

# Interaction Terms and Trajectory Analysis

*Advanced Biostatistics*

Dean Adams

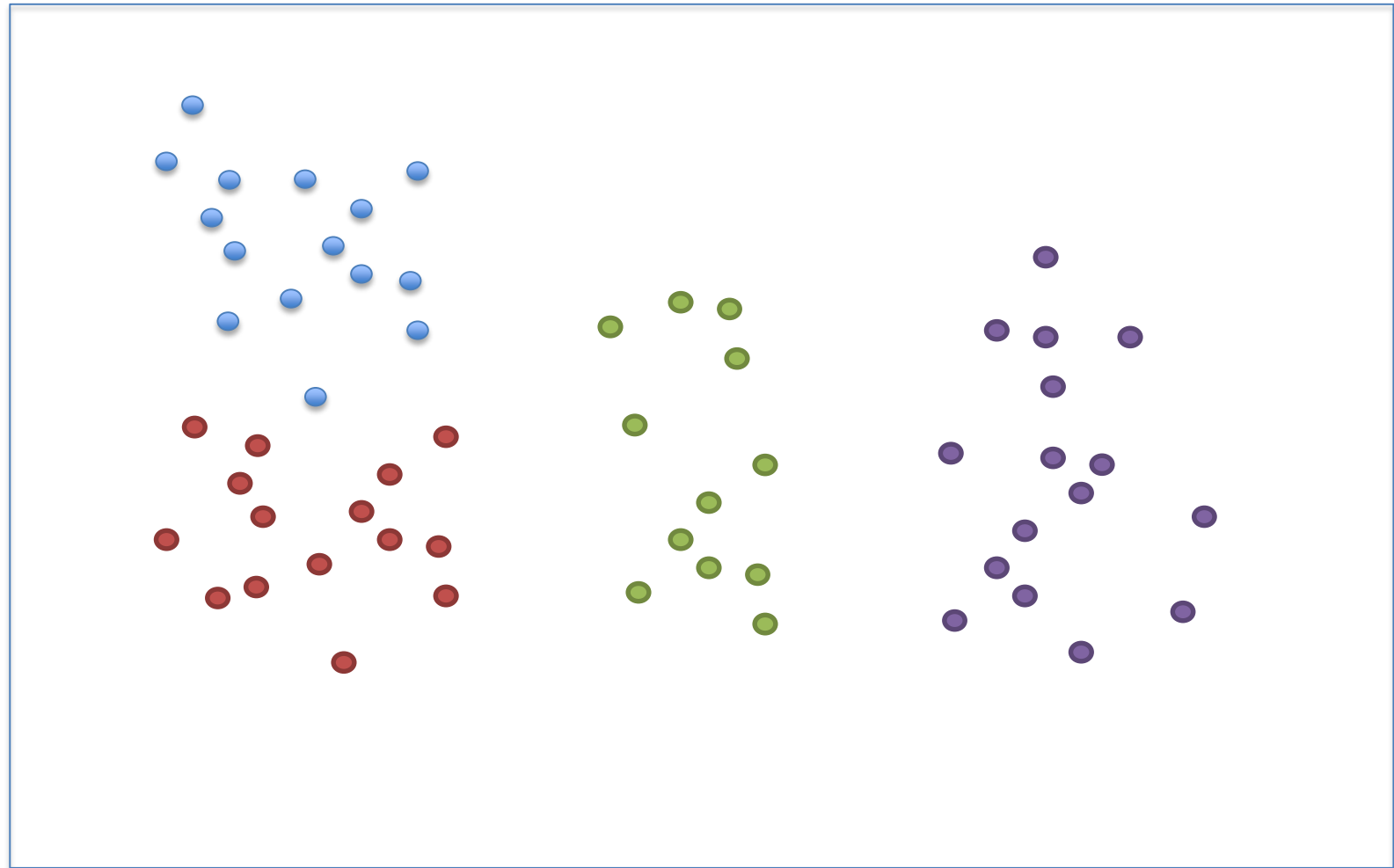
Lecture 7

EEOB 590C

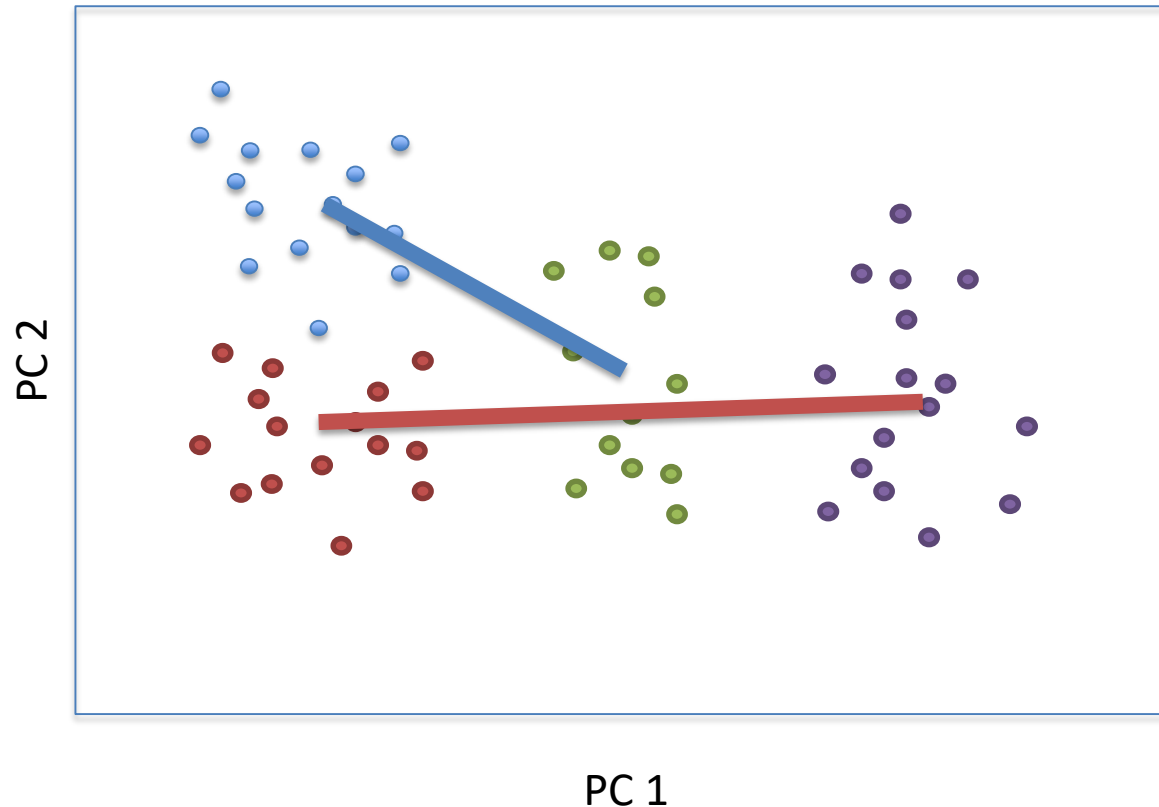
## Overview

- Factor interactions: why are they important?
- Factor interactions: understanding them
- Factor interactions in a multivariate sense
- Hypothesis tests
- Analysis of phenotypic change vectors: Examples
- Trajectories in multivariate data spaces = patterns of change
- Trajectory attributes as testable statistics
- Hypothesis tests
- Trajectory Analysis: Examples

PC 2



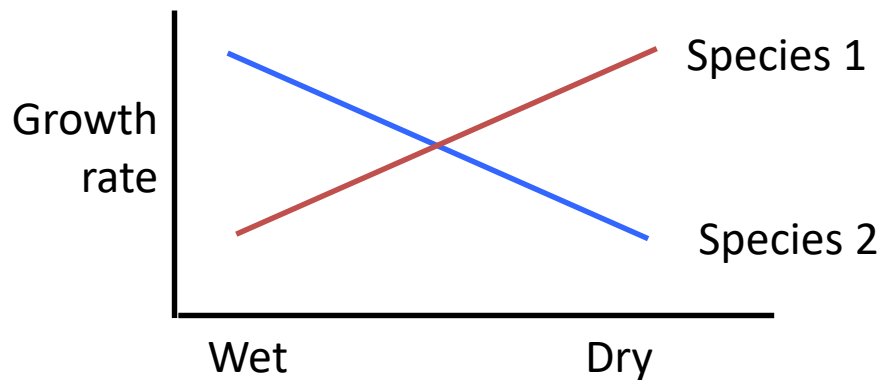
PC 1



We have dealt with ways to ask the question, are these groups different? We might use something like MANOVA to address this question. But what if we have more information that would make the question more informed? For example, if the blue and green dots were species 1 in environments A and B, and the red and purple dots were species 2 in the same environments, respectively, we might rather ask if the difference in locations for species 1 is different than the difference in locations for species 2, for the scatter of subject values in the multivariate data space.

- Many hypotheses really target patterns of change:
  - How does the phenotype change across environments? (plasticity)
  - How do traits change through evolutionary time? (comparative methods)
  - How do traits change through development? (ontogenetics)
  - Are patterns of variation constant across space or time? (e.g., spatial data)
- Can estimate coefficients via GLM to describe patterns
- However, MANOVA is limited in the types of hypotheses it can test (can only compare SSP matrices with the null expectation that they are not different)
- More on this in a moment
- How can GLM be used to model patterns of change?

- Interactions measure the *joint* effect of main effects A & B
- Identifies whether response to A dependent on level of B
- Are *VERY* common in biology
- Example: 2 species in 2 environments (Factors A & B), species 1 has higher growth rate in moist environment, while species 2 has higher growth rate in dry environment. This would be identified as an interaction between species & environment



Note: The study of trade-offs (reaction norms) in evolutionary ecology is based on the study of interactions

```

> Growth.Rate = c(30,32,30,36,35,15,14,17,18,13,14,14,13,19,17,28,38,34,33,30)
> Species = factor(c(rep("1",10),rep("2",10)))
> Environment = factor(rep(c(rep("Wet",5),rep("Dry",5)),2))
>
> fit = lm(Growth.Rate~Species*Environment)
>
> interaction.plot(Environment,Species,Growth.Rate)
> anova(fit)

```

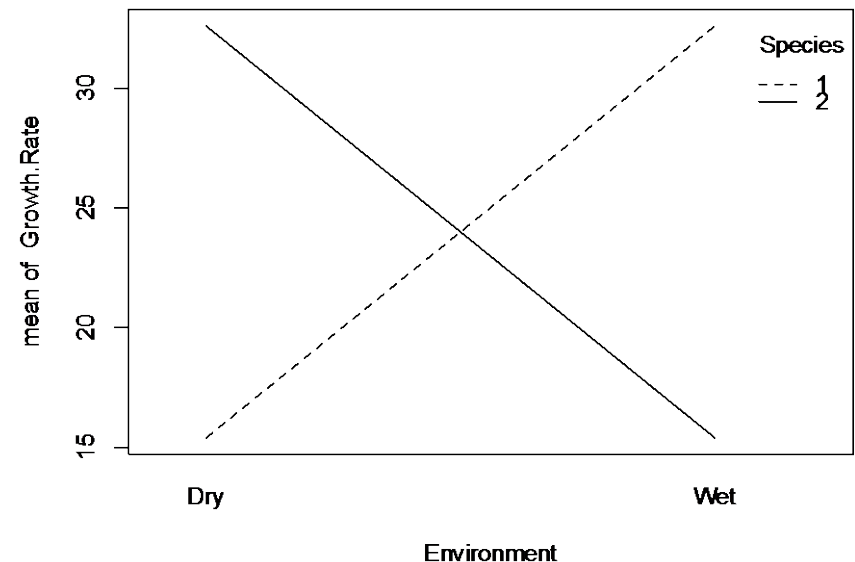
### Analysis of Variance Table

Response: Growth.Rate

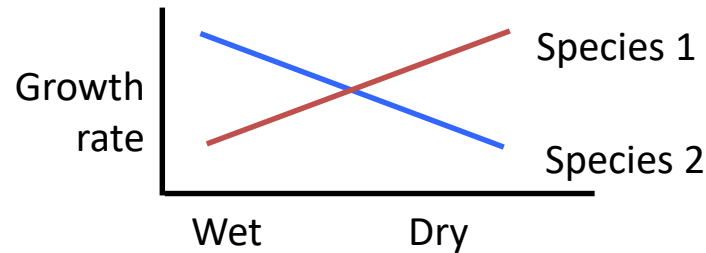
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Species	1	0.0	0.0	0.00	1
Environment	1	0.0	0.0	0.00	1
Species:Environment	1	1479.2	1479.2	178.22	4.328e-10 ***
Residuals	16	132.8	8.3		

