FW 599 Special Topics: Multivariate Analysis of Ecological Data in R

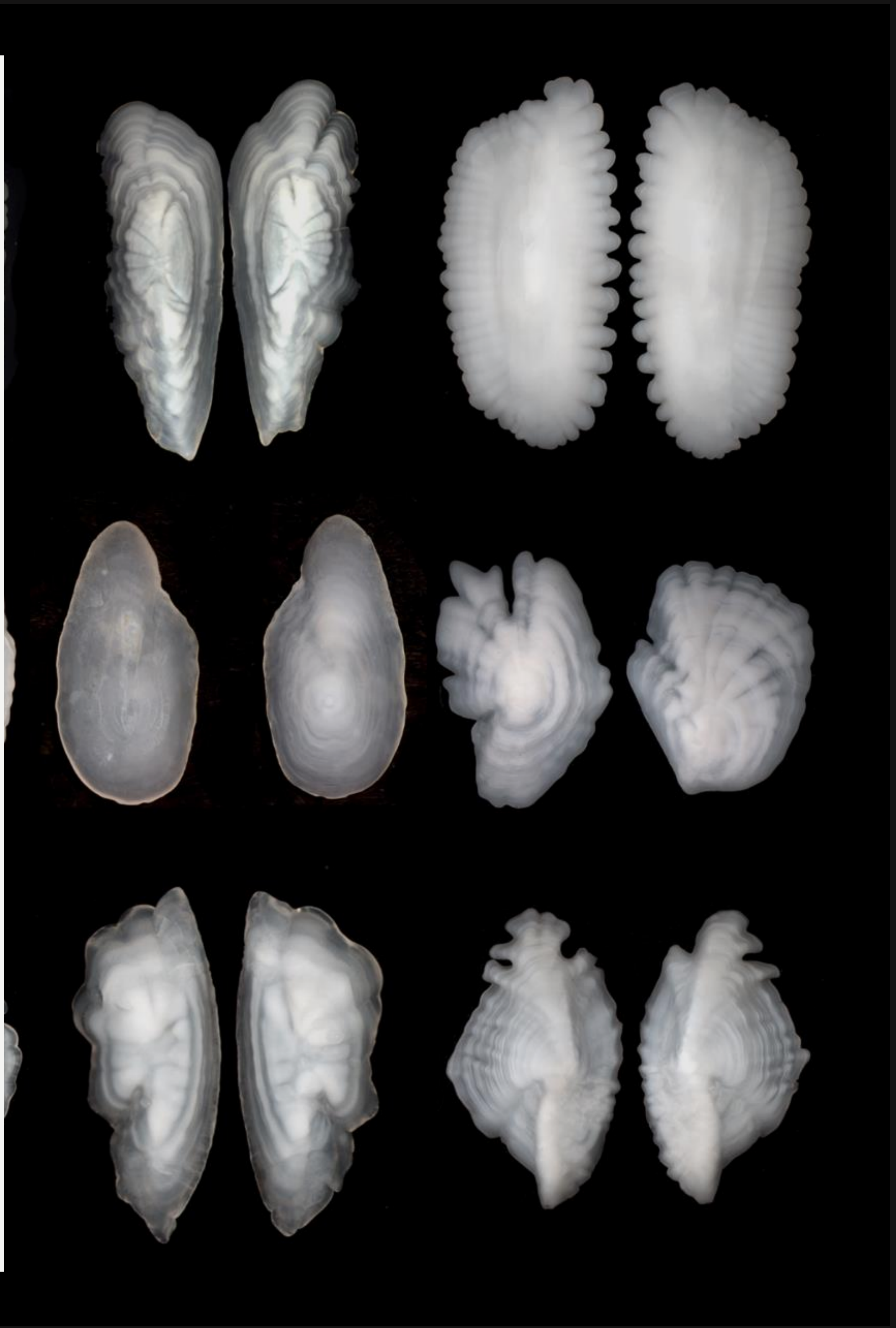
Lecture 9: Constrained Ordination

Tuesday, October 29, 2024

Lecture 9: Constrained Ordination

- Redundancy Analysis (RDA)
- Canonical Correspondence Analysis (CCA)
- Co-inertia Analysis (CoIA)

Recap: Indirect vs. Direct Comparison



Making Inferences from Ordination: Objectives

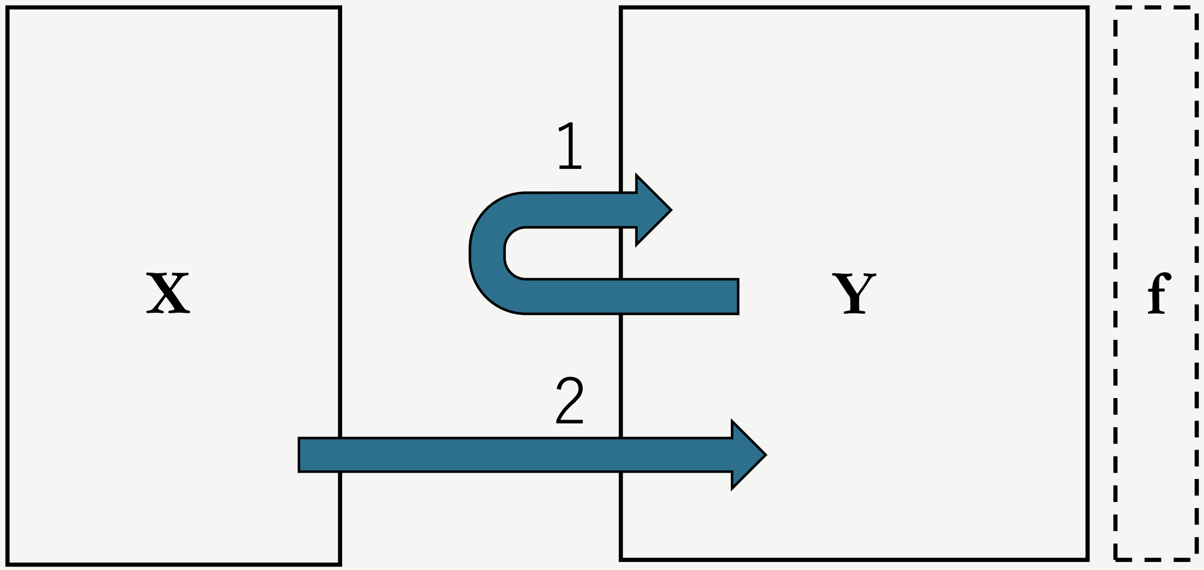
How do we translate our results into ecologically meaningful insights?

Interpretation: links patterns to ecological processes. Can be exploratory *or* inferential.

Inference: draws conclusions from patterns in complex datasets, usually <u>to test hypotheses or identify key explanatory variables</u>.

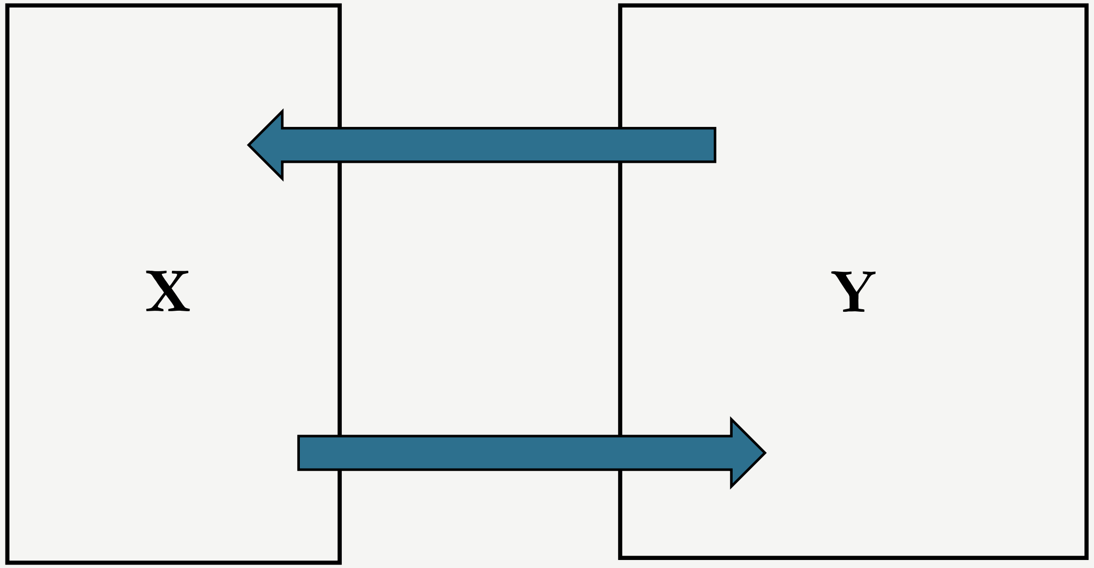
Making Inferences from Ordination: Indirect Gradient Analysis

The goal of **indirect comparison** is to interpret the structure of the descriptors (response variables) using either the descriptors themselves or another set of descriptors.



Making Inferences from Ordination: Direct Gradient Analysis

The goal of **direct comparison** is to simultaneously analyze the response and explanatory data matrices.

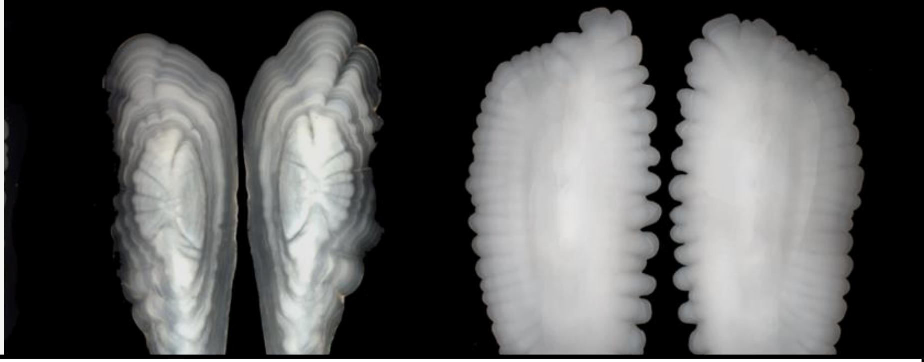


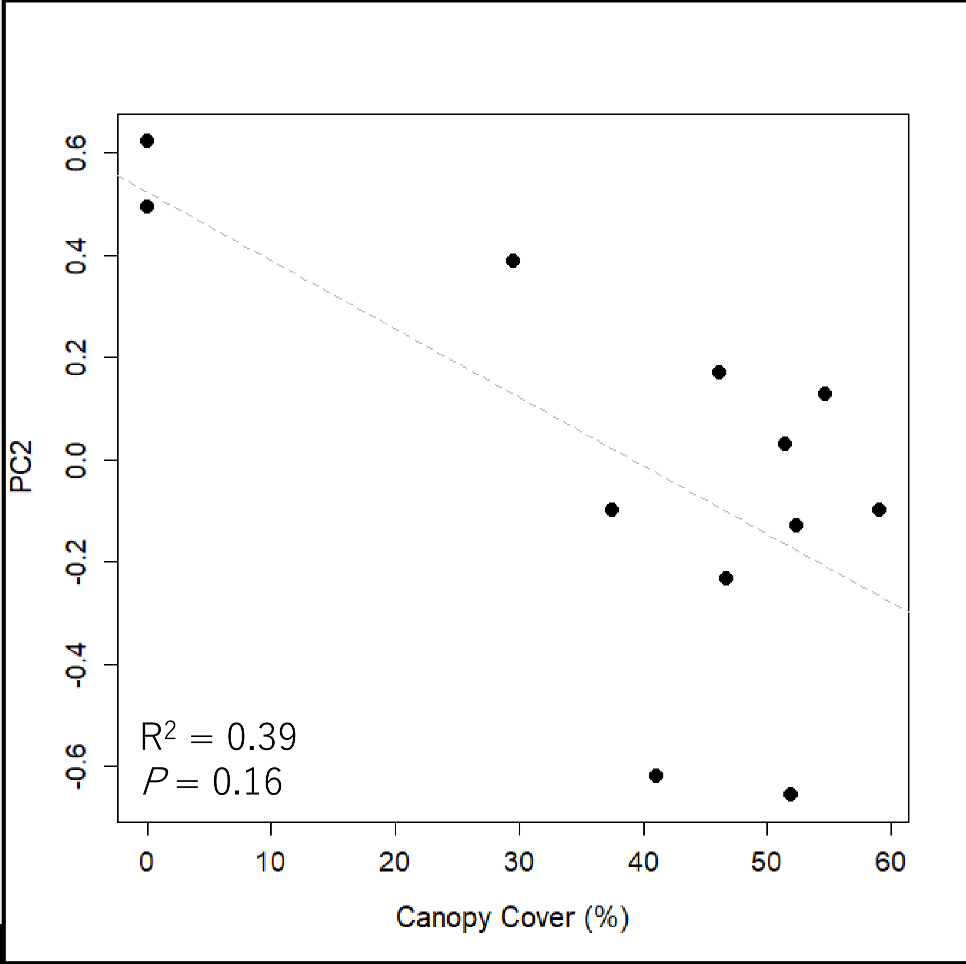
Making Inferences from Ordination: Explanatory

Explanatory data analysis looks for underlying relationships, patterns, and trends within a dataset.

1) **Indirect Comparison:** Treat principal axes/coordinates or clustering partitions as response variables in a regression analysis.

2) **Direct Comparison:** Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA).



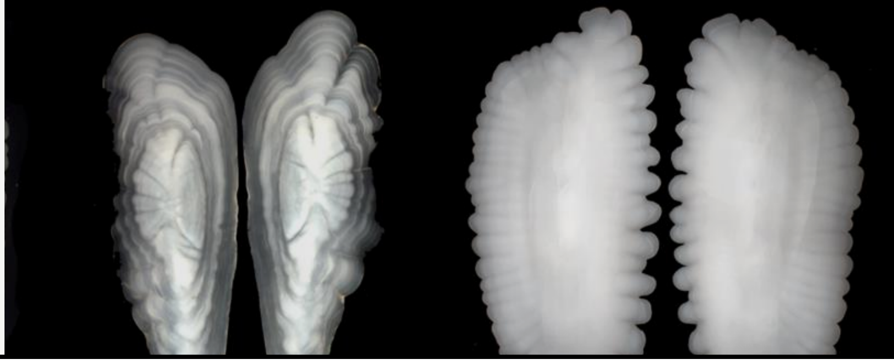


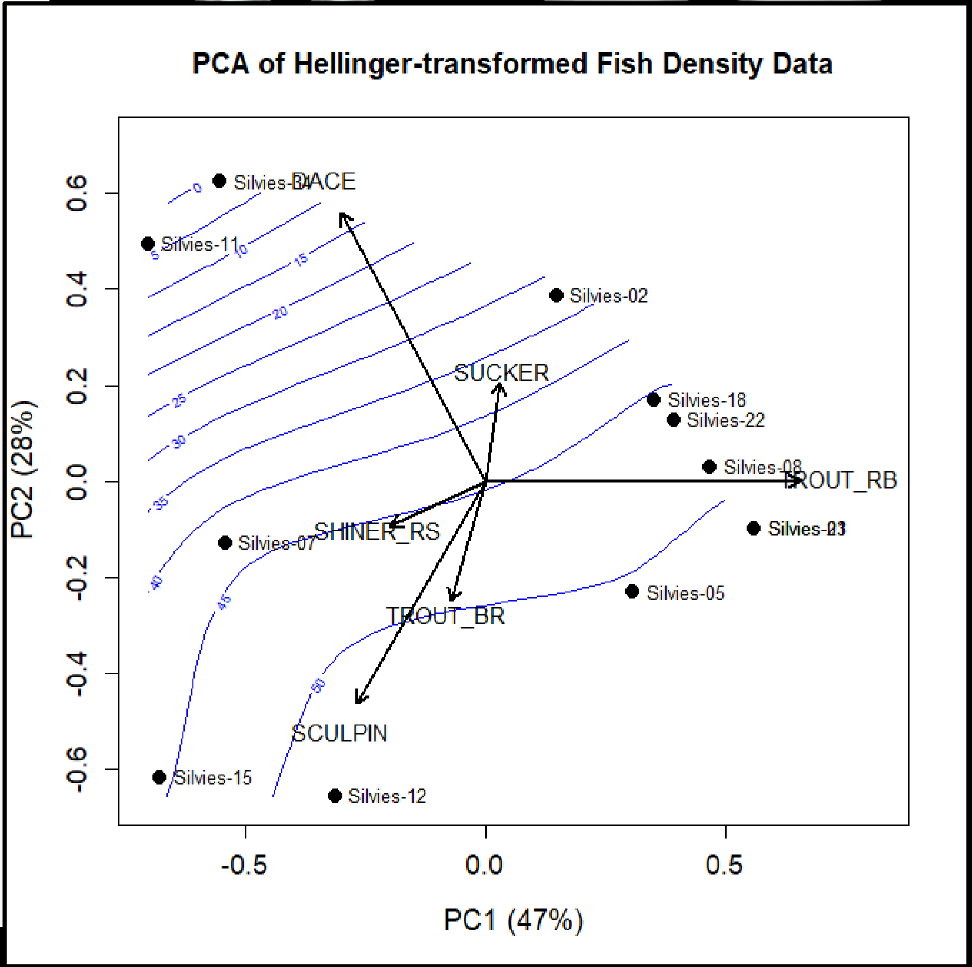
Making Inferences from Ordination: Explanatory

Explanatory data analysis looks for underlying relationships, patterns, and trends within a dataset.

1) **Indirect Comparison:** Treat principal axes/coordinates or clustering partitions as response variables in a regression analysis.

2) **Direct Comparison:** Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA).





Making Inferences from Ordination: Explanatory

Explanatory data analysis looks for underlying relationships, patterns, and trends within a dataset.

1) **Indirect Comparison:** Treat principal axes/coordinates or clustering partitions as response variables in a regression analysis.

2) **Direct Comparison:** Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA).

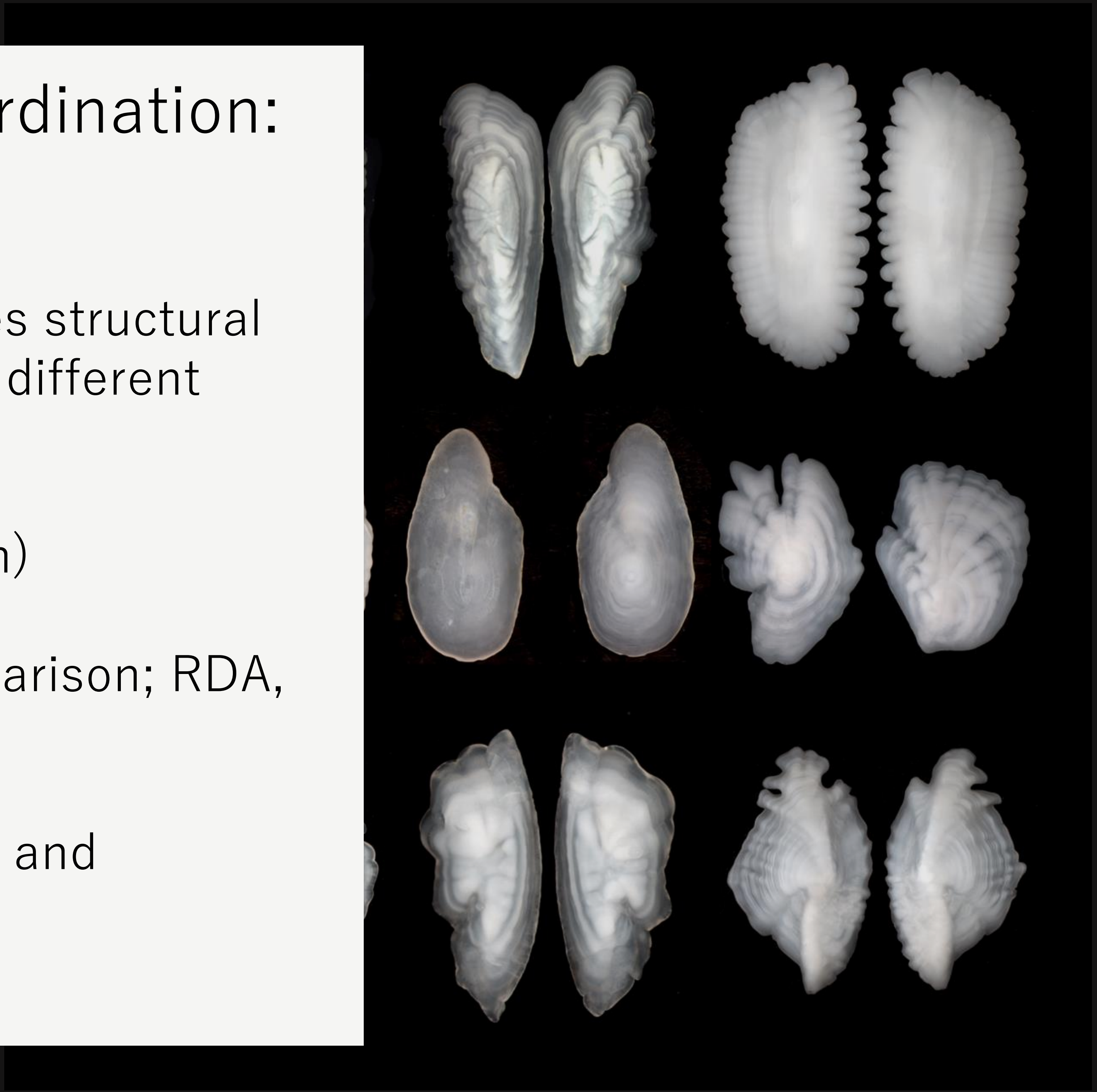
Making Inferences from Ordination: Forecasting

Ecological forecasting extrapolates structural relationships among descriptors to different sites, time periods, etc.

1) Regression (indirect comparison)

2) Canonical analysis (direct comparison; RDA, CCA)

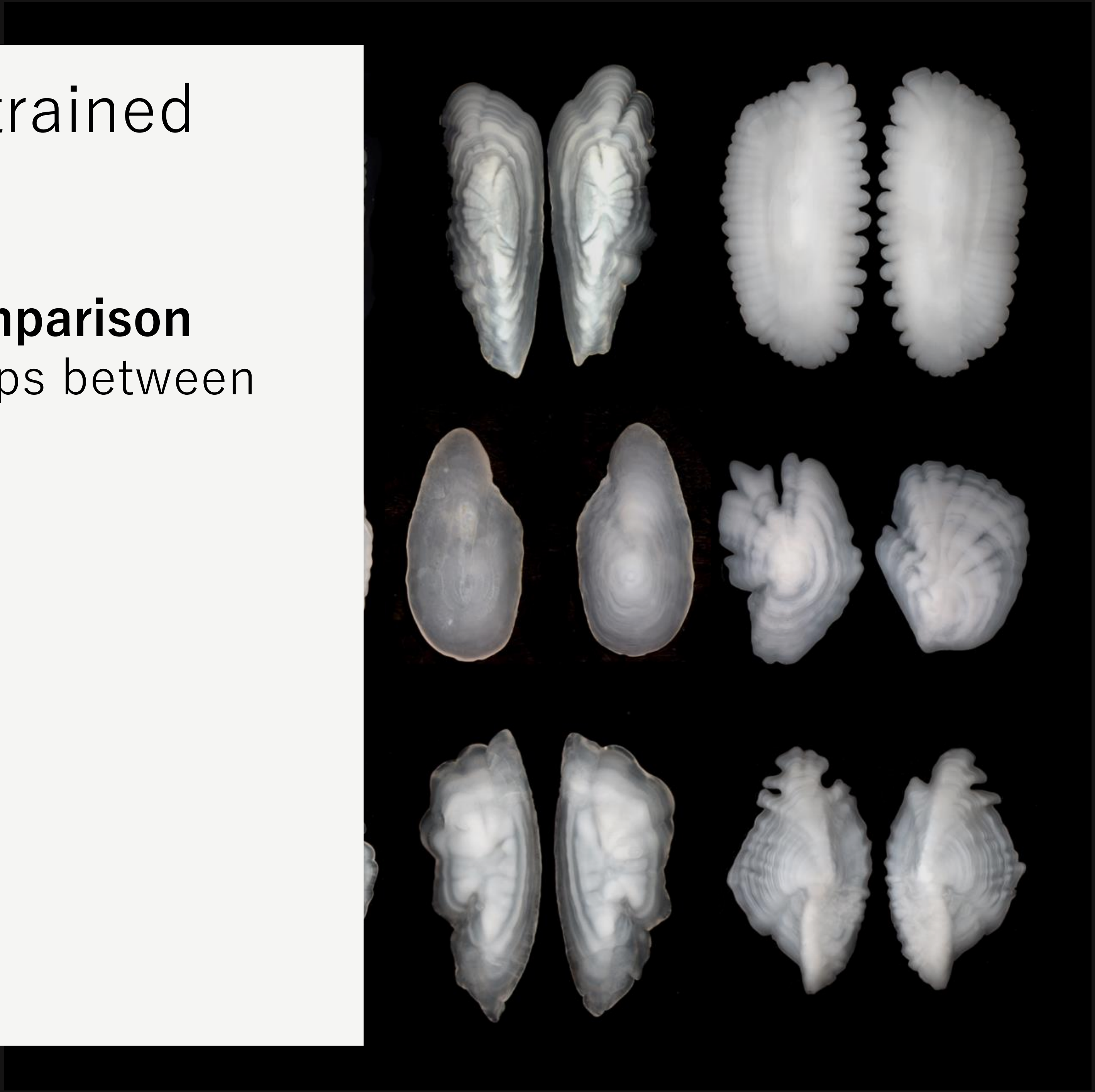
3) Decision analysis: classification and regression trees



Canonical Methods: Constrained Ordination

Canonical Methods: Constrained Ordination

Canonical analysis is a **direct comparison** method used to explore relationships between two sets of variables.



Canonical Methods: Constrained Ordination

Canonical analysis is a **direct comparison** method used to explore relationships between two sets of variables.

