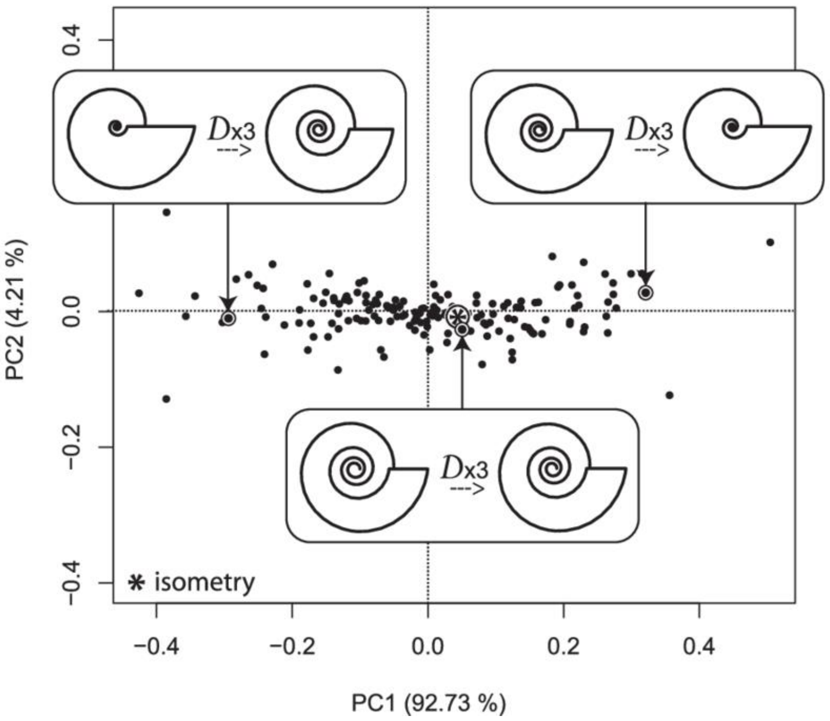
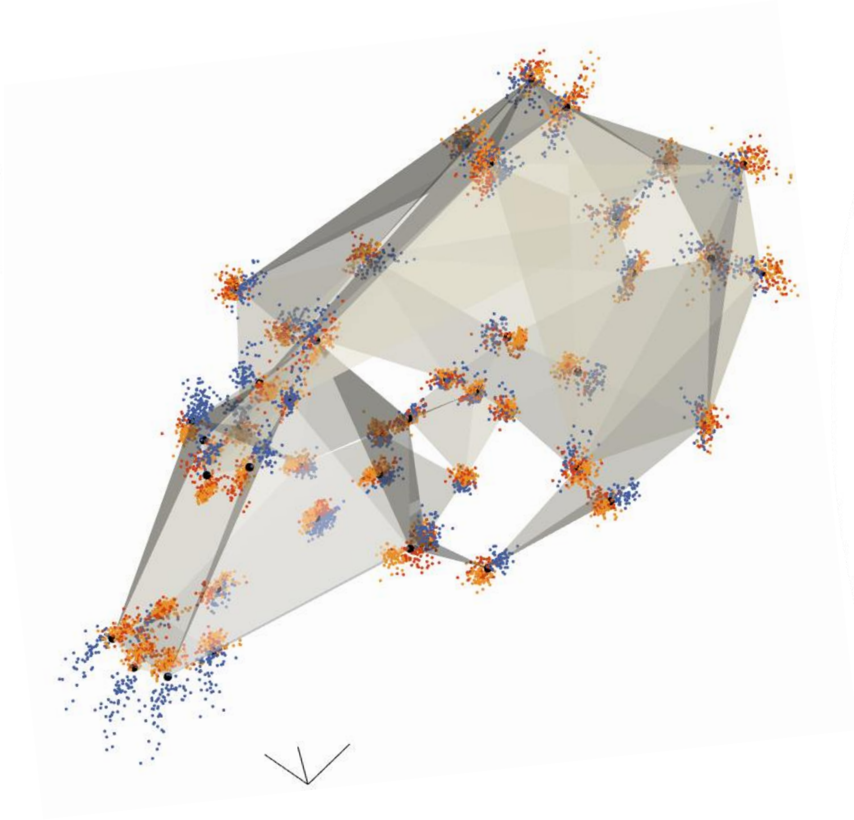
Modeling evolutionary accessibility in phenotype space