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

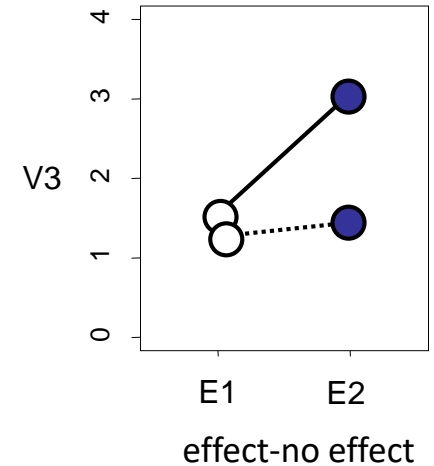
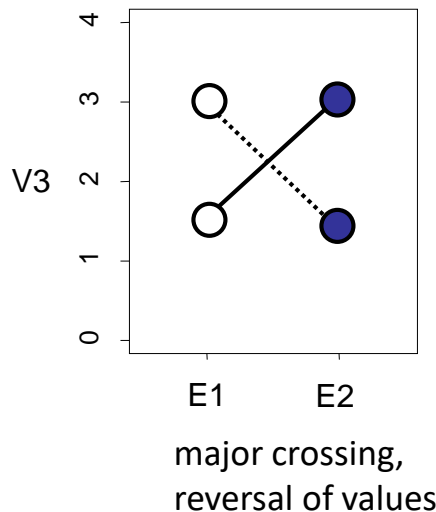
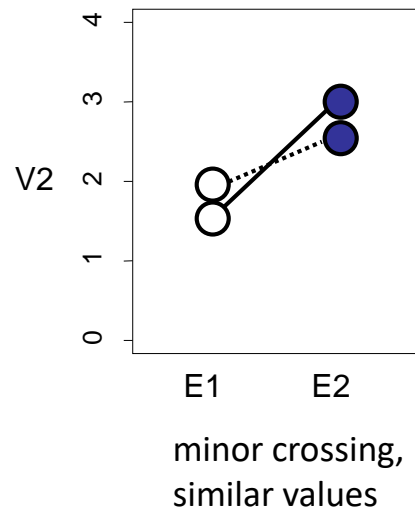
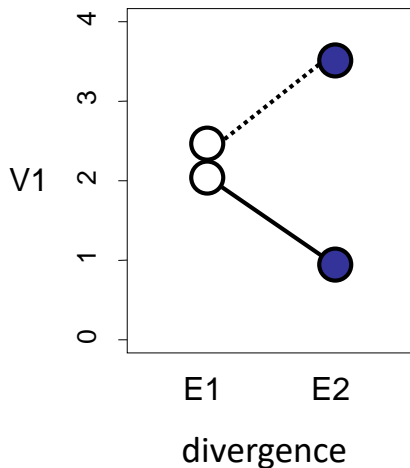
This interaction term tells us the species doesn't matter, nor the environment, but which species you are in a particular environment.



- Significant interactions identify a joint response of factors (response to Factor B depends on level in Factor A)

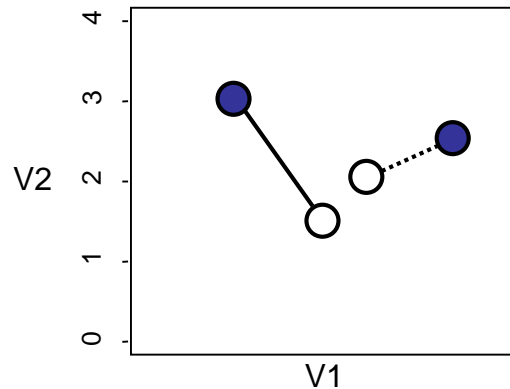


- Interpreting interactions for univariate data is straightforward

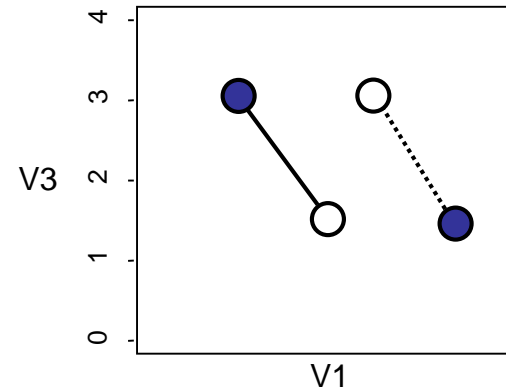




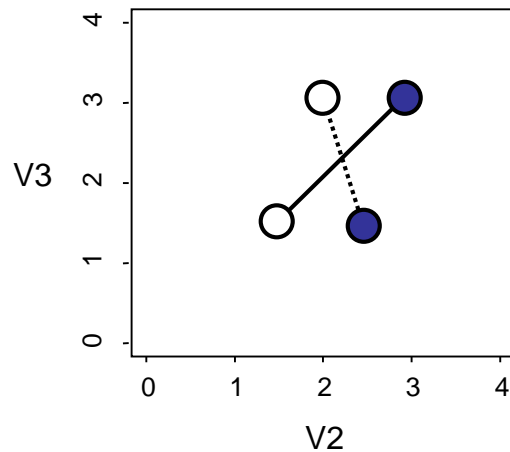
- For two traits, more complicated variants are possible



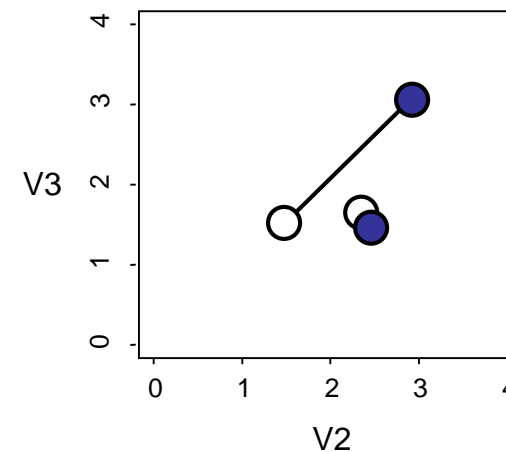
direction change



direction change:  
rank-order



direction change:  
crossing



effect-no effect

- Magnitude: amount of change
- Direction: orientation of change

Phenotypic Change Vector

$$\Delta \bar{\mathbf{y}} = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) = \begin{bmatrix} \bar{y}_{11} - \bar{y}_{21} \\ \bar{y}_{12} - \bar{y}_{22} \\ \bar{y}_{13} - \bar{y}_{23} \end{bmatrix}$$

$$\bar{\mathbf{Y}} = \begin{bmatrix} \bar{\mathbf{y}}_1^t \\ \bar{\mathbf{y}}_2^t \\ \vdots \\ \bar{\mathbf{y}}_k^t \end{bmatrix}$$

Magnitude

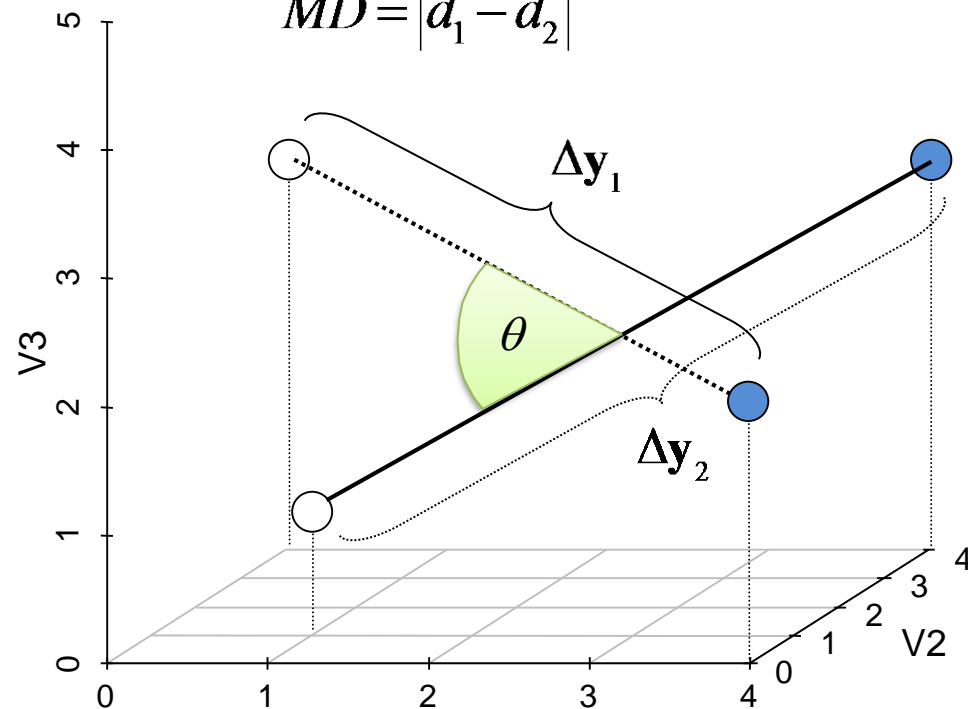
$$d = (\Delta \bar{\mathbf{y}}^t \Delta \bar{\mathbf{y}})^{1/2}$$

$$MD = |d_1 - d_2|$$

Direction

$$\theta = \cos^{-1} r$$

$$r = \frac{\Delta \bar{\mathbf{y}}_1^t \Delta \bar{\mathbf{y}}_2}{d_1 d_2}$$



- Patterns of change assessed using *residual randomization*
- Protocol

1. Define Model

$\mathbf{X}_f$

Design matrix with factors A, B, and A×B

2. Estimate coefficients

$$\hat{\mathbf{B}}_f = (\mathbf{X}_f^t \mathbf{X}_f)^{-1} \mathbf{X}_f^t \mathbf{Y}$$

3. Estimate LS means

$$\bar{\mathbf{Y}} = \bar{\mathbf{X}} \hat{\mathbf{B}}_f = E[\mathbf{Y} | \mathbf{X}_f, \hat{\mathbf{B}}_f]$$

Design matrix coded to find means

4. Calculate vector attributes and statistic

Magnitude

Direction

$$MD = |d_1 - d_2|$$

$$\theta = \cos^{-1} r$$

- Patterns of change assessed using *residual randomization*
- Protocol

1. Define Model  $\mathbf{X}_f$ 

Design matrix with factors A, B, and A×B
2. Estimate coefficients  $\hat{\mathbf{B}}_f = (\mathbf{X}_f^t \mathbf{X}_f)^{-1} \mathbf{X}_f^t \mathbf{Y}$
3. Estimate LS means  $\bar{\mathbf{Y}} = \bar{\mathbf{X}} \hat{\mathbf{B}}_f = E[\mathbf{Y} | \mathbf{X}_f, \hat{\mathbf{B}}_f]$ 

Design matrix coded to find means
4. Calculate vector attributes and statistic
5. Define “reduced” model  $\mathbf{X}_r$ 

Design matrix with factors A and B only
6. Estimate coefficients  $\hat{\mathbf{B}}_r = (\mathbf{X}_r^t \mathbf{X}_r)^{-1} \mathbf{X}_r^t \mathbf{Y}$
7. Estimate all values of  $\mathbf{Y}$   $\hat{\mathbf{Y}}_r = \mathbf{X}_r \hat{\mathbf{B}}_f$
8. Obtain residuals  $\mathbf{E}_r = \mathbf{Y} - \hat{\mathbf{Y}}_r$

- Patterns of change assessed using *RRPP*
- Protocol

9. Randomize residuals  $\mathbf{E}_r^*$  i.e., shuffle rows

10. Add randomized residuals to estimated values from reduced model

$$\mathbf{Y}^* = \hat{\mathbf{Y}}_r + \mathbf{E}_r^*$$

Repeat many times

Repeat steps 1 – 4 to obtain random statistics

Random value preserved the main effects of the reduced model

By creating random (sampling) distributions of the magnitude difference and angle between vectors, *P*-values for the observed values are described as the percentiles in the distributions. (I.e., the *P*-value is the probability of finding a greater or equal value by chance)

- Patterns of change assessed using *RRPP*
- Protocol

9. Randomize residuals  $\mathbf{E}_r^*$  i.e., shuffle rows

10. Add randomized residuals to estimated values from reduced model

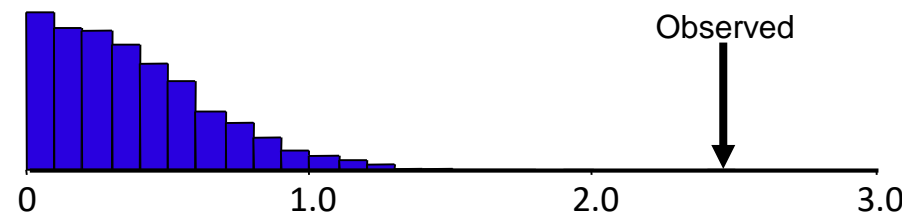
$$\mathbf{Y}^* = \hat{\mathbf{Y}}_r + \mathbf{E}_r^*$$