- **Symmetric methods** treat both sets of variables equally without assuming a response/predictor relationship
- **Asymmetric methods** test the effect of explanatory variables **X** on response variables **Y**

Canonical Methods: Constrained Ordination

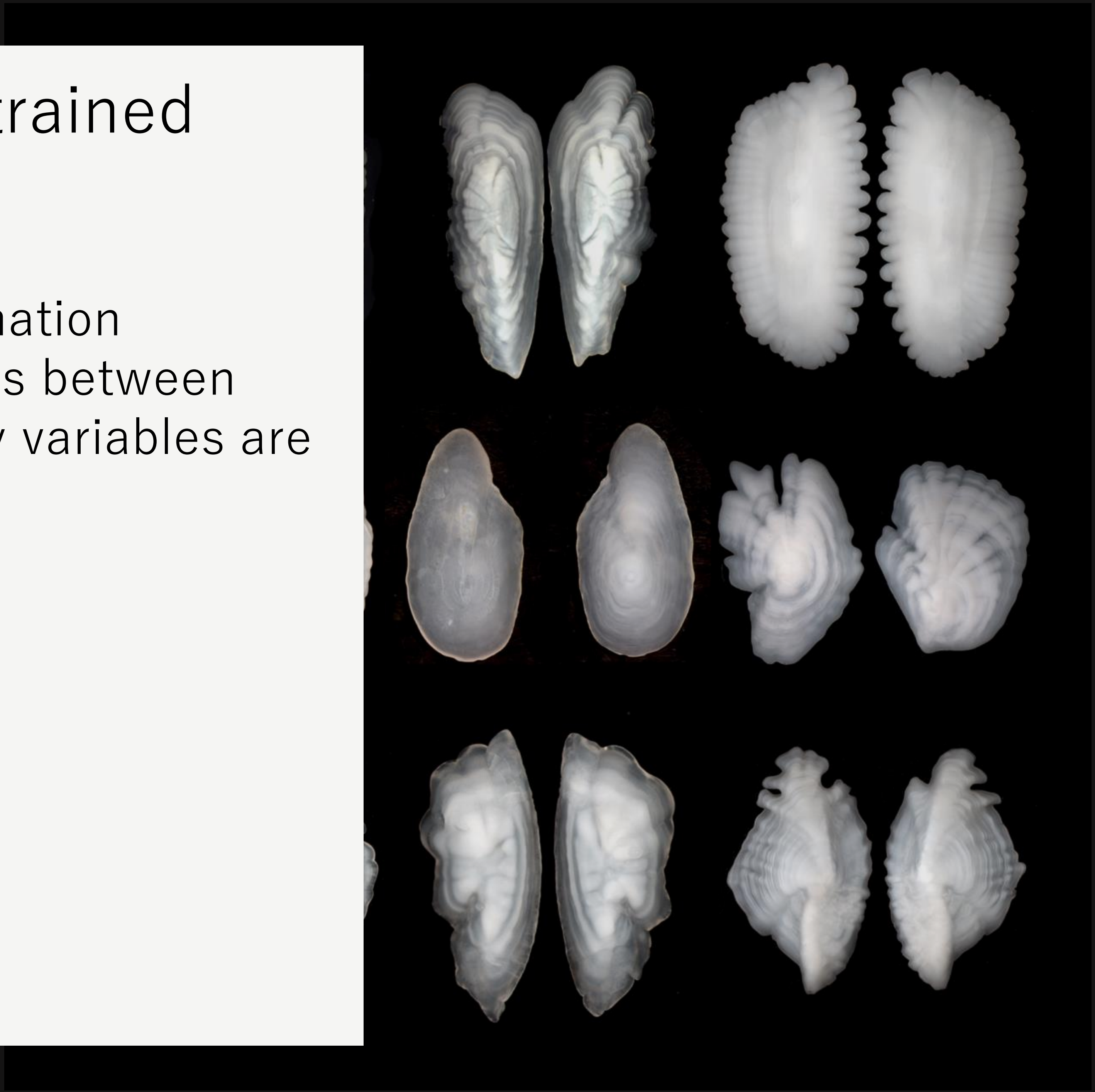
Canonical analysis is a **direct comparison** method used to explore relationships between two sets of variables.

- **Symmetric methods** treat both sets of variables equally without assuming a response/predictor relationship
- **Asymmetric methods** test the effect of explanatory variables **X** on response variables **Y**

Asymmetric methods are usually more appropriate in ecology, where we are testing clear hypotheses.

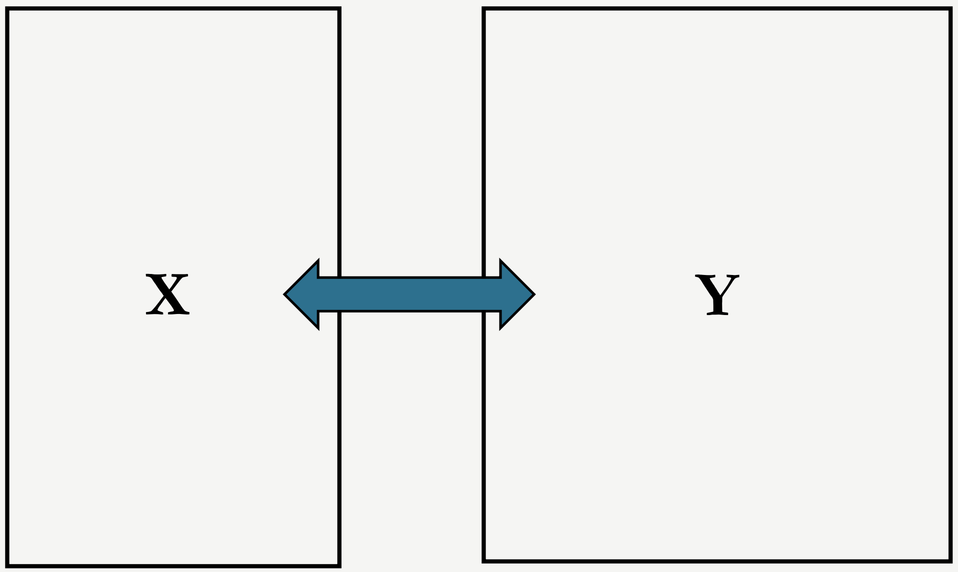
Canonical Methods: Constrained Ordination

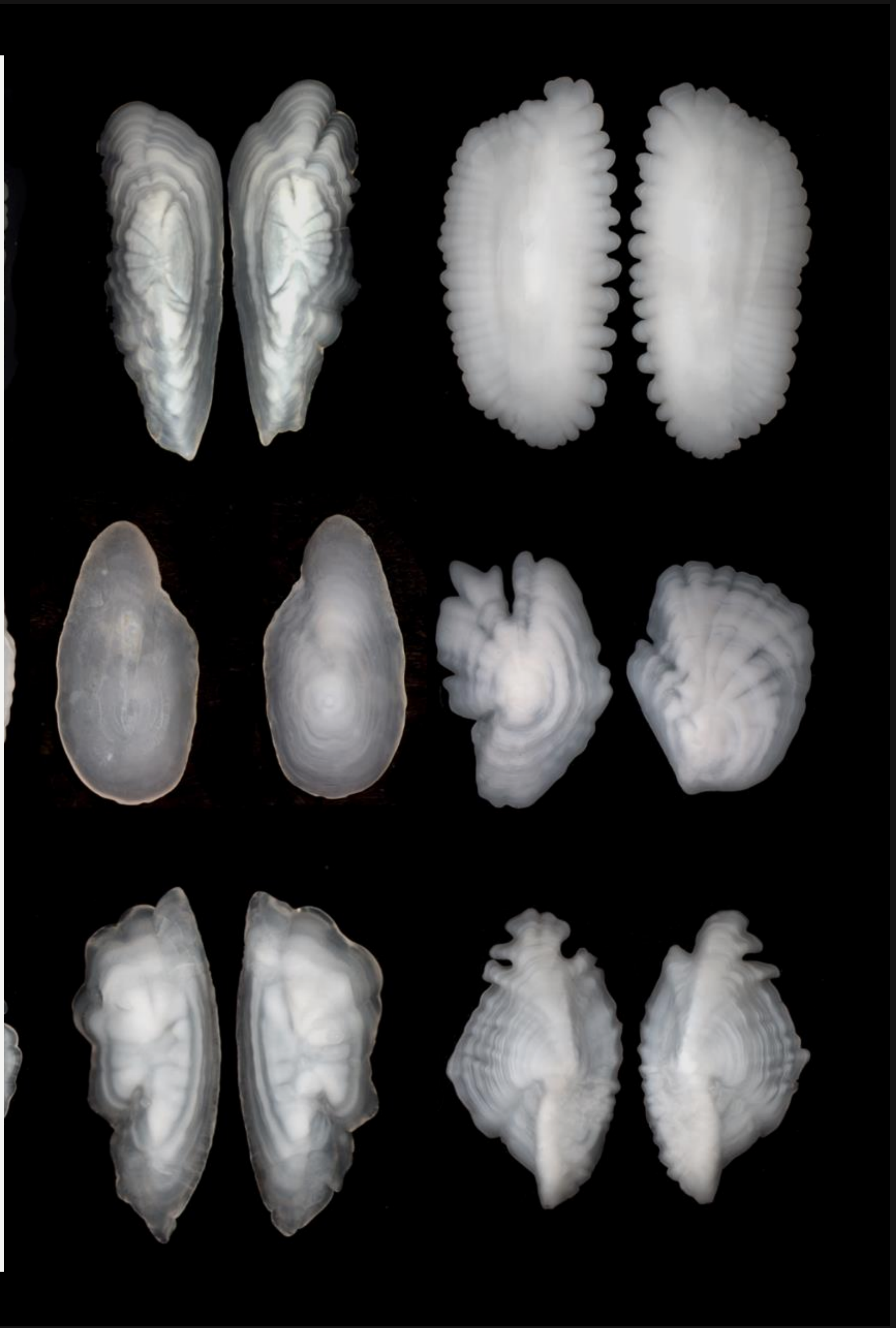
Constrained ordination is an ordination technique in which the relationships between response variables and explanatory variables are explored.



Canonical Methods: Constrained Ordination

Constrained ordination is an ordination technique in which the relationships between response variables and explanatory variables are explored.

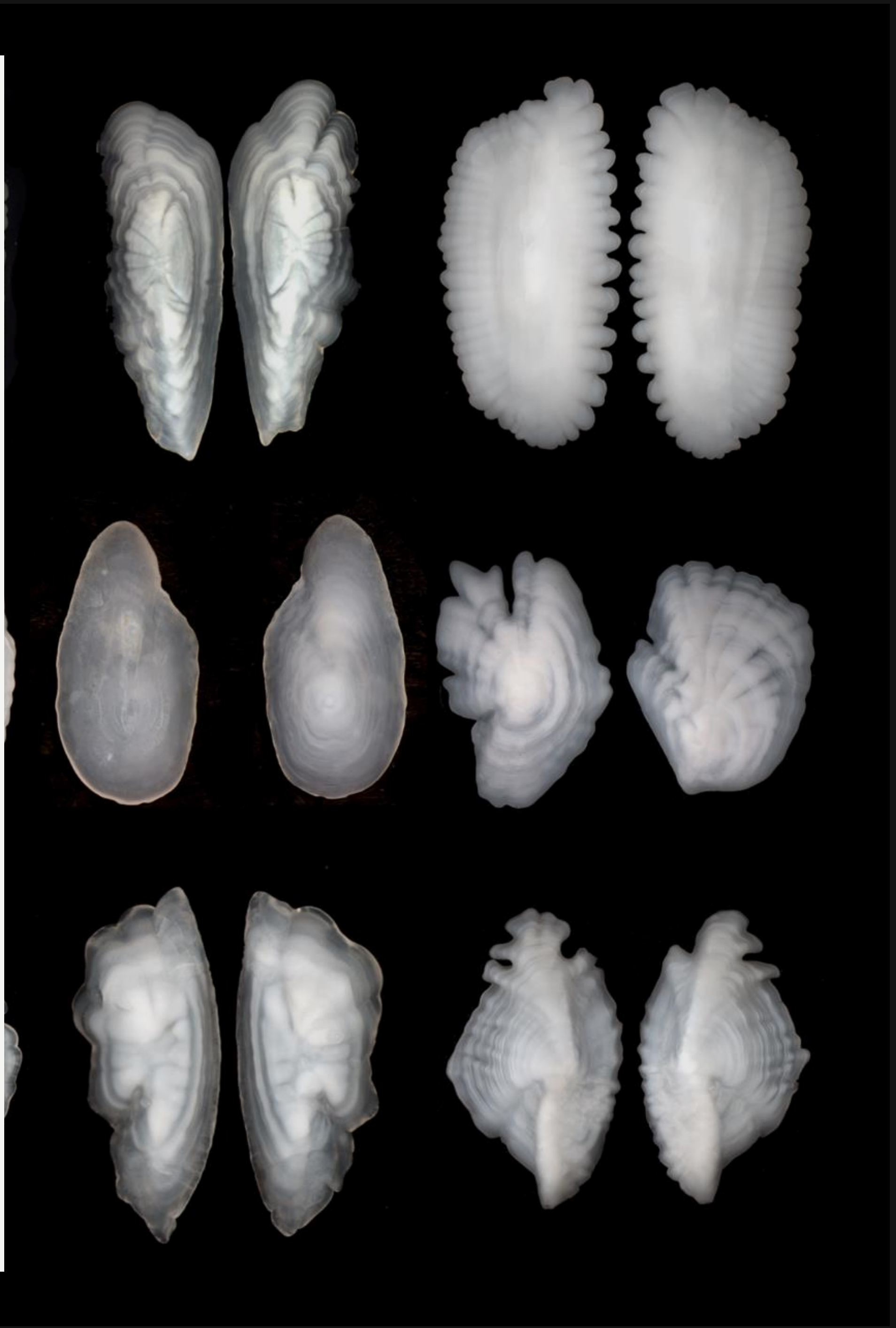




Canonical Methods: Constrained Ordination

Constrained ordination is an ordination technique in which the relationships between response variables and explanatory variables are explored.

Explanatory variables **constrain** or guide the ordination by asking: <u>how much of the variation in a multivariate dataset can be attributed to the explanatory variables?</u>



Redundancy Analysis

Redundancy Analysis: Introduction

Redundancy Analysis (RDA) is the constrained form of principal component analysis (PCA).

Redundancy Analysis: Introduction

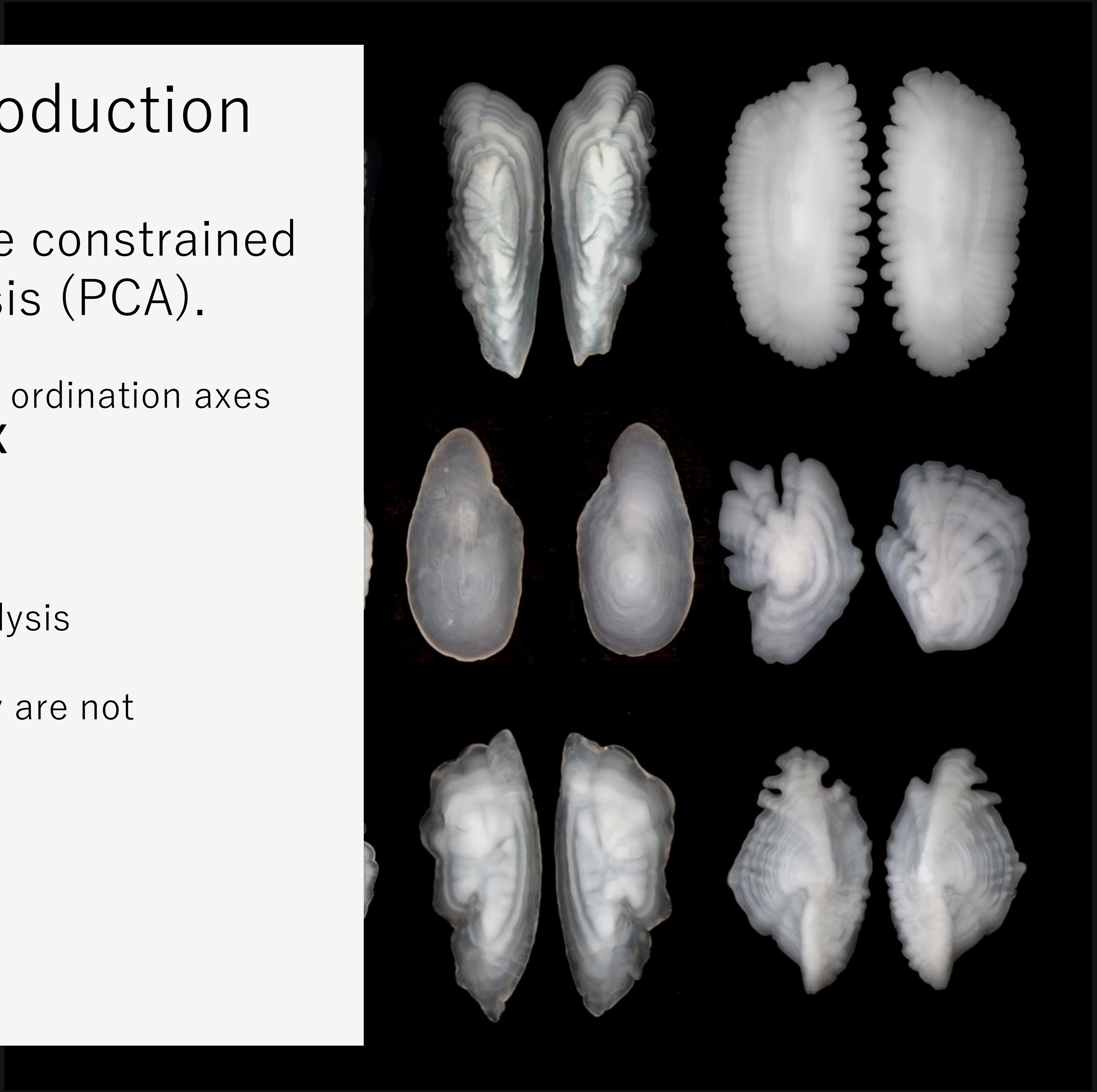
Redundancy Analysis (RDA) is the constrained form of principal component analysis (PCA).

“Redundancy” = “explained variance”

Redundancy Analysis: Introduction

Redundancy Analysis (RDA) is the constrained form of principal component analysis (PCA).

- The ordination of **Y** is constrained such that ordination axes are **linear** combinations of the variables in **X**
- Preserves Euclidean distances
- The data in **Y** must be centered prior to analysis
- The data in **Y** should be standardized if they are not dimensionally homogeneous



Redundancy Analysis: Introduction

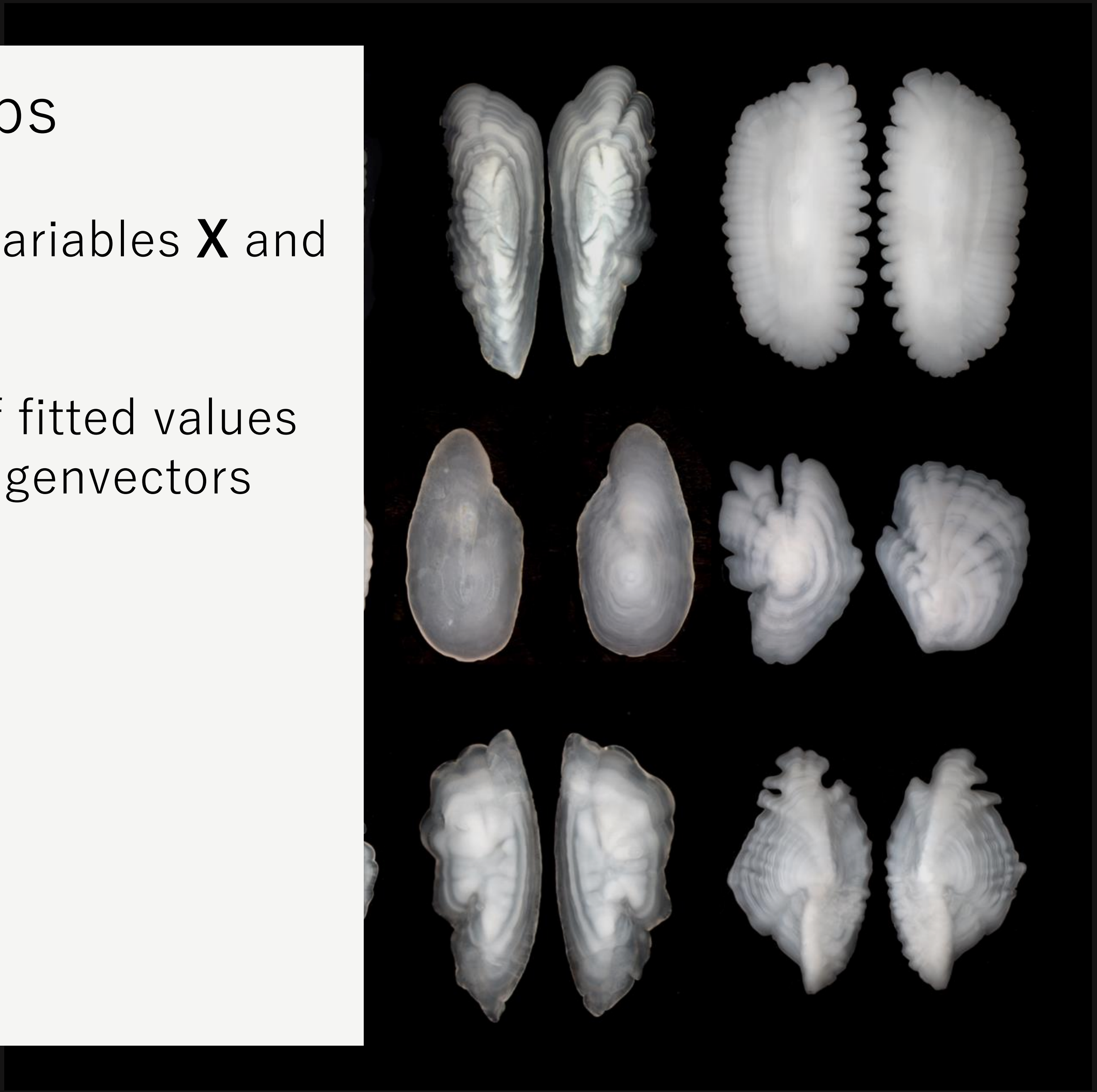
Redundancy Analysis (RDA) is the constrained form of principal component analysis (PCA).

- The ordination of **Y** is constrained such that ordination axes are **linear** combinations of the variables in **X**
- Preserves Euclidean distances
- The data in **Y** must be centered prior to analysis
- The data in **Y** should be standardized if they are not dimensionally homogeneous
- Matrix **X** can be any data type, but variables should be checked for collinearity first!

