved?
    - o What is tangent space?
    - o Does it matter?
- Morphometric transformations vs biological transformations

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

Quantitative evolution of morphology

Brownian motion of a single trait

1. Most likely outcome = starting value
2. Variance of the outcomes = number of step * (rate parameter)^2
3. Outcomes are normally distributed (reason is Central Limit Theorem: each step adds a random variable, sum of many random variables forms a normal distribution)



Brownian motion function for 2 uncorrelated traits

```
#################################################################
#
#    This function generates a Brownian-motion random walk
#    in two traits for n number of generations.  The default step
#    variance is 1.  Written by David Polly, 2008.
#
#################################################################

randomwalk <- function(n,r=1) {
scores <- matrix(ncol=3, nrow=n)
scores[1,] <- c(1,0,0)
for (i in 2:n) {
scores[i,1]=i
scores[i,2]=scores[i-1,2]+rnorm(1, mean=0, sd=sqrt(r))
scores[i,3]=scores[i-1,3]+rnorm(1, mean=0, sd=sqrt(r))
}
return(as.data.frame(scores))
}
```

Quantitative evolutionary theory

Change in phenotype

$$\Delta\bar{z} = \beta G$$

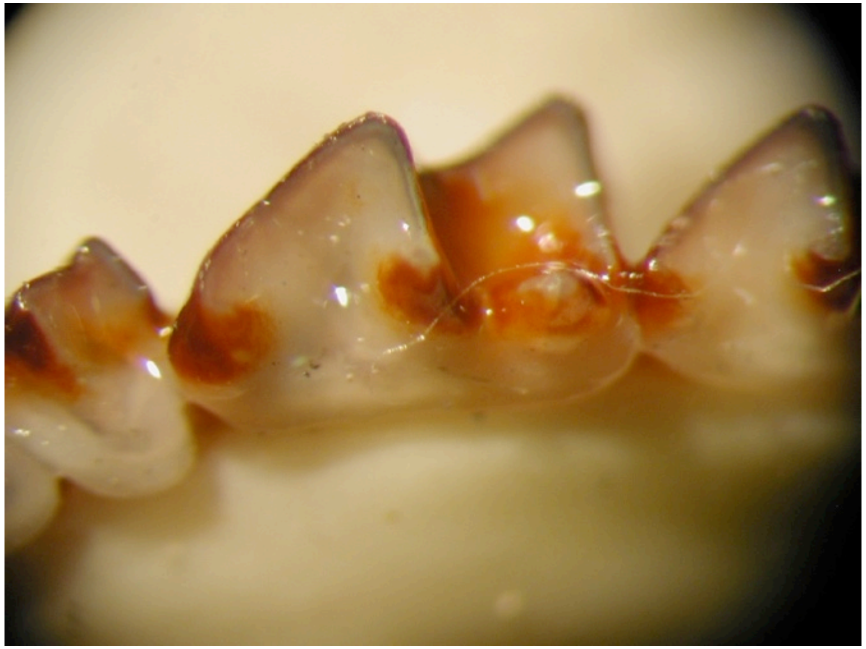
Additive genetic variance – covariance matrix

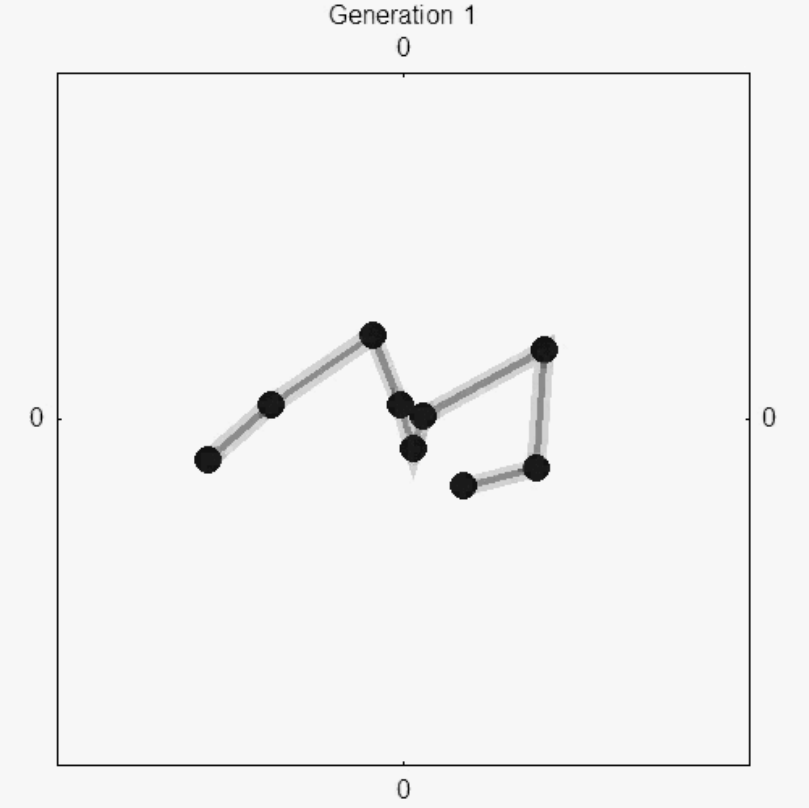
Selection coefficients

Selection coefficients can be:
Random
Directional
Stabilizing
Etc.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, **33:** 402-416.

Brownian motion on landmarks does not take into account biological covariance





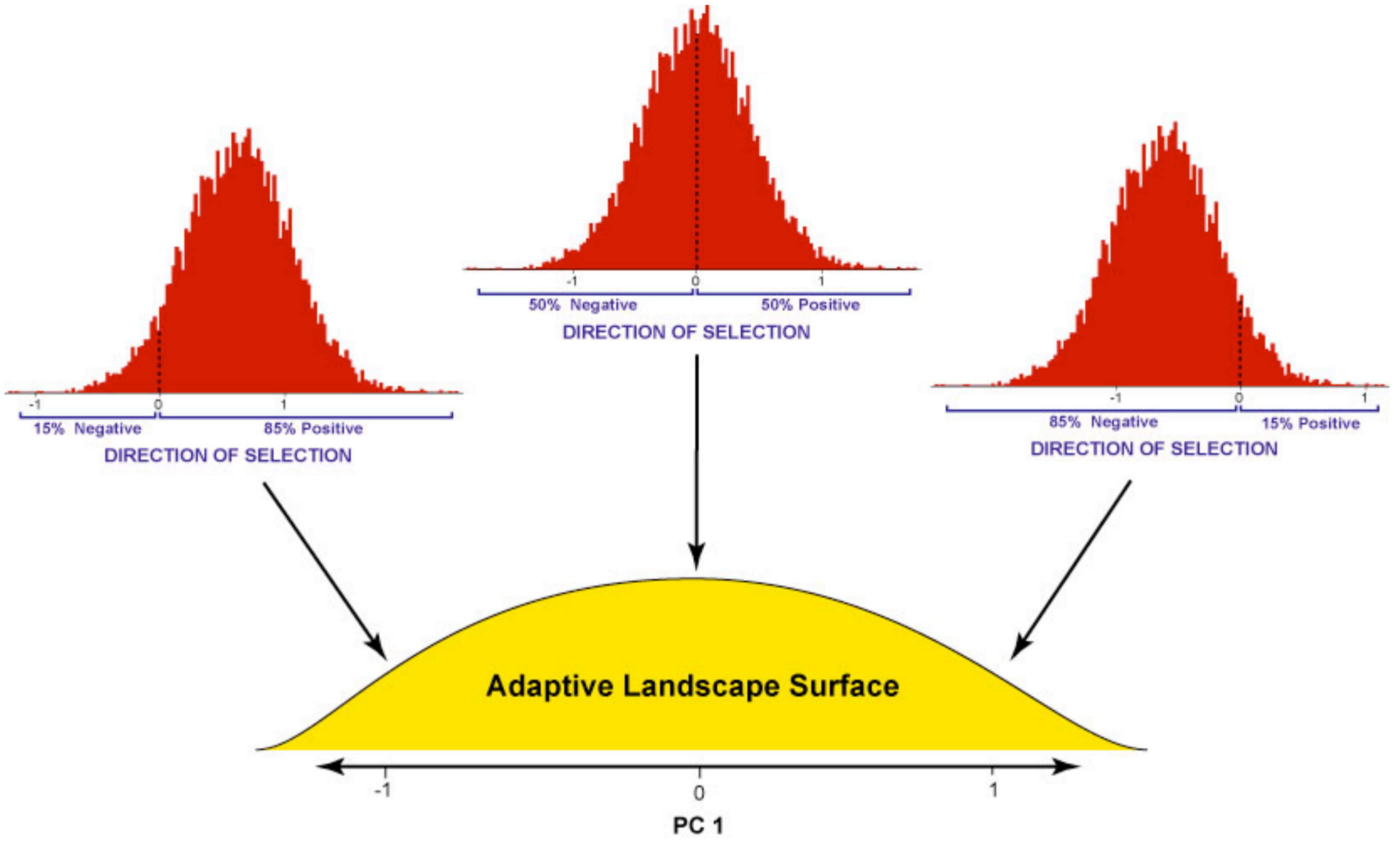
Polly, P. D. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica, 7.2.7A: 28pp, 2.3MB. http://palaeo-electronica.org/2004_2/evo/issue2_04.htm

To more realistically model evolution:

1. Estimate the additive genetic covariance matrix of traits for a <u>single</u> species
    a. Normally this is estimated from parent-offspring data
    b. Phenotypic covariance matrix (for a single species) can arguably be substituted
    c. Don't use covariance matrix based on multiple species because this confounds phenotypic covariances and phylogenetic covariances
2. Use this covariance matrix to construct a morphospace. Its PCs are the axes of genetic variation.
3. Estimate step rates from a phylogeny
4. Simulate evolution using desired model on the PC axes (which are independent) based on step rate
5. Reconstruct shapes using scores, eigenvectors, and consensus shape

Adaptive landscape

Wright, 1932 (original concept for allele frequency and reproductive fitness)
Simpson 1944 (phenotypic concept for macro evolution)
Lande, 1976 (quantitative theory for phenotypes)

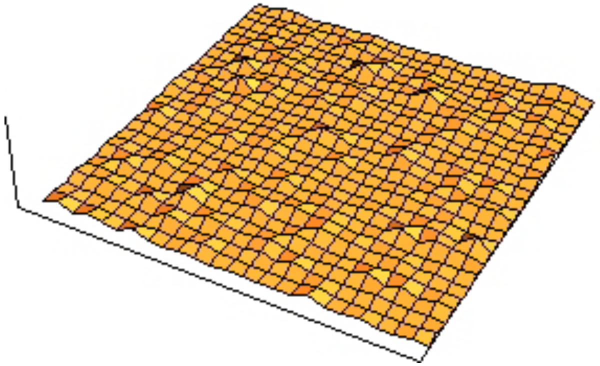


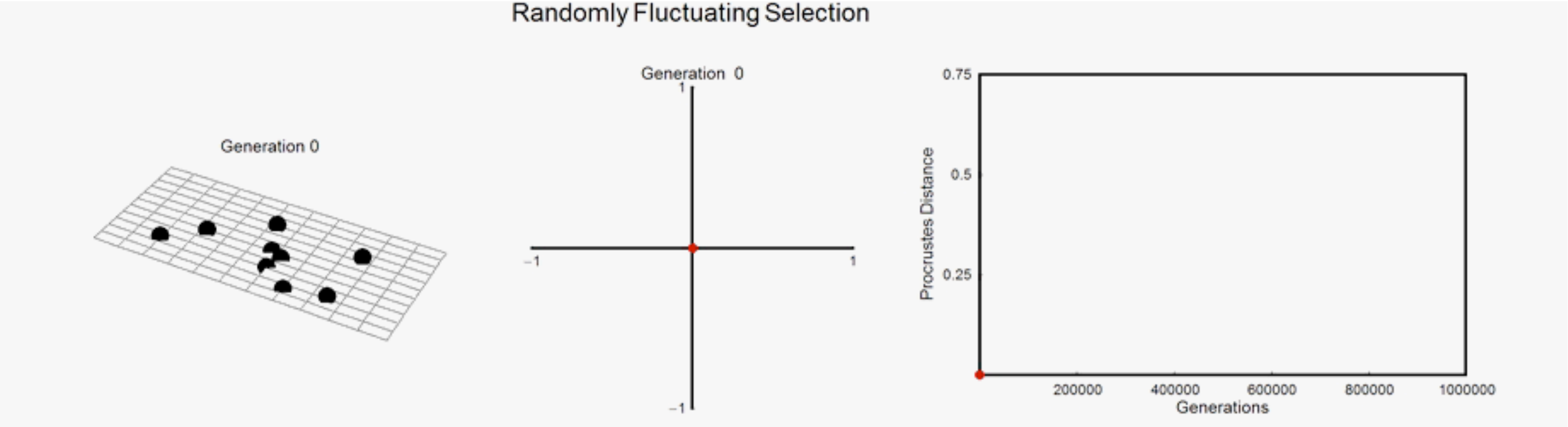
Polly, P. D. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica, 7.2.7A: 28pp, 2.3MB. http://palaeo-electronica.org/2004_2/evo/issue2_04.htm

Brownian motion

analogous to evolution on a flat adaptive landscape where random bumps appear and disappear

```
rnorm(1, mean=0, sd=sqrt(r))
```