Repeat many times

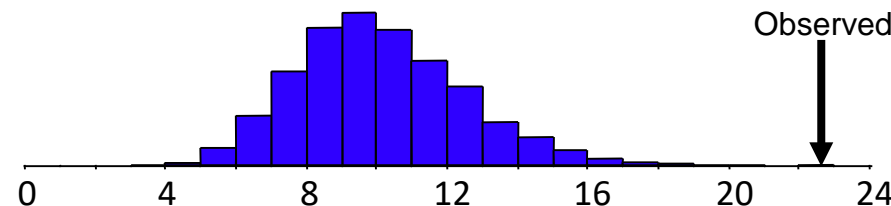
Repeat steps 1 – 4 to obtain random statistics

Random value preserved the main effects of the reduced model

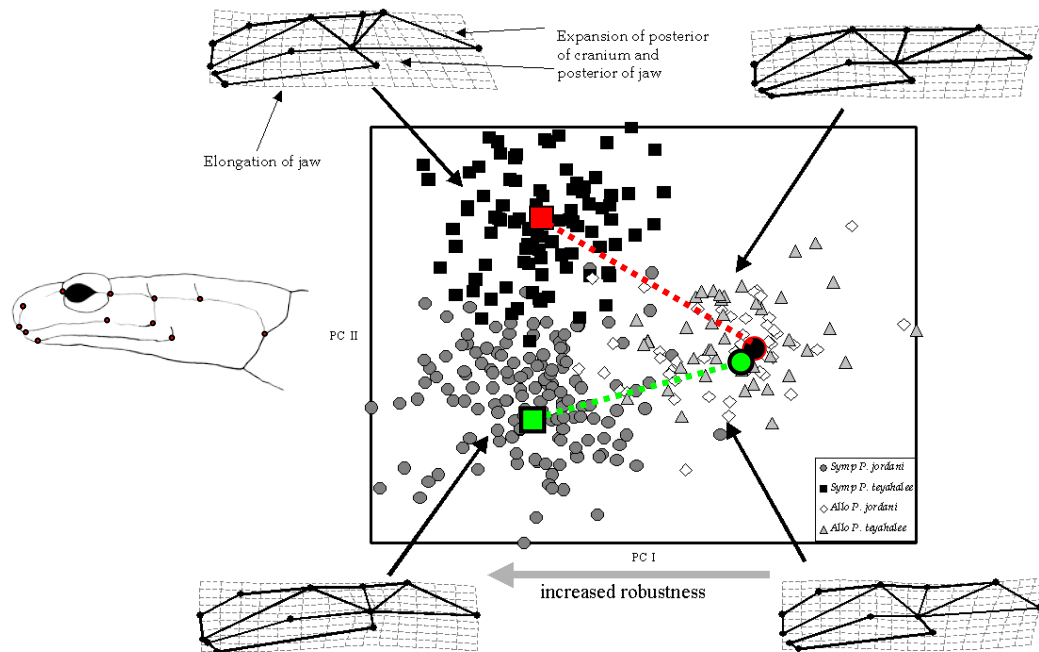
$|d_1 - d_2| \times 100$



Angle,  $\theta$



- Ecological character displacement (*P. jordani* vs. *P. teyahalee*)
- Significant species, site, species×site



$$D_{Jord} = 0.087, D_{Teh} = 0.099, P = 0.172 \text{ NS}$$

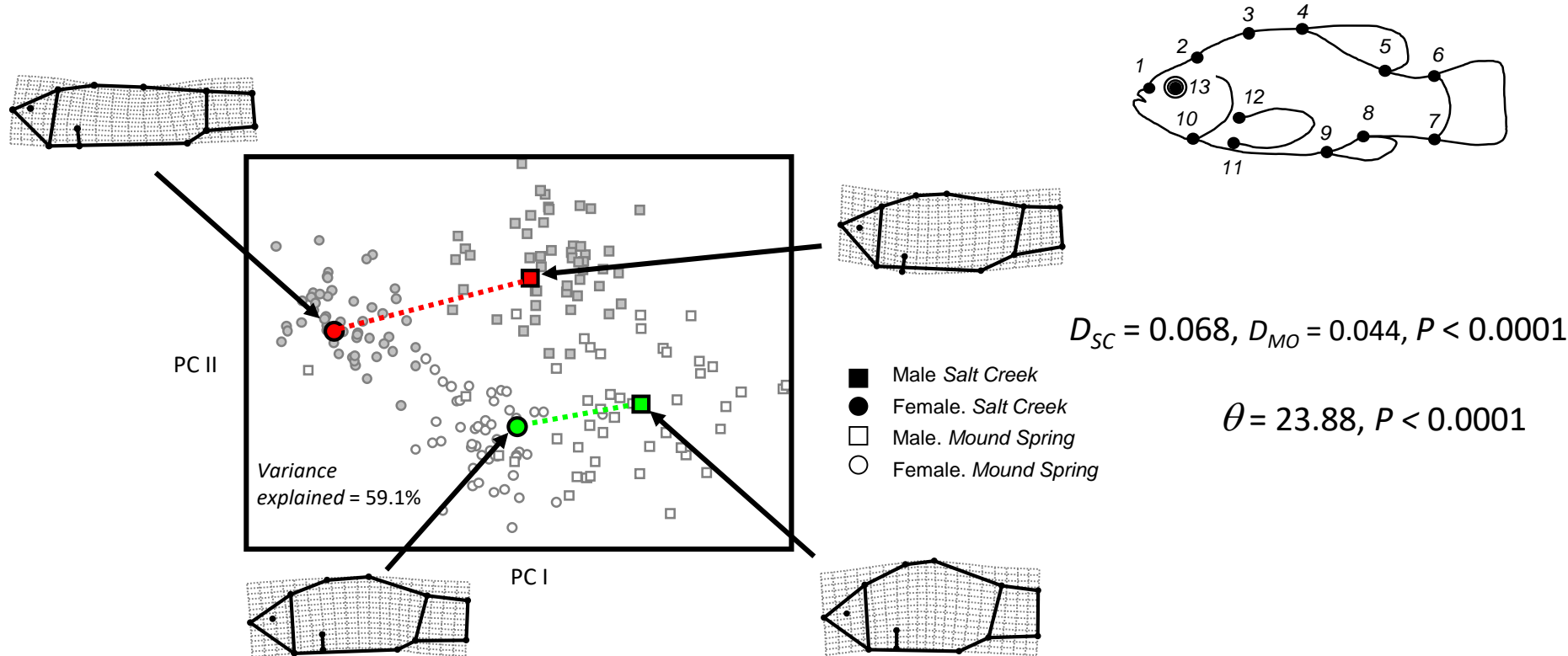
$$\theta = 47.71, P < 0.0001$$

- Conclusion: species differ in *way* they diverge, not how much change they exhibit from allopatry to sympatry

Data From Adams. (2004). *Ecology*. 85:2664-2670.

From Collyer and Adams. (2007). *Ecology*. 88:683-692.

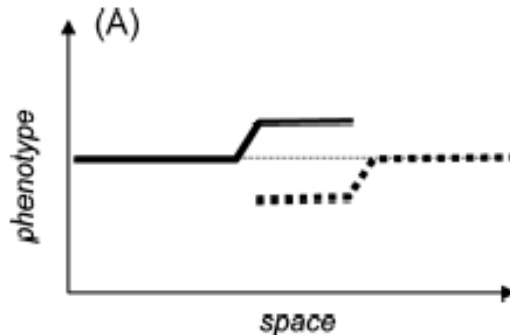
- Sexual dimorphism in white sands pupfish (*C. tularosa*)
- Significant population, sex, population×sex



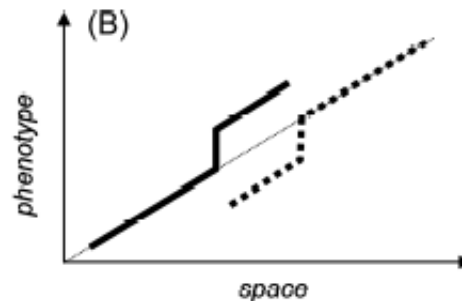
- Conclusion: populations display different *amounts* of sexual dimorphism *and* different directions of dimorphism



- For many hypotheses, we must account for covariate terms while assessing patterns of change
- Example: Character displacement tests: *Dsymp* > *Dallo*



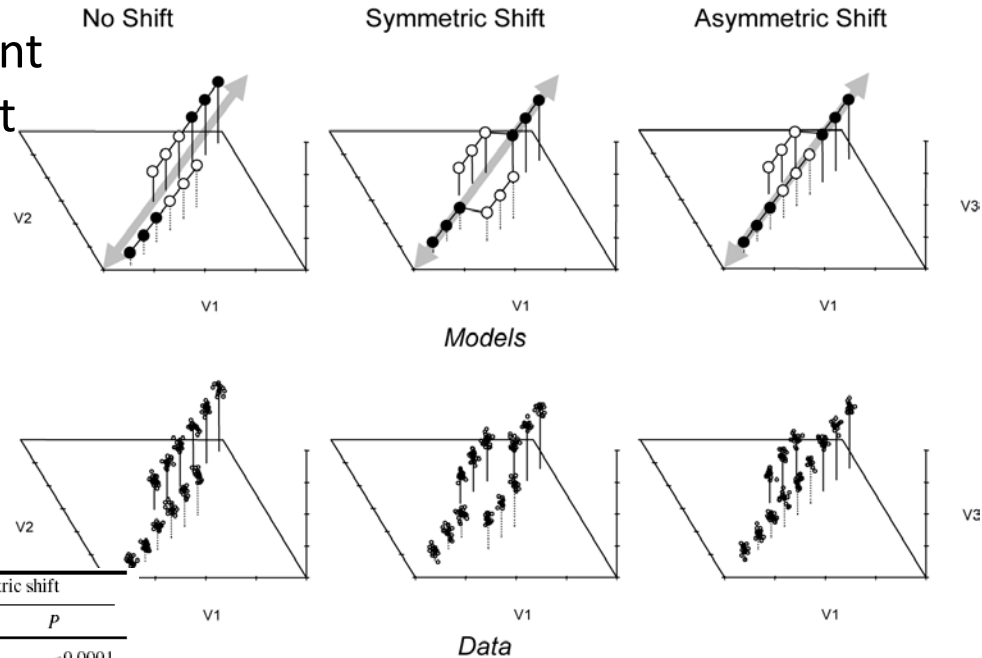
- If phenotype varies along environmental gradient, must account for it



- Incorporate covariate in X; rest of protocol remains unchanged

- Character change along a gradient, 3 scenarios:

- 1.No character displacement (CD)
- 2.Asymmetric character displacement
- 3.Symmetric character displacement

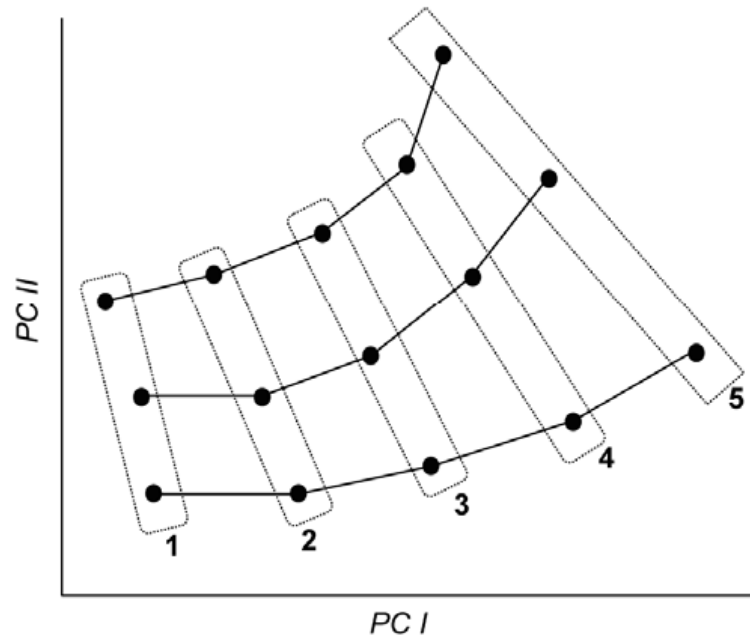


Factor	df	No shift		Symmetric shift		Asymmetric shift	
		$\Lambda$	$P$	$\Lambda$	$P$	$\Lambda$	$P$
Species	1	0.0207	<0.0001	0.0300	<0.0001	0.0800	<0.0001
Community	1	1.0000	0.3633	1.0000	0.9568	0.1300	<0.0001
Species $\times$ community	1	<b>0.2000</b>	<b>&lt;0.0001</b>	0.0300	<0.0001	0.0900	<0.0001
Error	235						
$(D_{\text{symp}} - D_{\text{allo}})$		-12.46	<b>0.0001<sup>1</sup></b>	-9.64	0.0001	-11.58	0.0001
Gradient	1	0.0012	<0.0001	0.0016	<0.0001	0.0017	<0.0001
Species	1	0.0277	<0.0001	0.0161	<0.0001	0.0624	<0.0001
Community	1	1.0000	0.3580	1.0000	0.7929	0.1130	<0.0001
Species $\times$ community	1	1.0000	<b>0.4764</b>	0.1157	<0.0001	0.3573	<0.0001
Error	235						
$(D_{\text{symp}} - D_{\text{allo}})$		-0.064	<b>0.4580</b>	3.895	<0.0001	1.774	<0.0001