Redundancy Analysis: Steps

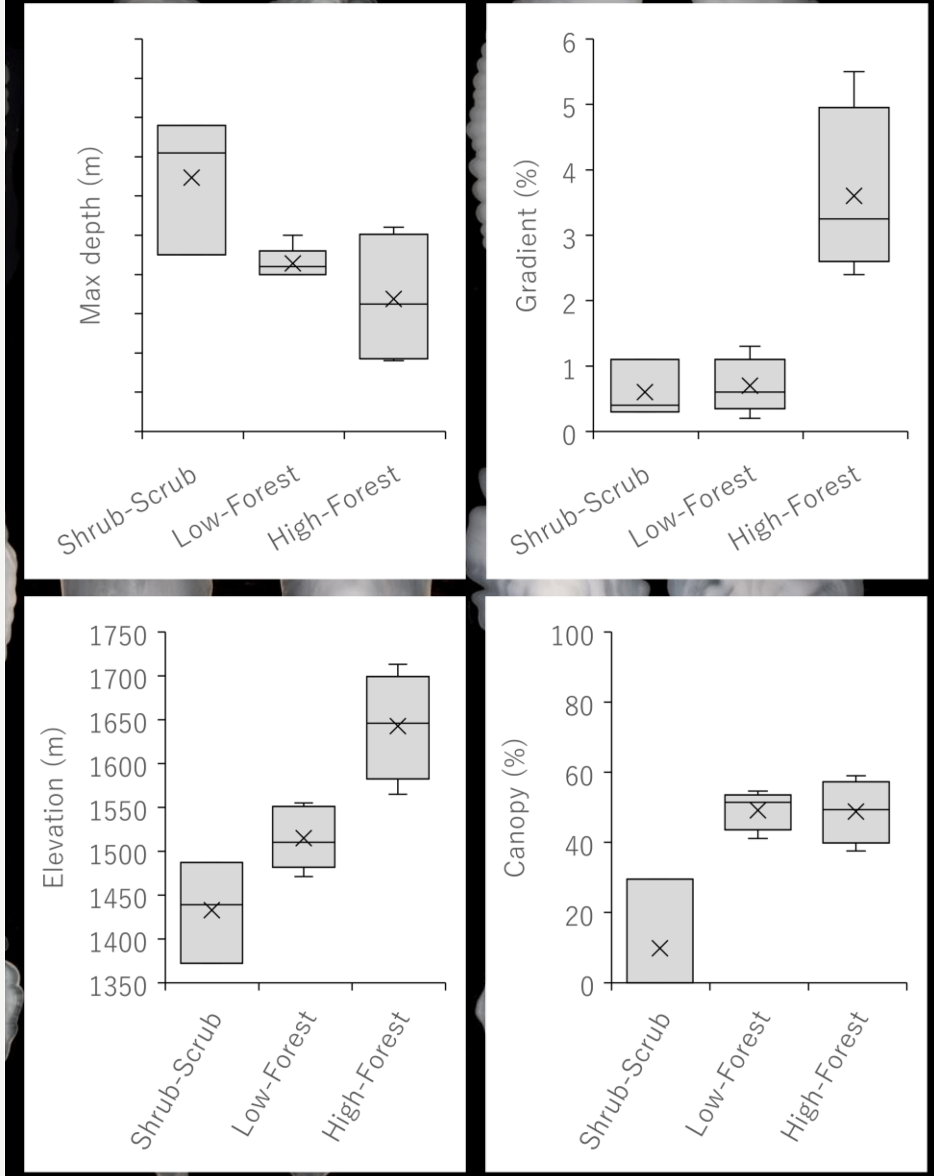
1) Regress each variable **Y** on all variables **X** and compute fitted values

2) Carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors



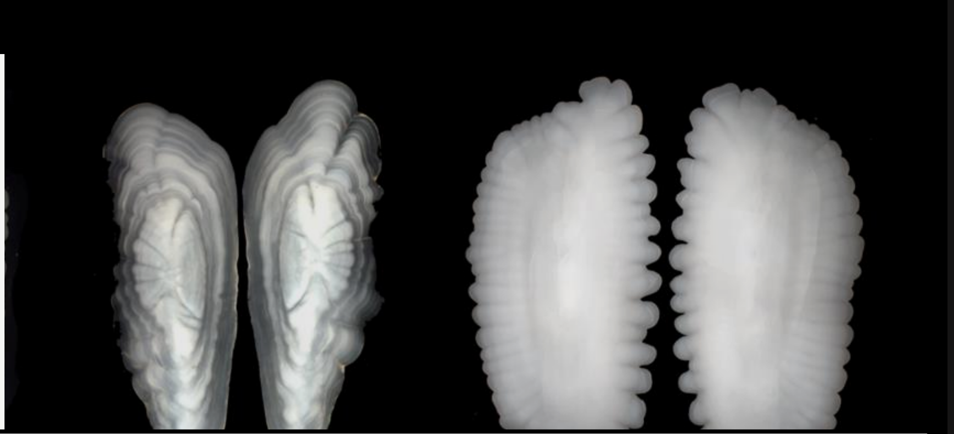
Redundancy Analysis: Steps

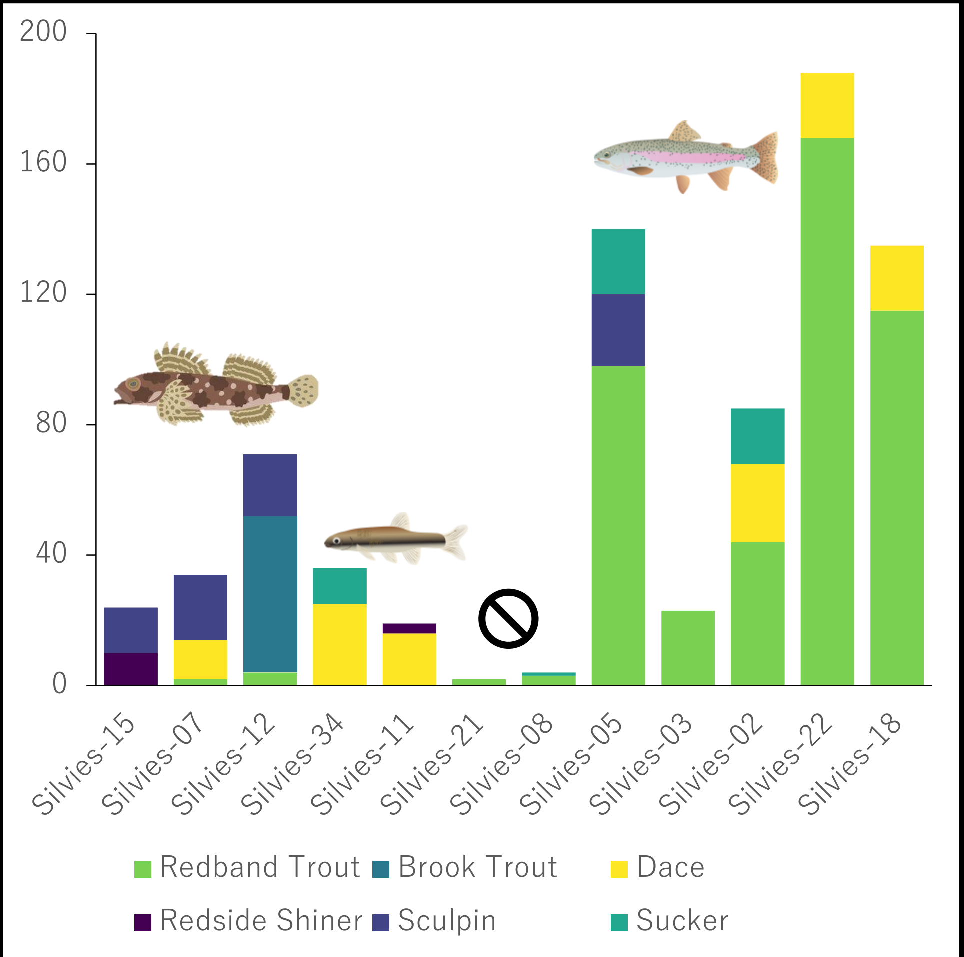
Site ID	Max Depth (m)	Gradient (%)	Elevation (m)	Canopy (%)
Silvies-11	0.45	0.3	1439	0.0
Silvies-34	0.78	1.1	1487	0.0
Silvies-02	0.71	0.4	1372	29.6
Silvies-15	0.40	0.2	1471	41.1
Silvies-07	0.50	1.3	1547	52.3
Silvies-08	0.40	0.6	1492	51.4
Silvies-22	0.42	0.9	1555	54.7
Silvies-18	0.42	0.5	1510	46.2
Silvies-12	0.52	3.2	1658	51.9
Silvies-21	0.18	2.4	1713	37.5
Silvies-05	0.45	5.5	1565	46.7
Silvies-03	0.20	3.3	1634	59.0



Redundancy Analysis: Steps

Site ID	Redband Trout	Brook Trout	Dace	Redside Shiner	Sculpin	Sucker
Silvies-15	0	0	0	10	14	0
Silvies-07	2	0	12	0	20	0
Silvies-12	4	48	0	0	19	0
Silvies-34	0	0	25	0	0	11
Silvies-11	0	0	16	3	0	0
Silvies-21	2	0	0	0	0	0
Silvies-08	3	0	0	0	0	1
Silvies-05	98	0	0	0	22	20
Silvies-03	23	0	0	0	0	0
Silvies-02	44	0	24	0	0	17
Silvies-22	168	0	20	0	0	0
Silvies-18	115	0	20	0	0	0

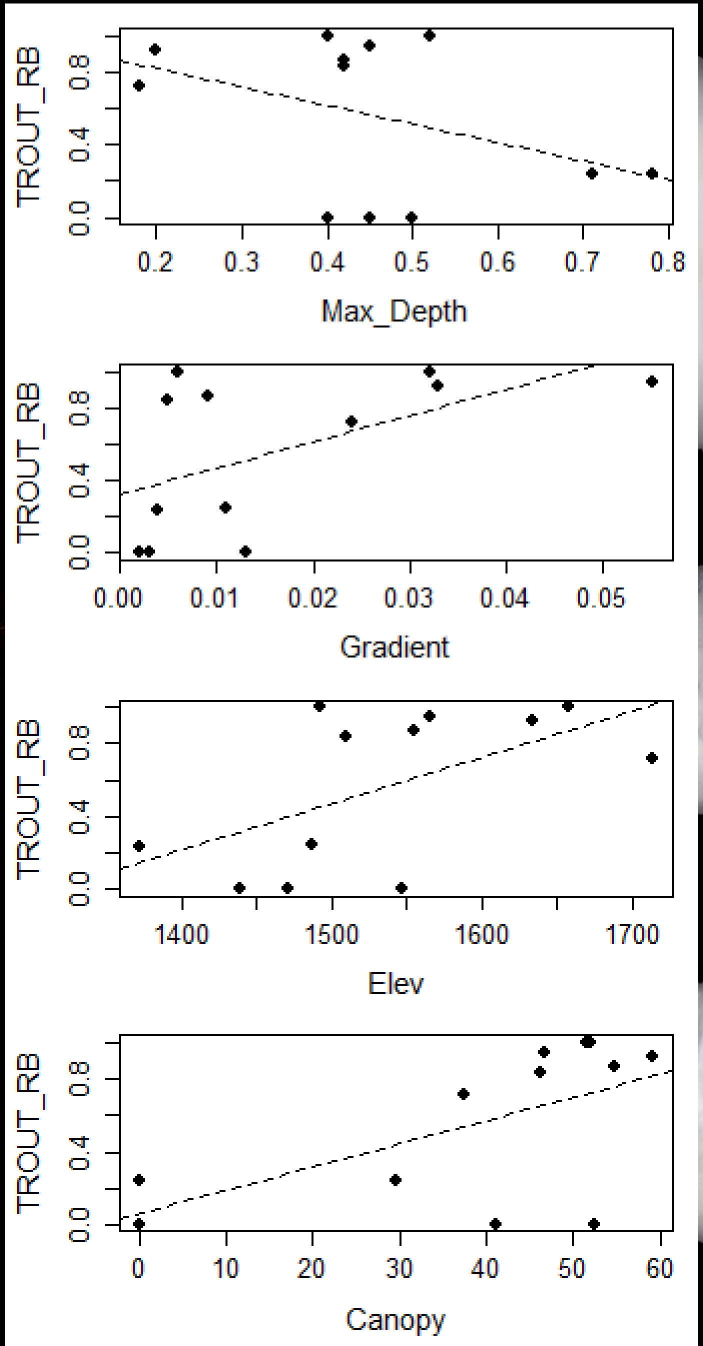




Redundancy Analysis: Steps

1) Regress each variable **Y** on all variables **X** and compute fitted values

$$\mathbf{y}_i = b_0 + b_1\mathbf{x}_{Depth} + b_2\mathbf{x}_{Grad} + b_3\mathbf{x}_{Elev} + b_4\mathbf{x}_{Canopy} + \varepsilon_i$$

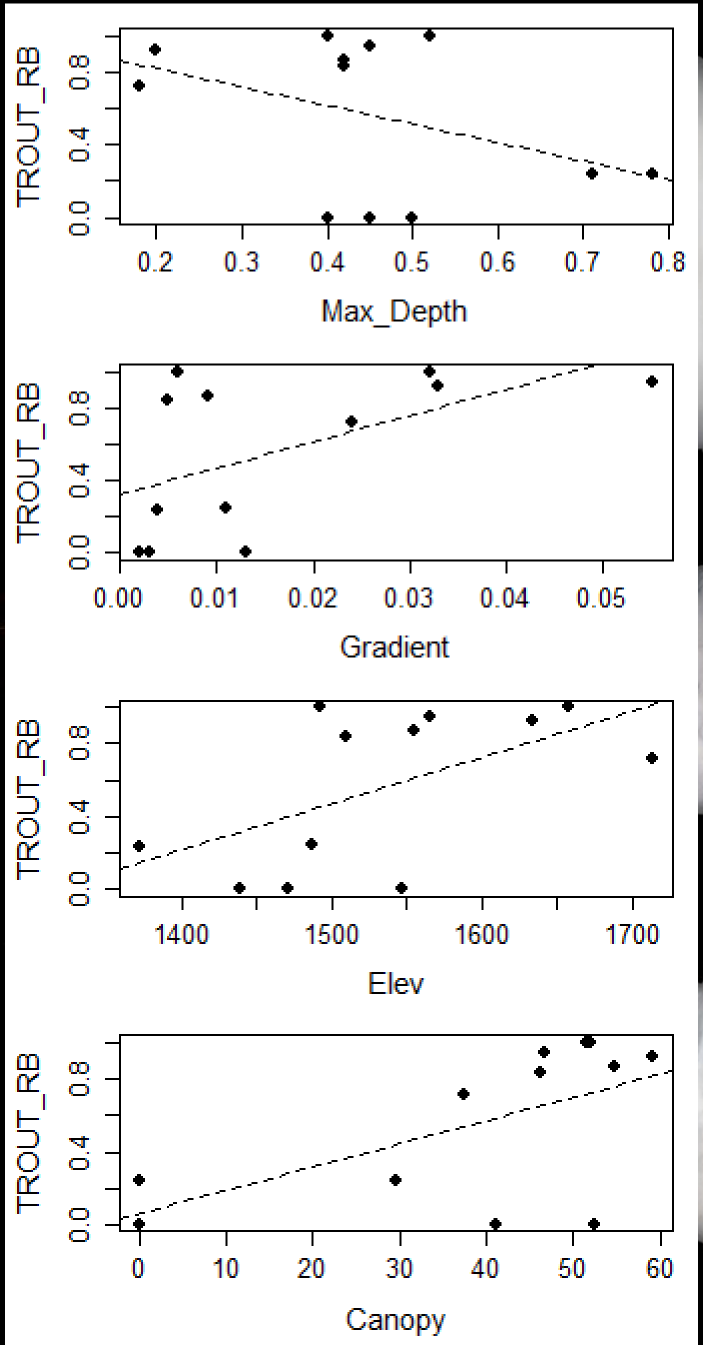


Redundancy Analysis: Steps

1) Regress each variable **Y** on all variables **X** and compute fitted values

$$\mathbf{y}_i = b_0 + b_1\mathbf{x}_{Depth} + b_2\mathbf{x}_{Grad} + b_3\mathbf{x}_{Elev} + b_4\mathbf{x}_{Canopy} + \varepsilon_i$$

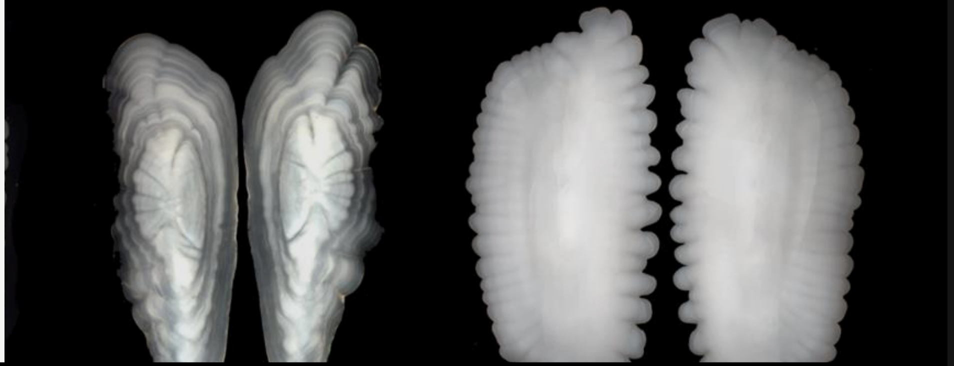
$$\hat{\mathbf{Y}} = \mathbf{X}[\mathbf{X}^{`}\mathbf{X}]^{-1}\mathbf{X}^{`}\mathbf{Y}$$

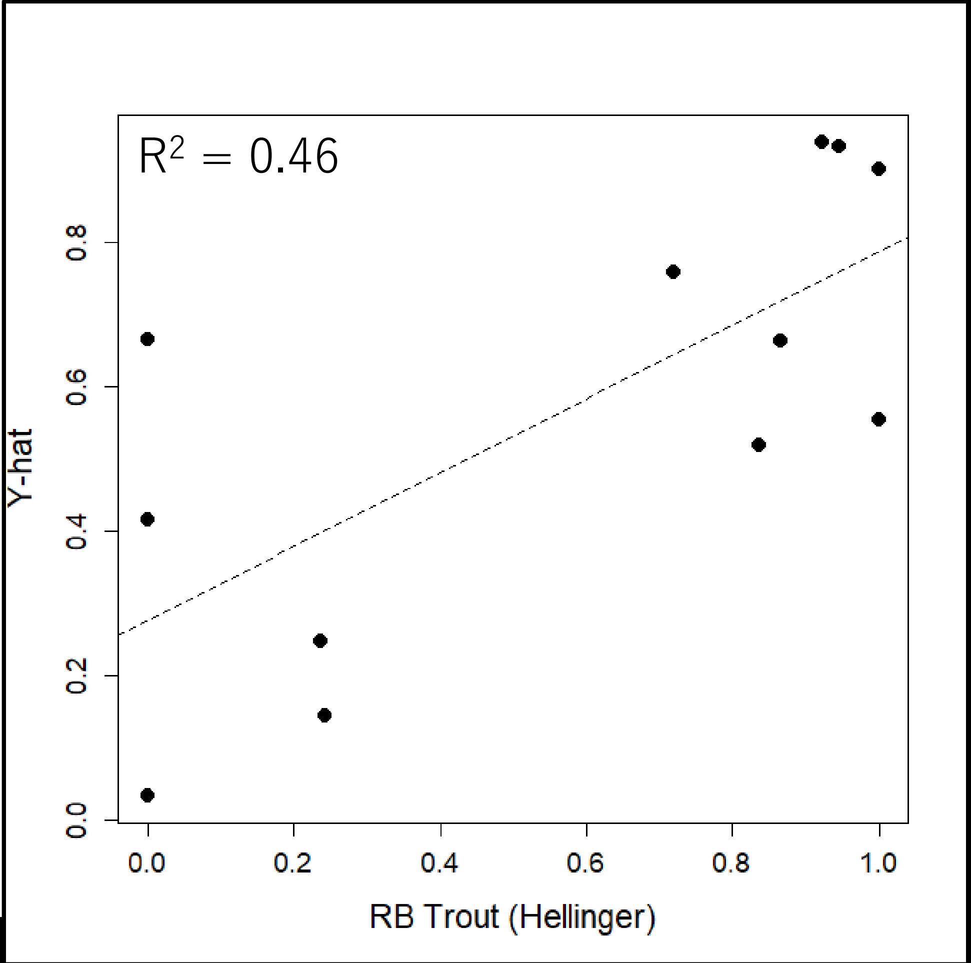


Redundancy Analysis: Steps

1) Regress each variable **Y** on all variables **X** and compute fitted values

Site ID	**RB Trout (Hellinger)**	**RB Trout ($\hat{Y}$)**
Silvies-15	**0.00**	**0.03**
Silvies-07	**0.24**	**0.15**
Silvies-12	**0.24**	**0.25**
Silvies-34	**0.00**	**0.42**
Silvies-11	**0.00**	**0.67**
Silvies-21	**1.00**	**0.55**
Silvies-08	**0.87**	**0.66**
Silvies-05	**0.84**	**0.52**
Silvies-03	**1.00**	**0.90**
Silvies-02	**0.72**	**0.76**
Silvies-22	**0.95**	**0.93**
Silvies-18	**0.92**	**0.94**





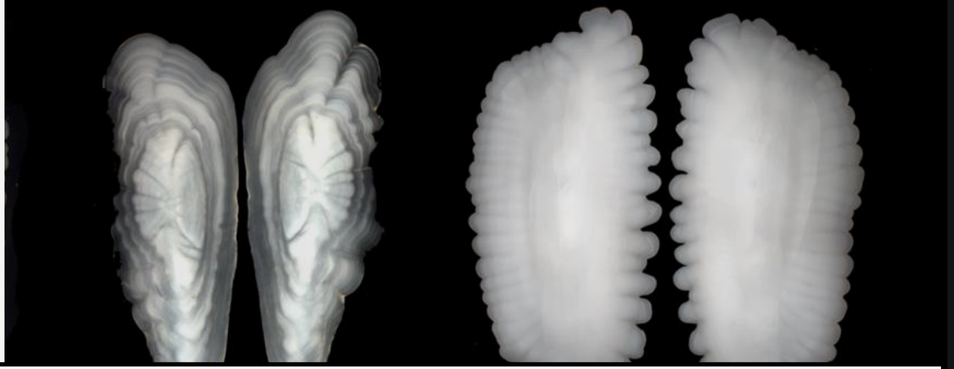
Redundancy Analysis: Steps

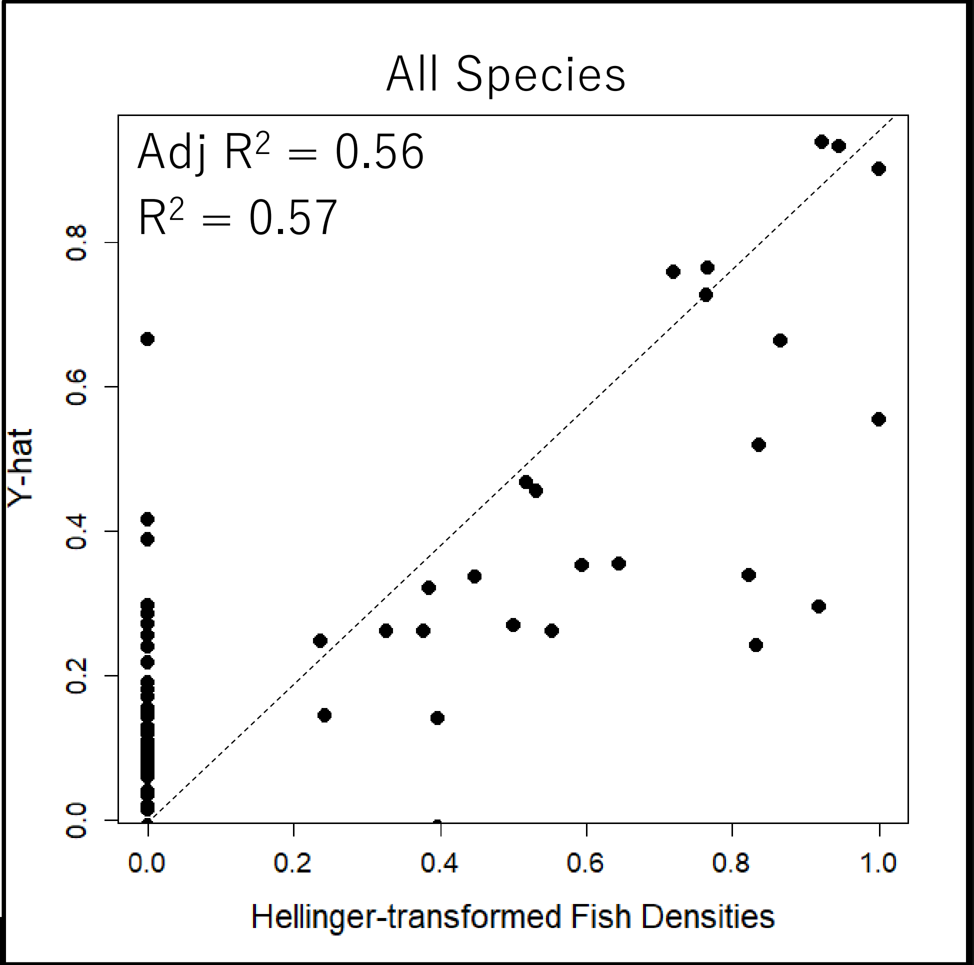
1) Regress each variable **Y** on all variables **X** and compute fitted values

The relationship between **Y** and **Ŷ** is called the **canonical R²** or the **bimultivariate redundancy statistic**.

$$\mathbf{R^2} = SS(\hat{\mathbf{Y}})/SS(\mathbf{Y})$$

It is <u>the proportion of the variation in **Y** explained by a linear model of the variables in **X**</u>.