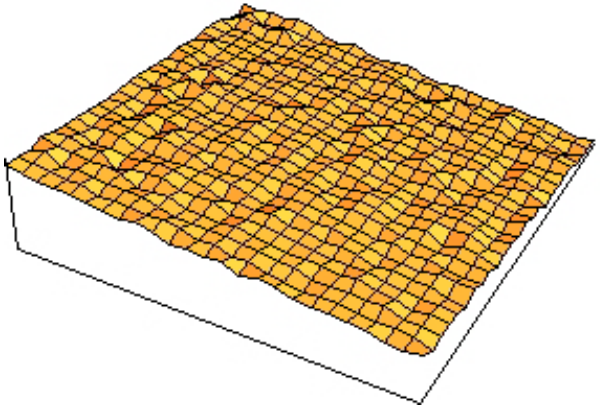
Shape model in landmark space

PC scores in shape space

Procrustes distance from ancestral (consensus) shape

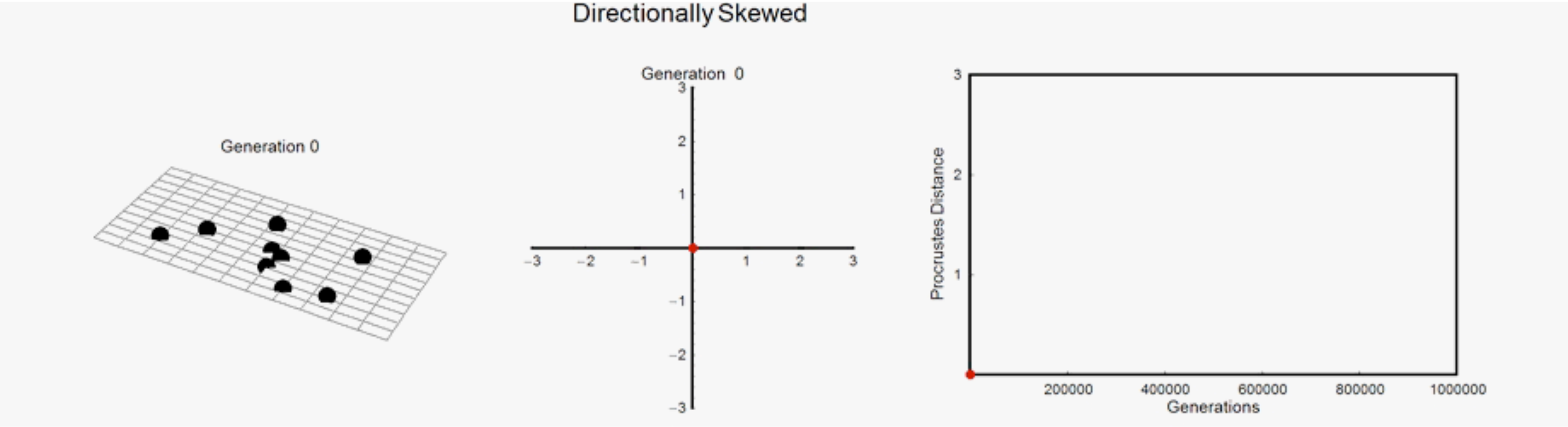
Polly, P. D. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica, 7.2.7A: 28pp, 2.3MB. http://palaeo-electronica.org/2004_2/evo/issue2_04.htm

Directional selection



analogous to a flat adaptive landscape that is tilted up in one direction

`rnorm(1, mean=-1, sd=sqrt(r))`

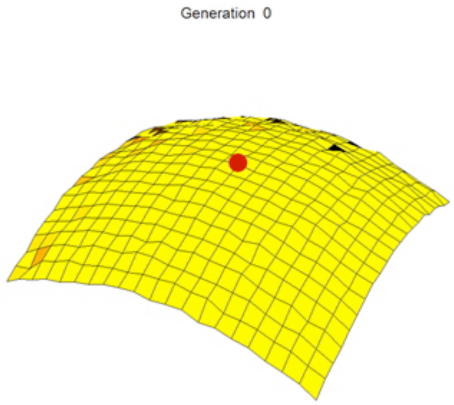


Shape model in landmark space

PC scores in shape space

Procrustes distance from ancestral (consensus) shape

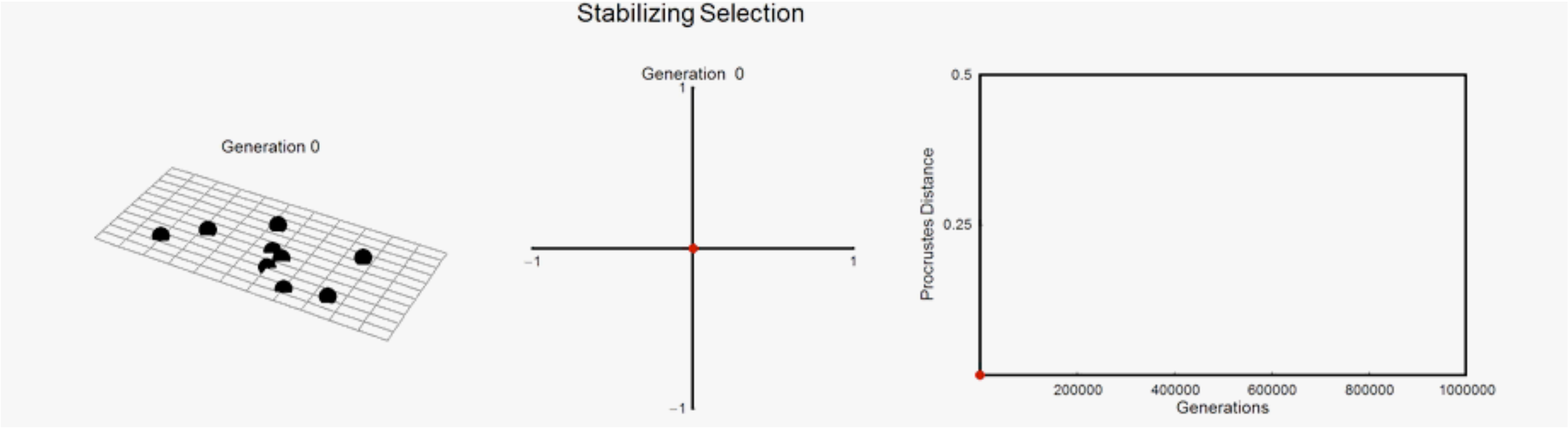
Polly, P. D. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica, 7.2.7A: 28pp, 2.3MB. http://palaeo-electronica.org/2004_2/evo/issue2_04.htm



Stabilizing selection

analogous to classic adaptive peak

`rnorm(1, mean=-1*score, sd=sqrt(r))`



Shape model in landmark space

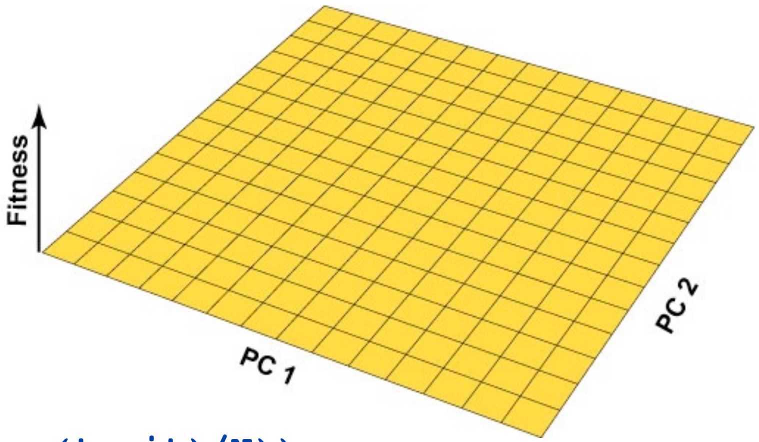
PC scores in shape space

Procrustes distance from ancestral (consensus) shape

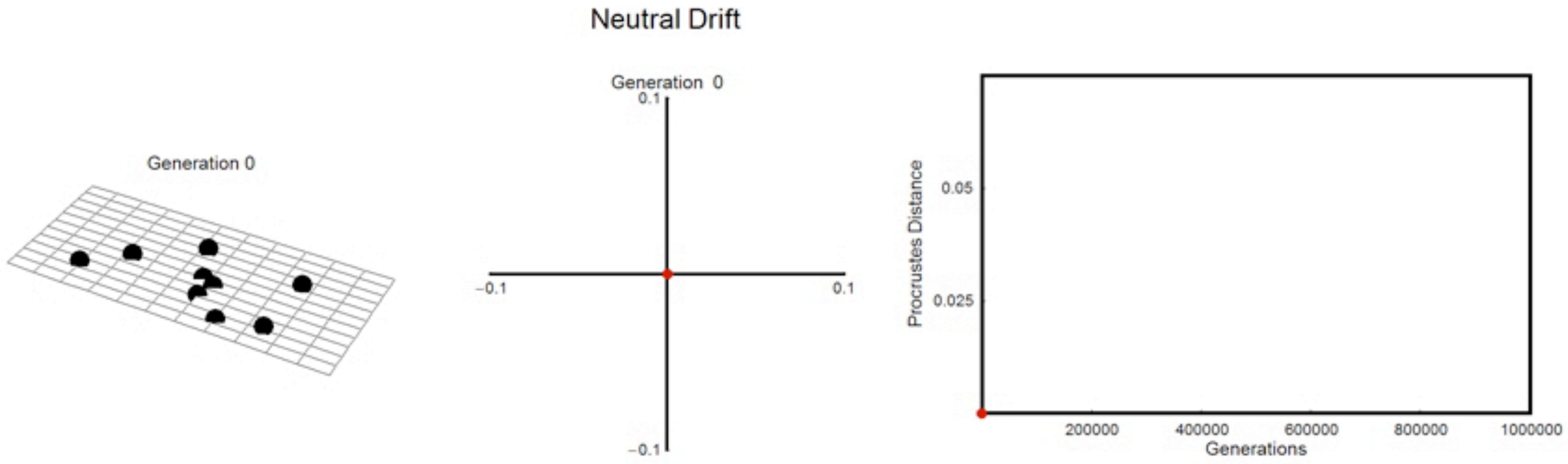
Polly, P. D. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica, 7.2.7A: 28pp, 2.3MB. http://palaeo-electronica.org/2004_2/evo/issue2_04.htm

Genetic drift (type of BM)

Perfectly flat landscape where change occurs by chance sampling from one generation to the next. Change is small and a function of population size



`rnorm(1, mean=0, sd=sqrt(var(trait)/N))`

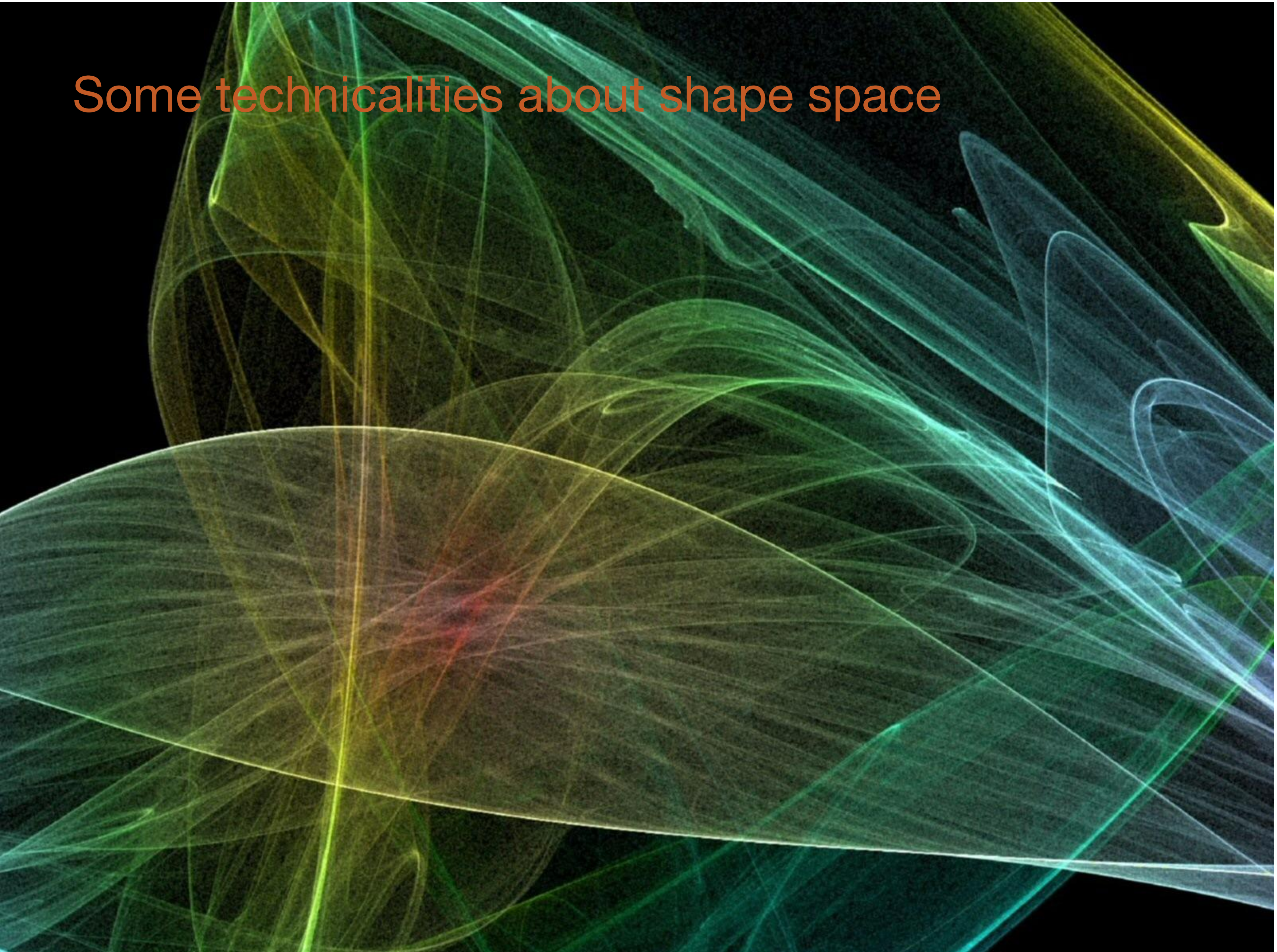

Neutral Drift

Shape model in landmark space

PC scores in shape space

Procrustes distance from ancestral (consensus) shape

Polly, P. D. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica, 7.2.7A: 28pp, 2.3MB. http://palaeo-electronica.org/2004_2/evo/issue2_04.htm



Some technicalities about shape space

Dimensions of shape space

Dimensions.– because size, rotation and translation have been removed from the data, the dimensions of morphospace are fewer than simply the number of landmark coordinates.