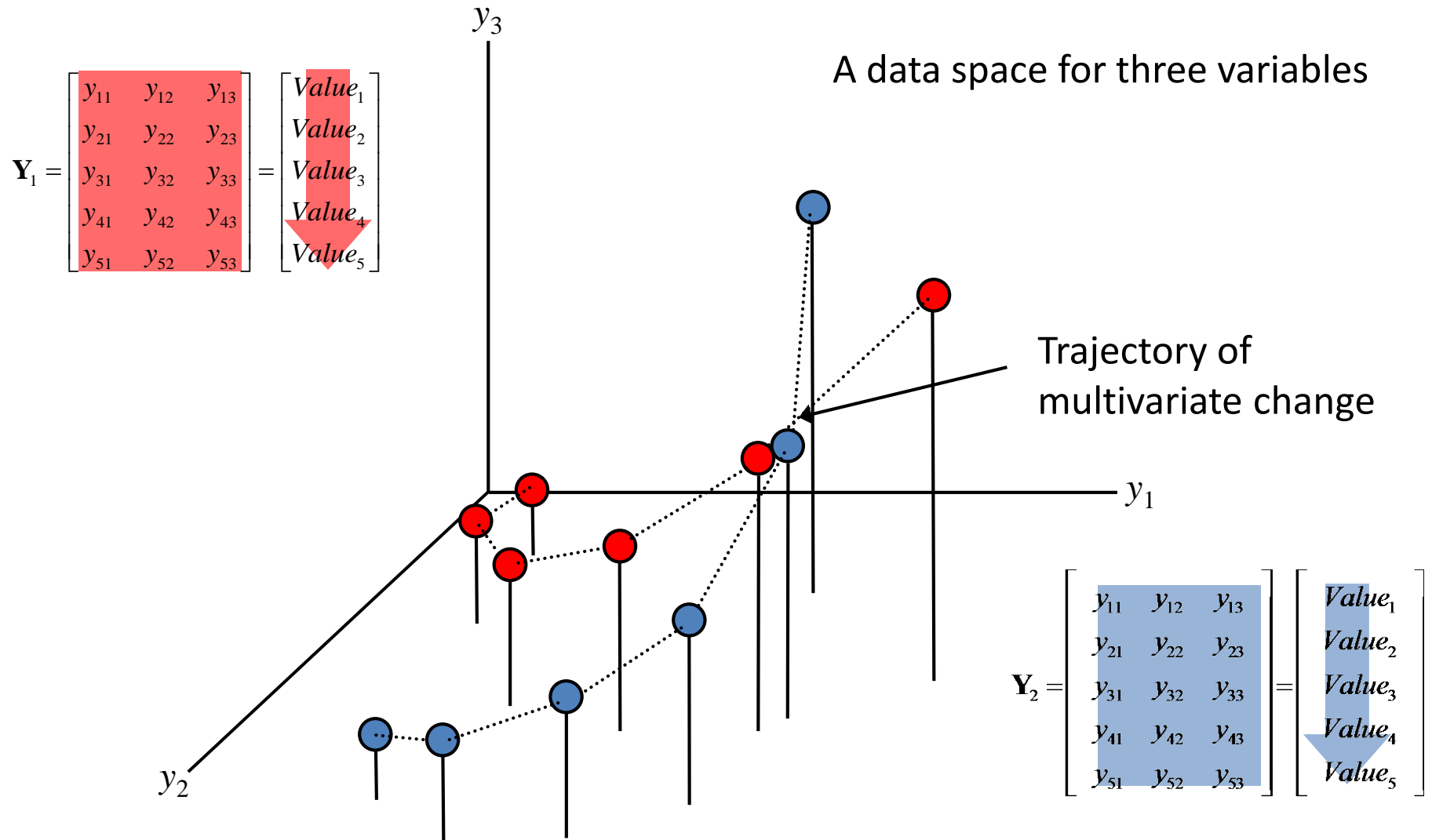
<sup>1</sup>  $P$ -values of direct tests are empirically derived from a permutation procedure. These values represent two-tailed considerations.

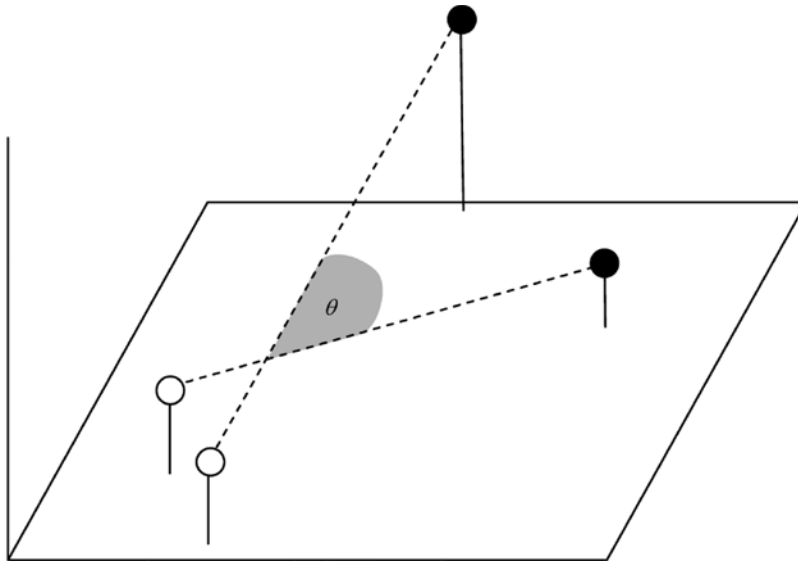
- This approach correctly identifies CD when it is present, and does not identify it when it is not present

- Method easily generalized for more than 2 groups
- Must do in pairwise fashion (1 vs. 2, 1 vs. 3, etc.)  
(e.g., Hollander et al. 2006. *J. Evol. Biol.* 19:1861-1872.)
- For  $> 2$  states (e.g., environments) phenotypic change vector is now a TRAJECTORY



Values represent sequential states (e.g., developmental stages, temporal points)



**A**

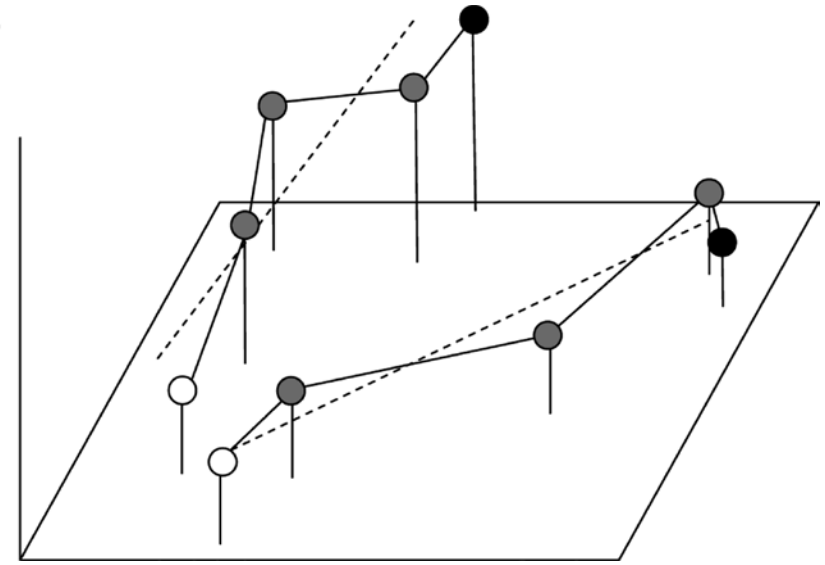
**Magnitude**  $d_i = \sqrt{\Delta \mathbf{y}_i^t \Delta \mathbf{y}_i}$

**Difference in Magnitude**  
 $\Delta d_{ij}$

**Difference in Direction**

$$r_{12} = \frac{\Delta \mathbf{y}_1^t \Delta \mathbf{y}_2}{d_1 d_2} \quad \theta_{ij} = \cos^{-1} r_{12}$$

$\mathbf{z}_i$  = vectors of matrices that have been scaled to unit size, centered, and rotated to minimize variation among them

**B**

**Magnitude**  $\sum d_i$  (Path distance)

**Difference in Magnitude**

$$\Delta \left( \sum d_{ij} \right)$$

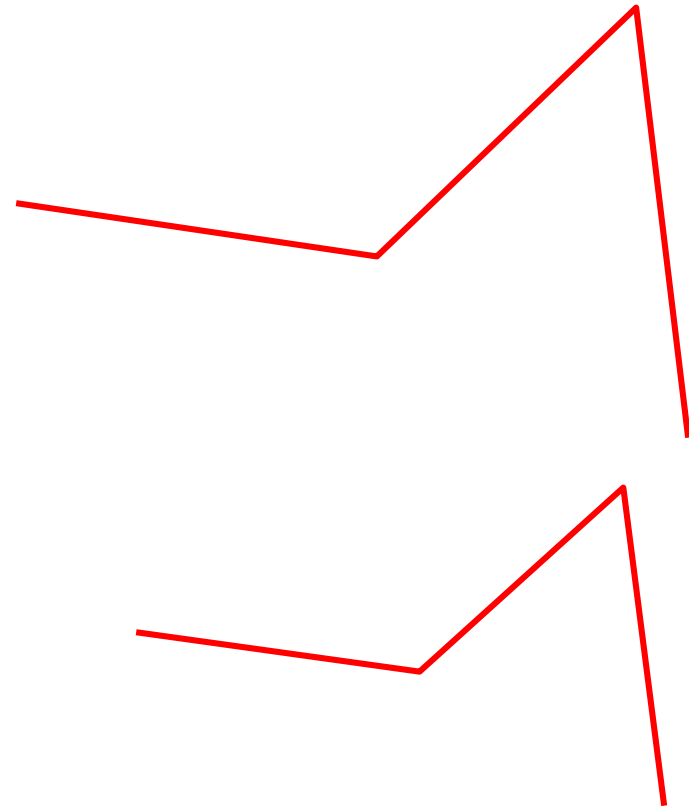
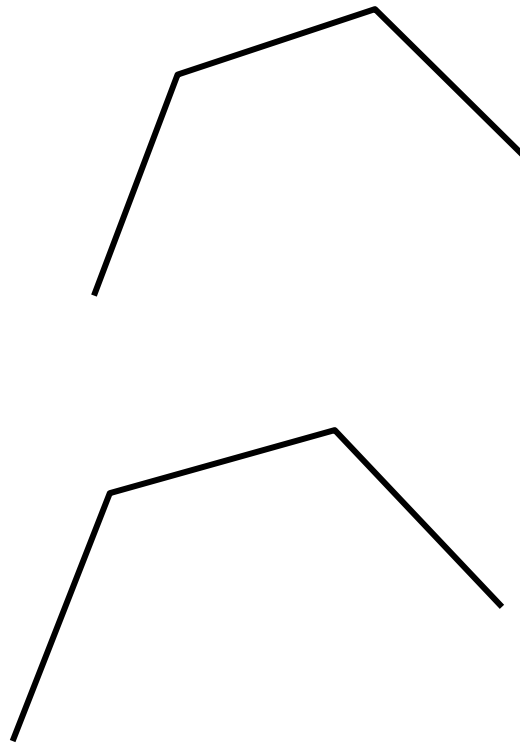
**Difference in Direction**

$$r_{12} = \mathbf{p}_1^t \mathbf{p}_2 \quad \theta_{ij} = \cos^{-1} r_{12}$$

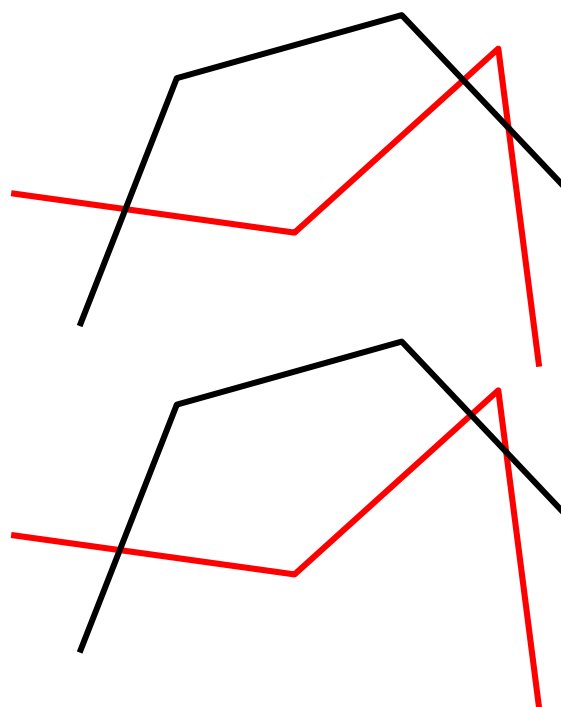
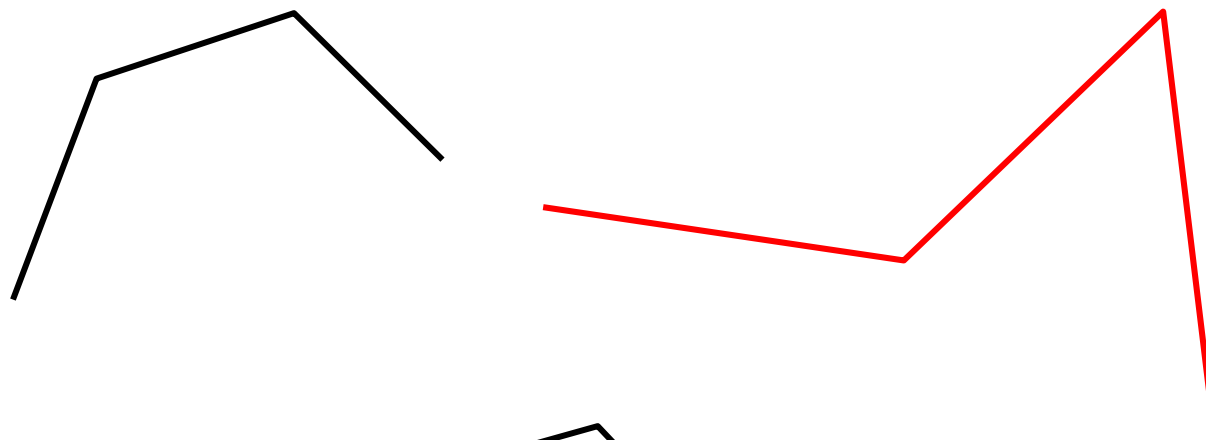
\* $\mathbf{p}$  = principal eigenvector scaled to unit size