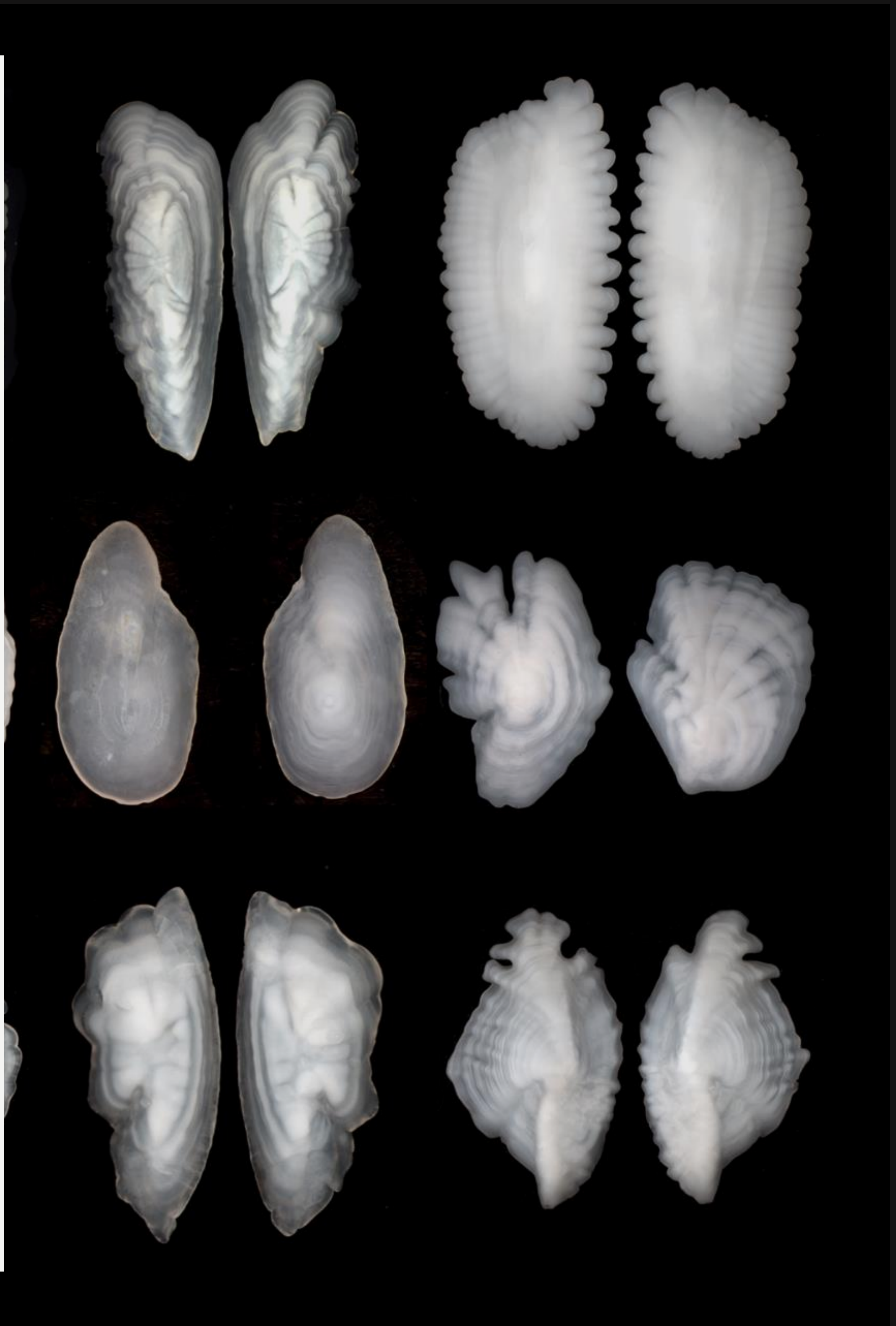




Redundancy Analysis: Steps

2) Carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors

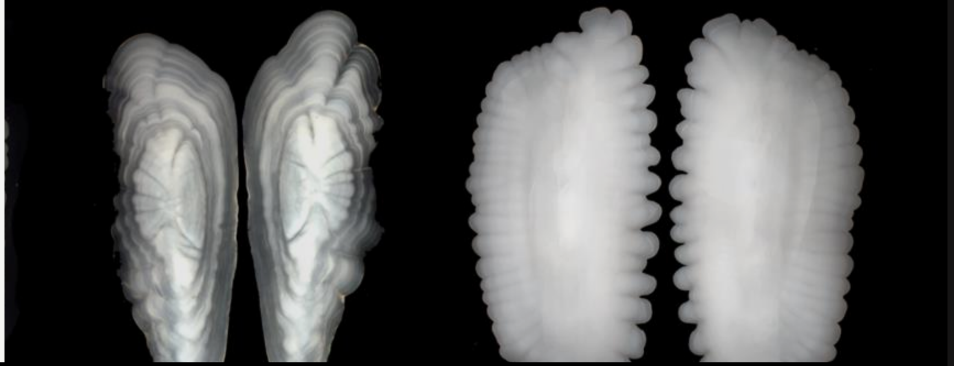
Site ID	Redband Trout	Brook Trout	Dace	Redside Shiner	Sculpin	Sucker
Silvies-15	0.03	0.08	0.28	0.35	0.73	0.15
Silvies-07	0.15	0.15	0.35	0.17	0.76	-0.03
Silvies-12	0.25	0.34	0.18	0.11	0.47	0.02
Silvies-34	0.42	0.12	0.24	0.14	0.22	0.26
Silvies-11	0.67	0.10	0.29	-0.01	0.06	0.19
Silvies-21	0.55	0.12	0.24	0.08	0.08	0.26
Silvies-08	0.66	0.06	0.30	0.01	0.01	0.27
Silvies-05	0.52	0.09	0.27	0.08	0.14	0.26
Silvies-03	0.90	-0.01	0.39	-0.08	-0.01	0.09
Silvies-02	0.76	-0.26	0.46	0.13	0.07	0.34
Silvies-22	0.93	0.11	0.26	0.01	0.06	-0.15
Silvies-18	0.94	-0.08	0.32	0.04	-0.15	0.22

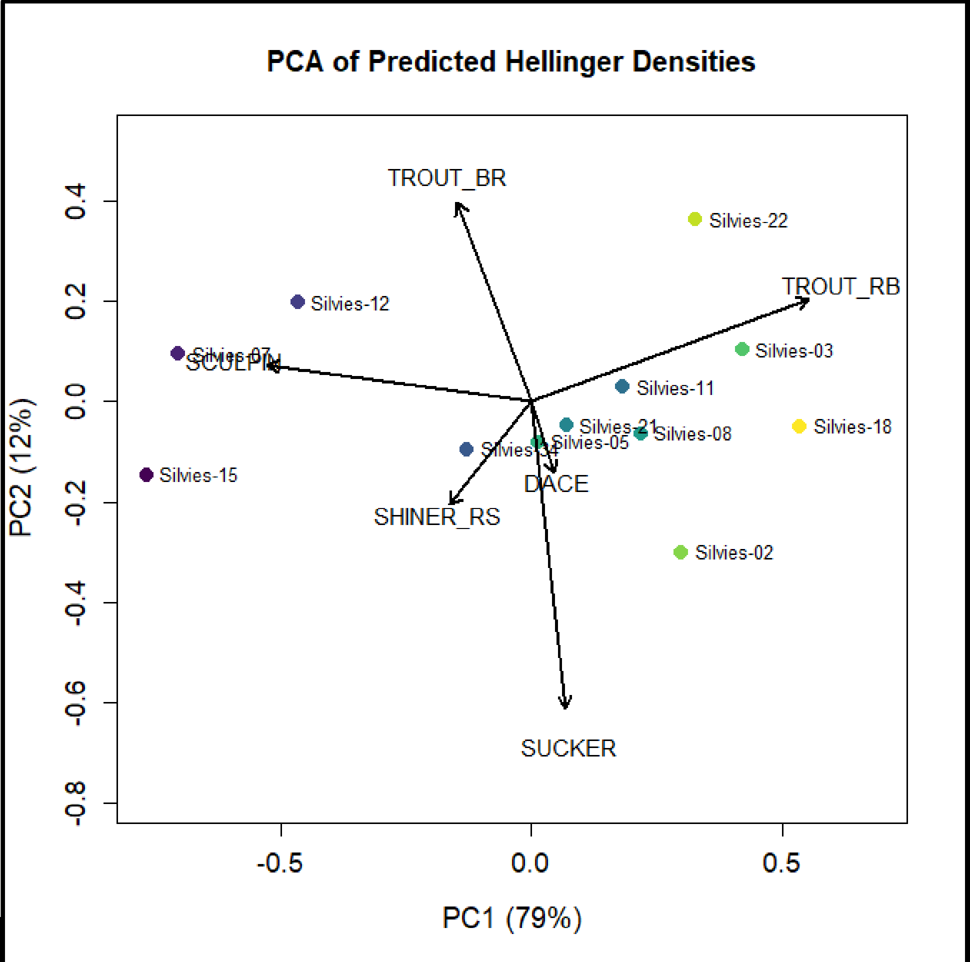


Redundancy Analysis: Steps

2) Carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors

Site ID	Redband Trout	Brook Trout	Dace	Redside Shiner	Sculpin	Sucker
Silvies-15	0.03	0.08	0.28	0.35	0.73	0.15
Silvies-07	0.15	0.15	0.35	0.17	0.76	-0.03
Silvies-12	0.25	0.34	0.18	0.11	0.47	0.02
Silvies-34	0.42	0.12	0.24	0.14	0.22	0.26
Silvies-11	0.67	0.10	0.29	-0.01	0.06	0.19
Silvies-21	0.55	0.12	0.24	0.08	0.08	0.26
Silvies-08	0.66	0.06	0.30	0.01	0.01	0.27
Silvies-05	0.52	0.09	0.27	0.08	0.14	0.26
Silvies-03	0.90	-0.01	0.39	-0.08	-0.01	0.09
Silvies-02	0.76	-0.26	0.46	0.13	0.07	0.34
Silvies-22	0.93	0.11	0.26	0.01	0.06	-0.15
Silvies-18	0.94	-0.08	0.32	0.04	-0.15	0.22

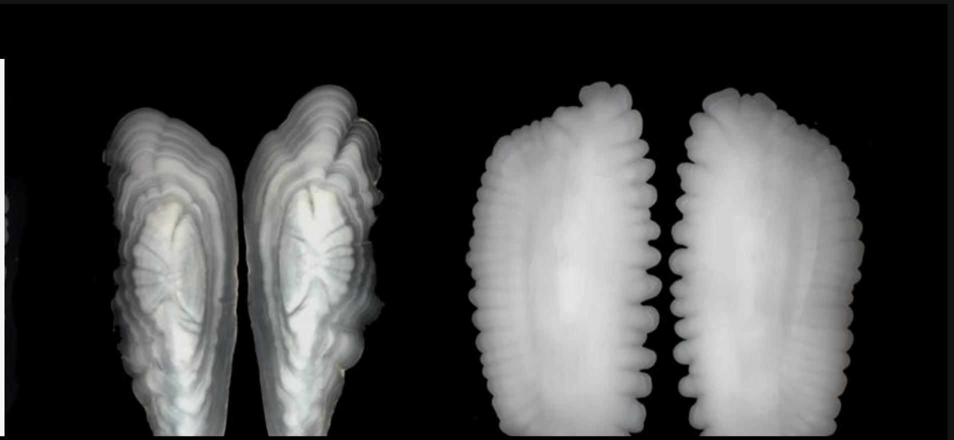


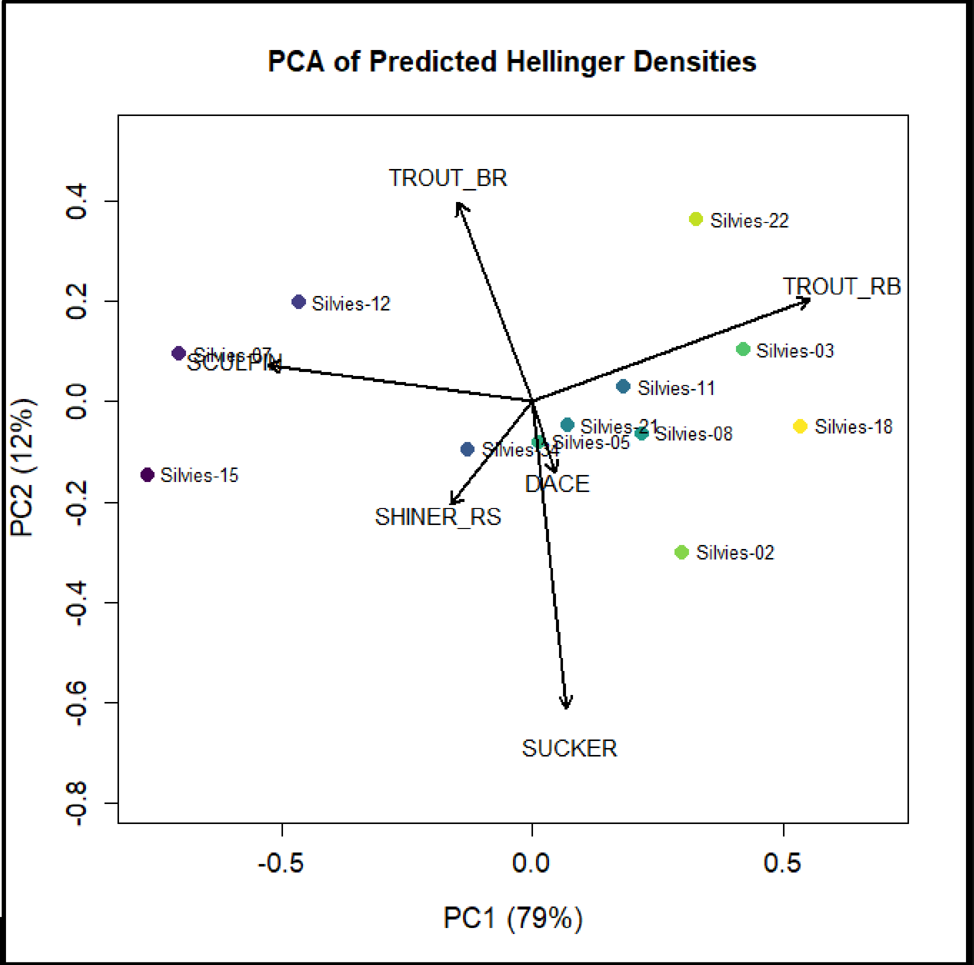


Redundancy Analysis: Steps

2) Carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors

This PCA produces the **canonical** eigenvalues and eigenvectors, and **canonical** axes.



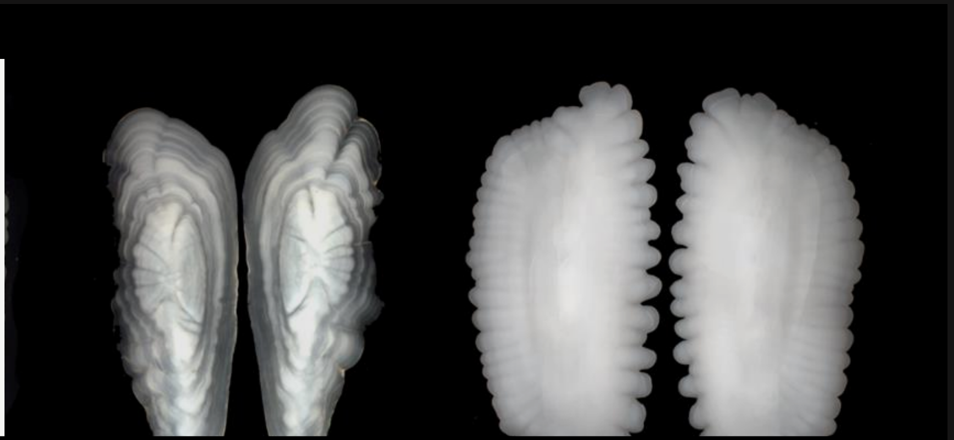


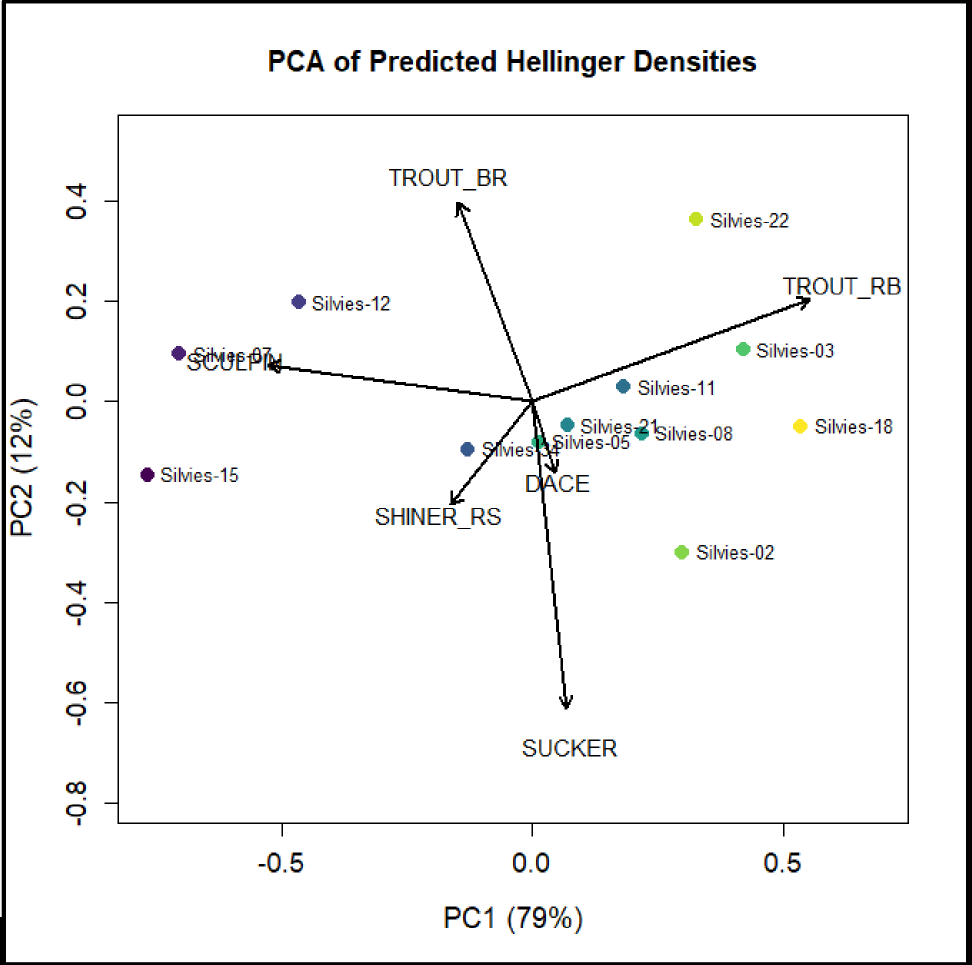
Redundancy Analysis: Steps

2) Carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors

This PCA produces the **canonical** eigenvalues and eigenvectors, and **canonical** axes.

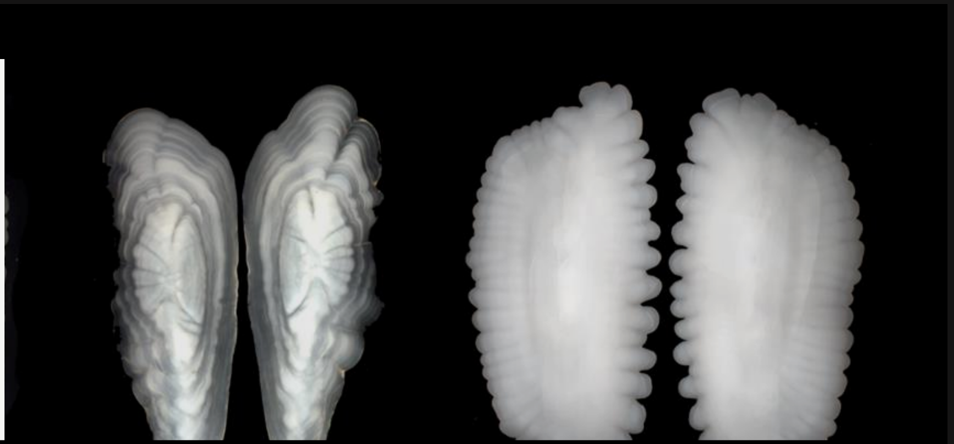
The canonical axes are linear combinations of the explanatory variables in **X**.

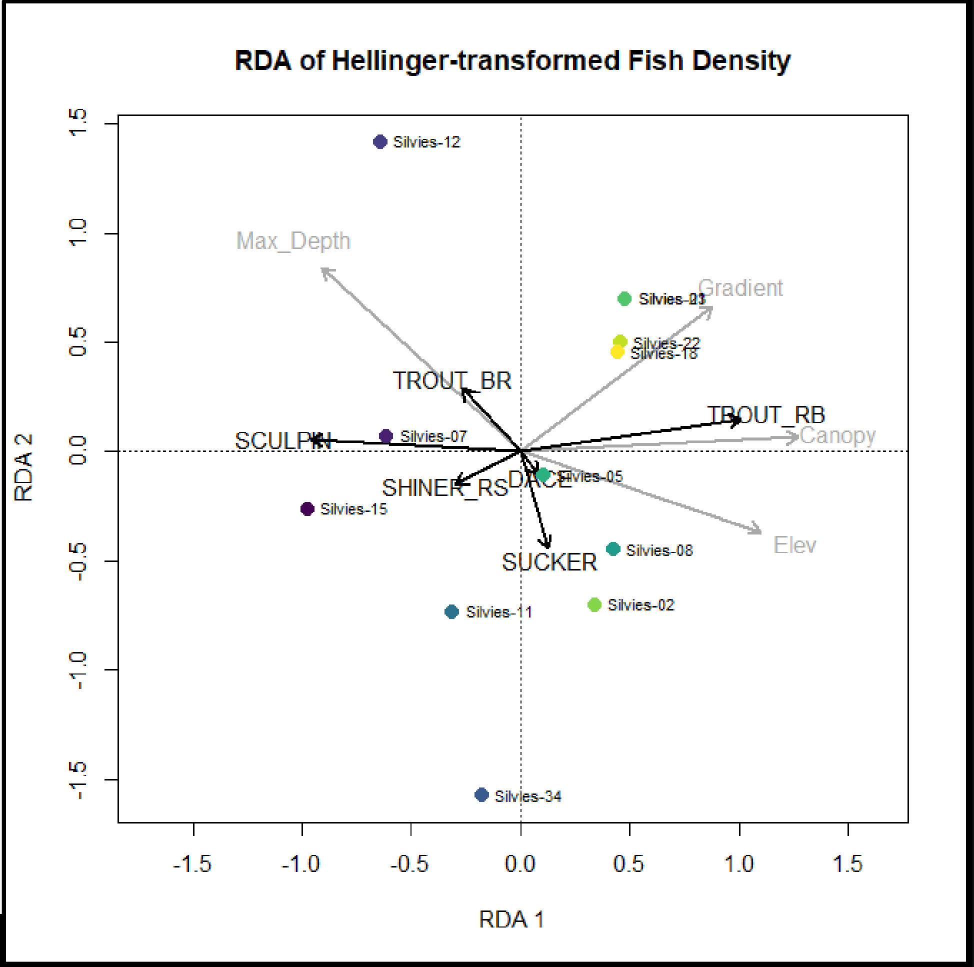




Redundancy Analysis: Interpretation

RDA **triplots** include: response variables, objects, and explanatory variables

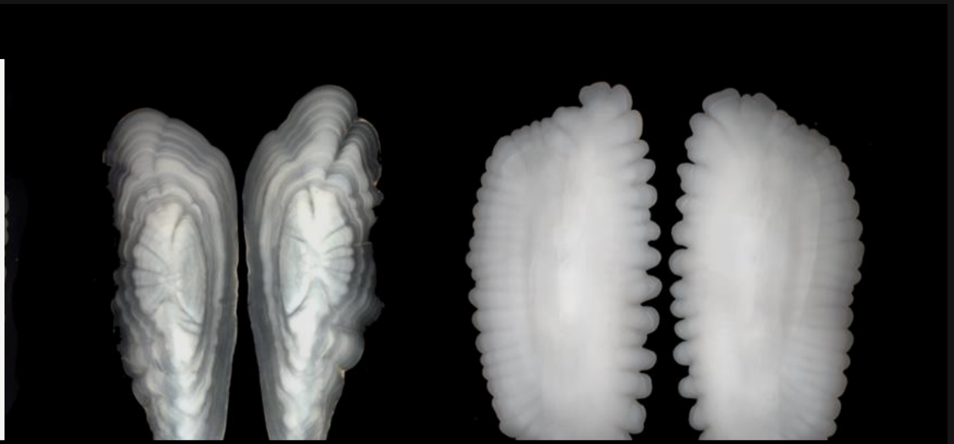


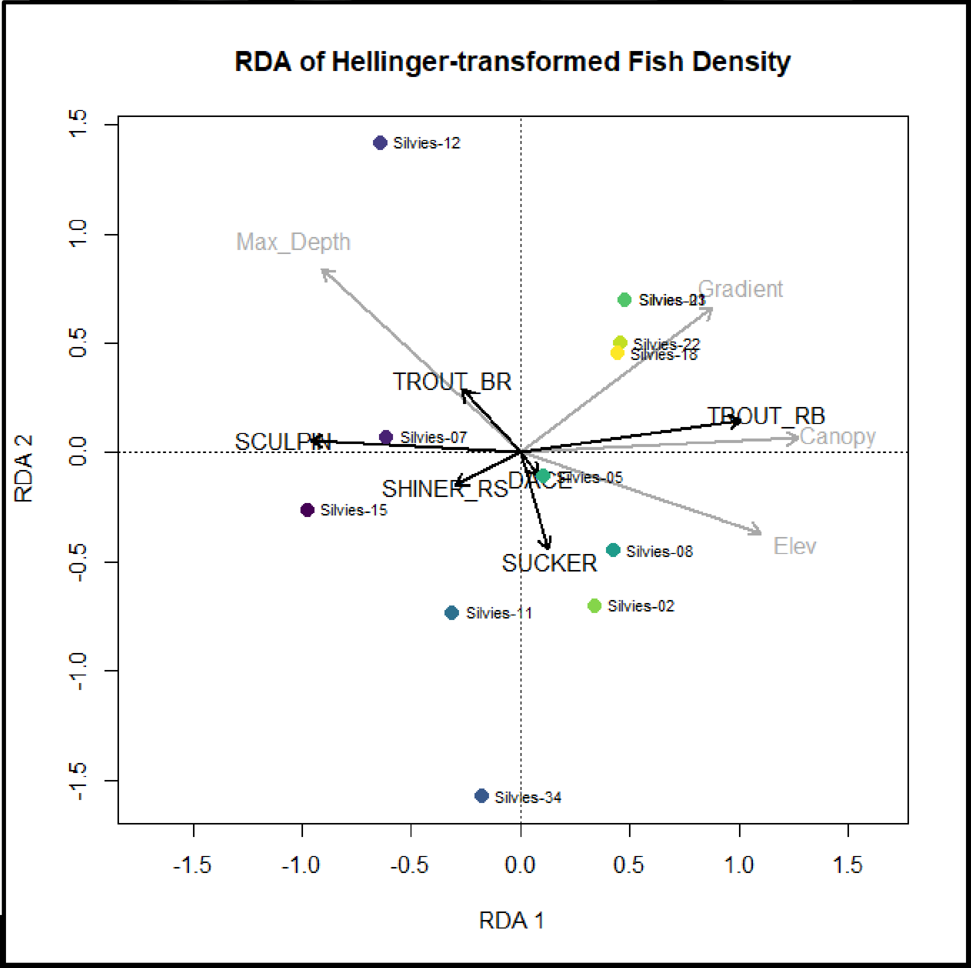


Redundancy Analysis: Interpretation

RDA **triplots** include: response variables, objects, and explanatory variables

Site scores can be obtained from the observed data OR the fitted data (**fitted site scores**).