For 2D:

Dimensions = 2K – 4

where K is the number of landmarks, and four dimensions are lost due to size (-1), translation (-2), and rotation (-1)

For 3D:

Dimensions = 3K – 7

where K is the number of landmarks and seven dimensions are lost due to size (-1), translation (-3), and rotation (-3)

For both:

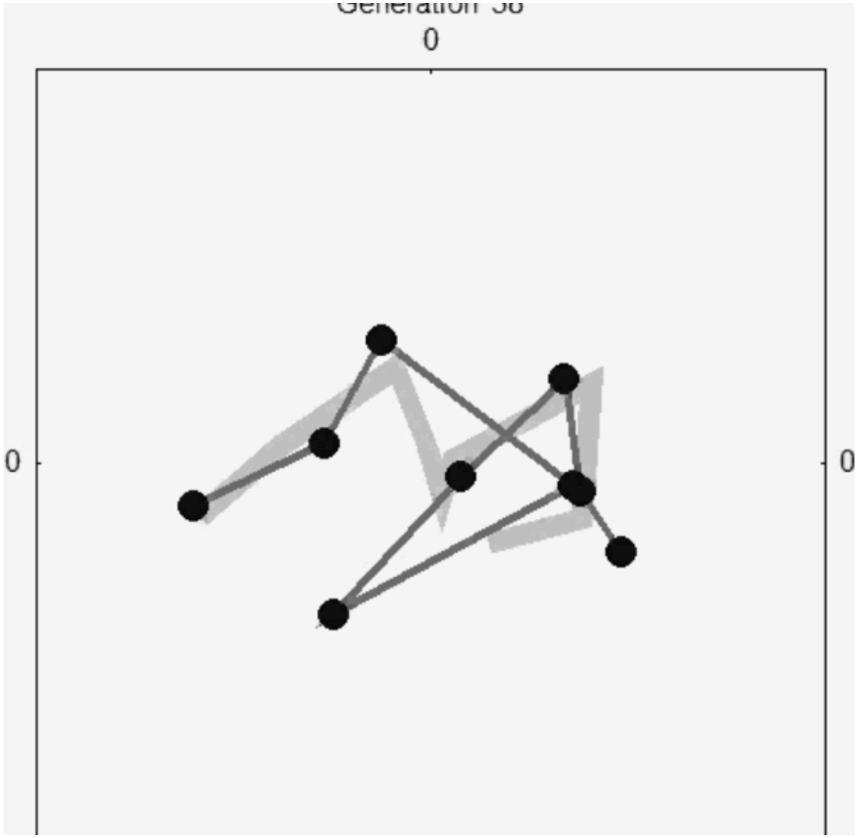
Dimensions = N-1

if N-1 < above dimensions, where N is the sample size

Number of dimensions

Morphospace is defined by the number of landmarks and their dimensionality. **All shapes** with the same number of landmarks fit somewhere in that morphospace, regardless of how different they are.

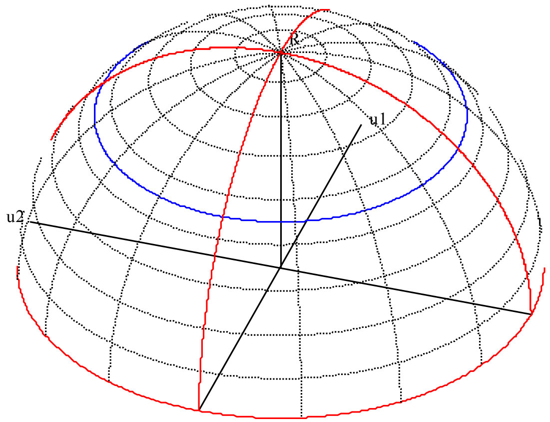
Important observation: Most biological shapes are similar to one another compared to random shapes with the same number of landmarks. Thus, the biological shapes cluster in a small region of shape space.



Morphospace is curved

Shape space is not Euclidean because of the dimensions lost to translation, scaling, and rotation

Shape space belongs to class of spaces called Reimannian manifolds, the two dimensional version of which is a sphere



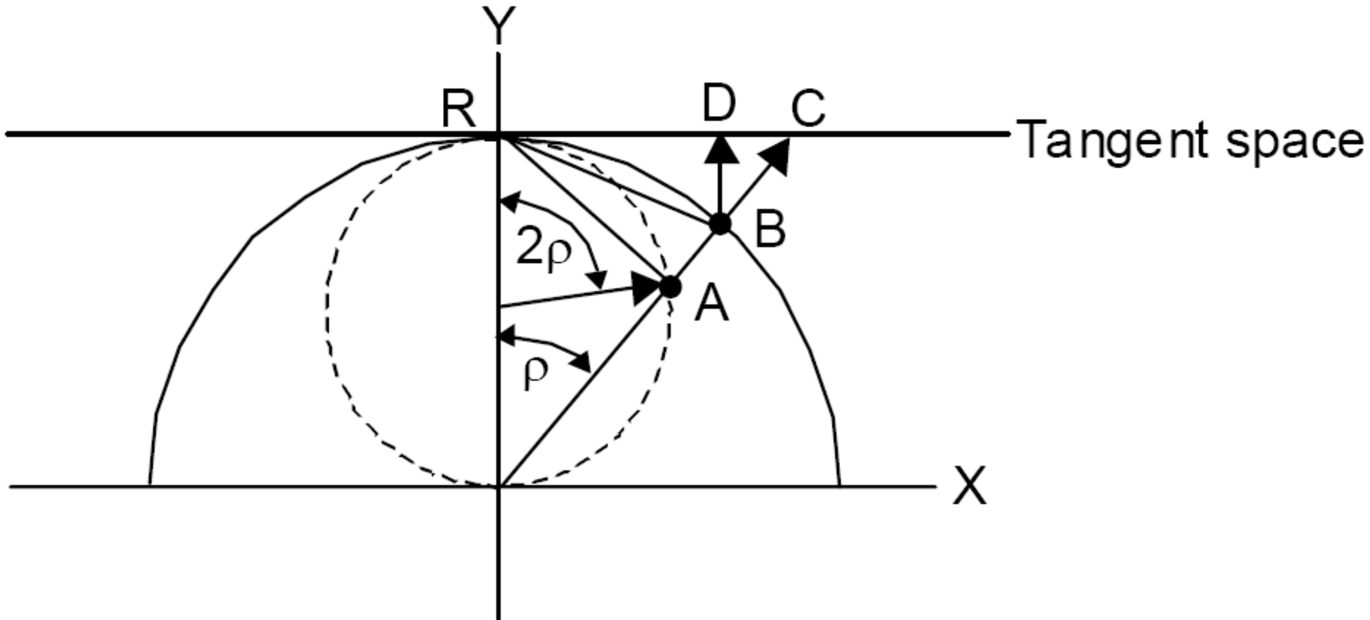
Morphospace for triangles

(what is the dimensionality of this space?)

Tangent space

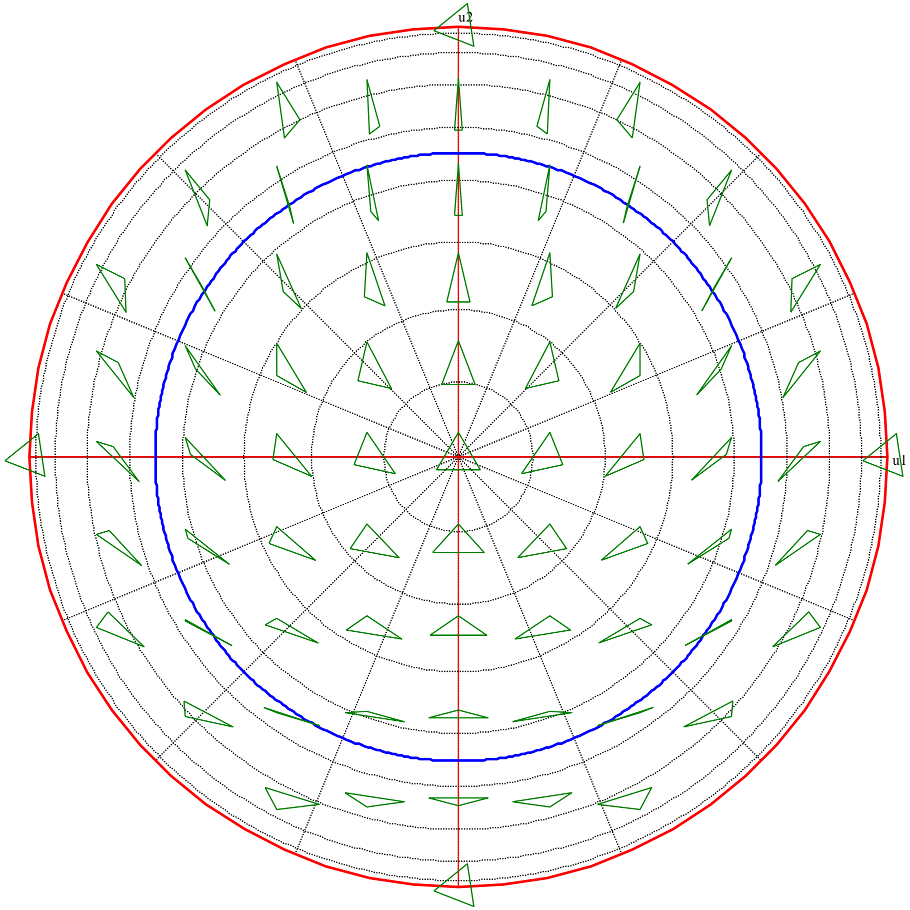
Tangent space is a projection of curved shape space onto a plane (or into an uncurved Euclidean space), much like a map projection is the projection of a curved surface onto a flat paper.

Geometric analysis normally takes place on a tangent plane because most statistical methods.



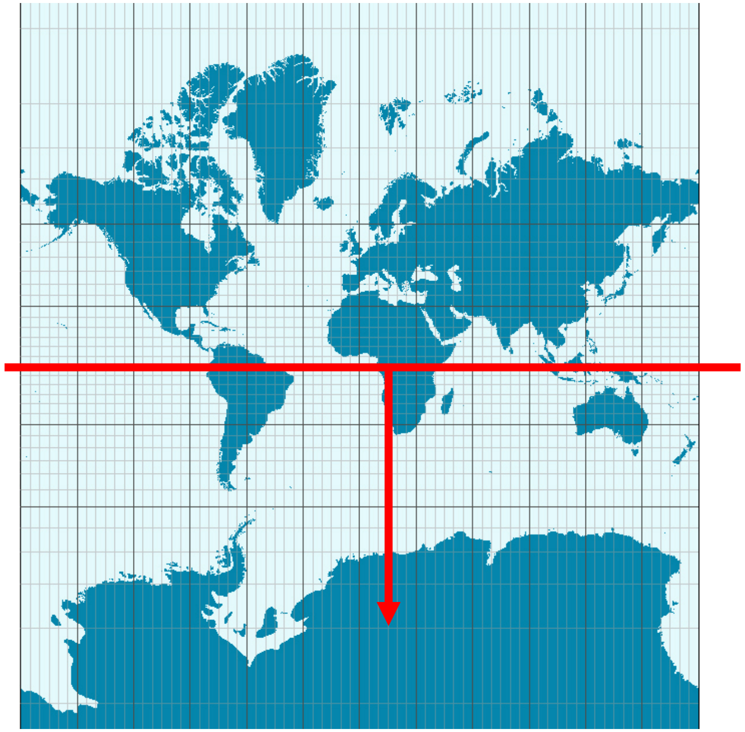
D = Orthogonal projection of B onto Tangent Space.
C = Stereographic projection onto Tangent Space

Triangle shape space projected onto a plane



Like with a globe, distortion is greater toward the edges

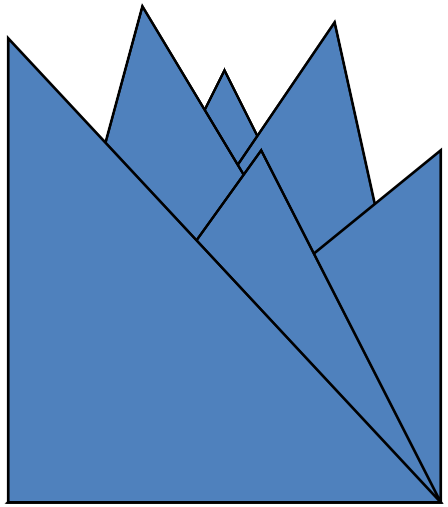
Distortion introduced by projection



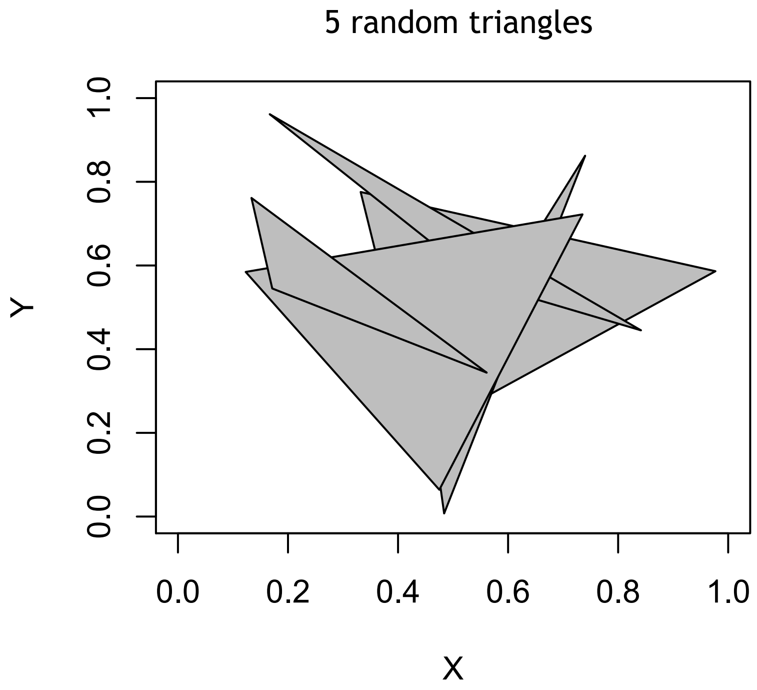
An experiment to prove that shape space is curved