**Difference in Shape**

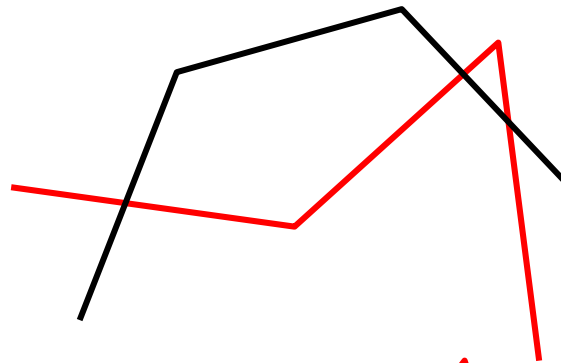
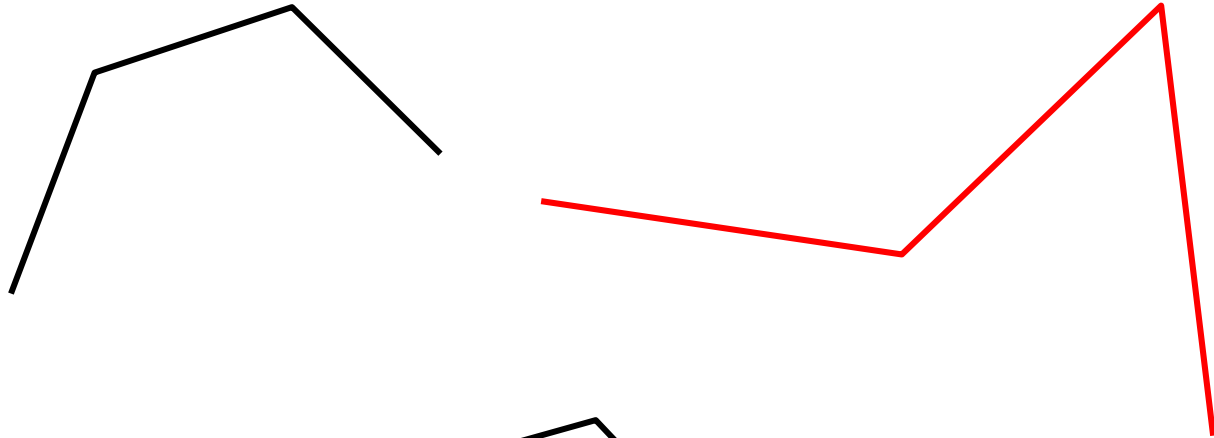
$$D_{ij} = \sqrt{\Delta \mathbf{z}_i^t \Delta \mathbf{z}_i}$$



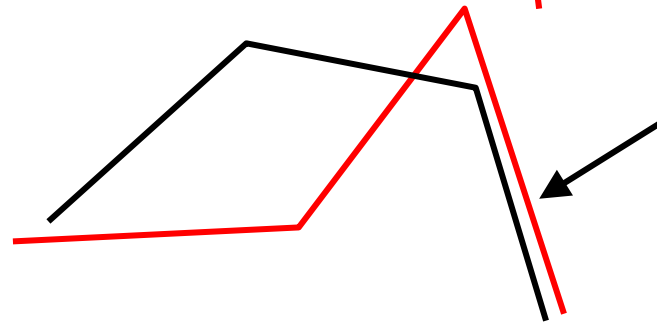
Scaled and Centered



Scaled and Centered



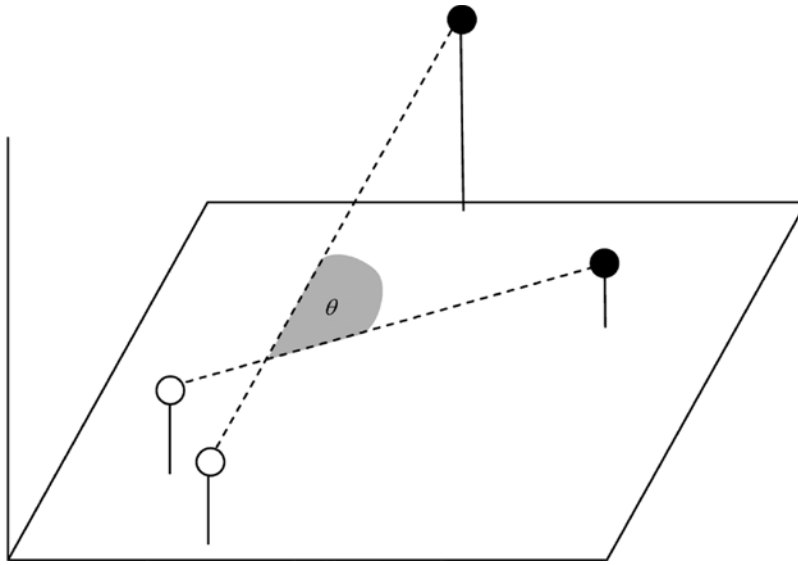
Scaled and Centered  
Rotated



Shape difference = square  
root of summed squared  
differences between  
corresponding  
“landmarks”



A



Magnitude  $d_i = \sqrt{\Delta \mathbf{y}_i^t \Delta \mathbf{y}_i}$

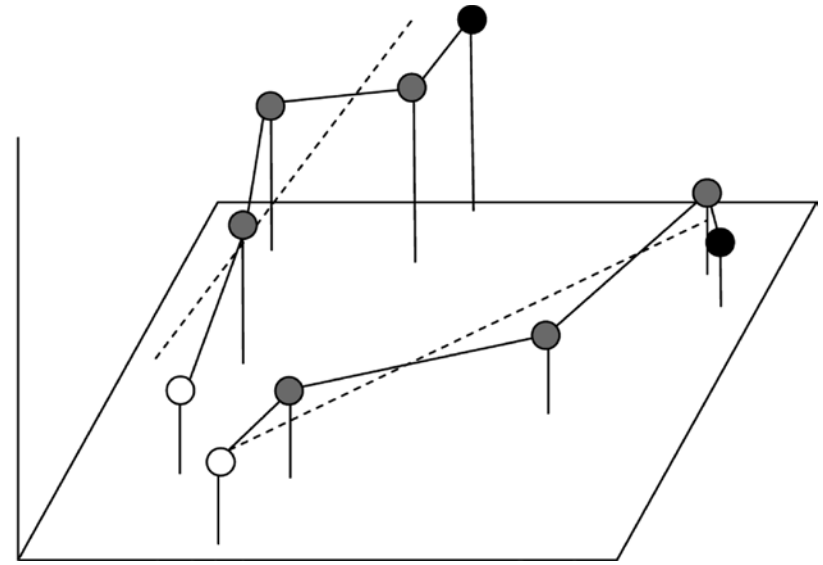
Difference in Magnitude  
 $\Delta d_{ij}$

Difference in Direction

$$r_{12} = \frac{\Delta \mathbf{y}_1^t \Delta \mathbf{y}_2}{d_1 d_2} \quad \theta_{ij} = \cos^{-1} r_{12}$$

Can RRPP be used to test null hypotheses for these statistics?

B



Magnitude  $\sum d_i$  (Path distance)