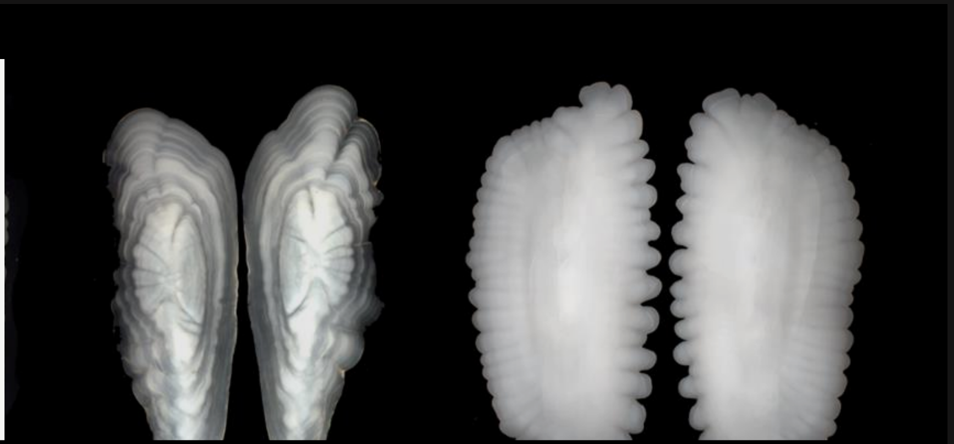


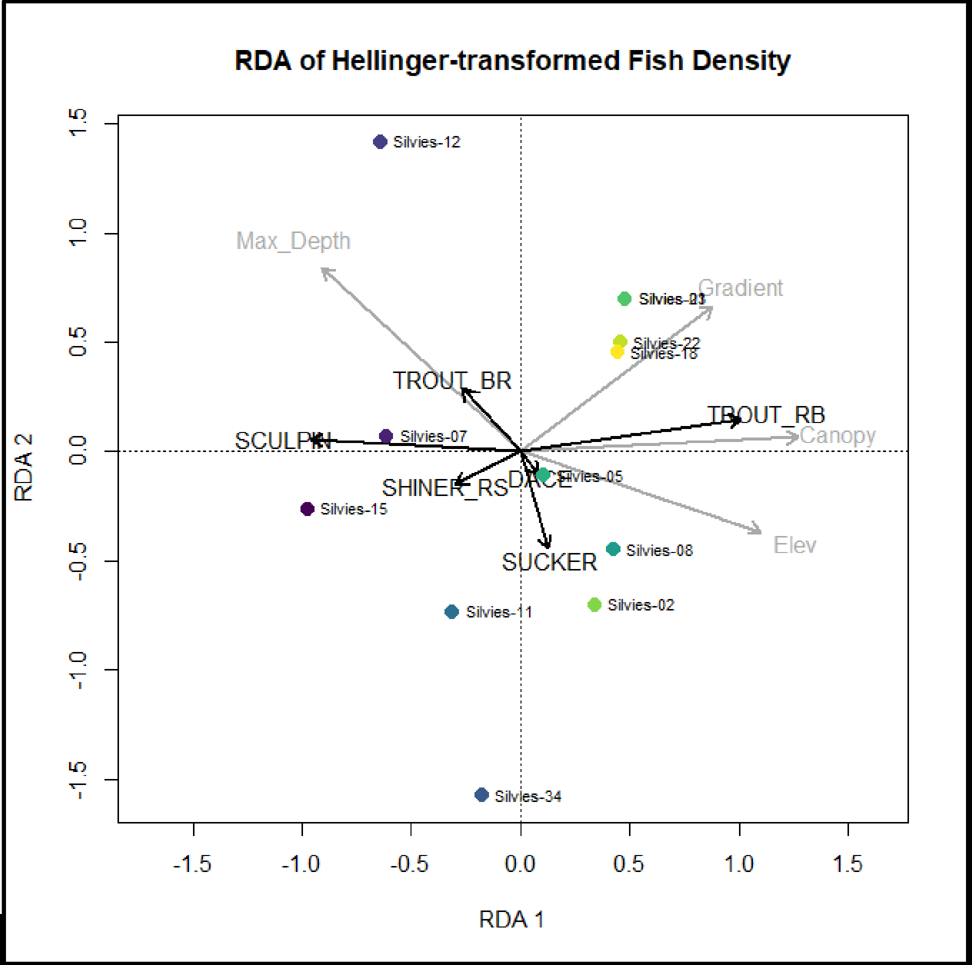


Redundancy Analysis: Interpretation

Scaling 1) Shows similarities among objects in the response matrix, a.k.a "**distance triplot**"

Scaling 2) Shows the effects of explanatory variables, a.k.a "**correlation triplot**"



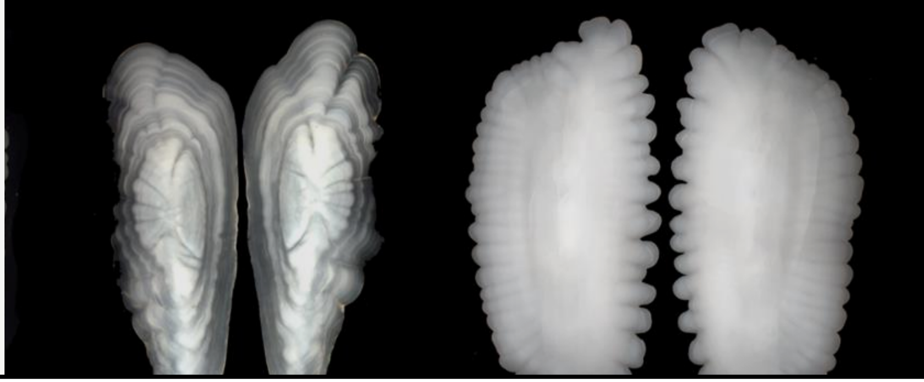


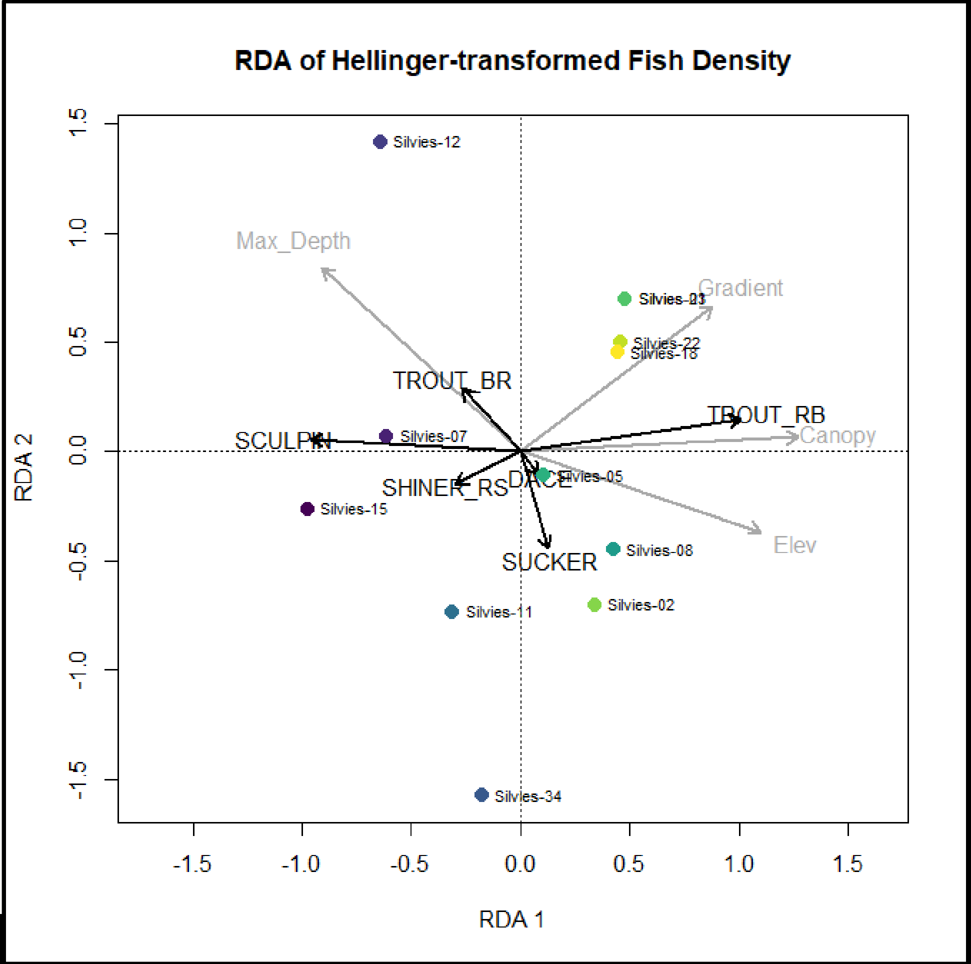
Redundancy Analysis: Interpretation

Scaling 1) Shows similarities among objects in the response matrix, a.k.a "**distance triplot**"

Scaling 2) Shows the effects of explanatory variables, a.k.a "**correlation triplot**"

Scaling 2 is the default option in R and is almost always what we're interested in.

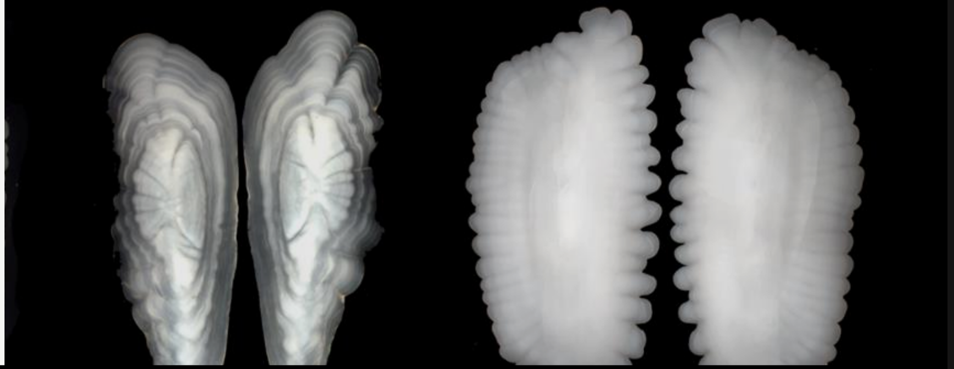


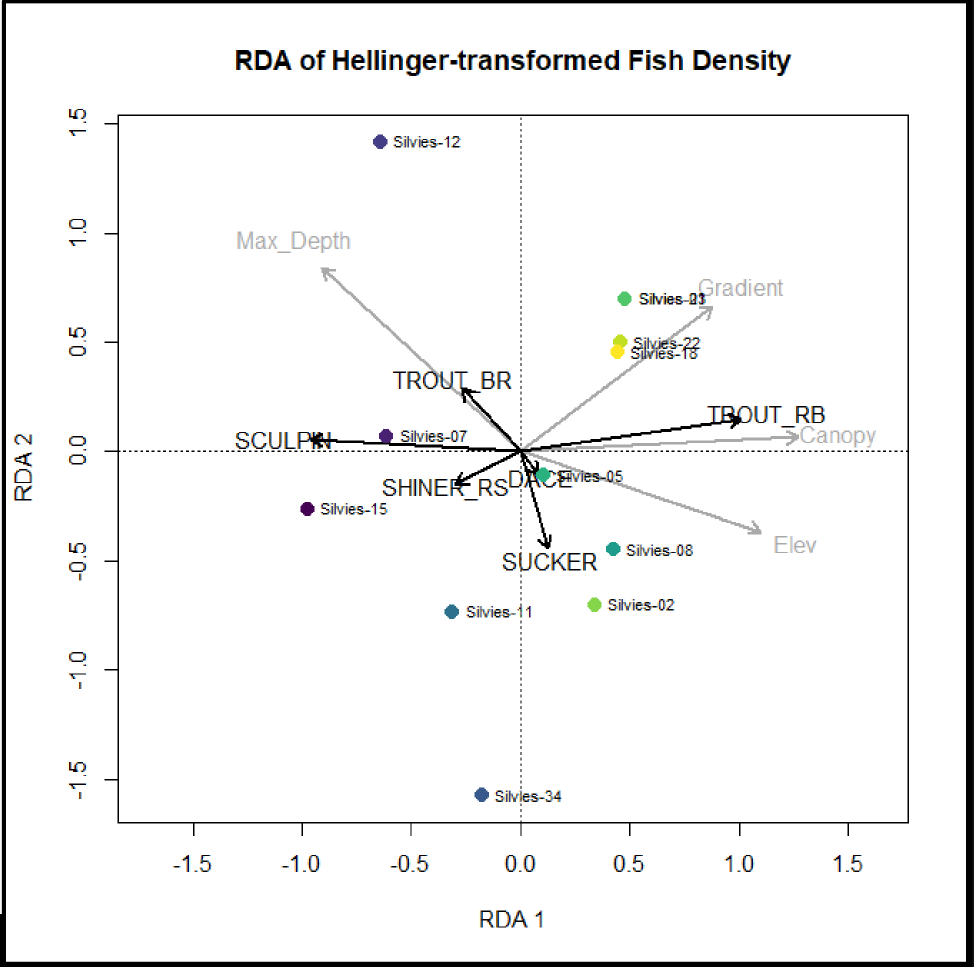


Redundancy Analysis: Interpretation

Partitioning of variance:

	Inertia	Proportion
Total	0.5608	1.0000
Constrained	0.2380	0.4245
Unconstrained	0.3228	0.5755





Redundancy Analysis: Interpretation

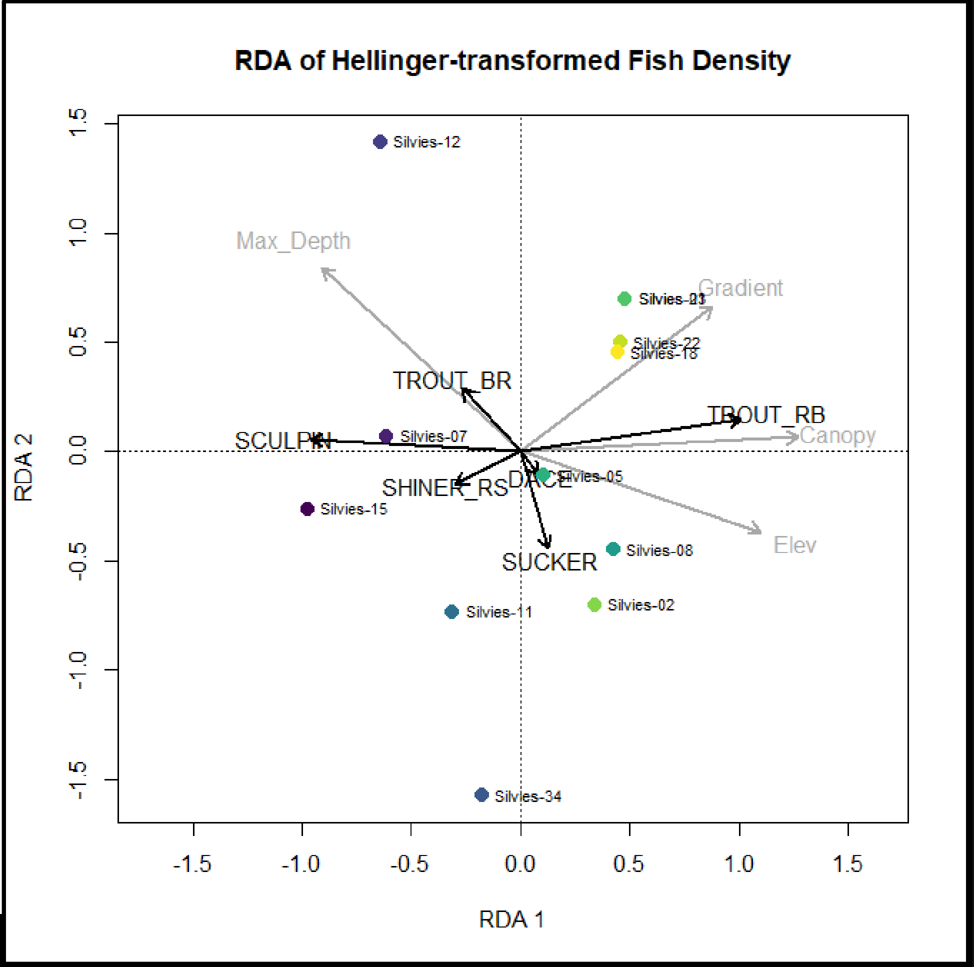
Partitioning of variance:

	Inertia	Proportion
Total	0.5608	1.0000
Constrained	0.2380	0.4245
Unconstrained	0.3228	0.5755

Importance of components:

	RDA1	RDA2	RDA3	RDA4
Eigenvalue	0.187	0.030	0.017	0.003
Proportion Explained	0.335	0.053	0.031	0.005
Cumulative Proportion	0.335	0.388	0.419	0.424

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	0.196	0.055	0.039	0.021	0.009	0.001
Proportion Explained	0.349	0.098	0.070	0.037	0.017	0.002
Cumulative Proportion	0.774	0.872	0.942	0.980	0.997	1.000



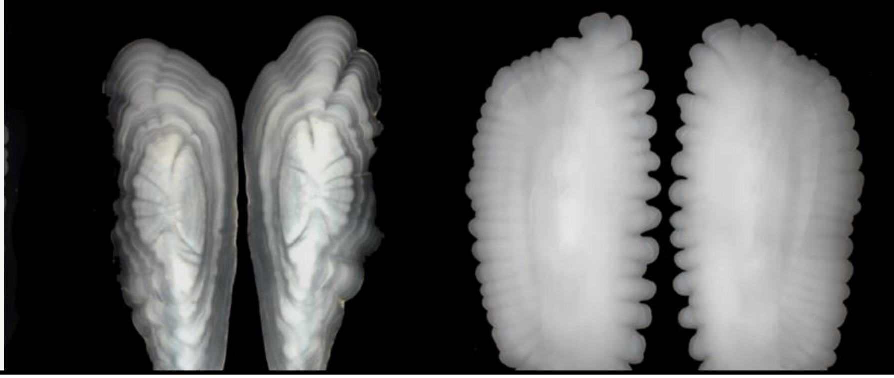
Redundancy Analysis: Interpretation

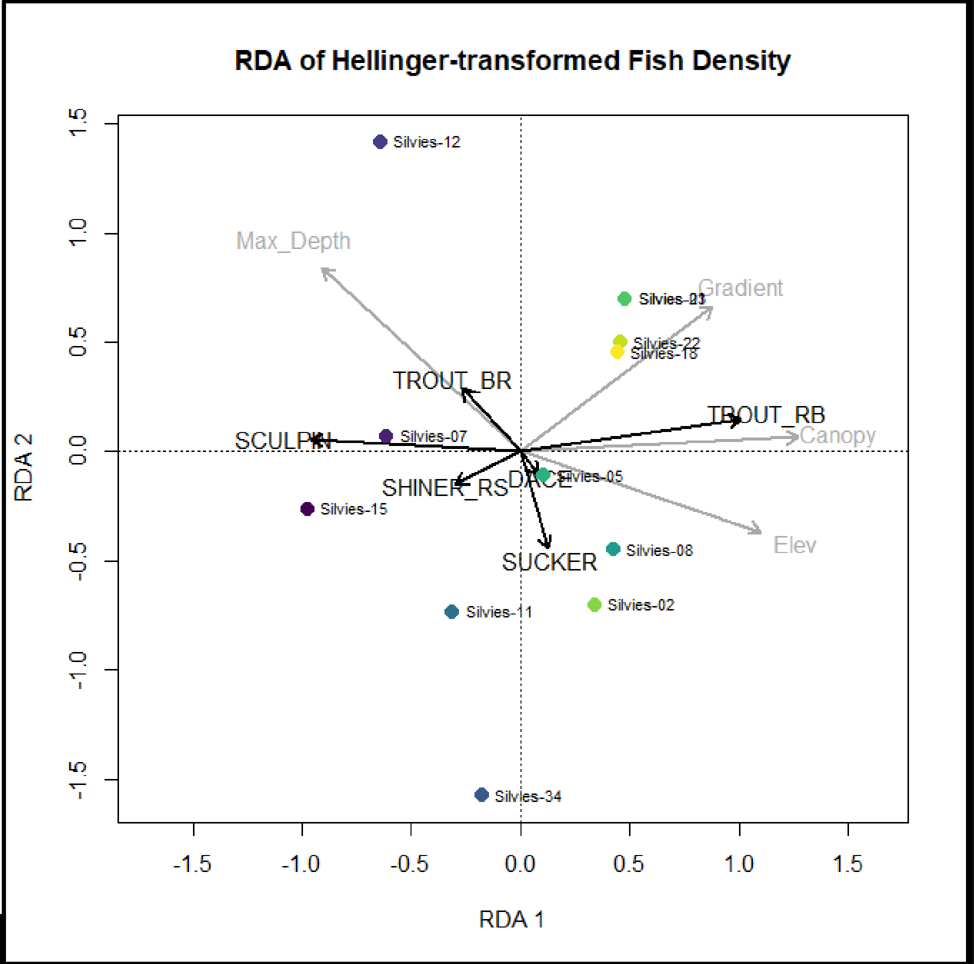
Partitioning of variance:

	Inertia	Proportion
Total	0.5608	1.0000
Constrained	0.2380	0.4245
Unconstrained	0.3228	0.5755

Accumulated constrained eigenvalues
Importance of components:

	RDA1	RDA2	RDA3	RDA4
Eigenvalue	0.187	0.029	0.017	0.003
Proportion Explained	0.789	0.124	0.073	0.012
Cumulative Proportion	0.789	0.914	0.987	1.000

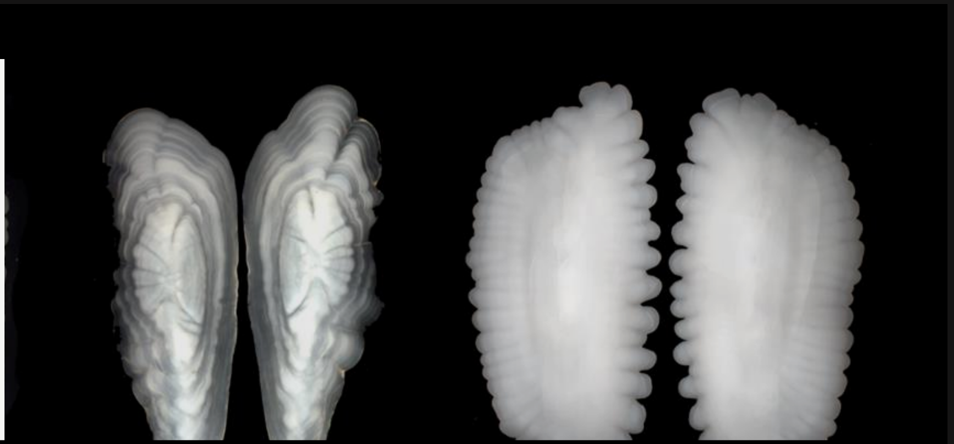


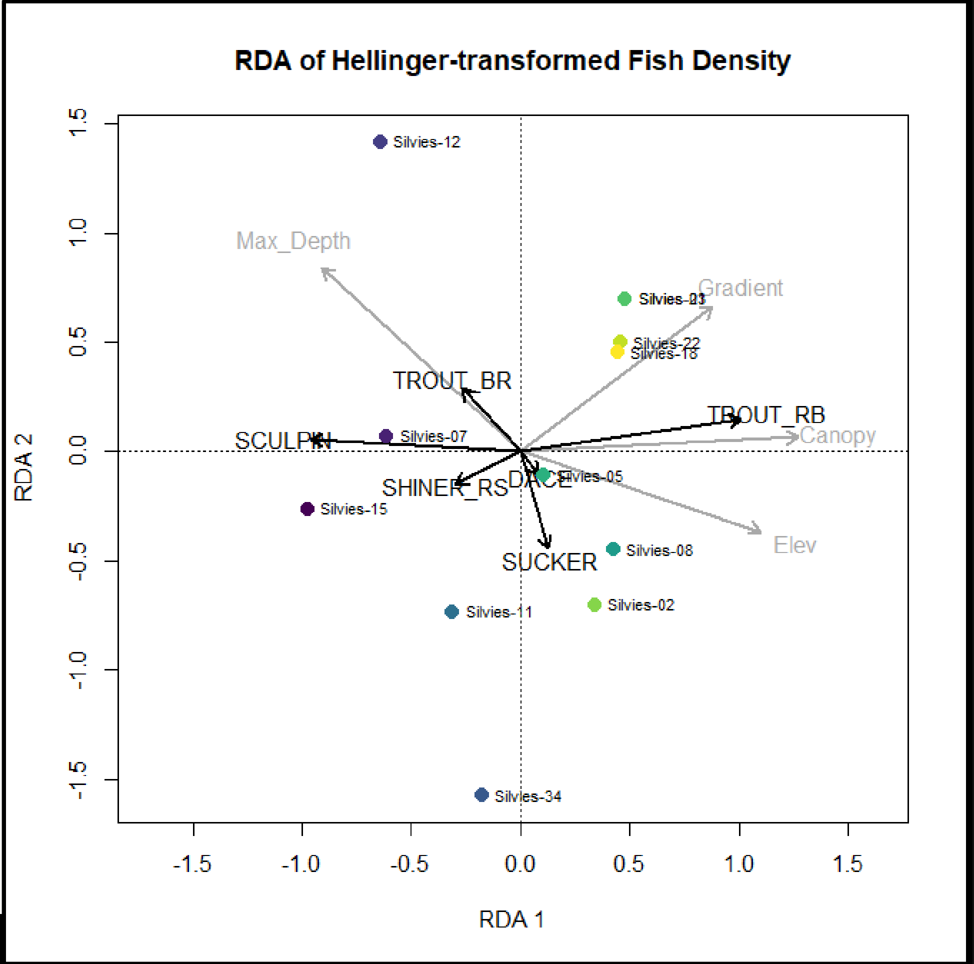


Redundancy Analysis: Interpretation

Regression Coefficients:

	RDA1	RDA2	RDA3	RDA4
Max_Depth	-0.089	1.148	-0.041	2.192
Elev	0.001	-0.002	-0.002	0.004
Gradient	3.239	16.791	-8.259	-16.038
Canopy	0.010	0.005	0.014	0.005





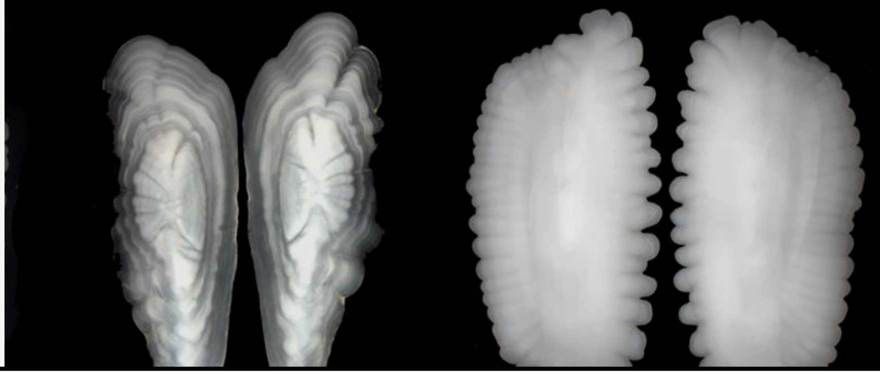
Redundancy Analysis: Interpretation

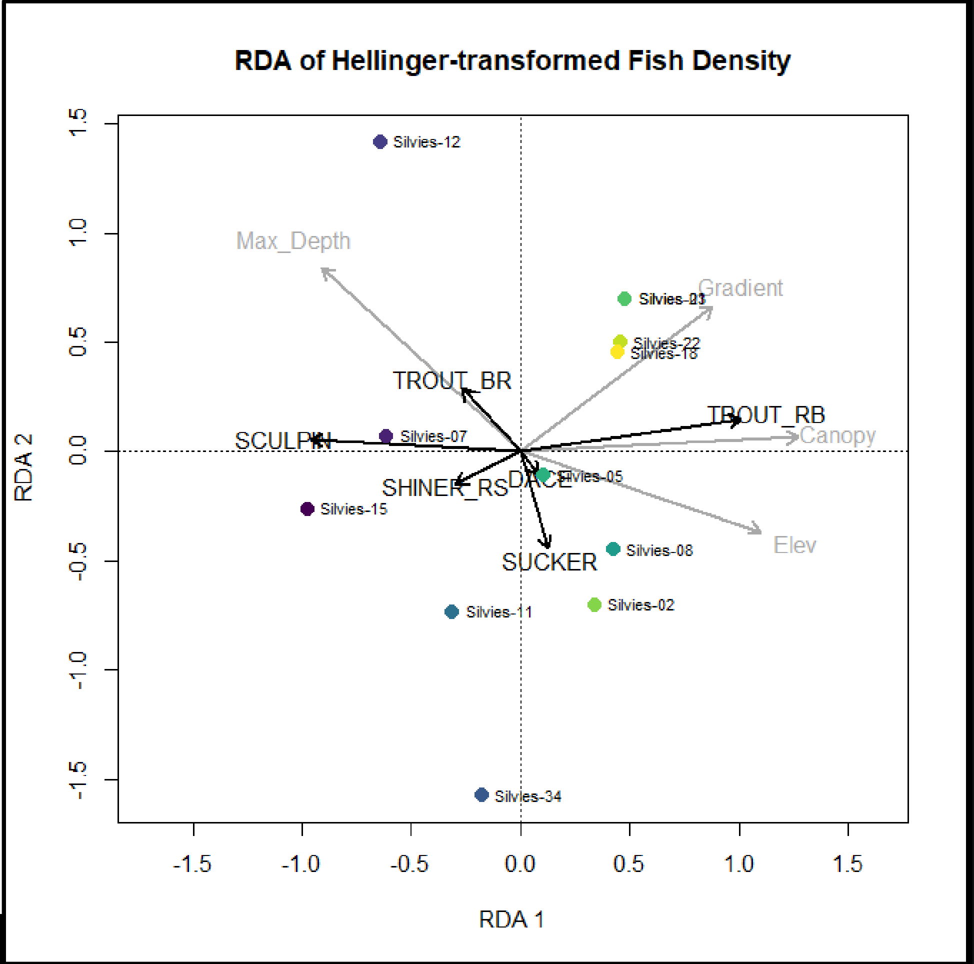
Regression Coefficients:

	RDA1	RDA2	RDA3	RDA4
Max_Depth	-0.089	1.148	-0.041	2.192
Elev	0.001	-0.002	-0.002	0.004
Gradient	3.239	16.791	-8.259	-16.038
Canopy	0.010	0.005	0.014	0.005

So a site's position on RDA 1 would be predicted using the coefficients in column 1!