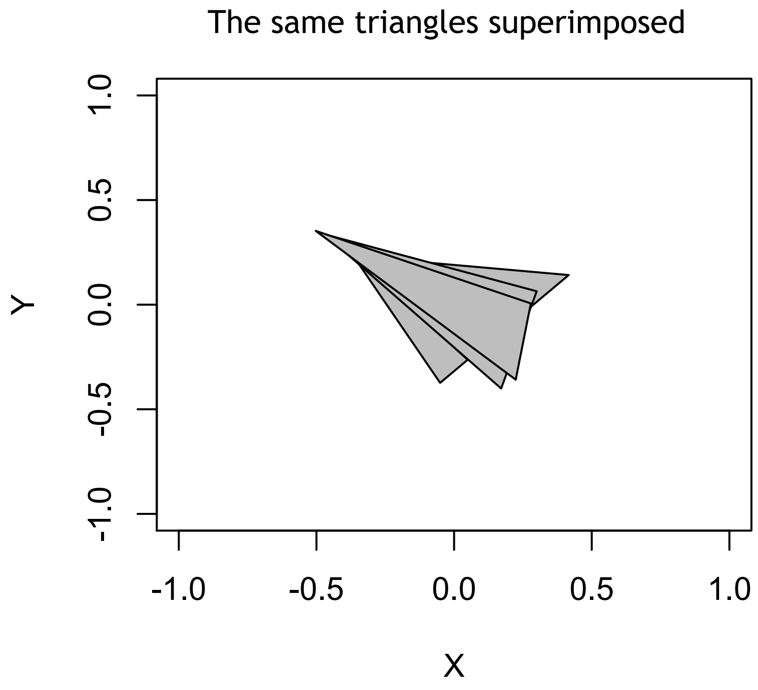
Take the simple example of triangles. Three 2D landmarks. The dimensionality of the shape space is 2 (2K = 6. 2K – 4 =2).

One can imagine this by considering that the alignment of triangles could be made by lining up one side (two point) exactly, which leaves only the apex to move in two dimensions...



Random triangles



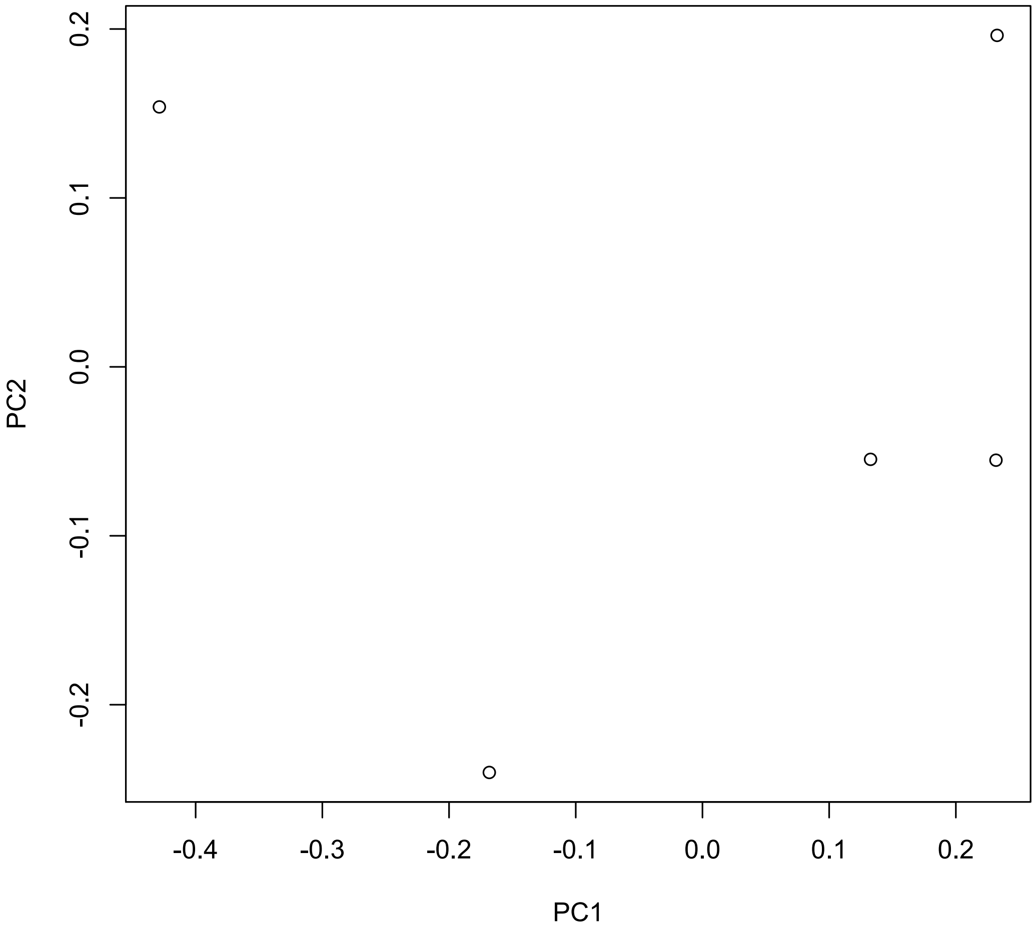


```
> tris=array(runif(5*6,0,1),dim=c(3,2,5))
> plot(c(0,1),c(0,1), xlab=“X”, ylab=“Y”, type=“n”)
> for(i in 1:5) polygon(tris[,,i],col=ceiling(runif(1,1,657)))
```

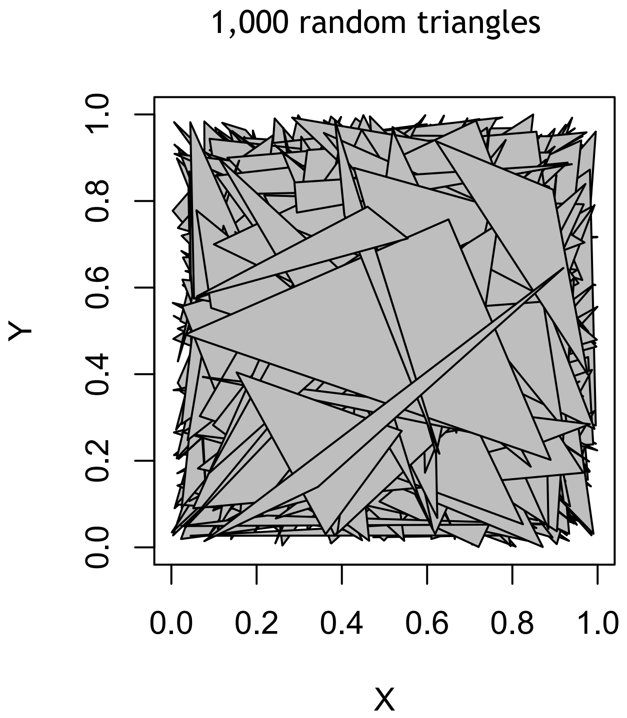
```
> resultTri <- procGPA(tris)
> plot(c(-1,1),c(-1,1),xlab="X",ylab="Y",type="n")
> for(i in 1:5) polygon(resultTri$rotated[,,i],col="Grey”)
```

Random triangles are as different as any three-landmark shape can possibly be, thus they span the entire range of three-landmark (2D) shape.

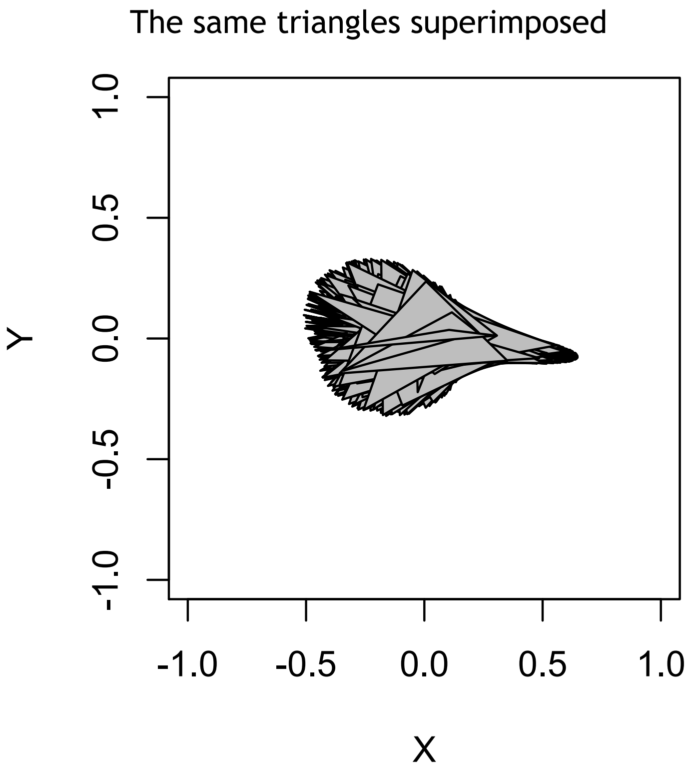
PCA of 5 random triangles



Random triangles



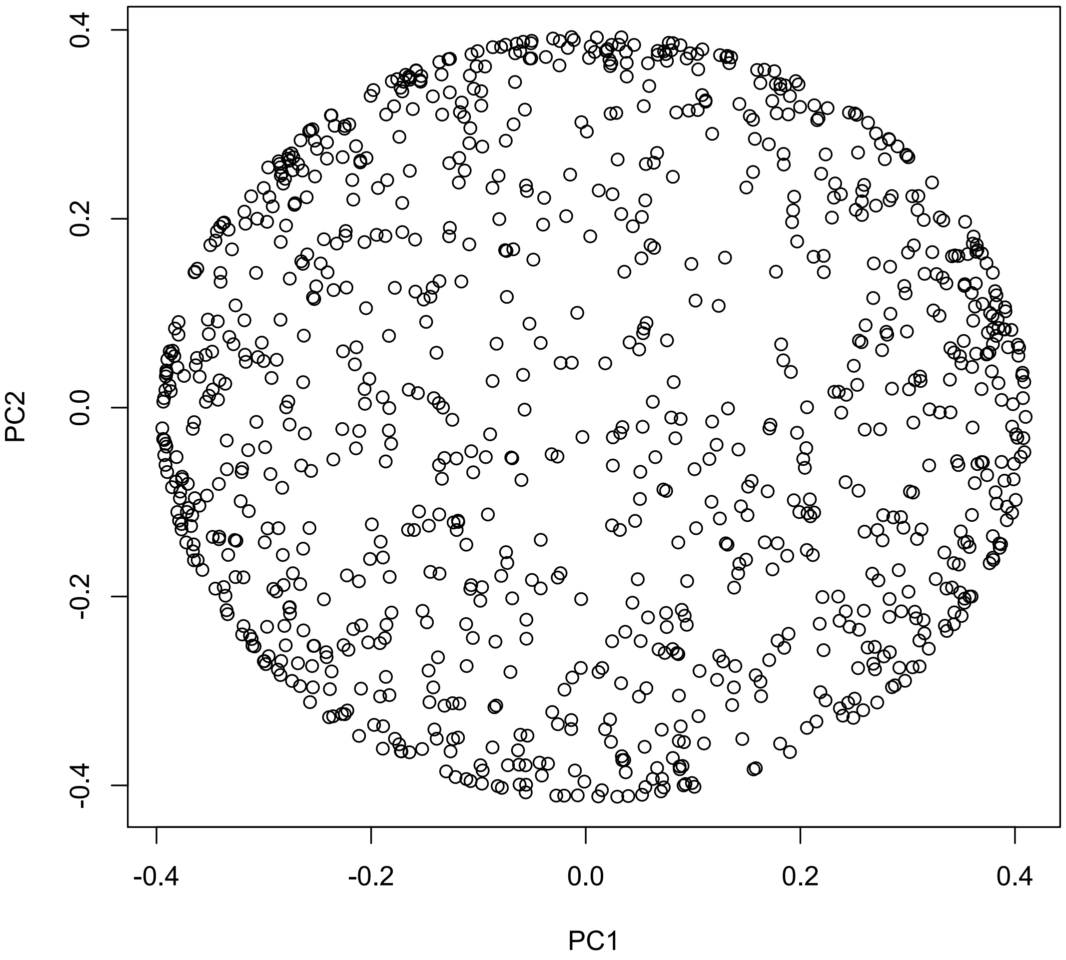
```
> tris=array(runif(5*6,0,1),dim=c(3,2,5))
> plot(c(0,1),c(0,1), xlab=“X”, ylab=“Y”, type=“n”)
> for(i in 1:5) polygon(tris[,,i],col=ceiling(runif(1,1,657)))
```