Difference in Magnitude

$$\Delta \left( \sum d_{ij} \right)$$

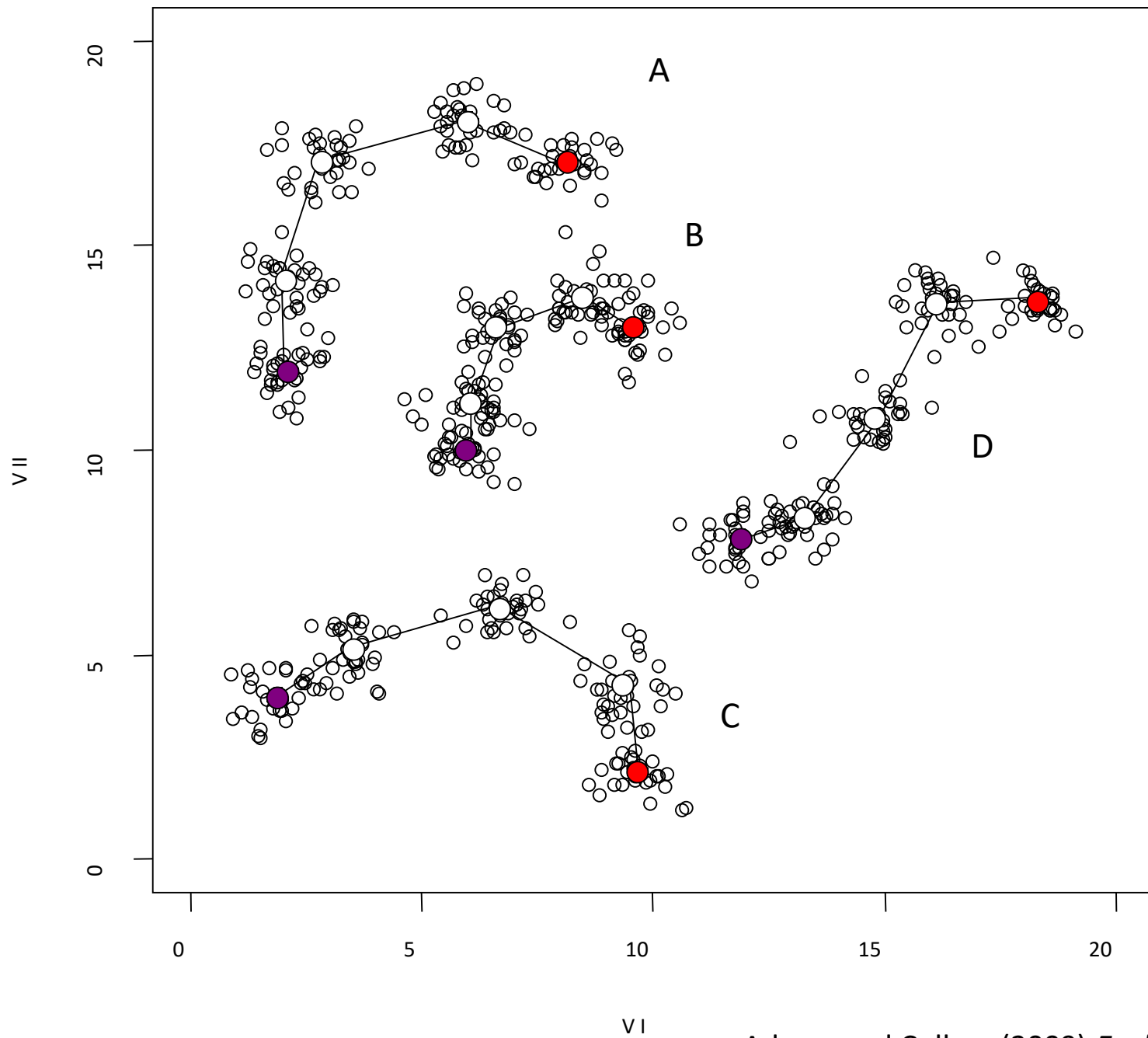
Difference in Direction

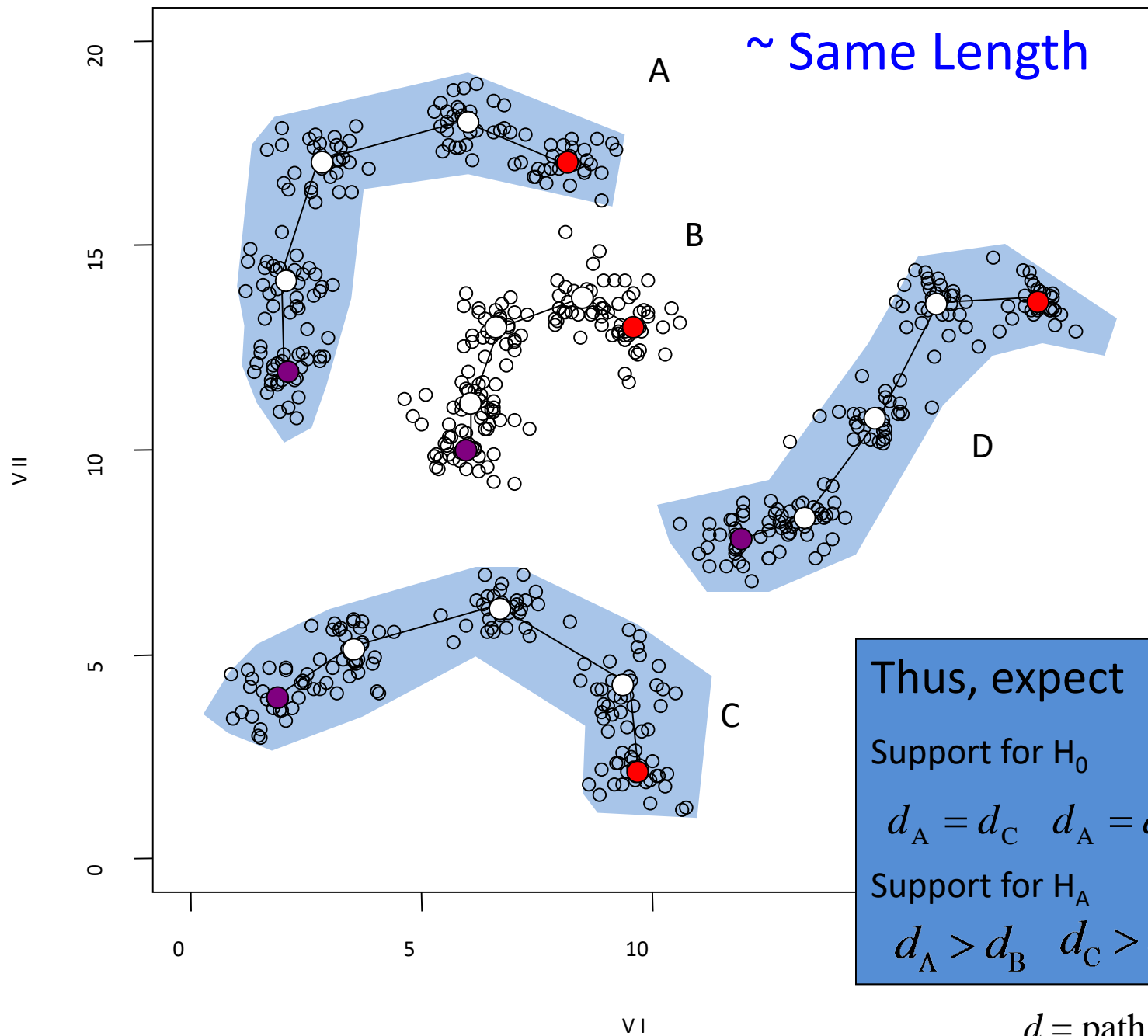
$$r_{12} = \mathbf{p}_1^t \mathbf{p}_2 \quad \theta_{ij} = \cos^{-1} r_{12}$$

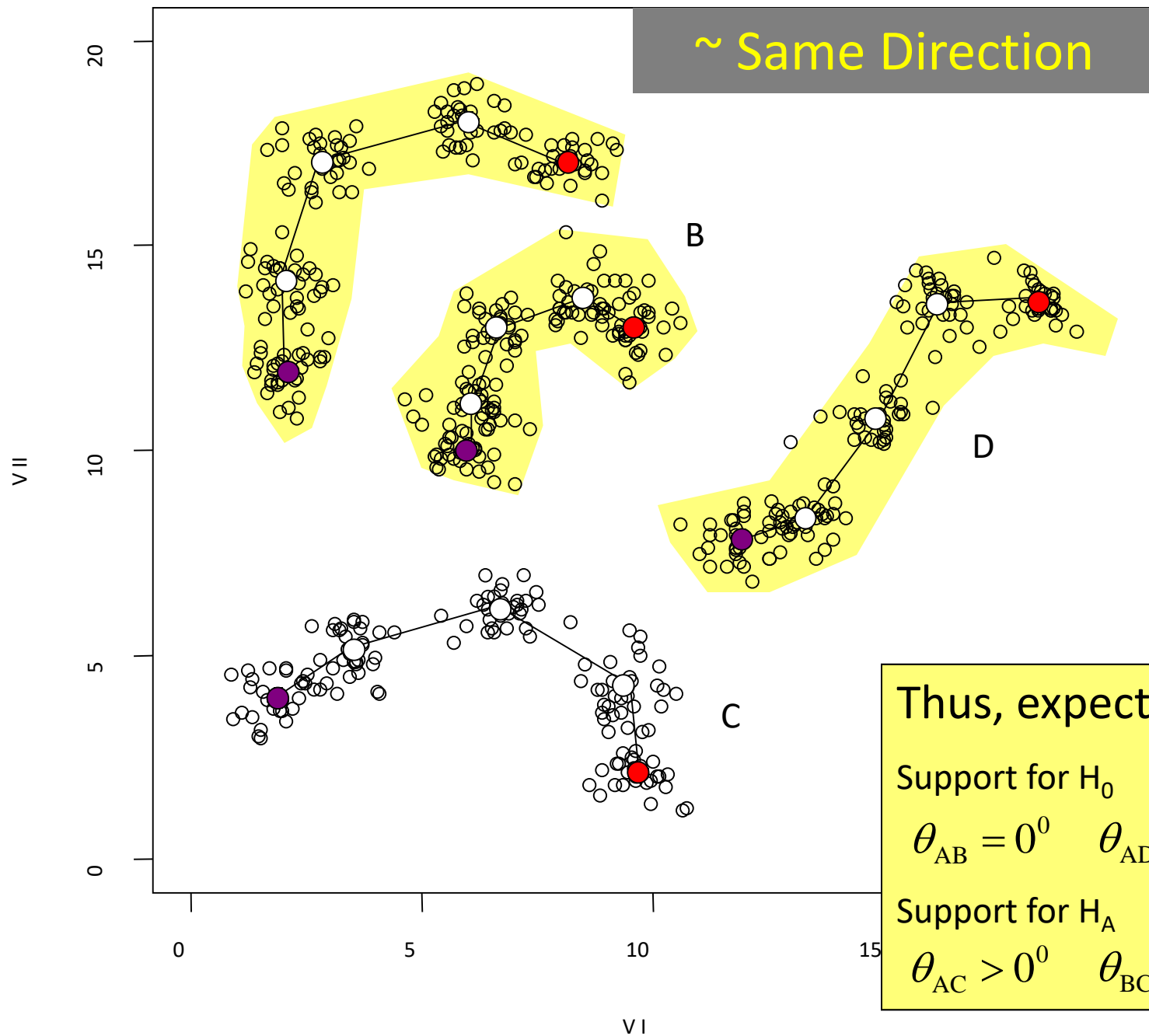
\* $\mathbf{p}$  = principal eigenvector scaled to unit size