It follows that a species distribution model projecting RDA 1 on the landscape would reflect redband trout and sculpin distributions.



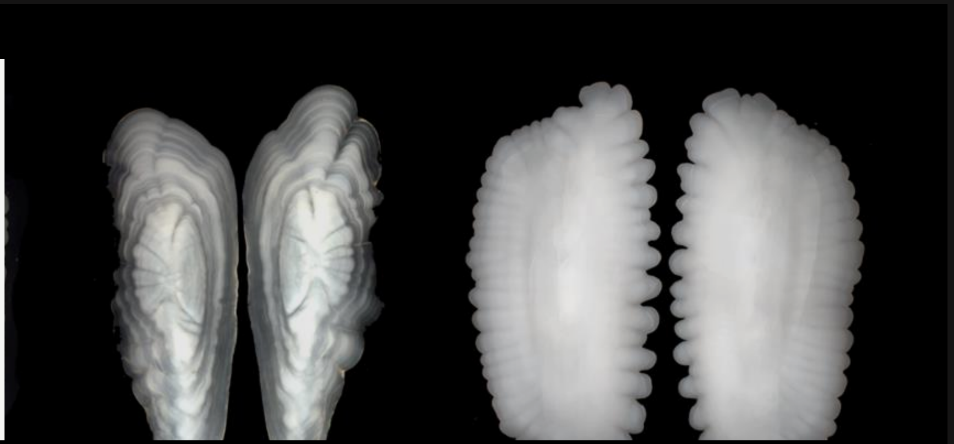


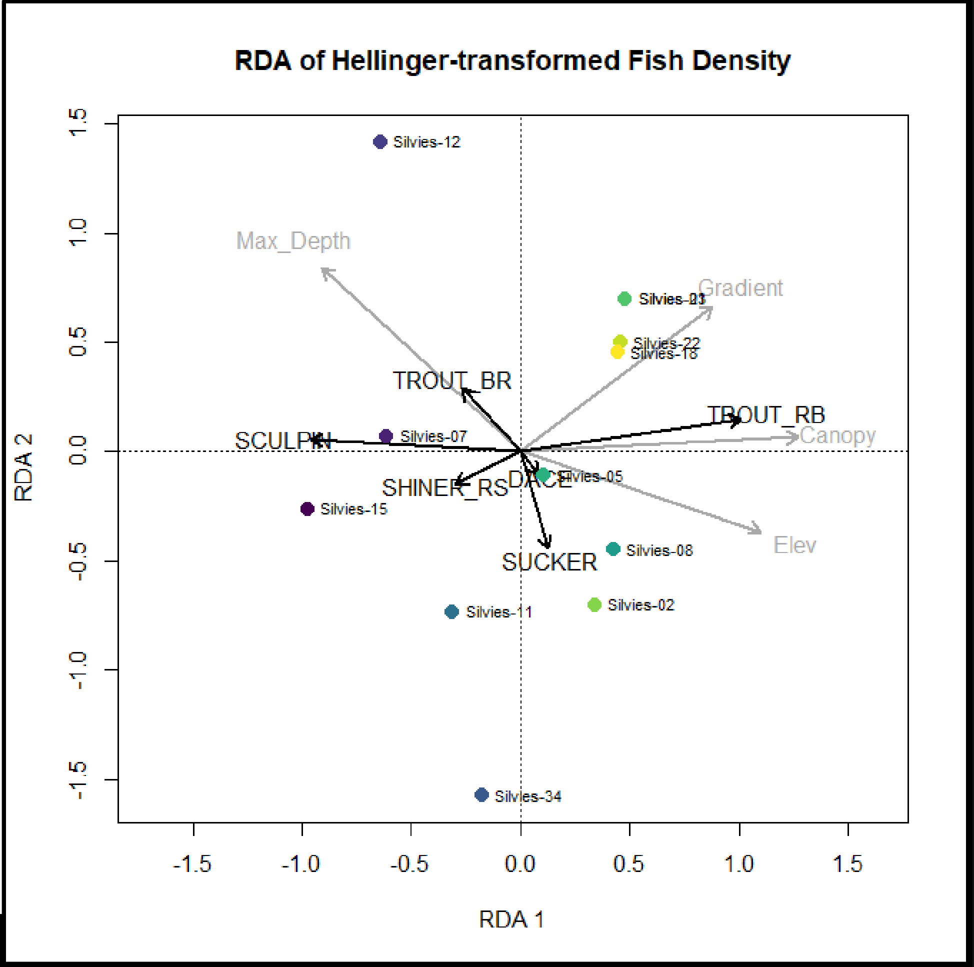
Redundancy Analysis: Interpretation

Permutational tests of significance:

	Simple RDA	Partial RDA
	Without covariables	With matrix **W** of covariables
	[a] Explained by **X** / [d] Unexplained variation	[a] [b] Explained by **X** / [b] [c] Explained by **W** / [d] Unexplained variation
Permute raw data	Permute [a + d]	Permute [a + b + c + d]
Permute residuals:		
• reduced model	Equivalent to permuting raw data	Permute [a + d]
• full model	Permute [d]	Permute [d]

Legendre & Legendre Table 11.6



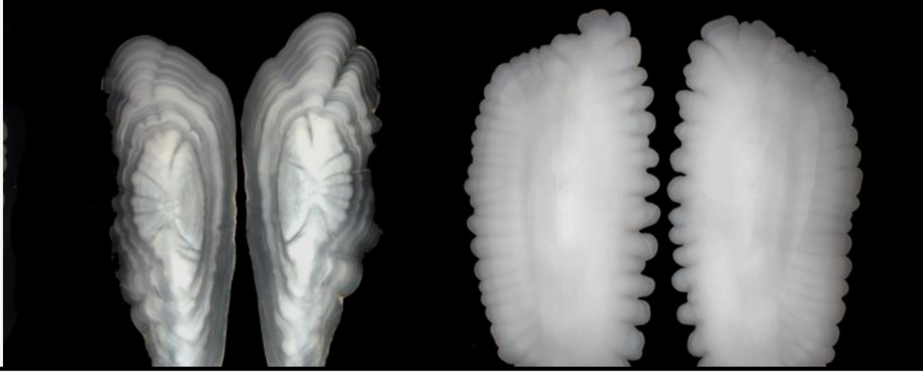


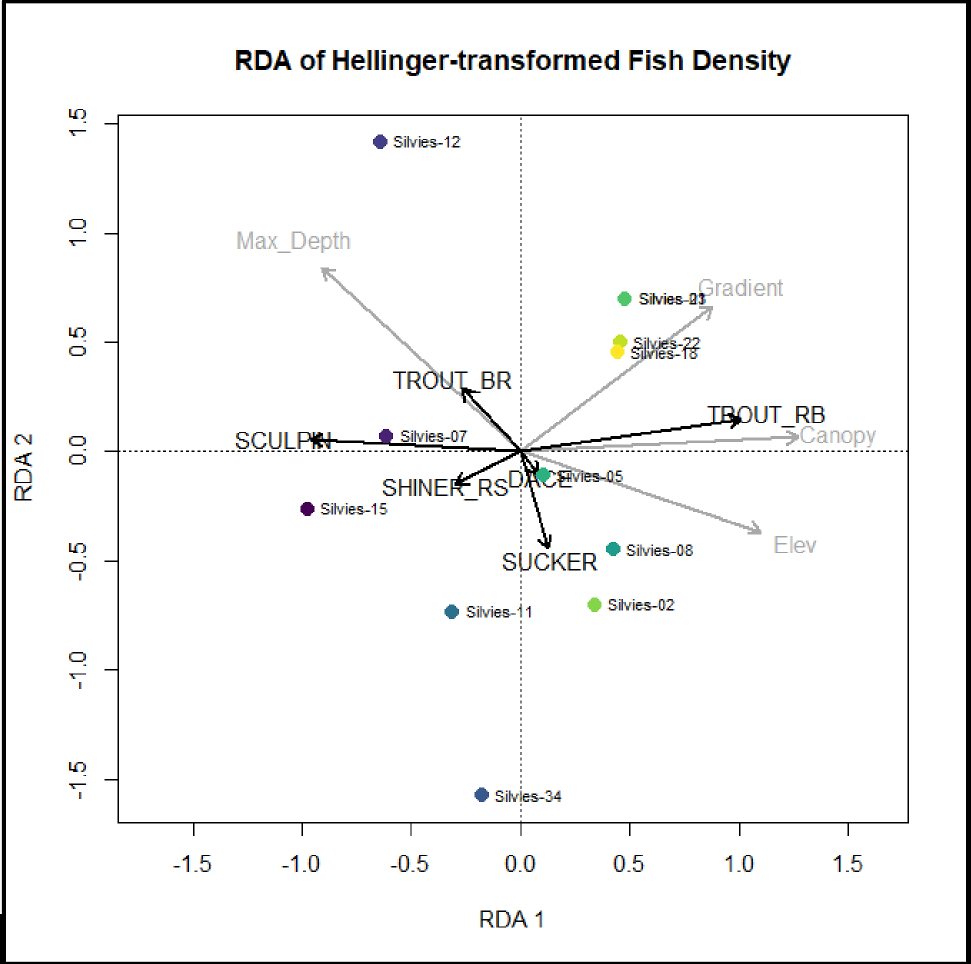
Redundancy Analysis: Interpretation

Permutational tests of significance:

	Simple RDA	Partial RDA
	Without covariables	With matrix **W** of covariables
	[a] [d] Explained by **X** / Unexplained variation	[a] [b] [c] [d] Explained by **X** / Explained by **W** / Unexplained variation
Permute raw data	Permute [a + d]	Permute [a + b + c + d]
Permute residuals:		
• reduced model	Equivalent to permuting raw data	Permute [a + d]
• full model	Permute [d]	Permute [d]

Legendre & Legendre Table 11.6



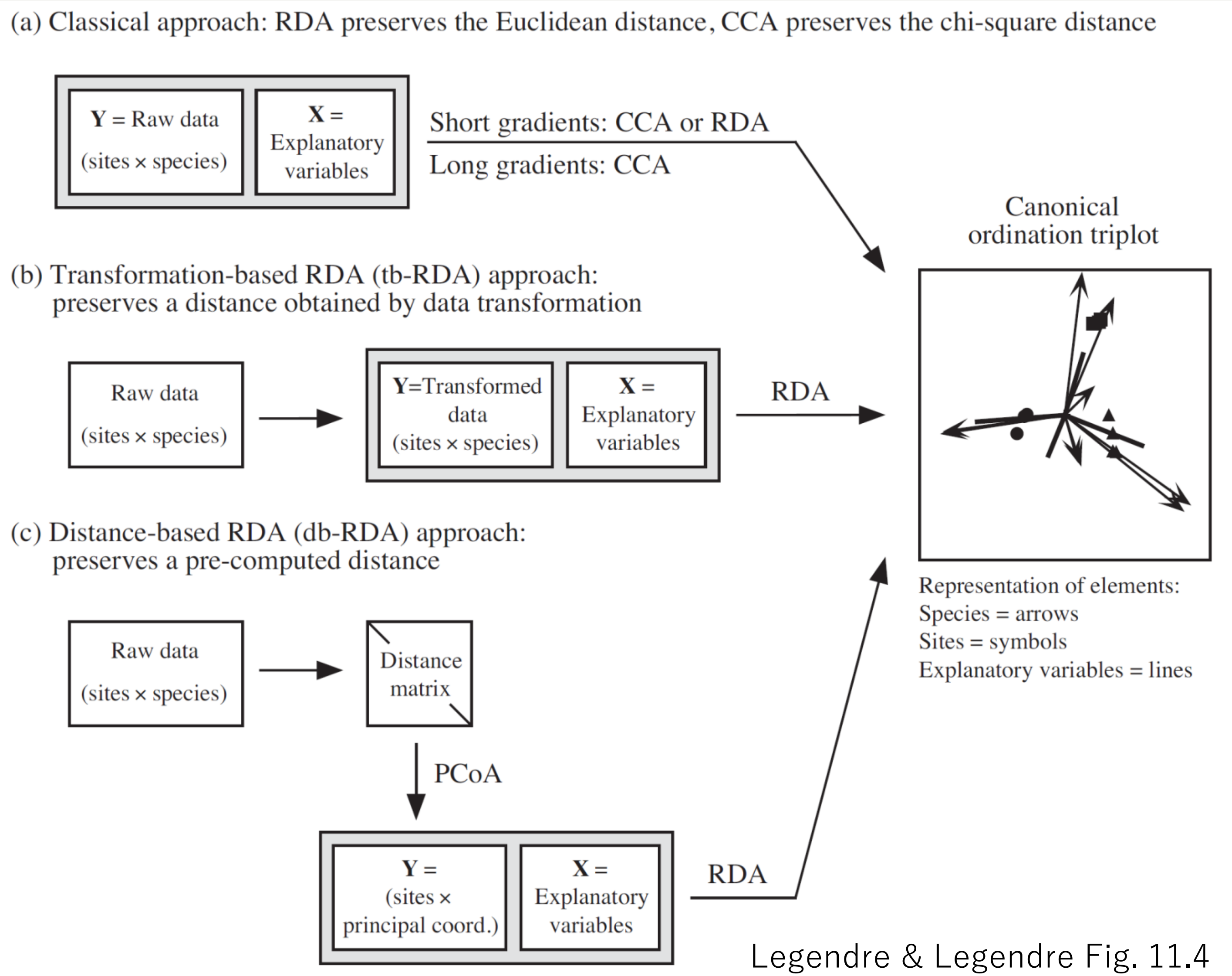


Redundancy Analysis: Flavors

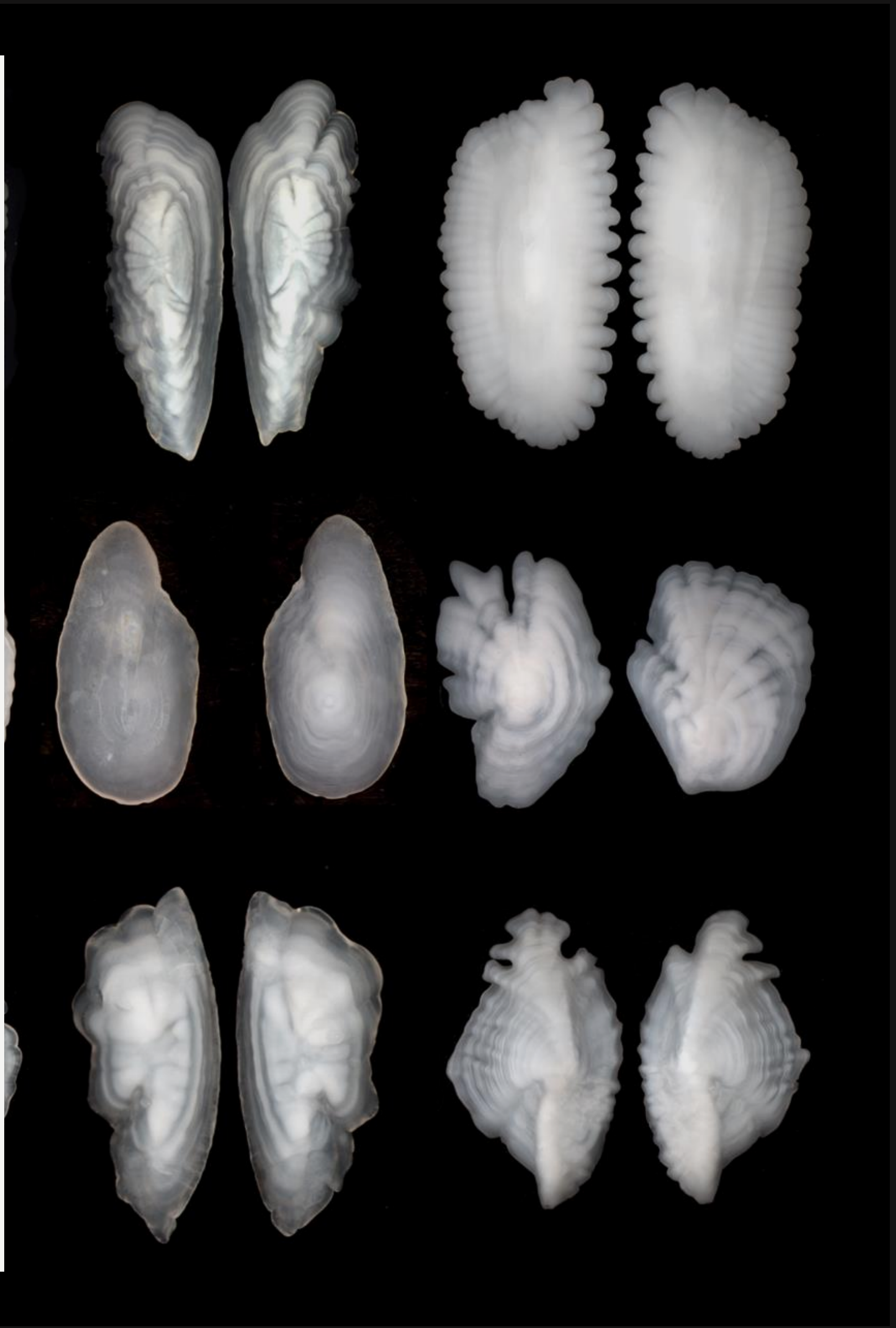
Transformation-based RDA: Did you notice we used Hellinger-transformed fish density data in this example? RDA supports Hellinger, chord, and chi-square transformations just like PCA!

Distance-based RDA: PCoA eigenvectors can be used as input for the RDA instead of PCA eigenvectors.

Redundancy Analysis: Flavors



Legendre & Legendre Fig. 11.4



Redundancy Analysis: Flavors

Transformation-based RDA: Did you notice we used Hellinger-transformed fish density data in this example? RDA supports Hellinger, chord, and chi-square transformations just like PCA!

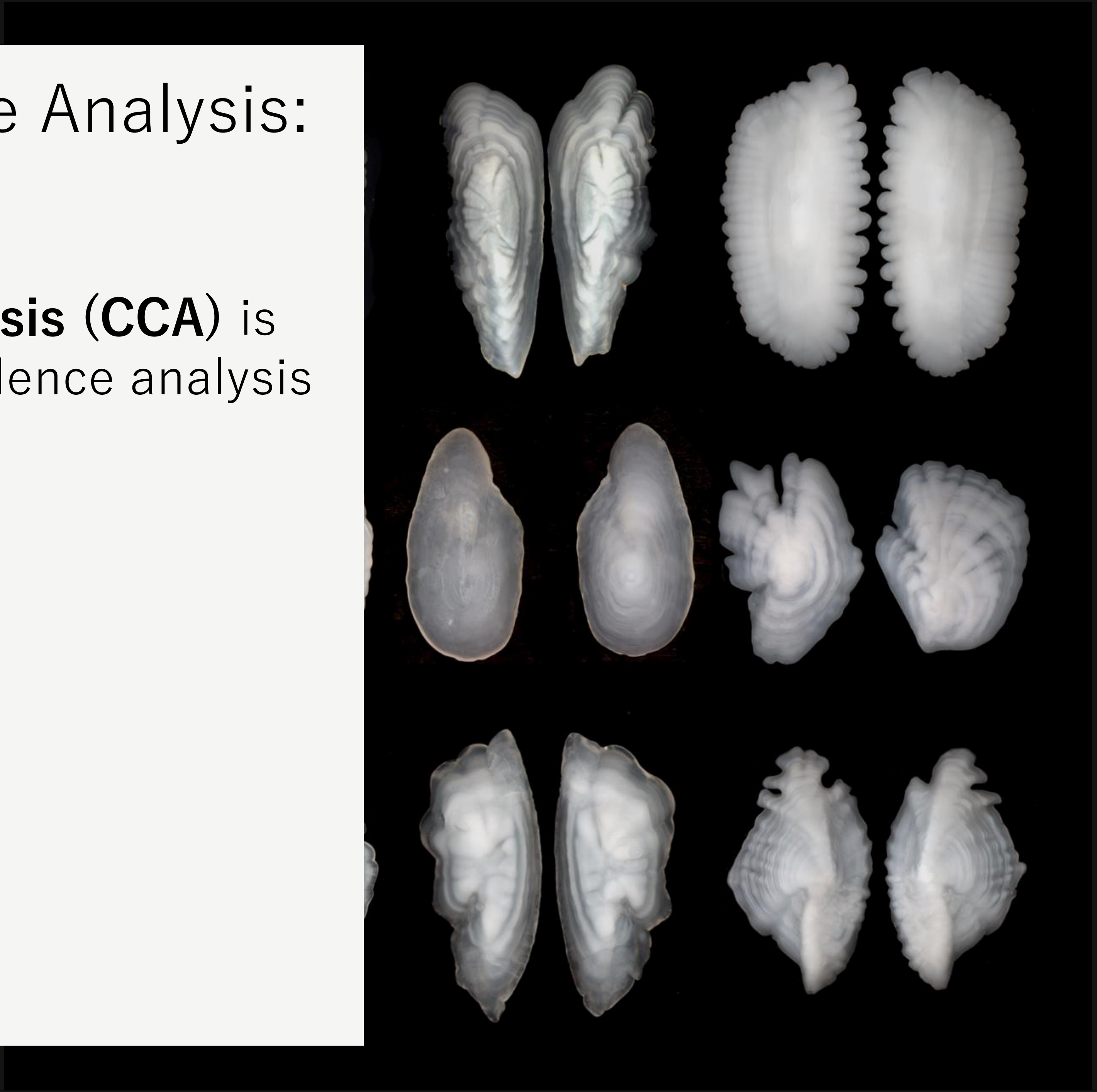
Distance-based RDA: PCoA eigenvectors can be used as input for the RDA instead of PCA eigenvectors.

Partial RDA: Analyzes response variables **Y**, predictor variables **X**, and covariates **W**.

Canonical Correspondence Analysis

Canonical Correspondence Analysis: Introduction

Canonical Correspondence Analysis (CCA) is the constrained form of correspondence analysis (CA).



Canonical Correspondence Analysis: Introduction

Canonical Correspondence Analysis (CCA) is the constrained form of correspondence analysis (CA).