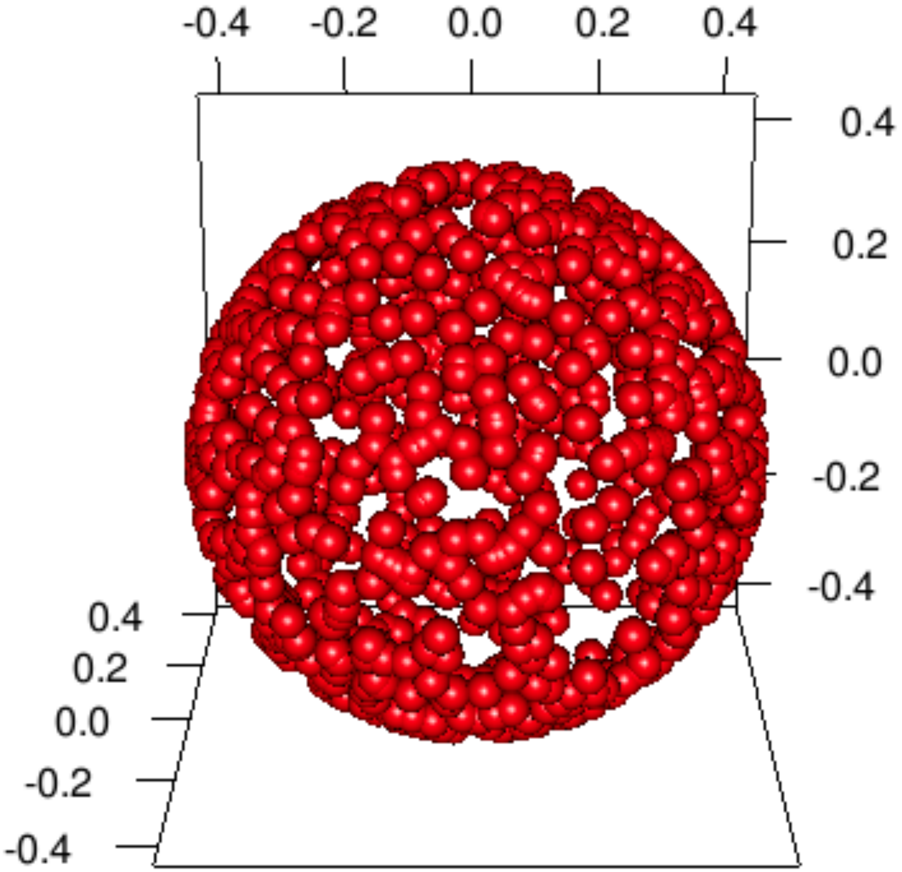
```
> resultTri <- procGPA(tris)
> plot(c(-1,1),c(-1,1),xlab="X",ylab="Y",type="n")
> for(i in 1:5) polygon(resultTri$rotated[,,i],col="Grey”)
```

Random triangles are as different as any three-landmark shape can possibly be, thus they span the entire range of three-landmark (2D) shape.

PCA of 1,000 random triangles



PCA of random triangles plotted with three dimensions

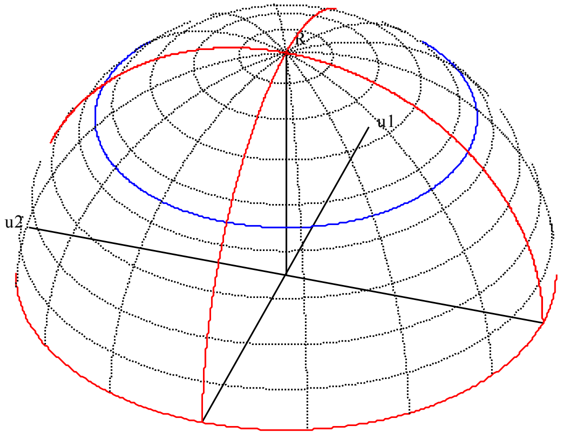


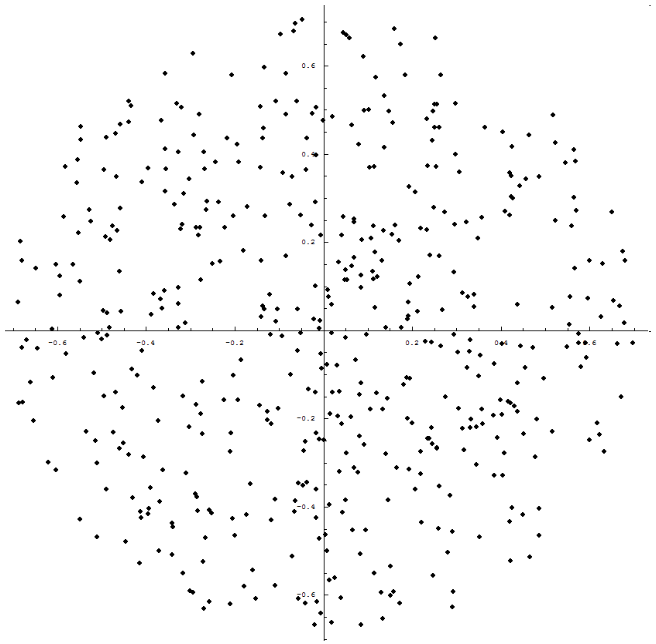
Minimizing the problem

Shape space should be centered at the mean of your sample so that objects receive least distortion possible

PCA accomplishes this by subtracting the mean from the Procrustes coordinates before calculating axes

Usually distortion is negligible for biological shapes because constraints make them comparatively like one another



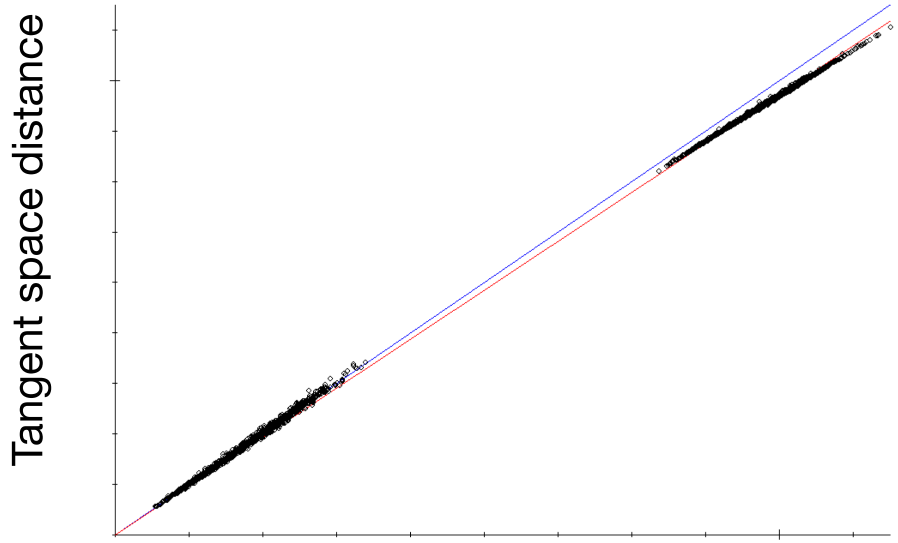


Verifying whether your data have a problem

The distortion caused by curvature of shape space can be tested by comparing Reimmanian distance

Regression through the origin for
distance in tangent
space, Y, regressed onto Procrustes
distance (in radians), X

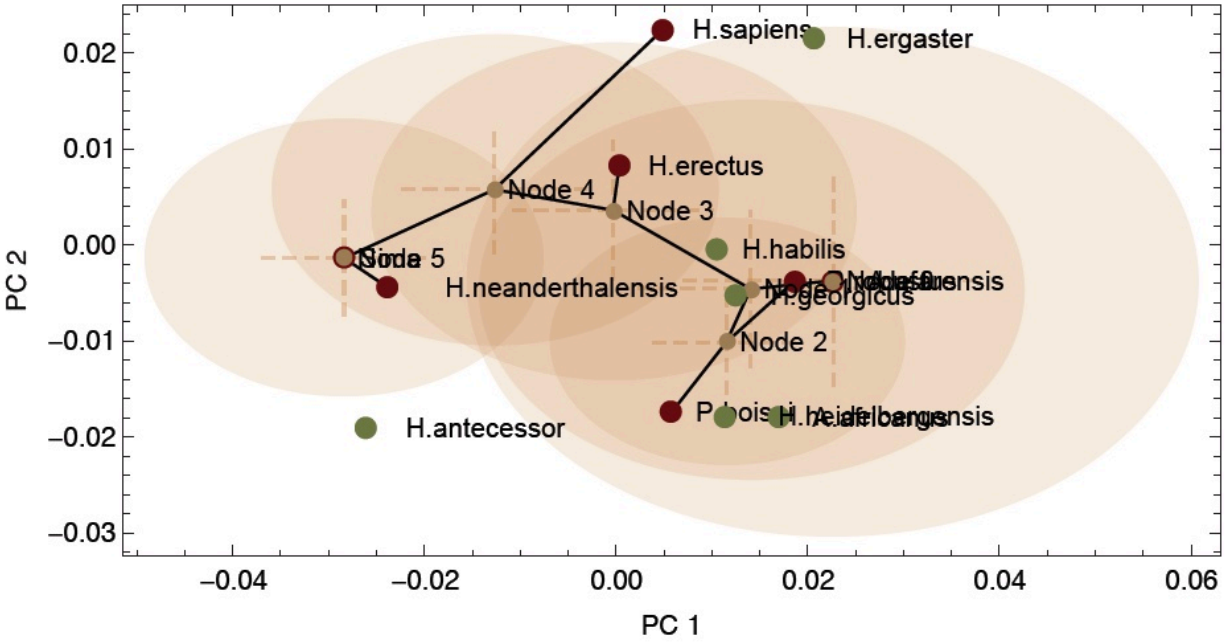
Slope: 0.964798
Correlation (uncentered): 0.999962

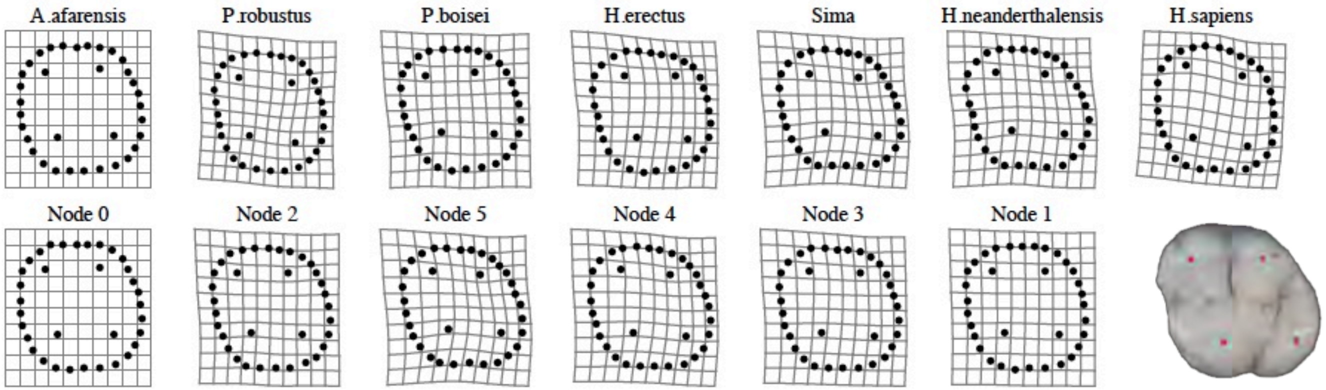


‘Real’ distance on the surface of curved shape space

Is shape space “evolutionary”?

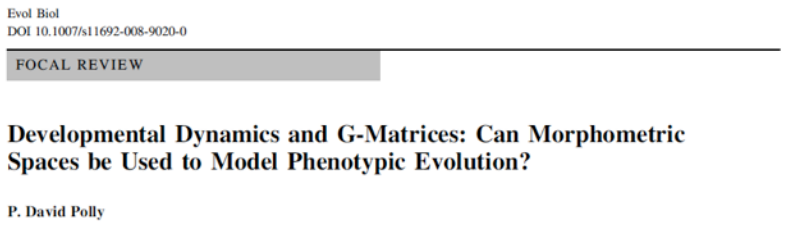
reconstruction of ancestors and evolutionary trajectories





Limitations of shape space

1. it cannot model the gain or loss of features
2. it assumes that trait covariances don’t change
3. it assumes that evolutionary transitions are continuous

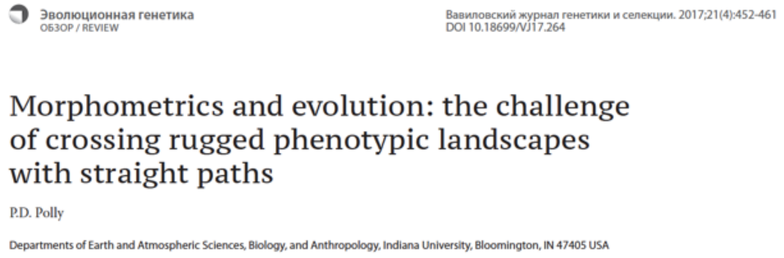


Evol Biol
DOI 10.1007/s11692-008-9020-0

FOCAL REVIEW

Developmental Dynamics and G-Matrices: Can Morphometric Spaces be Used to Model Phenotypic Evolution?