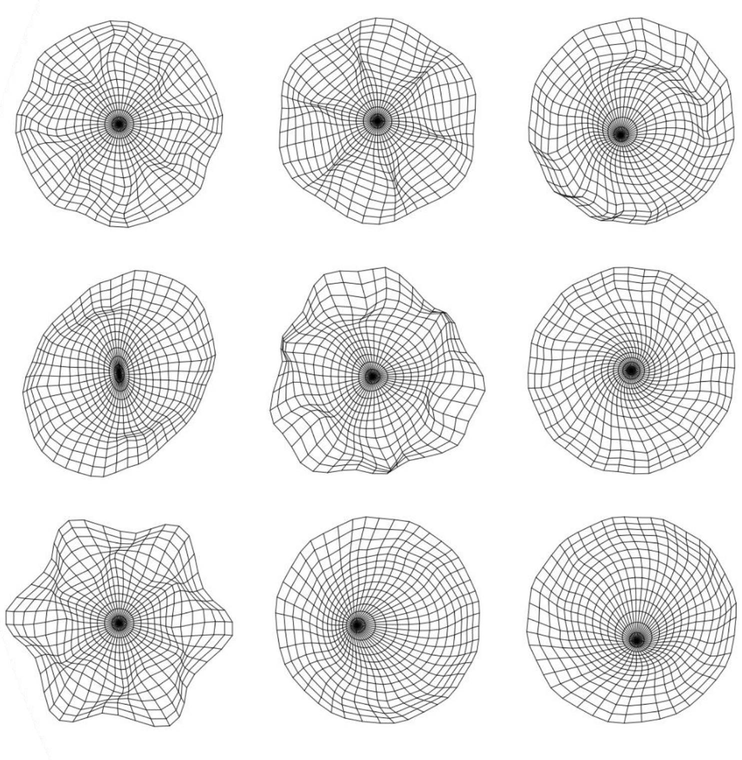
Sylvain Gerber
Muséum national d'Histoire naturelle
sylvain.gerber@mnhn.fr