- The ordination of **Y** is constrained such that ordination axes are **weighted averages** of the variables in **X**
- Preserves **chi-square** distances
- The data in **Y** can be **log-transformed** prior to analysis to avoid excessive influence of dominant species
- The data in **Y** should be standardized if they are not dimensionally homogeneous

Canonical Correspondence Analysis: Steps

1) Calculate weighted averages of each variable **Y** for all variables **X**

2) Carry out a CA of the matrix of **X**-**Y** relationships to obtain the eigenvalues and eigenvectors

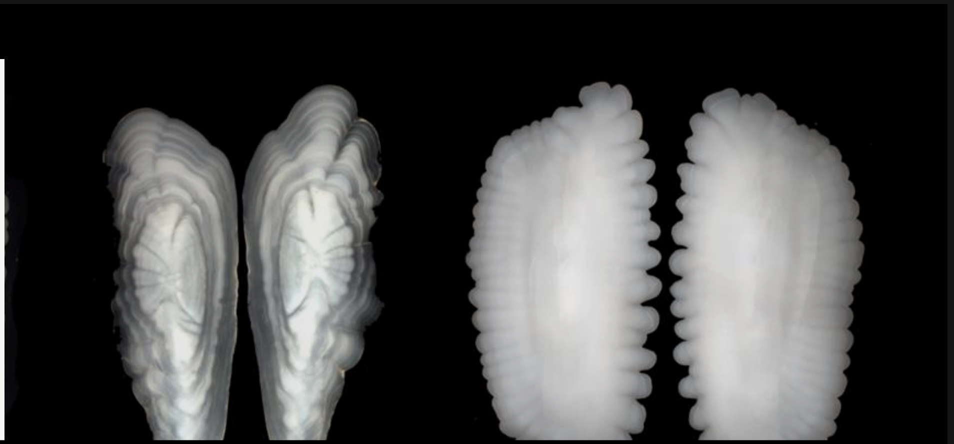
Canonical Correspondence Analysis: Steps

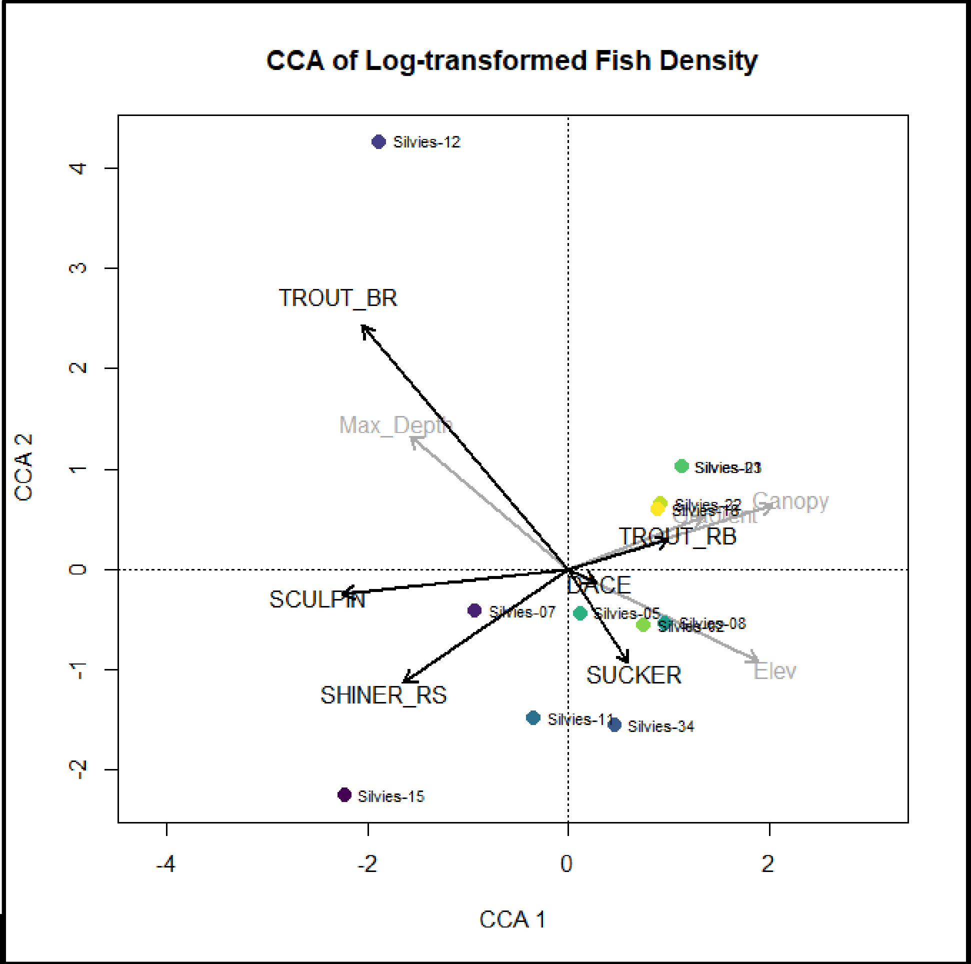
1) Calculate weighted averages of each variable **Y** for all variables **X**

   - Construct inflated data matrices **Q** and **X**
   - Apply weighted multiple regression
   - Compute fitted values
   - Center fitted values

Canonical Correspondence Analysis: Steps

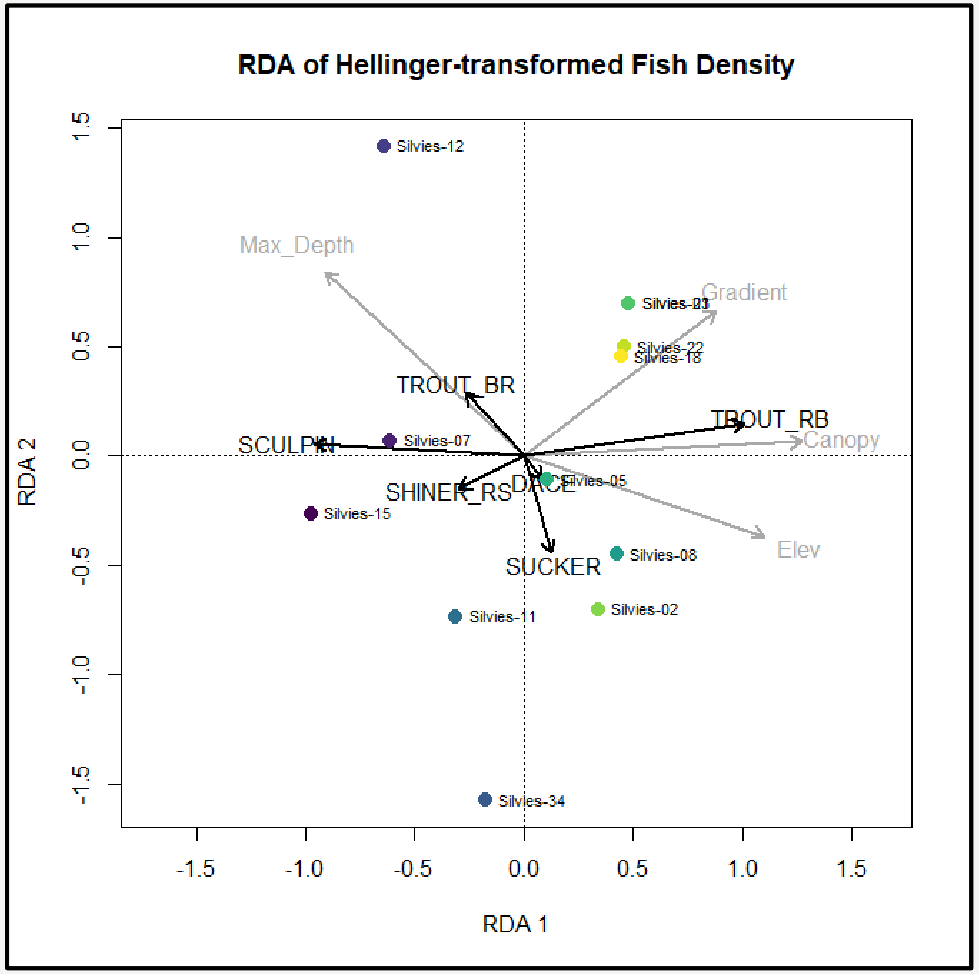
2) Carry out a CA of the matrix of **X-Y** relationships to obtain the eigenvalues and eigenvectors

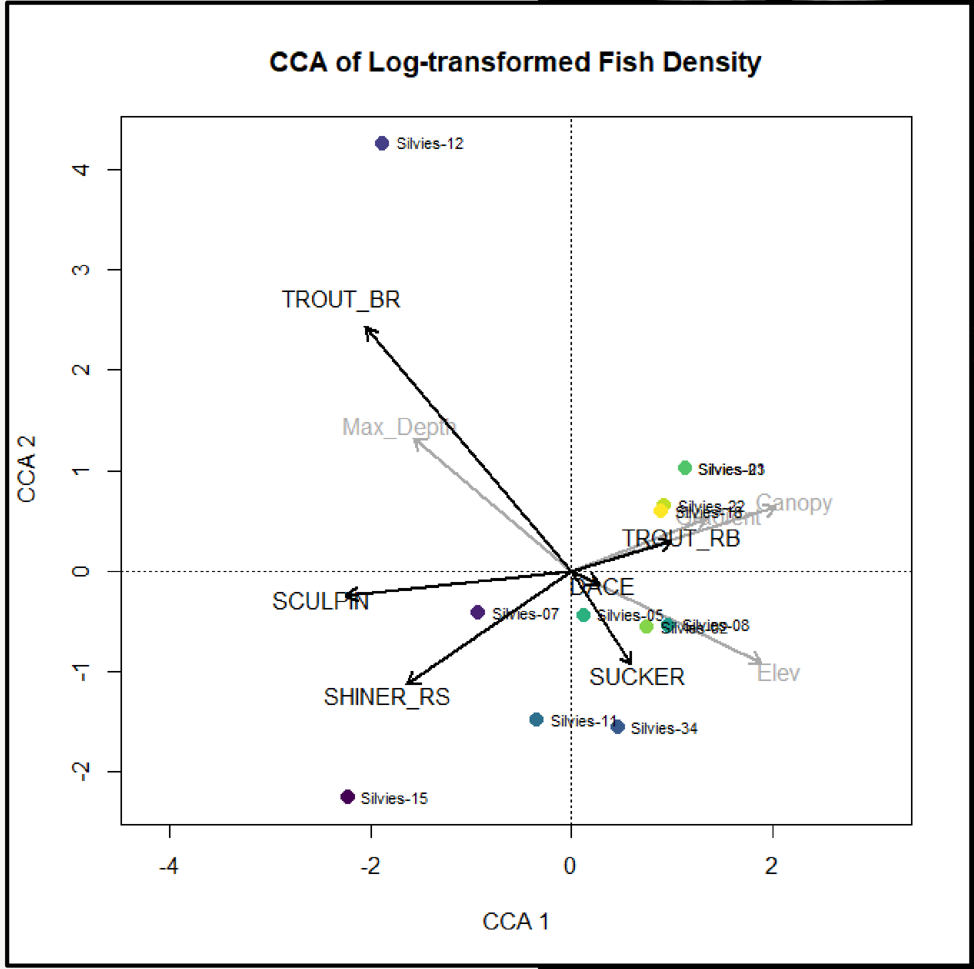


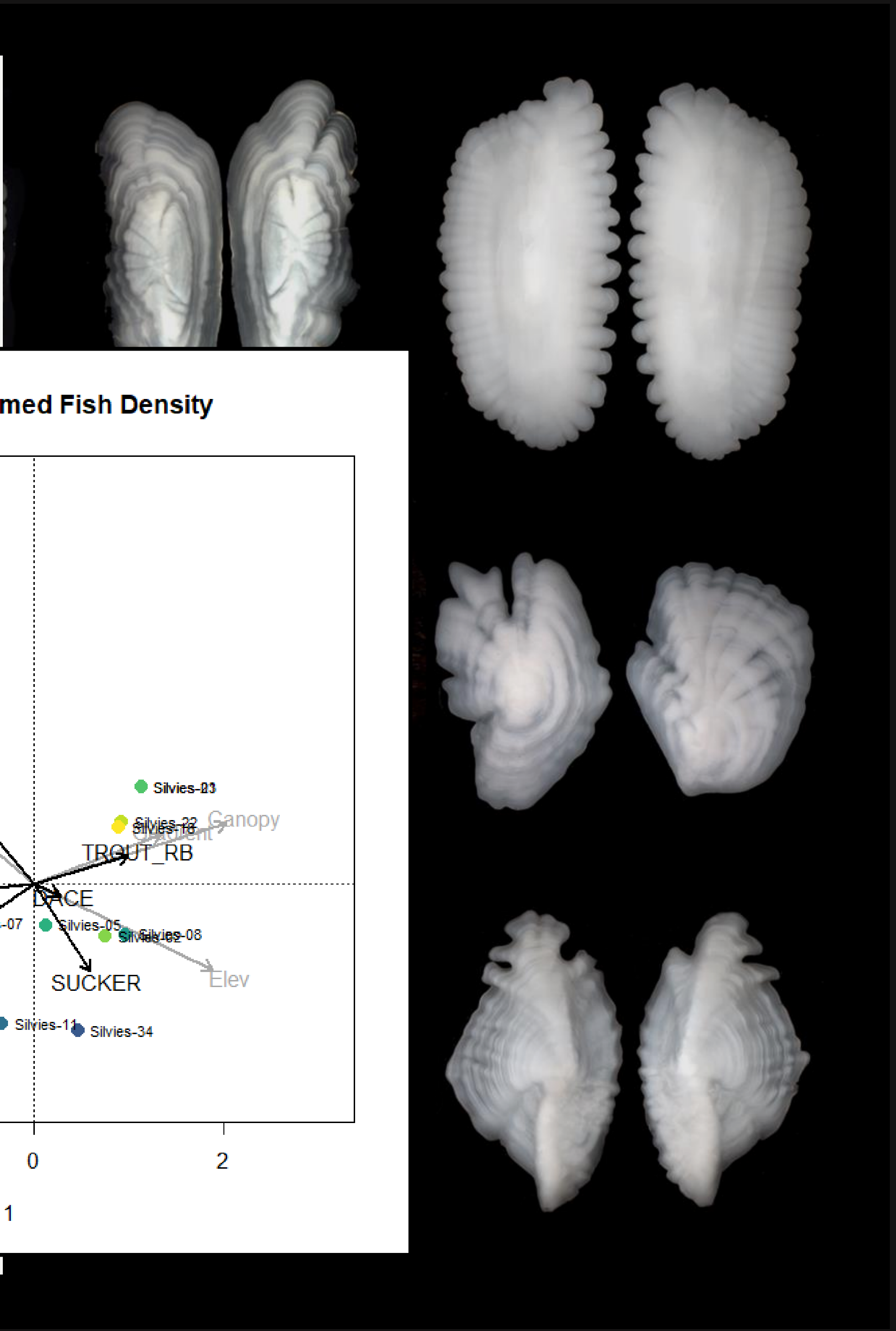


RDA or CCA?

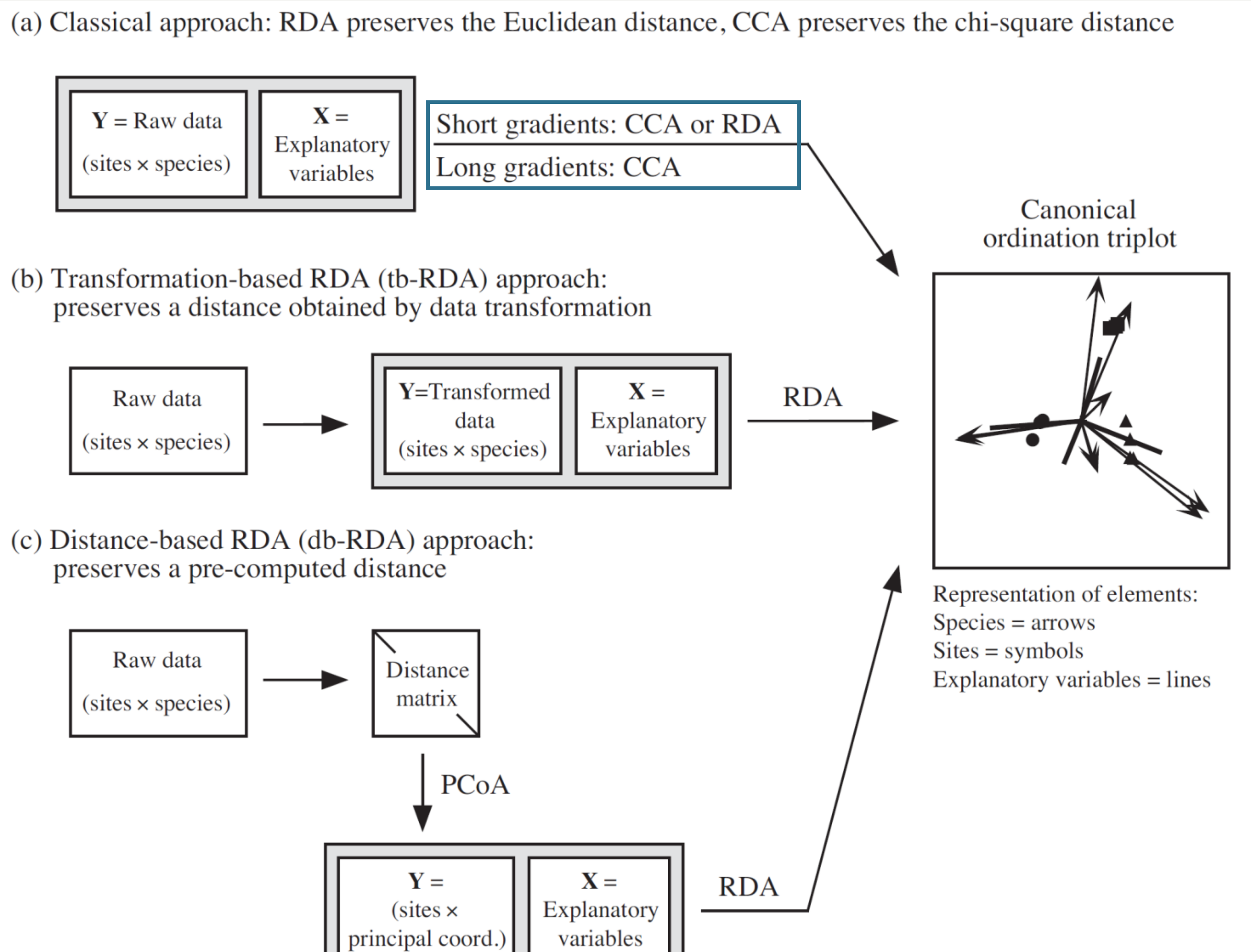
Which one to use?







RDA or CCA?



Legendre & Legendre Fig. 11.4

RDA or CCA?

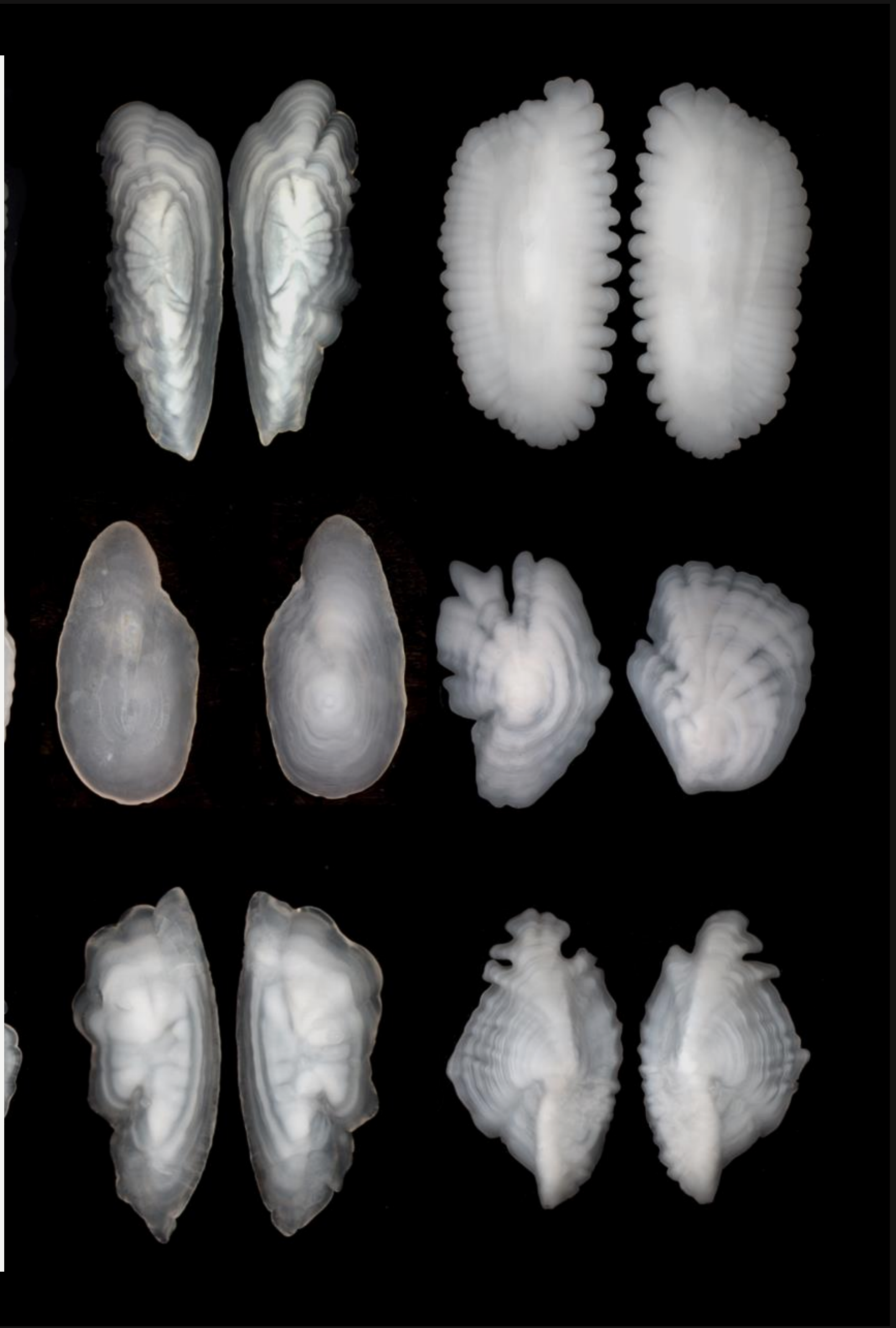
Use RDA when:

- Relationships between objects and descriptors are **linear**
- Descriptors are approximately **normally distributed**
- **Euclidean distance** can be accurately used as a measure of distance between objects and descriptors
- **Homoscedasticity of variance** applies

RDA or CCA?

Use CCA when:

- Relationships between objects and descriptors are **unimodal** and object-descriptor relations are **non-linear**
- Descriptors are **highly skewed**
- **Chi-square distance** can be accurately used as a measure of distance between objects and descriptors
- Environmental gradients are **long**



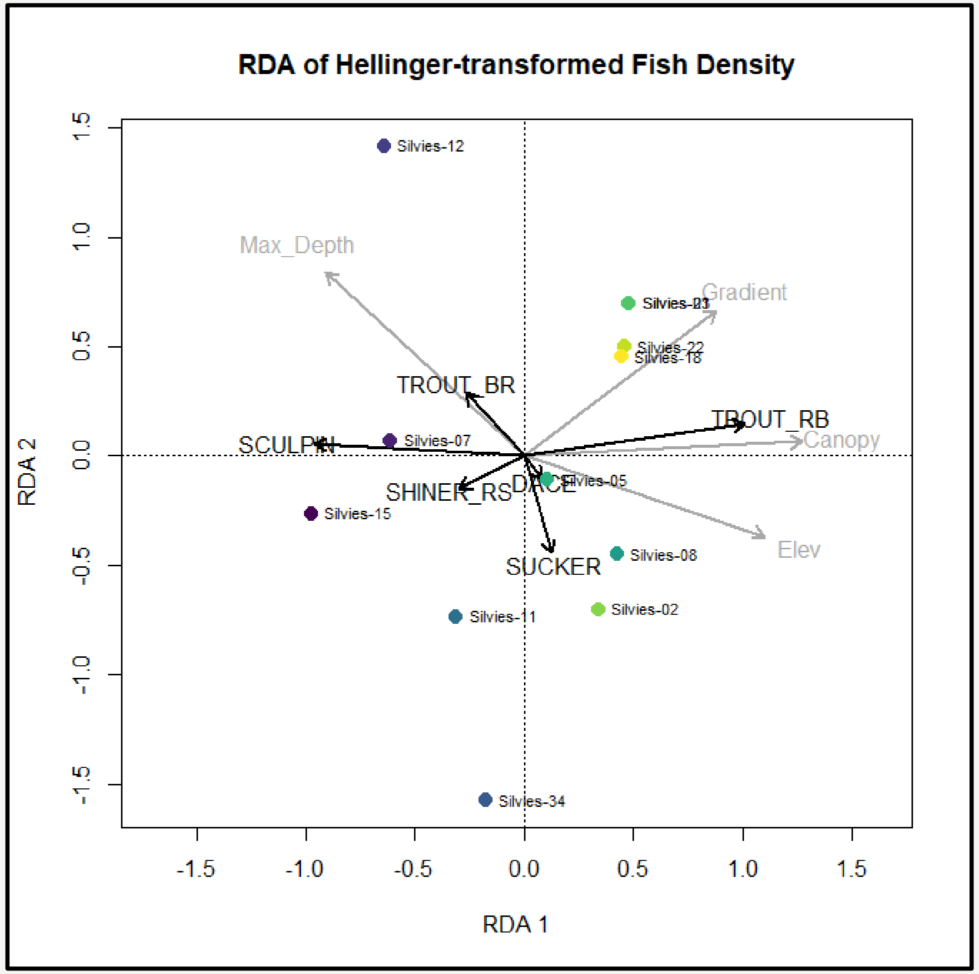
RDA or CCA?

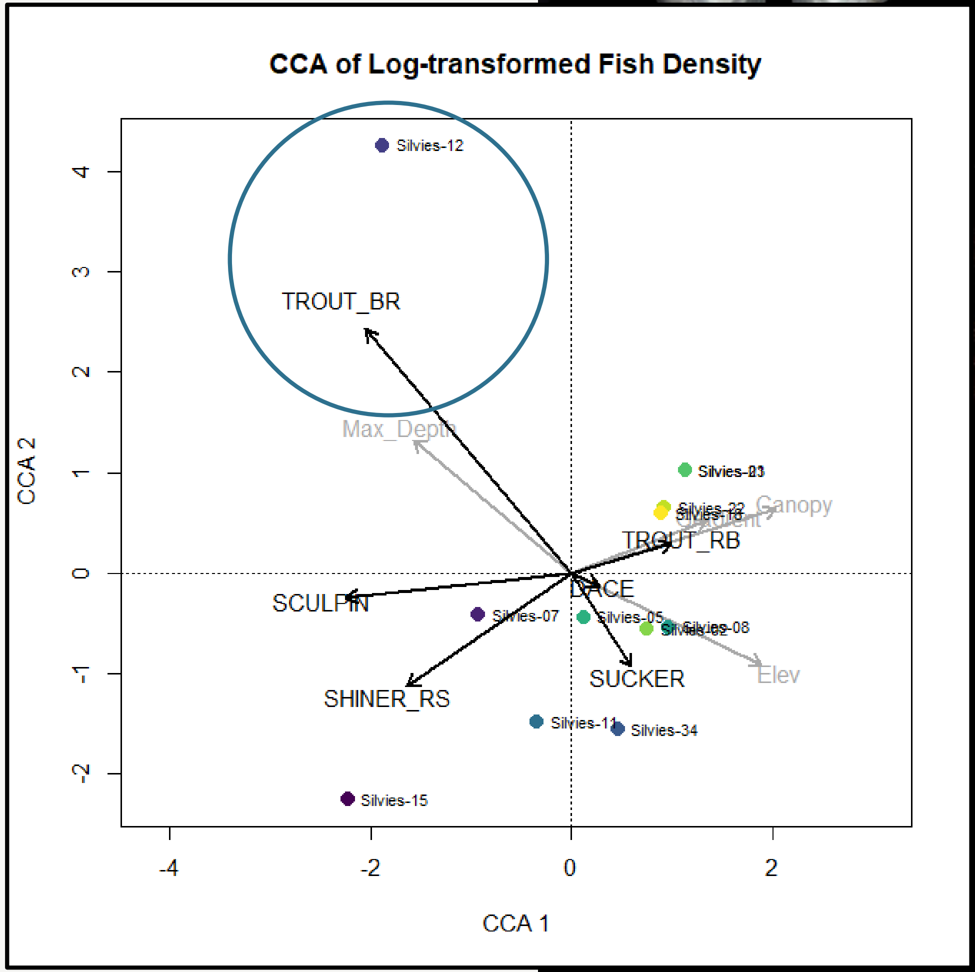
The same limitations apply to CCA as CA.