Difference in Shape

$$D_{ij} = \sqrt{\Delta \mathbf{z}_i^t \Delta \mathbf{z}_i}$$







Thus, expect

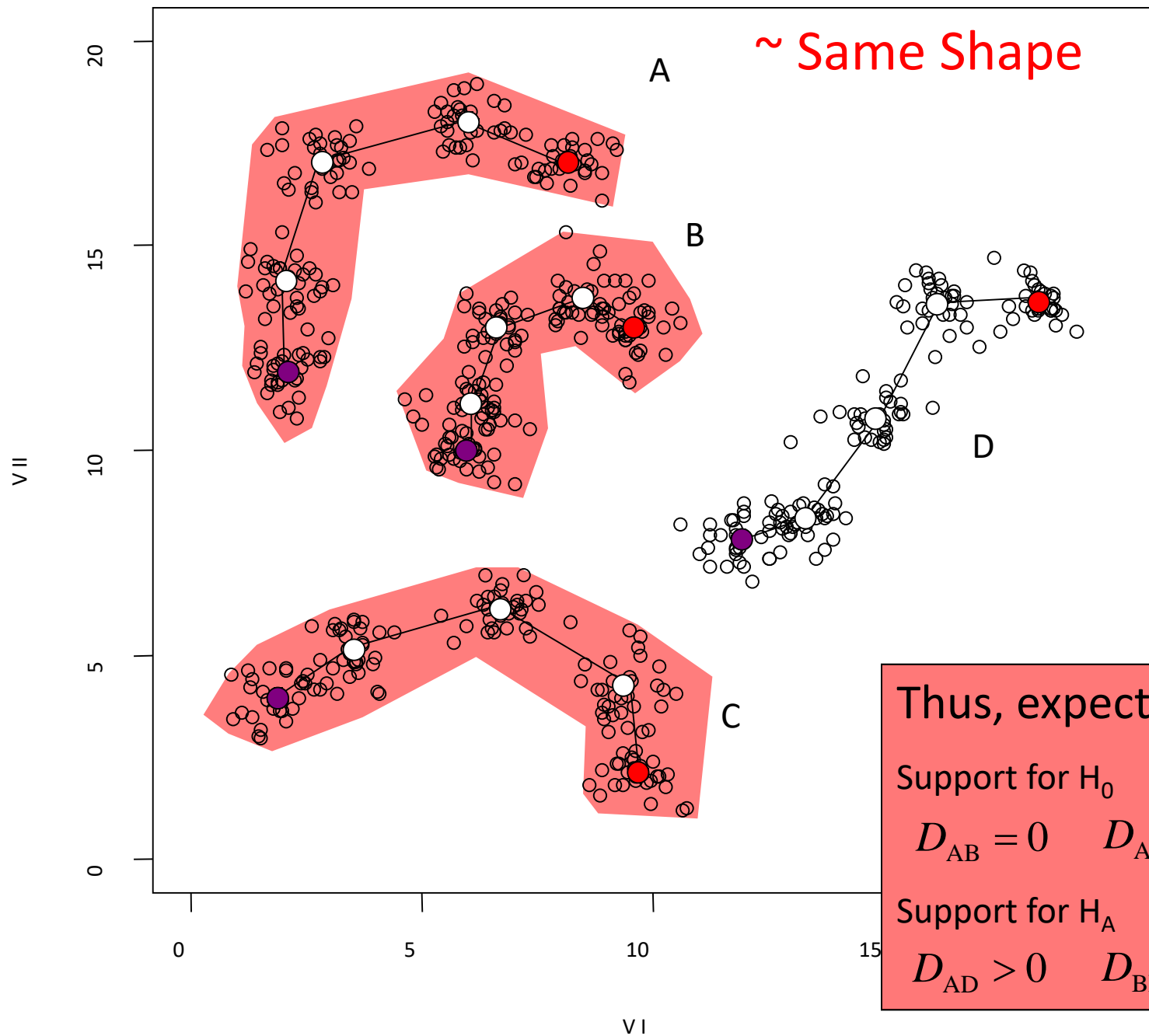
Support for  $H_0$

$$\theta_{AB} = 0^0 \quad \theta_{AD} = 0^0 \quad \theta_{BD} = 0^0$$

Support for  $H_A$

$$\theta_{AC} > 0^0 \quad \theta_{BC} > 0^0 \quad \theta_{DC} > 0^0$$

$\theta$  calculated between principal eigenvectors



Thus, expect

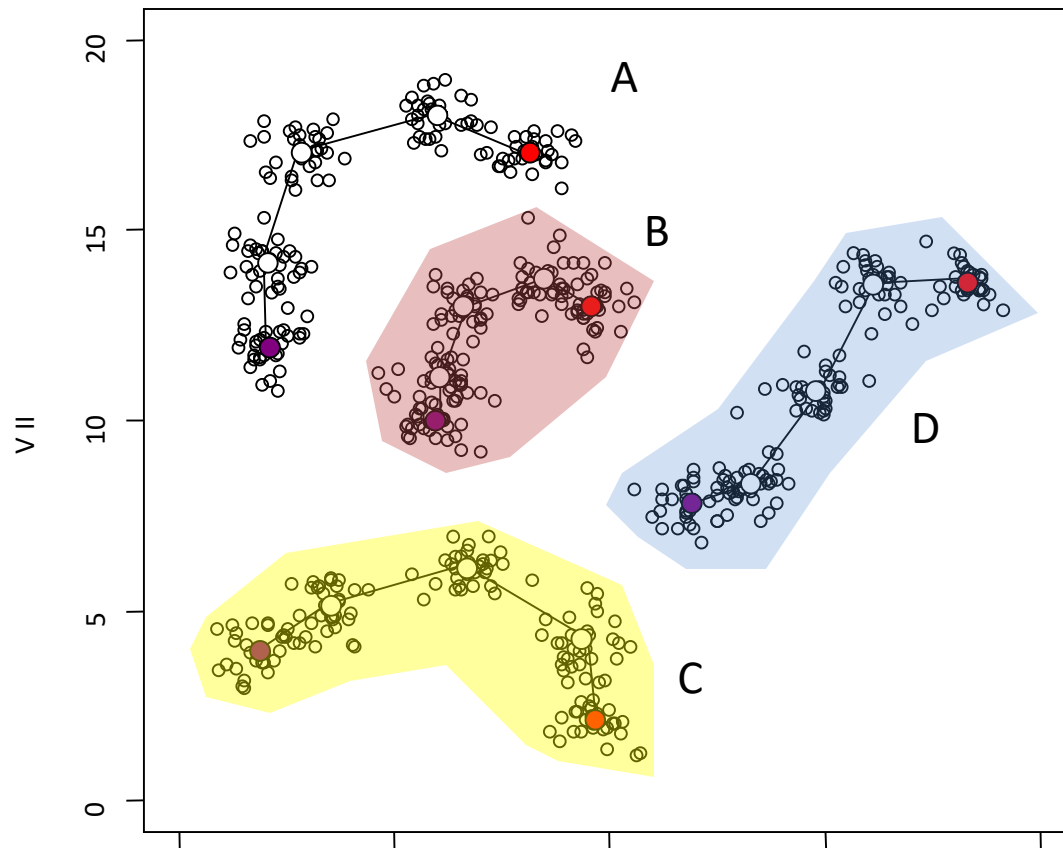
Support for  $H_0$

$$D_{AB} = 0 \quad D_{AC} = 0 \quad D_{BC} = 0$$

Support for  $H_A$

$$D_{AD} > 0 \quad D_{BD} > 0 \quad D_{CD} > 0$$

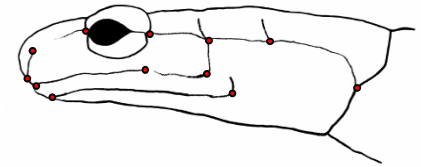
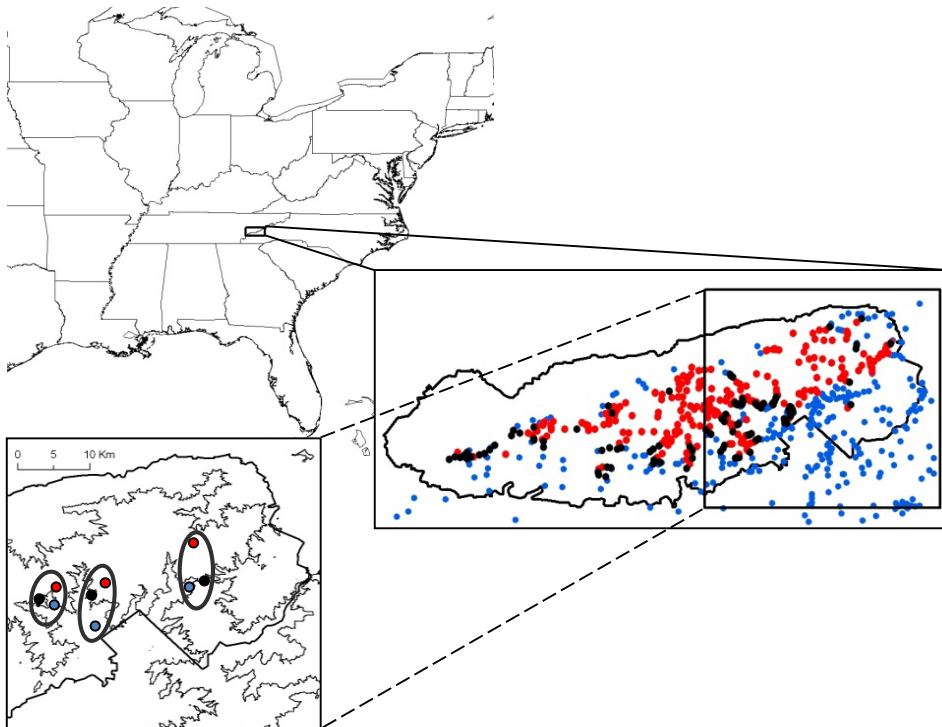
$D$  = Procrustes distance between shapes



**Table 1.** Statistical assessment of differences in phenotypic trajectory size ( $MD_{1,2}$ ), direction ( $\theta_{1,2}$ ), and shape ( $D_{Shape}$ ) differences of four phenotypic evolutionary trajectories (A, B, C, and D) from Figure 2B. Bolded values are those expected to be different based on the simulation of data. Observed significance levels ( $P$ -values) were empirically generated from 10,000 random permutations as described.

Comparison	$MD_{1,2}$	$P_{Size}$	$\theta_{1,2}$	$P_{\theta}$	$D_{Shape}$	$P_{Shape}$
A,B	3.7866	<b>0.0001</b>	0.0367	0.9927	0.0878	0.7491
A,C	0.0700	0.8953	<b>54.6515</b>	<b>0.0001</b>	0.0332	0.9982
A,D	0.2018	0.7120	1.1431	0.7649	<b>0.3011</b>	<b>0.0001</b>
B,C	<b>3.8566</b>	<b>0.0001</b>	<b>54.6883</b>	<b>0.0001</b>	0.0855	0.7784
B,D	<b>3.9884</b>	<b>0.0001</b>	1.1063	0.7735	<b>0.3576</b>	<b>0.0001</b>
C,D	0.1318	0.8131	<b>55.7946</b>	<b>0.0001</b>	<b>0.3065</b>	<b>0.0001</b>

- Ecological work demonstrates competition prevalent
- *Plethodon* biogeography: replicated communities across contact zones
- Are microevolutionary changes repeatable?



*Plethodon jordani*



*Plethodon teyahalee*

Measured head shape  
from 336 specimens  
across three mountain  
transects

(allopatry → sympatry)

- Phenotypic evolution is present

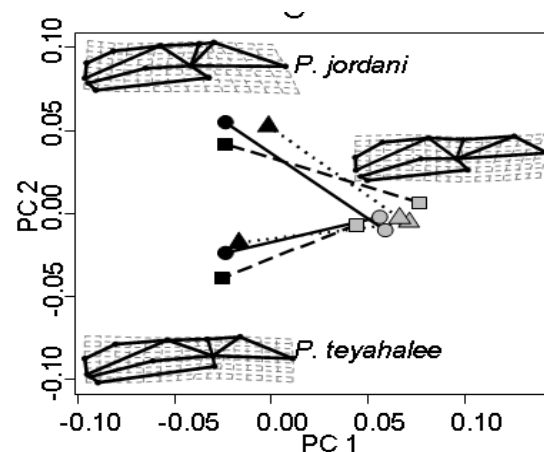
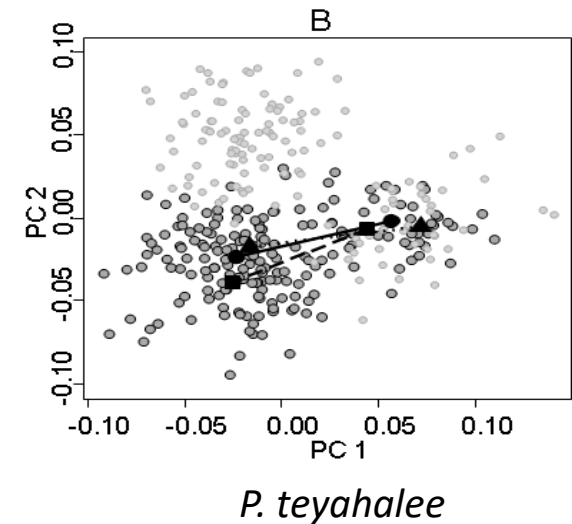
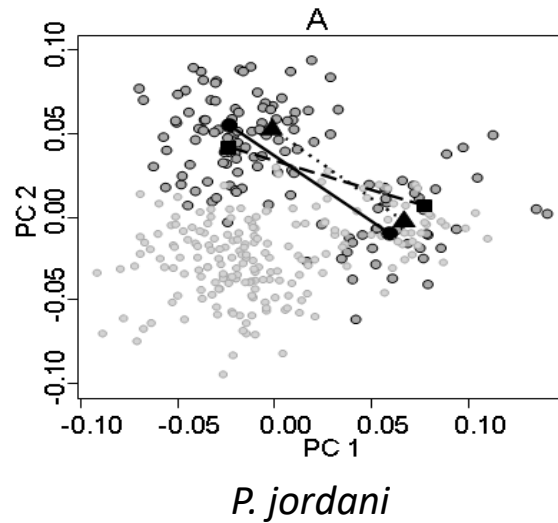
Factor	Df <sub>Factor</sub>	Pillai's Trace	Approx. F	df	P
Species	1	0.741	48.874	18, 307	< 0.0001
Locality Type	1	0.794	65.612	18, 307	< 0.0001
Geographic Transect	2	0.783	11.015	36, 616	< 0.0001
Species × Locality	1	0.519	18.373	18, 307	< 0.0001
Species × Transect	2	0.289	2.888	36, 616	< 0.0001
Locality × Transect	2	0.338	3.482	36, 616	< 0.0001
Species×Locality×Transect	2	0.161	1.499	36, 616	0.0327

- Patterns are REPEATABLE

	Vector Magnitude			Vector Orientation		
<b>A: <i>P. jordani</i></b>	<b>HR</b>	<b>KP</b>	<b>TC</b>	<b>HR</b>	<b>KP</b>	<b>TC</b>
<b>HR</b>		<b>0.1849 NS</b>	<b>0.3192 NS</b>		<b>0.6074 NS</b>	<b>0.3665 NS</b>
<b>KP</b>	<b>0.01689</b>		<b>0.0309 NS</b>	<b>26.785</b>		<b>0.4071 NS</b>
<b>TC</b>	<b>0.00871</b>	<b>0.02560</b>		<b>31.502</b>	<b>41.545</b>	
<b>B: <i>P. teyahalee</i></b>	<b>HR</b>	<b>KP</b>	<b>TC</b>	<b>HR</b>	<b>KP</b>	<b>TC</b>
<b>HR</b>		<b>0.3363 NS</b>	<b>0.8106 NS</b>		<b>0.7965 NS</b>	<b>0.5579 NS</b>
<b>KP</b>	<b>0.00871</b>		<b>0.3261 NS</b>	<b>19.506</b>		<b>0.5069 NS</b>
<b>TC</b>	<b>0.00253</b>	<b>0.01224</b>		<b>25.033</b>	<b>34.136</b>	

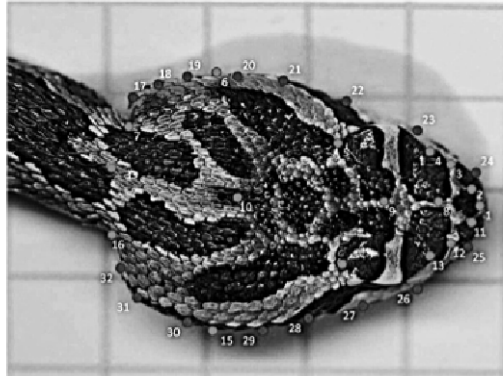


- NO difference in magnitude or direction of evolutionary changes among transects within species (i.e. common patterns found)

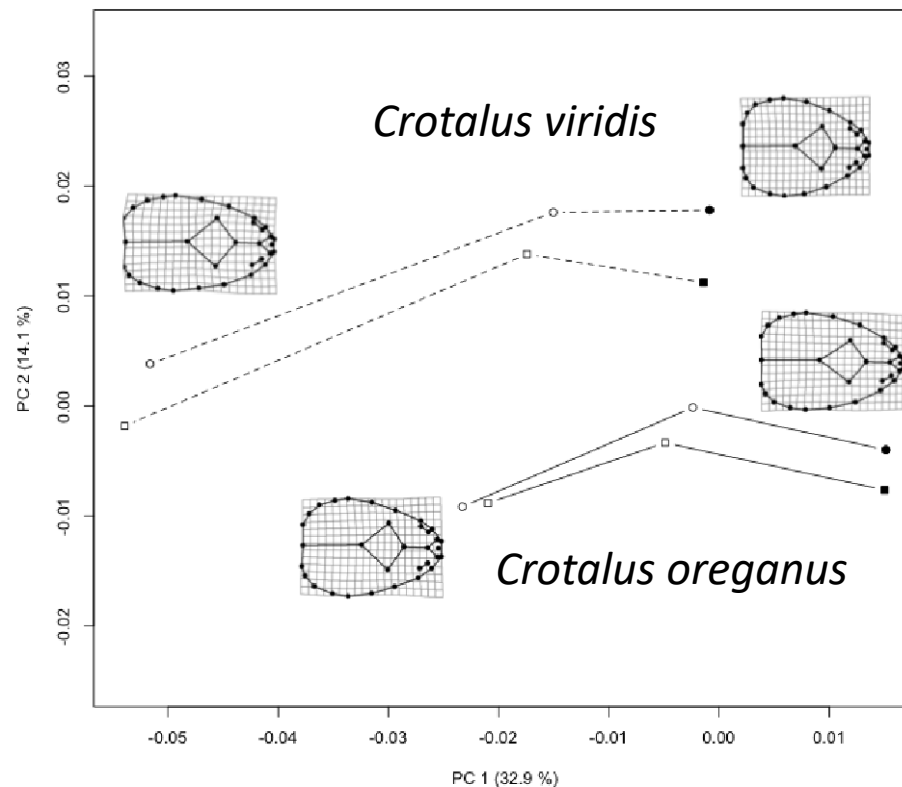


- Conclusion: Evolutionary response to competition repeatable in each species: parallel evolution of character displacement