P. David Polly



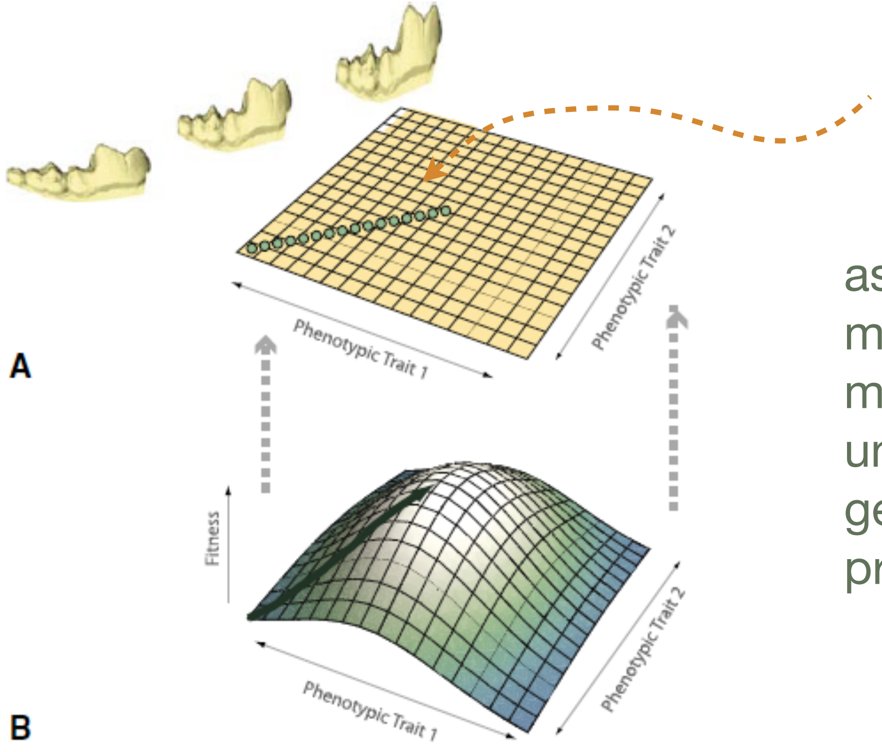
Эволюционная генетика
ОБЗОР / REVIEW

Вавиловский журнал генетики и селекции. 2017;21(4):452-461
DOI 10.18699/VJ17.264

Morphometrics and evolution: the challenge of crossing rugged phenotypic landscapes with straight paths

P.D. Polly

Departments of Earth and Atmospheric Sciences, Biology, and Anthropology, Indiana University, Bloomington, IN 47405 USA



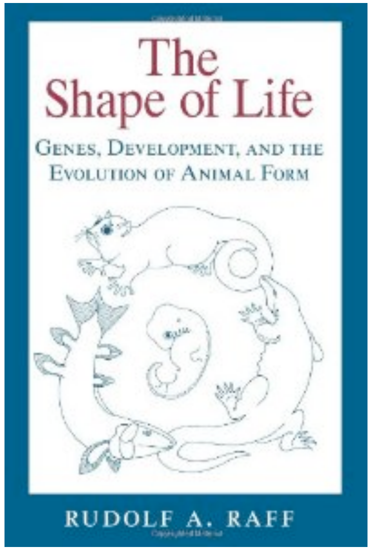
Paths minimize Procrustes distances, but are they the most probable biological transformations?

assumes a 1-to-1 linear mapping between morphospace and underlying phenotypic/ genetic/developmental processes

When methods and theory collide

Key question in evolutionary morphology: continuity or discontinuity?

(1) Evolutionary novelties



The Shape of Life

Genes, Development, and the Evolution of Animal Form

Rudolf A. Raff

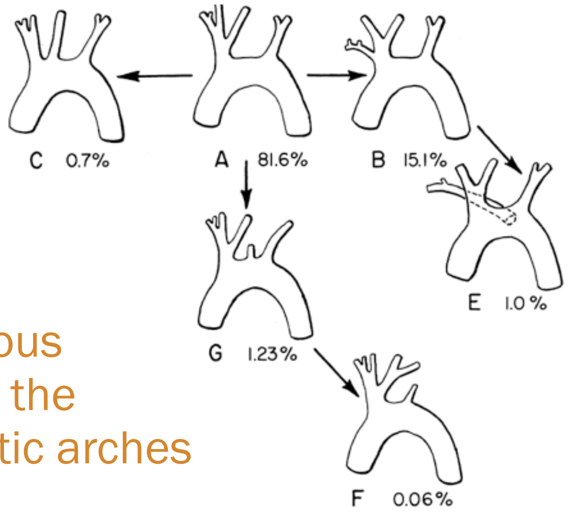
(2) The nature of evolutionary transformations

Amer. Zool., 20:653–667 (1980)

Ontogenesis and Morphological Diversification[1]

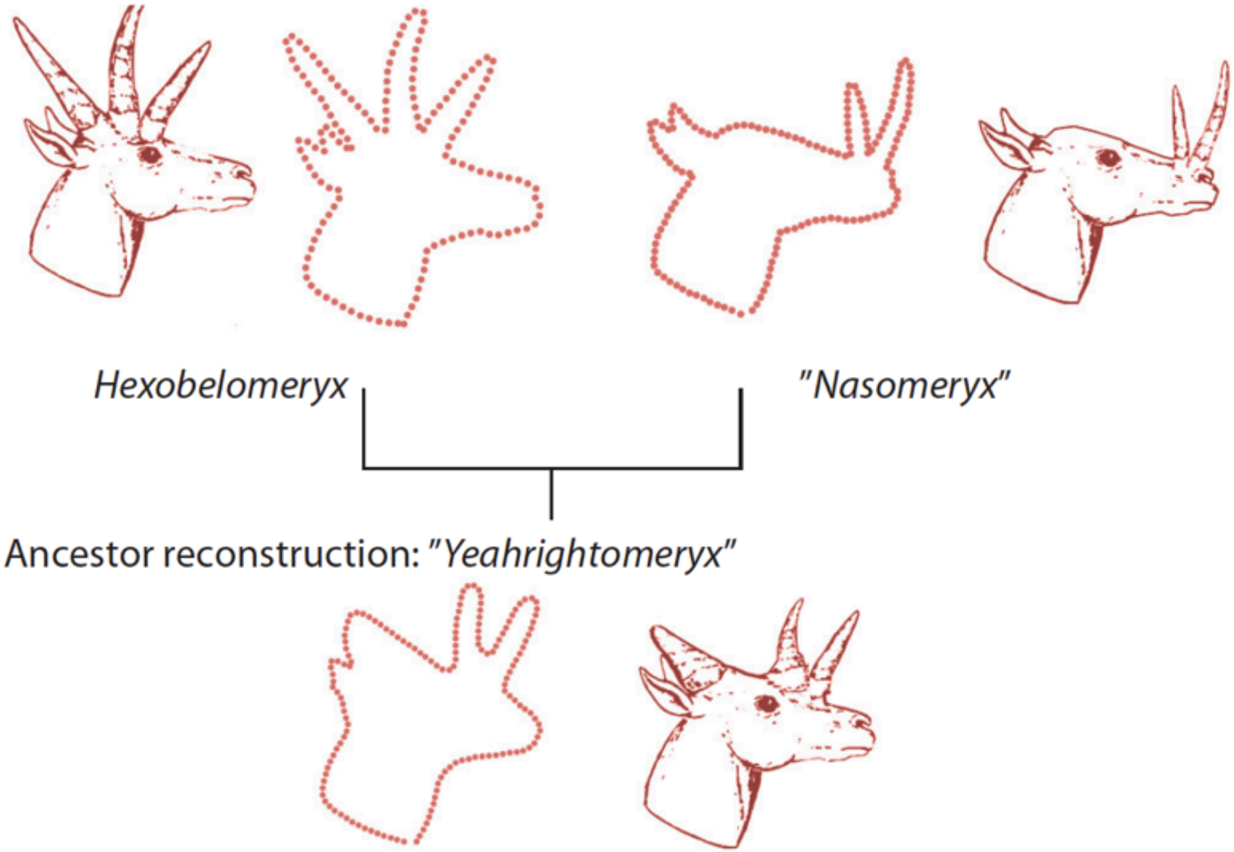
Pere Alberch[2]

Museum of Vertebrate Zoology and Department of Zoology, University of California, Berkeley, California 94720



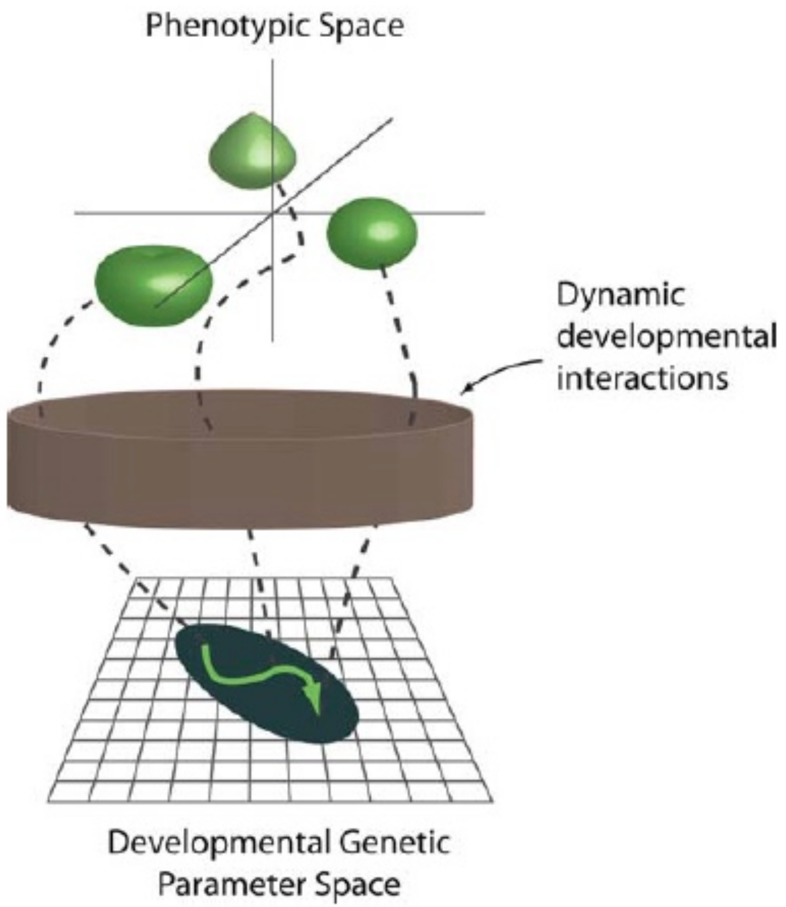
Discontinuous variation in the human aortic arches

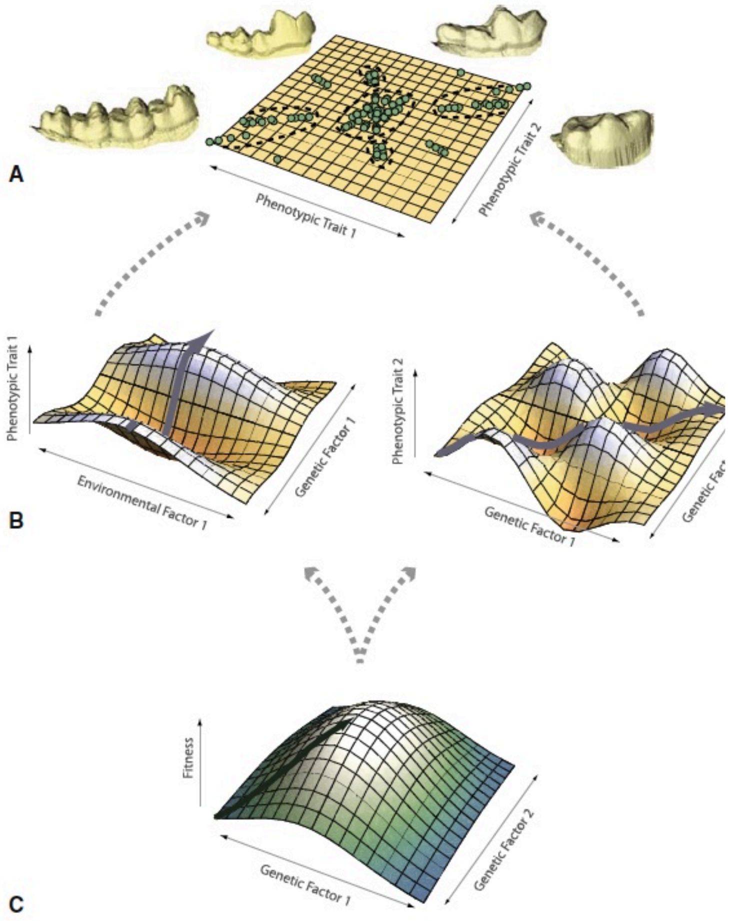
A simple example....



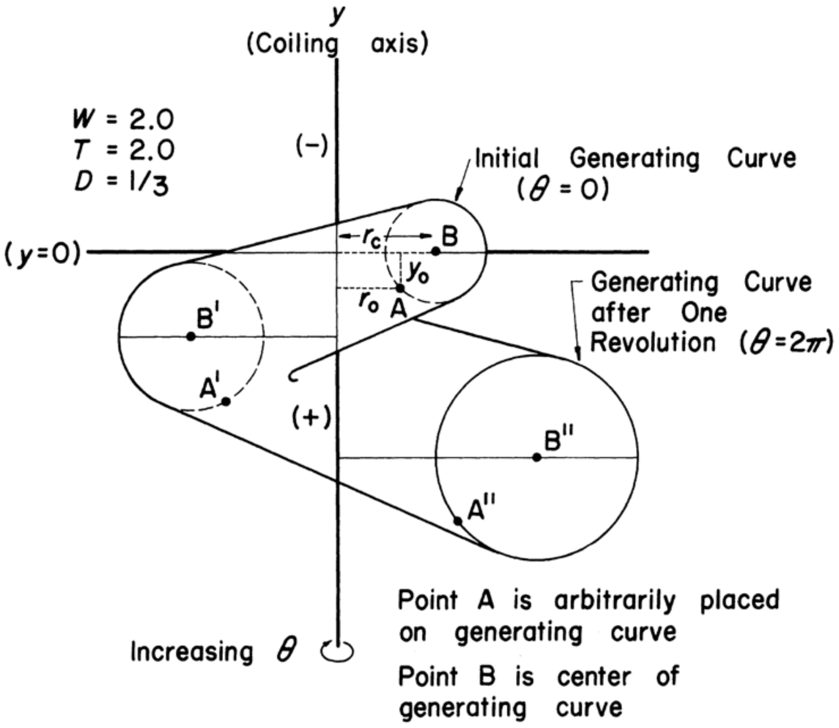
Polly P.D. 2017. Morphometrics and evolution: the challenge of crossing rugged phenotypic landscapes with straight paths. Vavilovskii Zhurnal Genetiki I Selektsii (=Vavilov Journal of Genetics and Breeding), 21: 452-461.