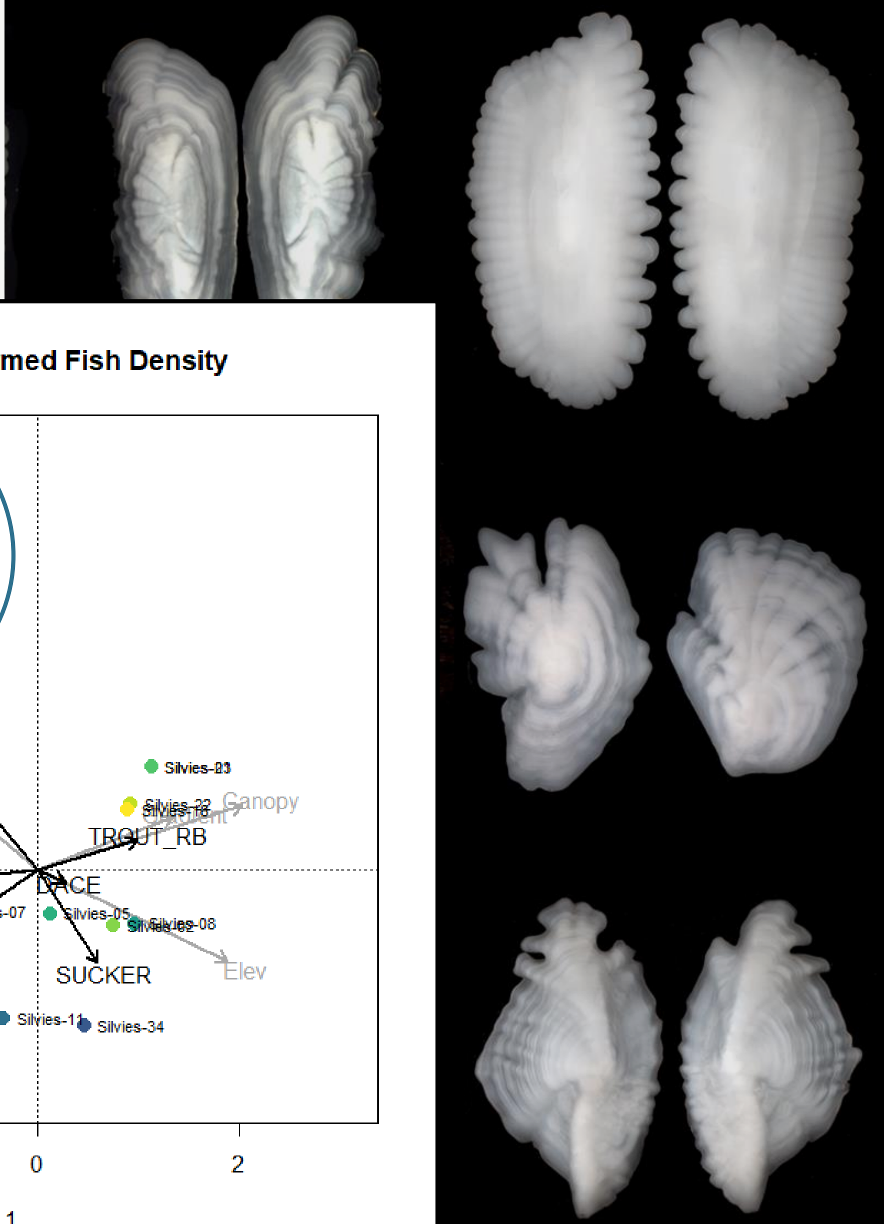
If you wouldn't use CA on a data set, don't use a CCA! Use tb-RDA or db-RDA instead!

RDA or CCA?

Which one to use?



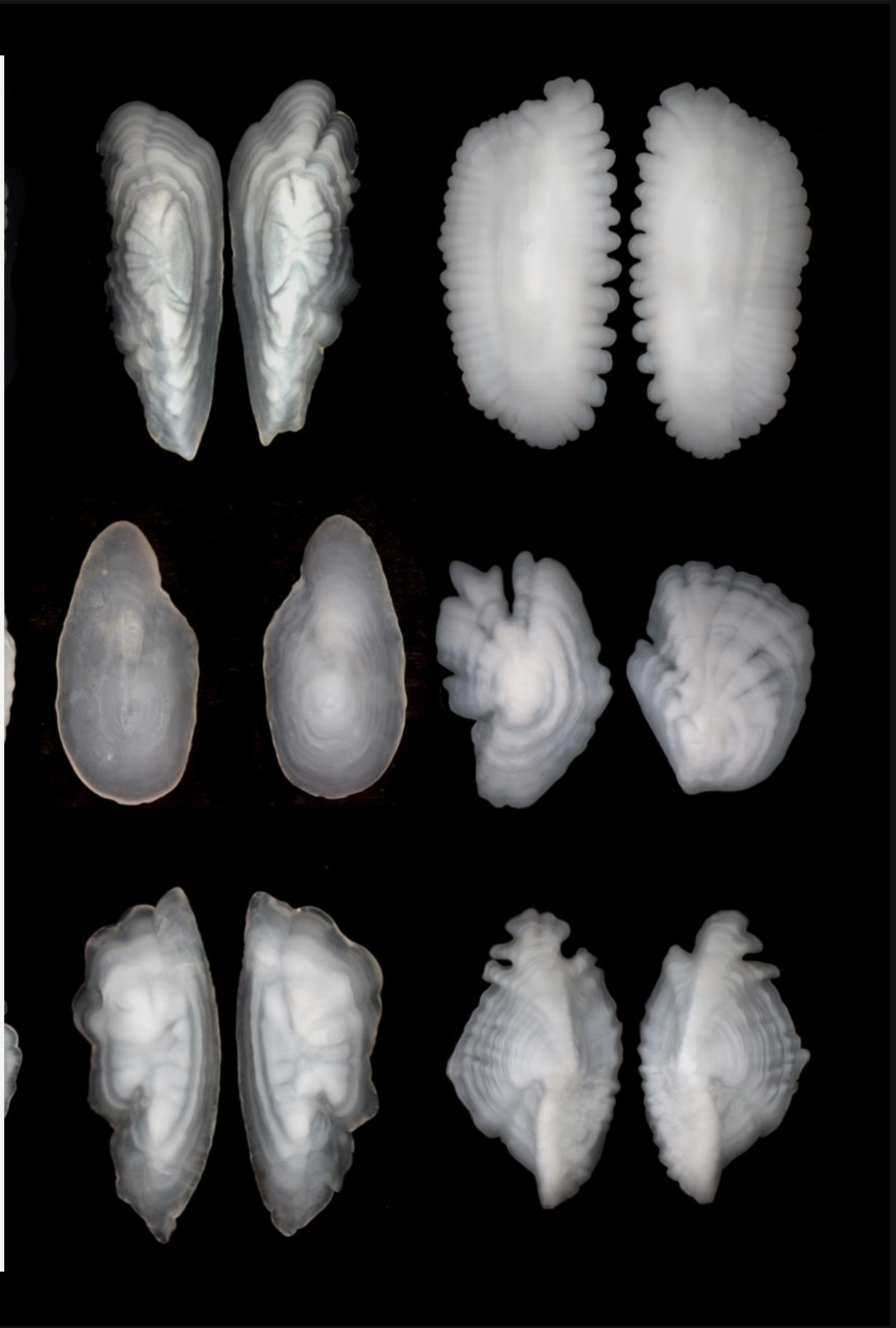




Co-Inertia Analysis

Co-Inertia Analysis: Introduction

Co-Inertia Analysis (CoIA) is a **symmetrical** alternative to CCA.



Co-Inertia Analysis: Introduction

Co-Inertia Analysis (CoIA) is a **symmetrical** alternative to CCA.

Quantifying common structures between $\mathbf{Y_1}$ **and** $\mathbf{Y_2}$.

Co-Inertia Analysis: Introduction

Co-Inertia Analysis (CoIA) is a **symmetrical** alternative to CCA.

Quantifying common structures between $\mathbf{Y_1}$ **and** $\mathbf{Y_2}$.

Co-inertia is the measure of shared structure between two datasets.

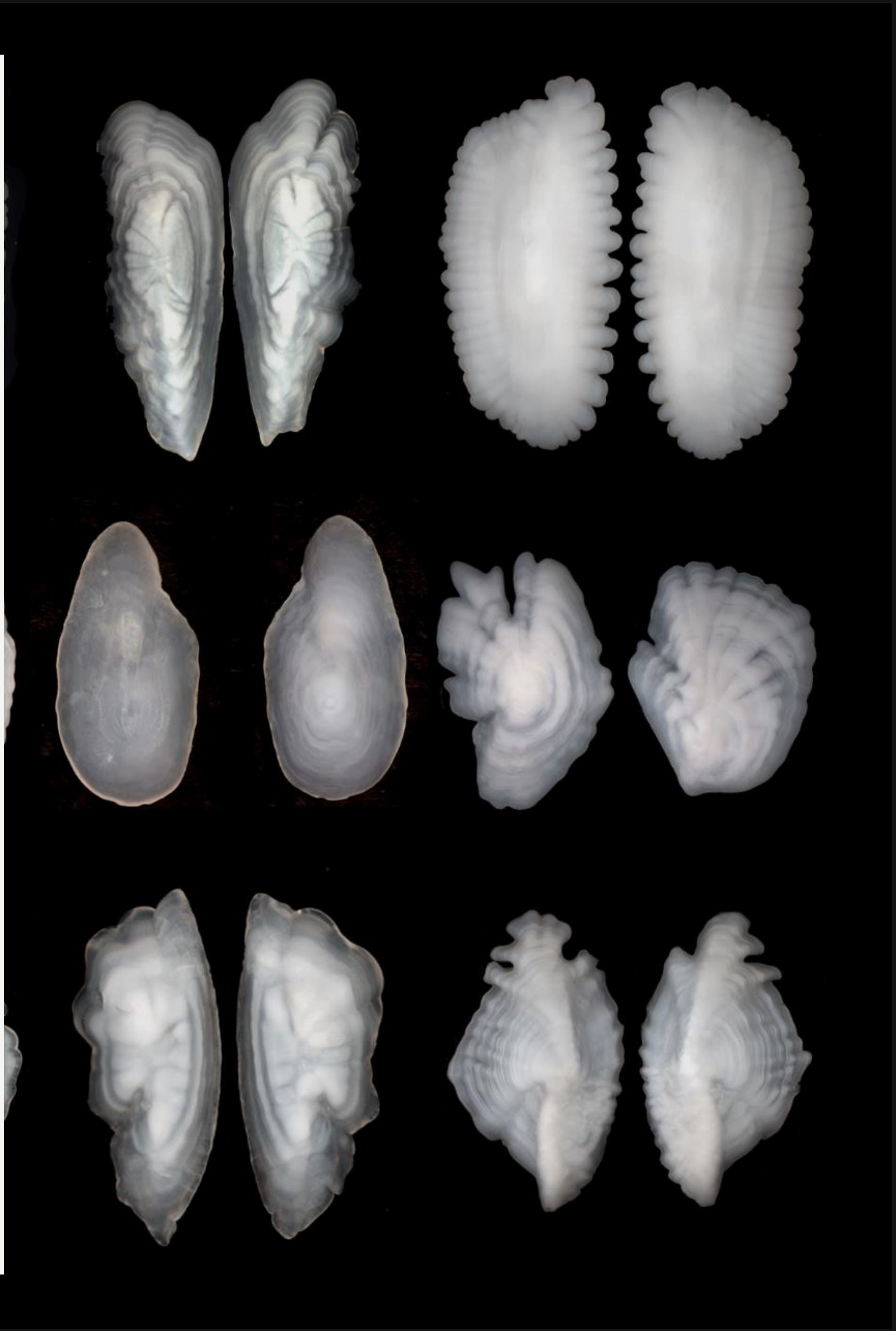
Co-Inertia Analysis: Steps

1) Perform separate ordinations of each dataset

Co-Inertia Analysis: Steps

1) Perform separate ordinations of each dataset

- Results in ordination scores (site scores) and eigenvalues that summarize the variance in each dataset



Co-Inertia Analysis: Steps

1) Perform separate ordinations of each dataset

2) Maximize co-inertia between the two ordinations

Co-Inertia Analysis: Steps

1) Perform separate ordinations of each dataset

2) Maximize co-inertia between the two ordinations

   - Co-inertia is the covariance between the ordination scores from the two datasets

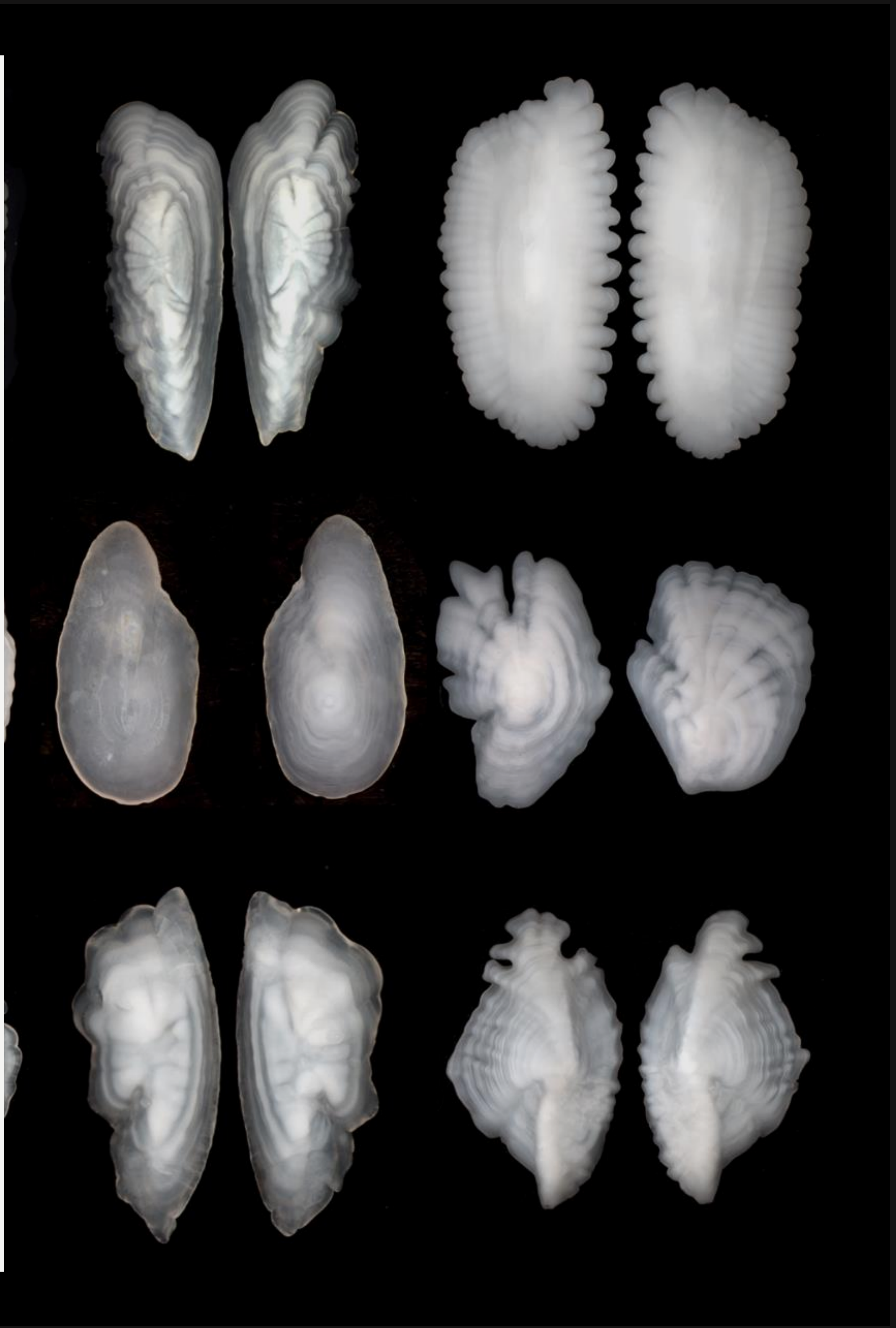
   - Finds linear combinations of the variables in both datasets that align their ordination axes as closely as possible

Co-Inertia Analysis: Steps

1) Perform separate ordinations of each dataset

2) Maximize co-inertia between the two ordinations

3) Produce RV coefficient and co-inertia scores



Co-Inertia Analysis: Steps

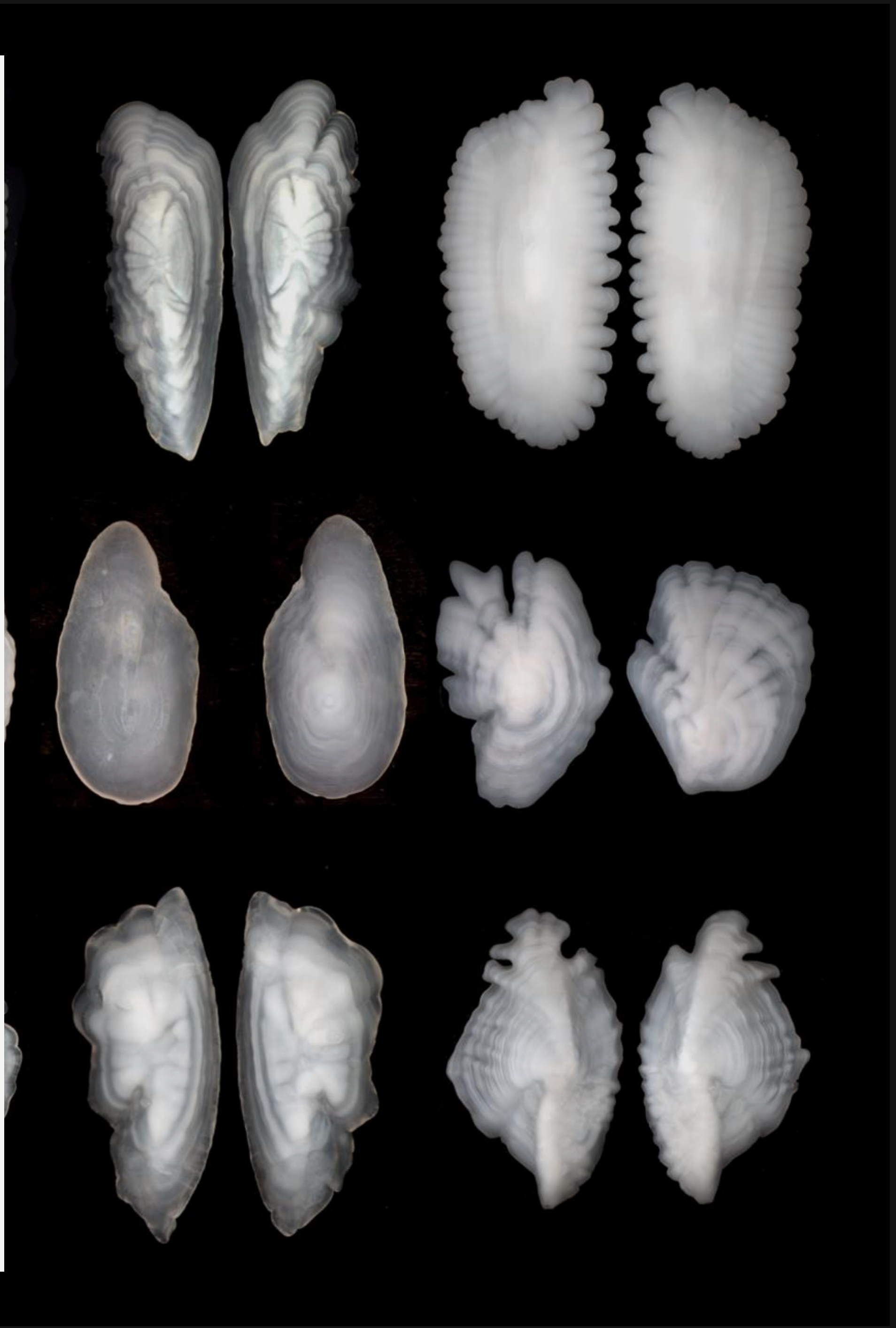
1) Perform separate ordinations of each dataset

2) Maximize co-inertia between the two ordinations

3) Produce RV coefficient and co-inertia scores

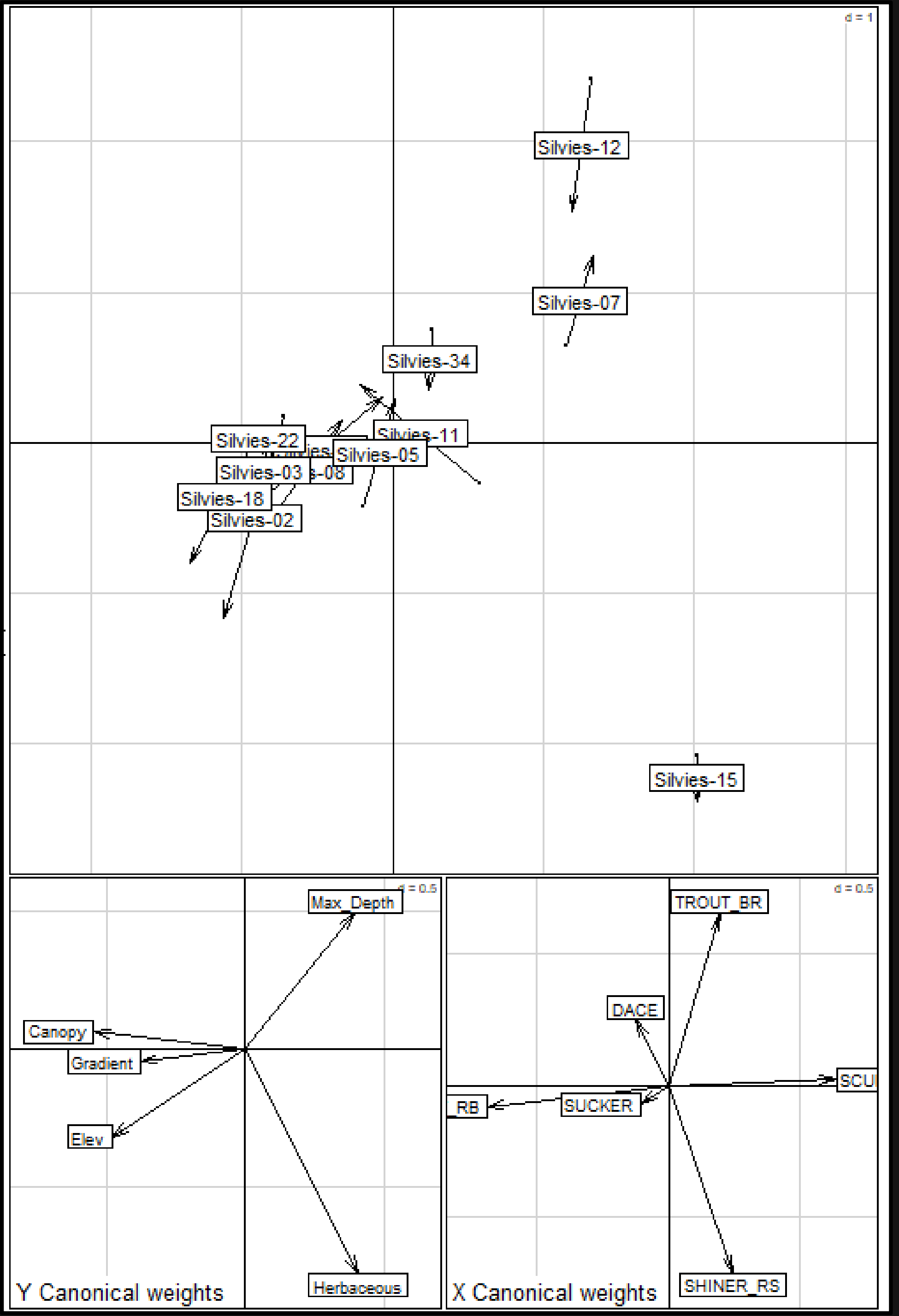
- The **RV coefficient** is a measure of how much two datasets share common structure (0-1)

- **Co-inertia scores** represent the positions of the objects in shared multidimensional space



Co-Inertia Analysis: Interpretation

- RV = 0.59, $p = 0.007$
- Axes 1 and 2 account for 95% of co-inertia
- The closer two points are, the more similar the patterns in the two datasets for that observation.
- The longer the arrow, the more that site contributes to multivariate dispersion



Conclusion: Summary of Key Points

- **Redundancy Analysis (RDA)**
  - Linear regression of Y (species) on X (environment)
  - Preserves Euclidean distances
  - Assumes linear relationships
- **Canonical Correspondence Analysis (CCA)**
  - Assumes unimodal responses
  - Preserves chi-square distances
- **Co-inertia Analysis (CoIA)**
  - Examines shared structure between Y_1 and Y_2
  - Symmetric
  - Maximizes covariance

Questions?