A

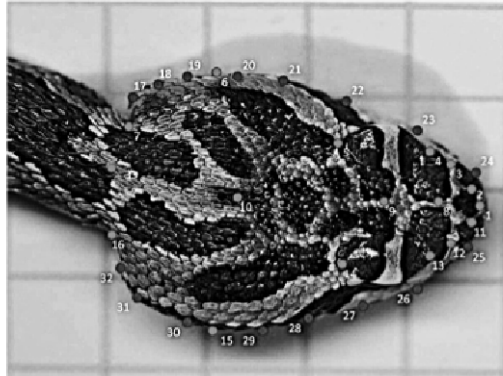


B

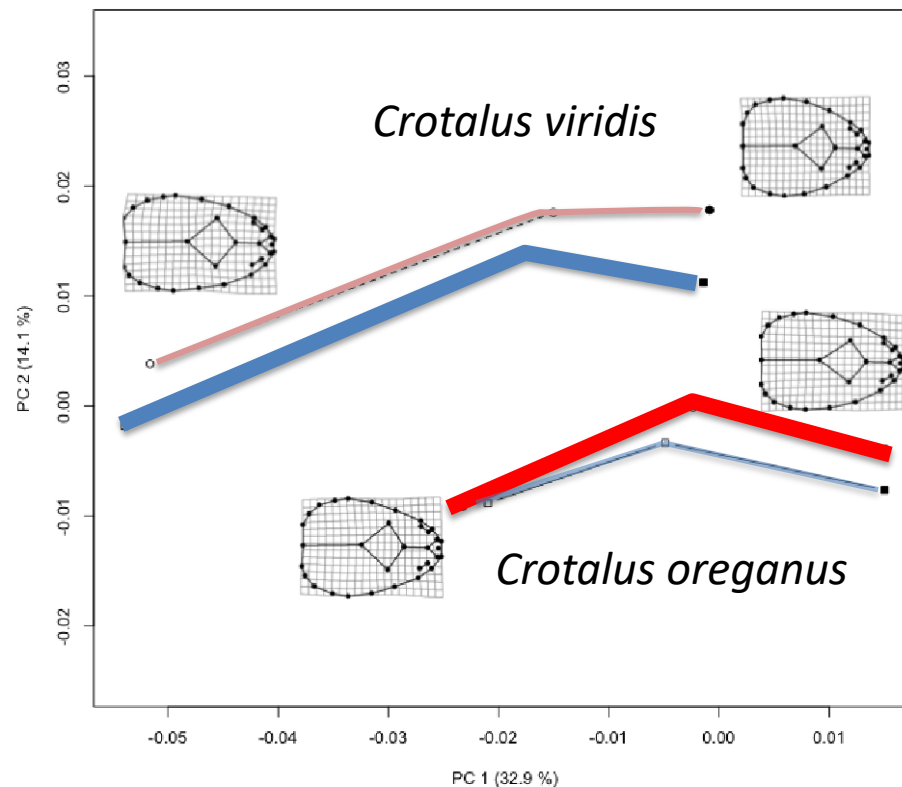


Study examined 3,107 rattlesnakes (*Crotalus*) from museum collections. Shown are two species with overlapping distributions and presumably different ecologies. The question is whether they have the same ontogenetic pattern of head shape change. The trajectories are neonate-juvenile-adult trajectories, for females and males, within species.

A



B



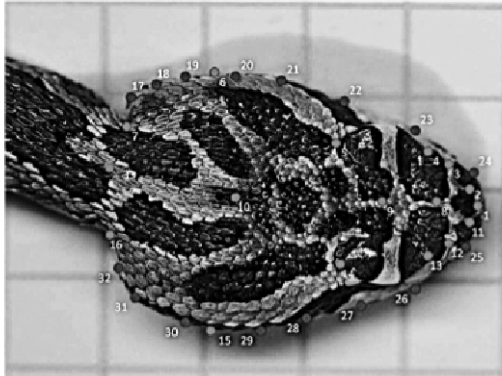
Study examined 3,107 rattlesnakes (*Crotalus*) from museum collections. Shown are two species with overlapping distributions and presumably different ecologies. The question is whether they have the same ontogenetic pattern of head shape change. The trajectories are neonate-juvenile-adult trajectories, for females and males, within species.

### *Sexual dimorphism*

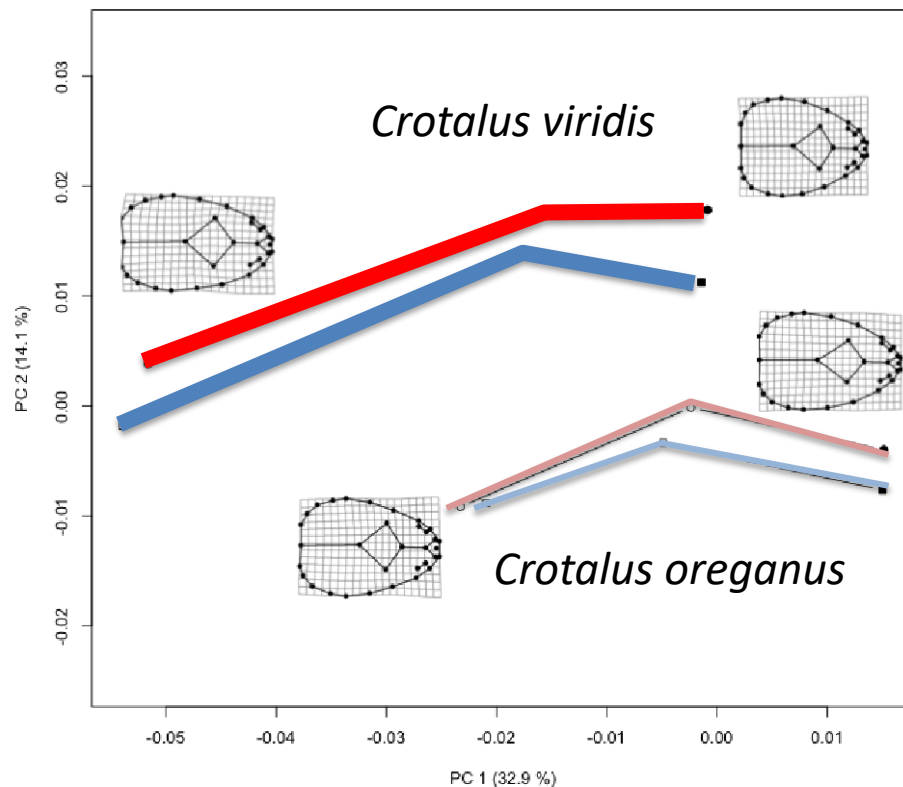
$MD = 0.0005$  *C. viridis*  $P = 0.0005$

$MD = 0.0060$  *C. oreganus*  $P = 0.0001$

A



B



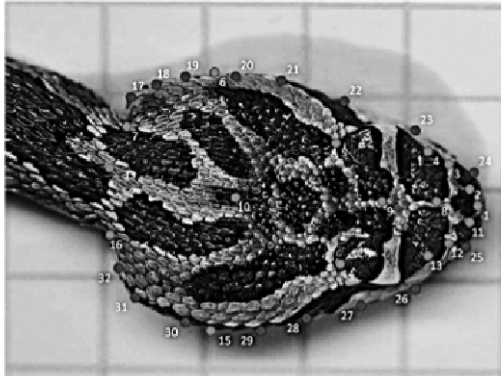
Study examined 3,107 rattlesnakes (*Crotalus*) from museum collections. Shown are two species with overlapping distributions and presumably different ecologies. The question is whether they have the same ontogenetic pattern of head shape change. The trajectories are neonate-juvenile-adult trajectories, for females and males, within species.

*Amount of ontogenetic shape change*

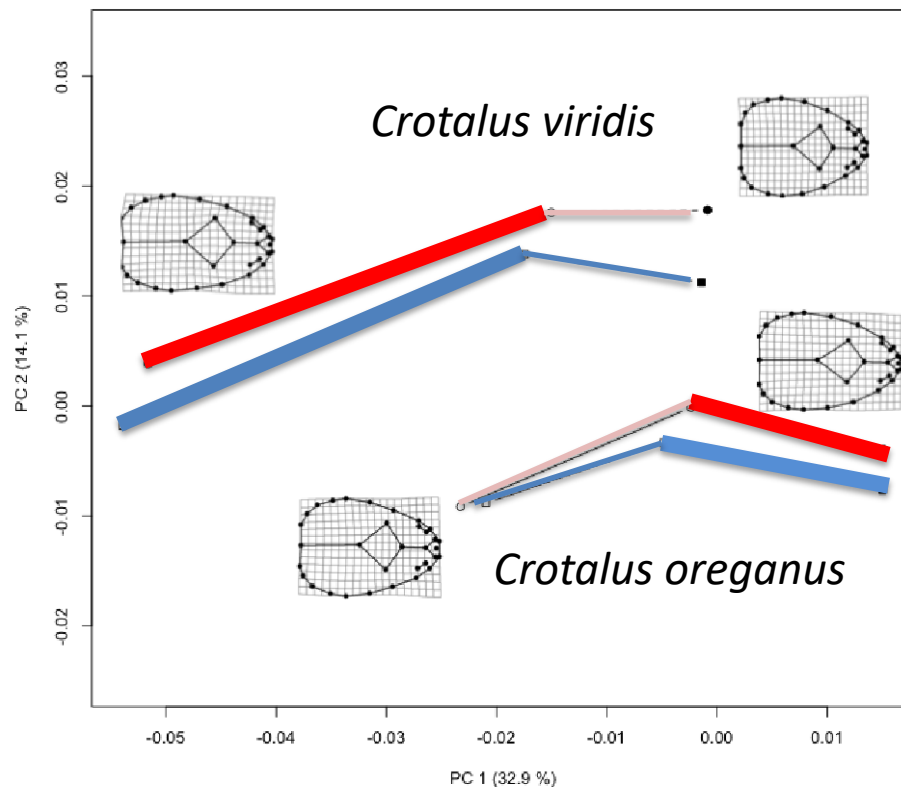
*MD = 0.0119 Females,  $P = 0.0069$*

*MD = 0.0184 Males,  $P = 0.0005$*

A



B



Study examined 3,107 rattlesnakes (*Crotalus*) from museum collections. Shown are two species with overlapping distributions and presumably different ecologies. The question is whether they have the same ontogenetic pattern of head shape change. The trajectories are neonate-juvenile-adult trajectories, for females and males, within species.

*Shape of ontogenetic shape change*

$D_p = 0.21$  Females,  $P = 0.0405$

$D_p = 0.21$  Males,  $P = 0.0048$

- For  $> 2$  trajectories, can compare attributes in pairwise fashion
- For overall test, calculate summary statistic for each attribute (e.g.,):

$$Var_{Size} = \frac{\sum_i^{m(m-1)/2} (MD_i - \overline{MD})^2}{(m(m-1)/2) - 1}$$

- Generate summary statistics in residual randomization procedure for statistical assessment
- Tests the ‘concordance’ of trajectory size, shape, or orientation for set of trajectories
- Provides a quantitative means of assessing the ‘repeatability’ of ecological or evolutionary events (Do trajectories have similar magnitude, orient similarly, etc.?)

- Significant interactions are the most interesting result biologically
- Tell us that response to factor A dependent on level of factor B
- Imply that the *change* across levels is not consistent
- Many biological questions are really interested in comparing phenotypic change, rather than comparing phenotypes
- Phenotypic plasticity, ontogenetics, species interactions, local adaptation, adaptive diversification, etc.
- Hypothesis tests of “effects” are not sufficient to determine *how* change has occurred and how patterns of change differ
- Attributes of change (magnitude, orientation, shape of change trajectory) function as test statistics for more interesting, and precise, hypotheses
- Provides more complete understanding of biological change