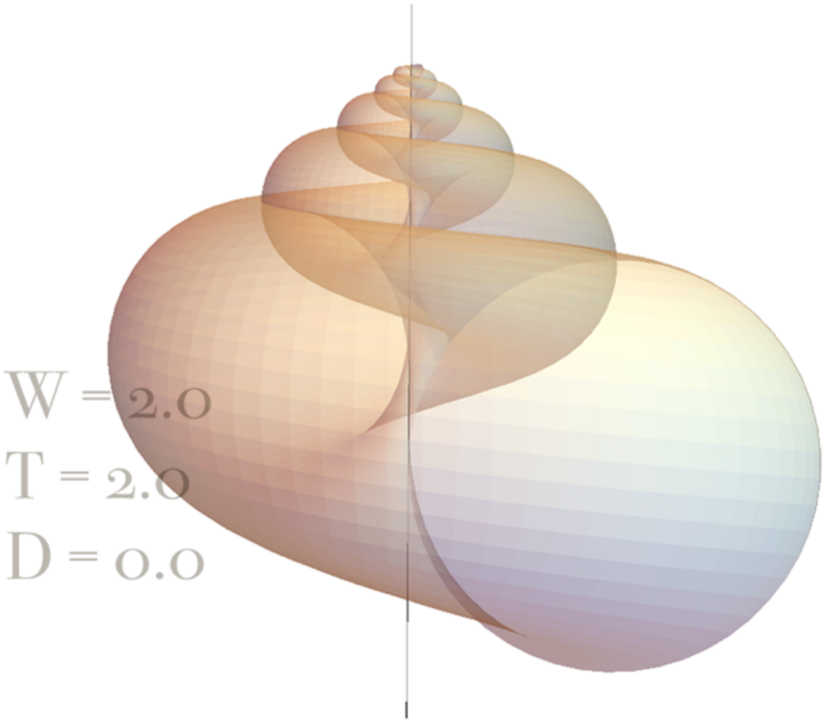
The map between morphospace and factors undergoing selection and drift may be non-linear and discontinuous





An example with Raup's shell coiling

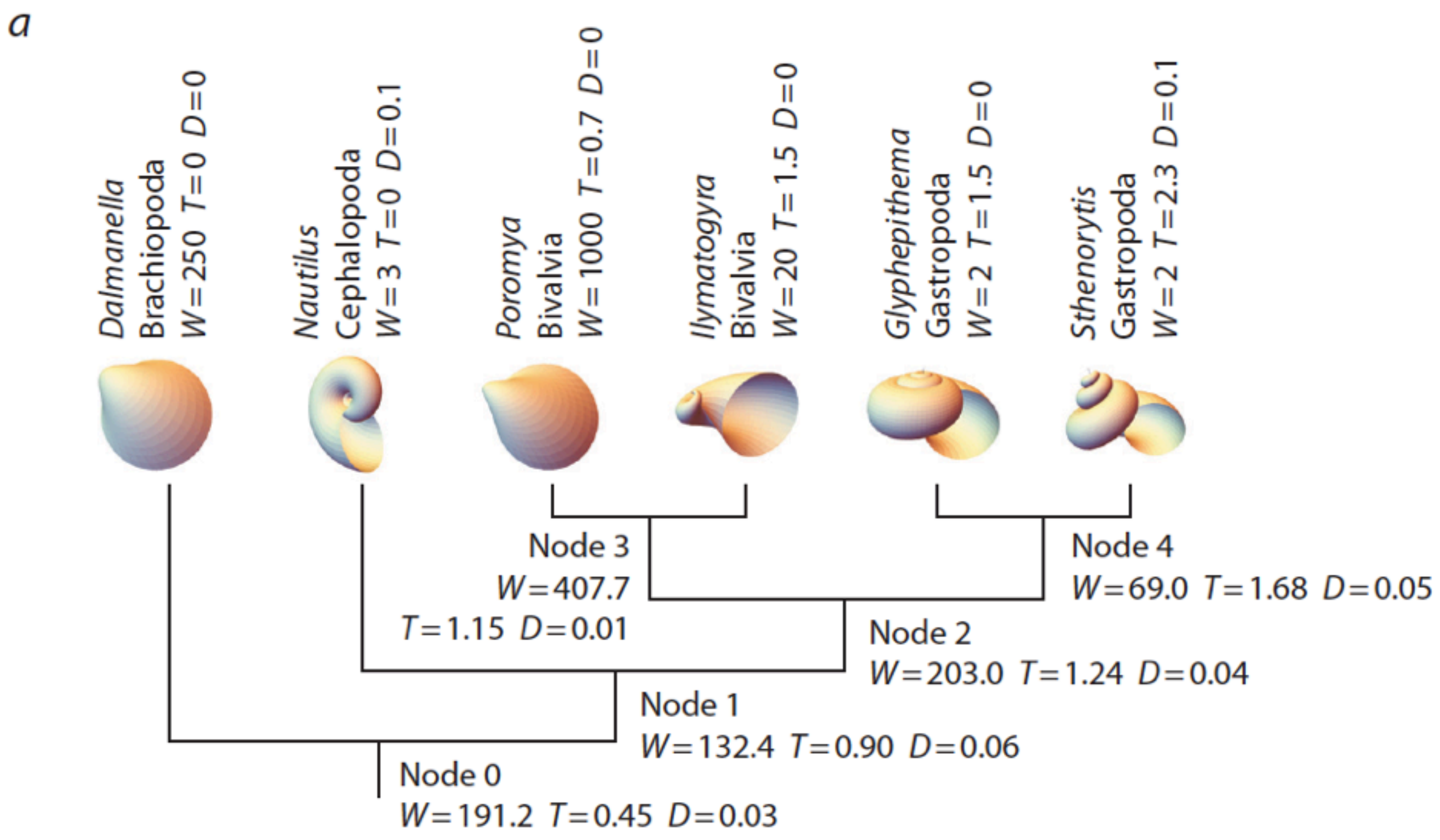




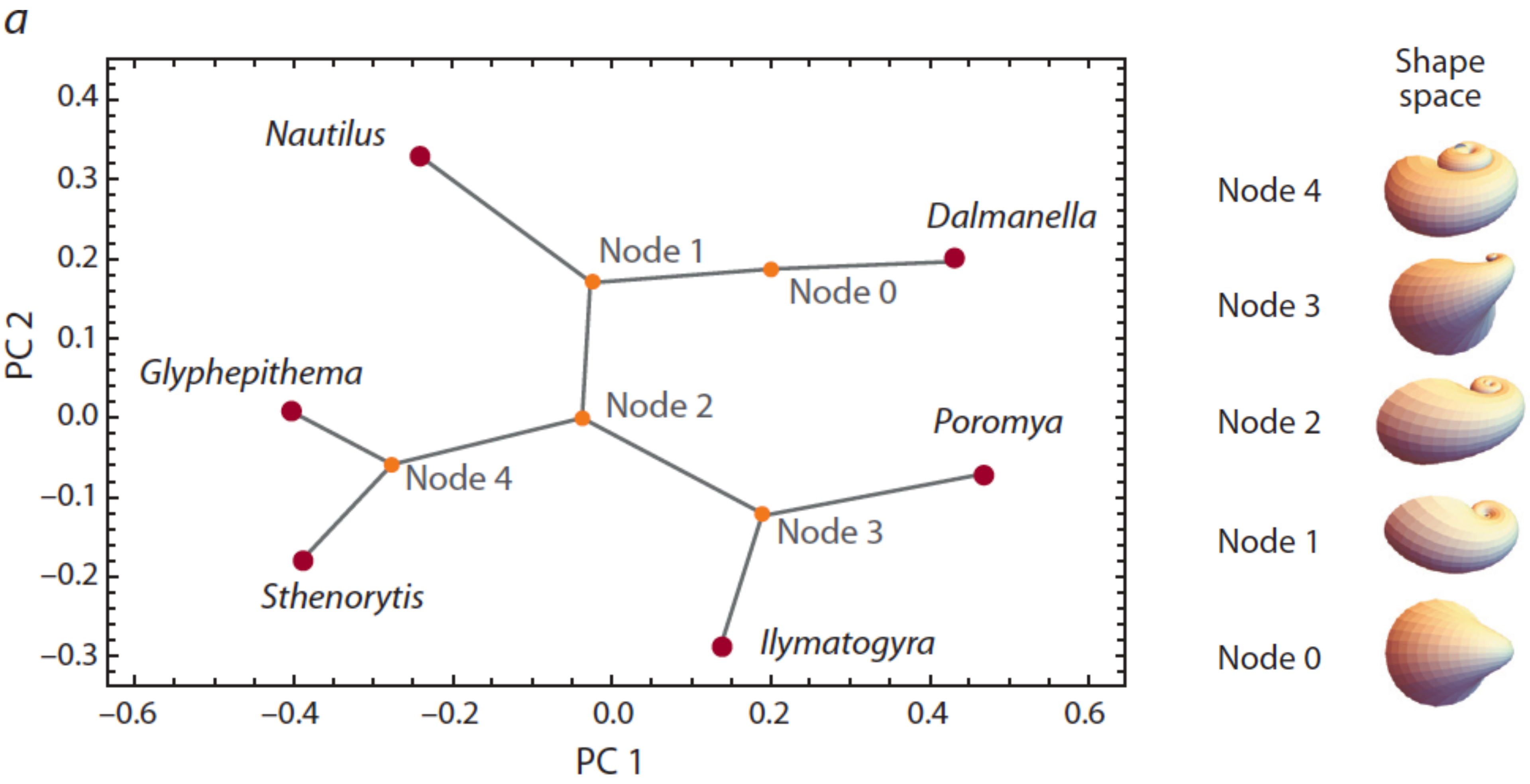
Raup, D. 1966. Geometric analysis of shell coiling: general problems. Journal of Paleontology, 40: 1178-1190.

Six simulated shells

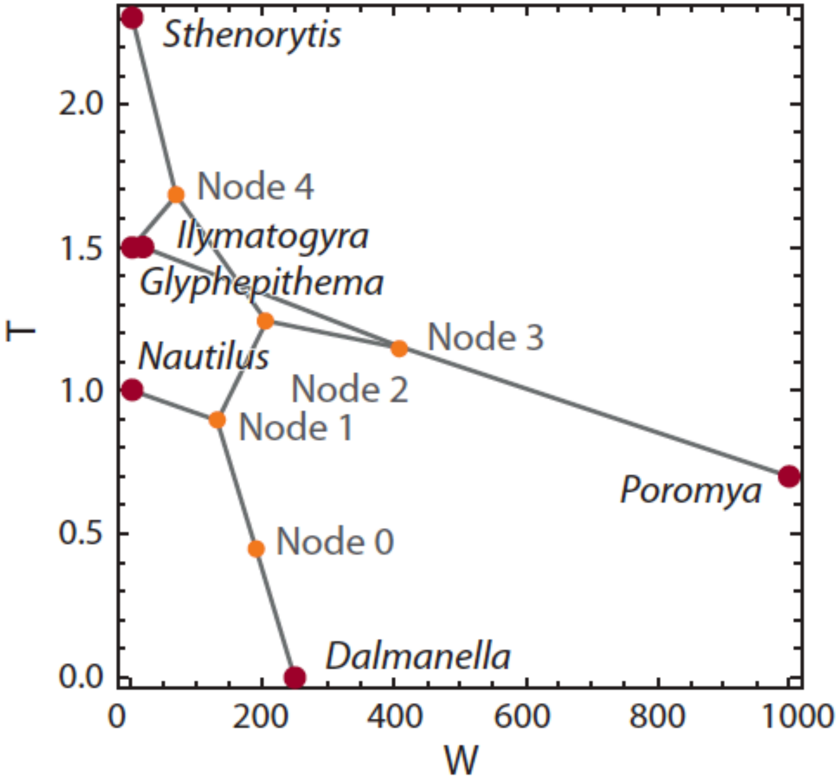
and their Raupian ("genetic") parameters

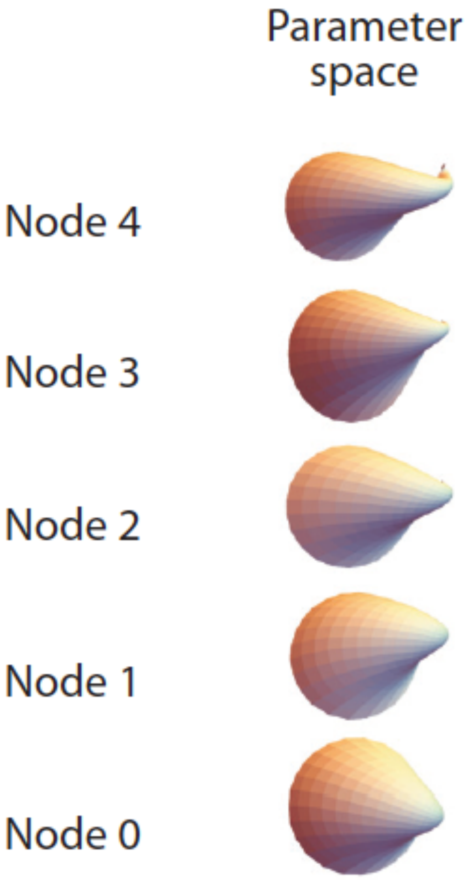


Shells in morphospace with ancestor reconstructions

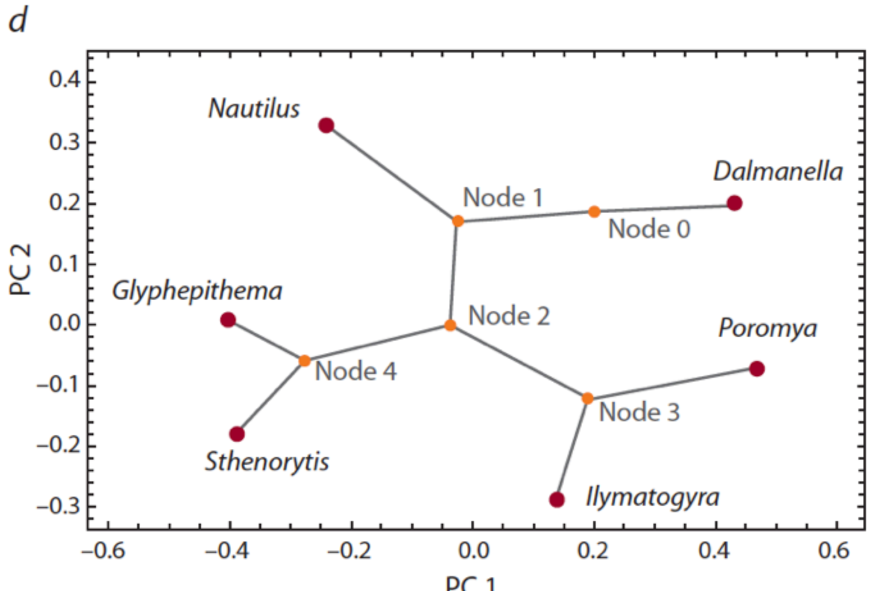


Shells in parameter space with ancestor reconstructions

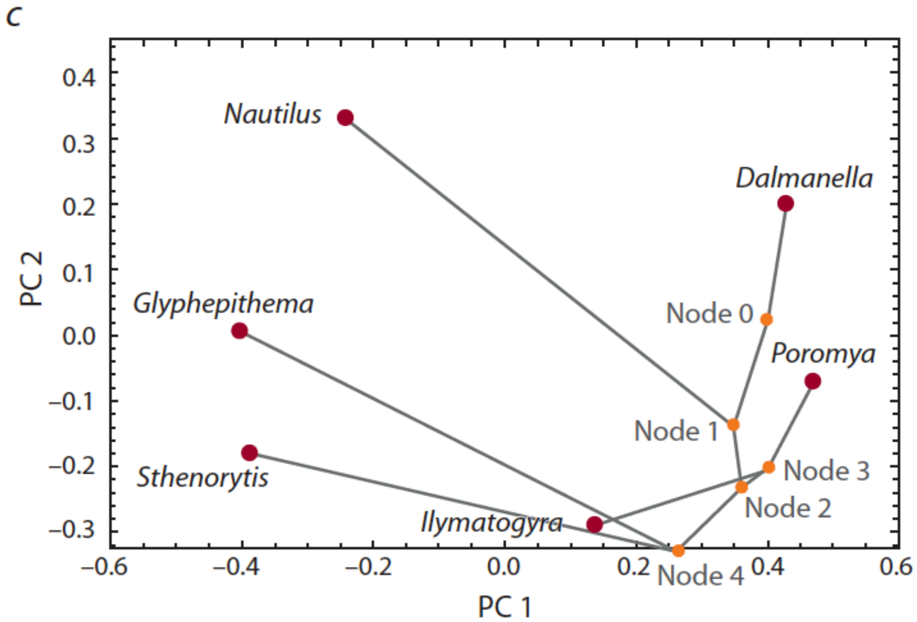




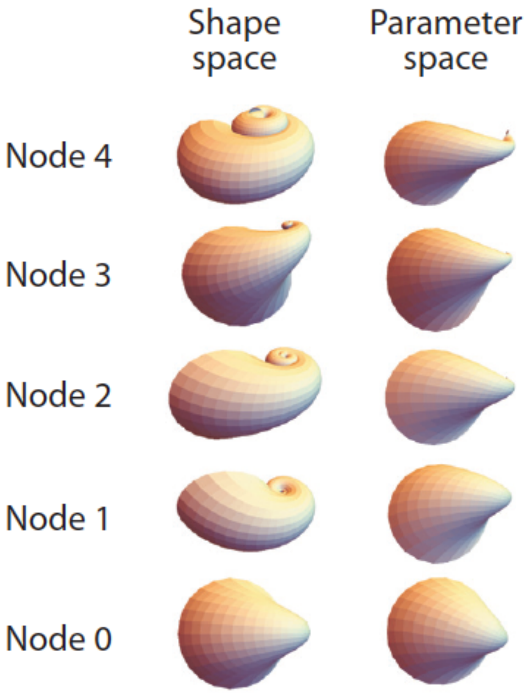
The two do not match...



phylomorphospace based on shape



phylomorphospace based on Raup parameters

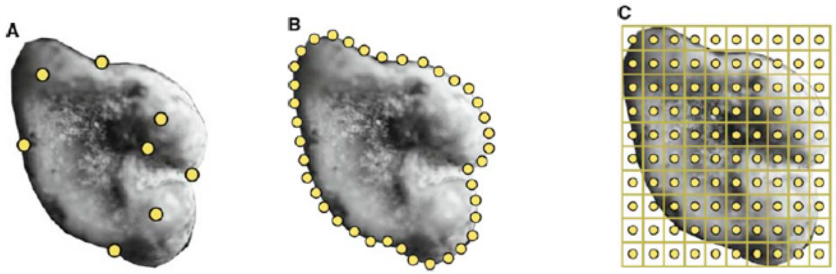


The problem: non-linear mappings and discontinuous transitions...

Solutions?

1. “Homology free” geometric methods that can accommodate gain and loss of features

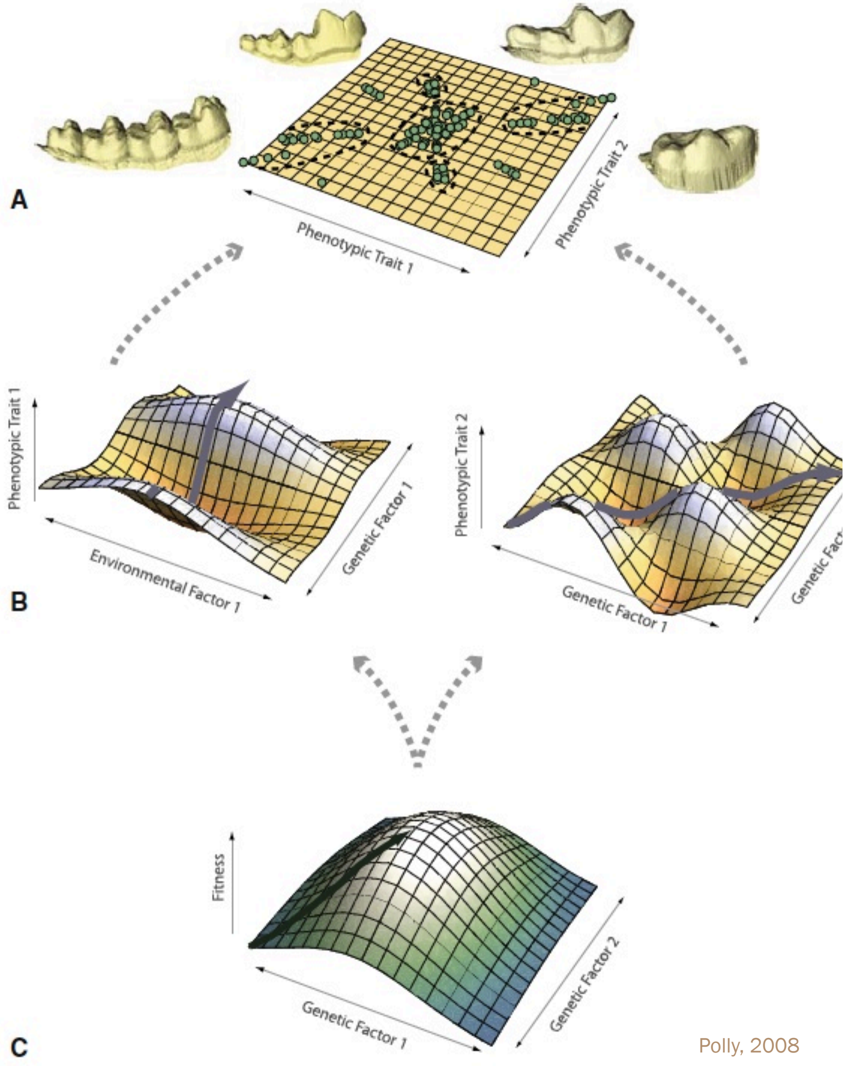
2. Non-linear shape spaces that can be used to model interactions of genetic, developmental, and environmental effect



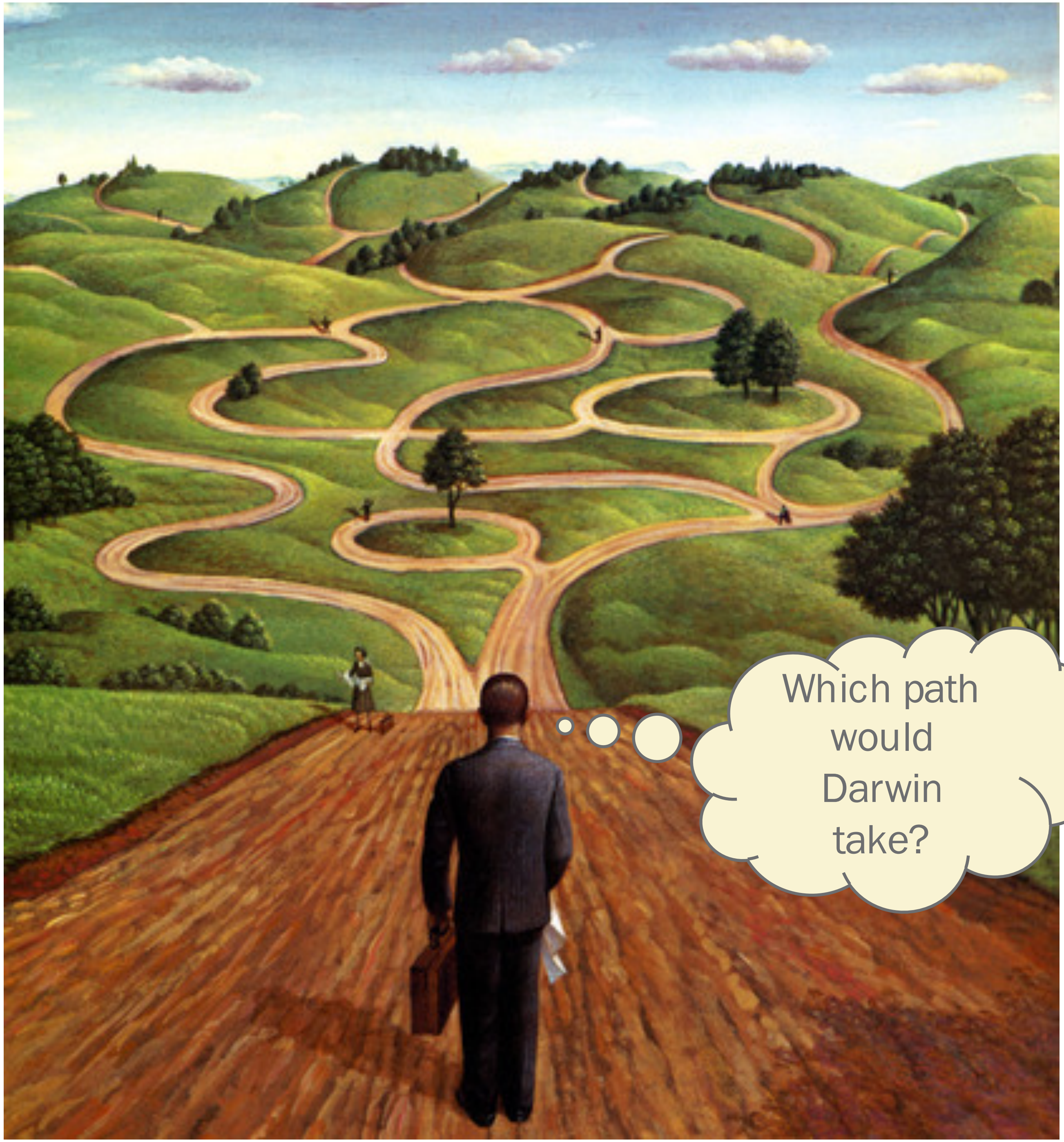
Homologous landmarks

“Homology free” outline semilandmarks

“Homology free” surface semilandmarks



Polly, 2008



Which path would Darwin take?

Fix the rotation on your faces

```
faces<-readShapes('facelands')
face.landmarks<-faces$landmarks.pixel

rotated.faces<-array(dim=c(9,2,36))
face.proc<-gpagen(face.landmarks)

for(i in 1:36){
rotated.faces[,,i]<-face.proc$coords[,,i]%*%matrix(c(cos(-pi/2),-sin(-pi/2),sin(-pi/
2),cos(-pi/2)),nrow=2,ncol=2)
}


facedf<-plotTangentSpace(rotated.faces)
face.consensus<-apply(rotated.faces,c(1,2), mean)
```

Turn your shape modelling code into a function

```
make.my.model <- function(score,PC,consensus,evectors) {
   modelscore<-c(0,0,0,0,0,0,0,0,0,0,0,0,0,0)
   modelscore[PC]<-score
   my.model <- matrix(modelscore%*%t(evectors),ncol=2,nrow=9,byrow=T)+consensus
   return(my.model)
}
```

```
make.my.model(0.05,1,consensus,eigenvectors)
```

```
for(i in seq(from=-0.1, to=0.1, by=0.05)) {
   plotRefToTarget(face.consensus,make.my.model(i,1,face.consensus,facedf$rotation))
}
```

Brownian motion function

```
randomwalk <- function(n,r=.01) {
  scores <- matrix(ncol=1, nrow=n)
  scores[1] <- 0
  for (i in 2:n) {
    scores[i]=scores[i-1]+rnorm(1, mean=0, sd=sqrt(r))
  }
  return(scores)
}

randomwalk(20)
```

Put the two together to simulate Brownian motion on one PC

```
my.randwalk<-randomwalk(30,r=0.005)

for(i in 1:length(my.randwalk)) {
   plotRefToTarget(face.consensus,make.my.model(my.randwalk[i],
   2,face.consensus,facedf$rotation))
}
```