Rencontre de la chaire MMB
5 décembre 2023

Morphological evolution and morphometrics

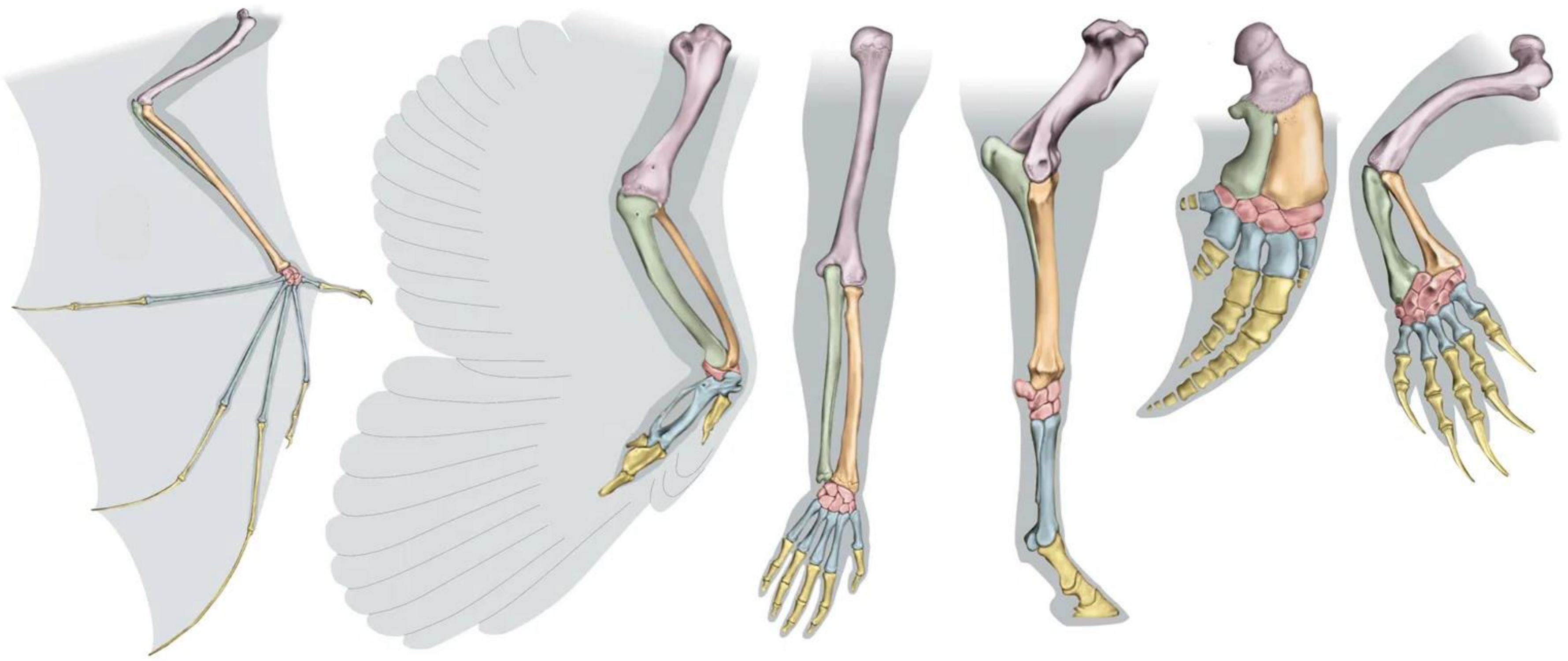






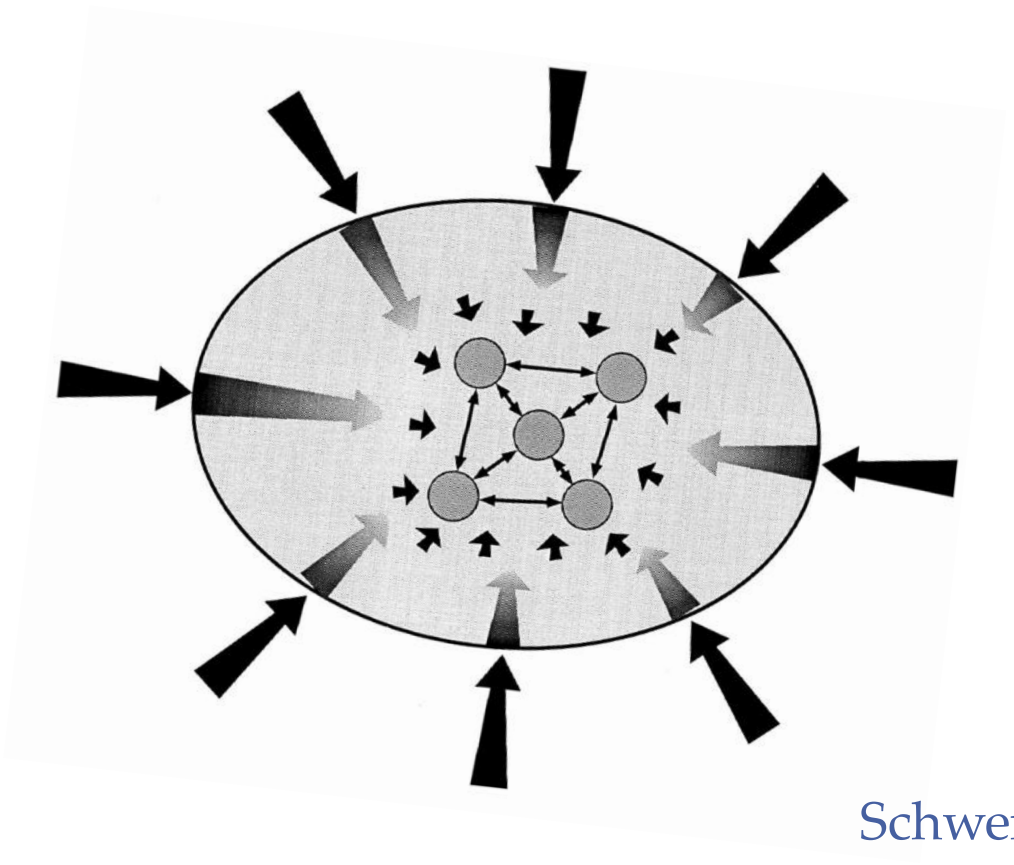
General context

An important but challenging aim of evolutionary biology and paleobiology is to determine the factors that have driven and structured past and current patterns of phenotypic evolution



General context

Within a lineage, the state of adaptation and the evolvability of a phenotype at a given life stage is determined by the interplay of direct and indirect forms of external selection and lineage-specific variational constraints (including internal selection).



Schwenk & Wagner (2000)

General context

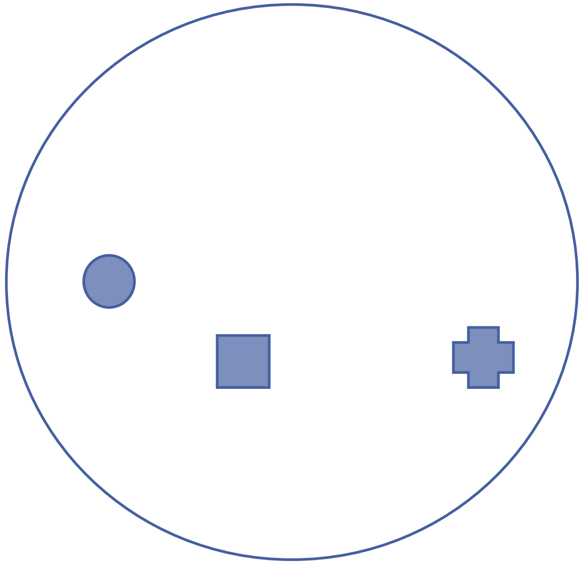
Variation *vs.* variability (Wagner & Altenberg 1996)

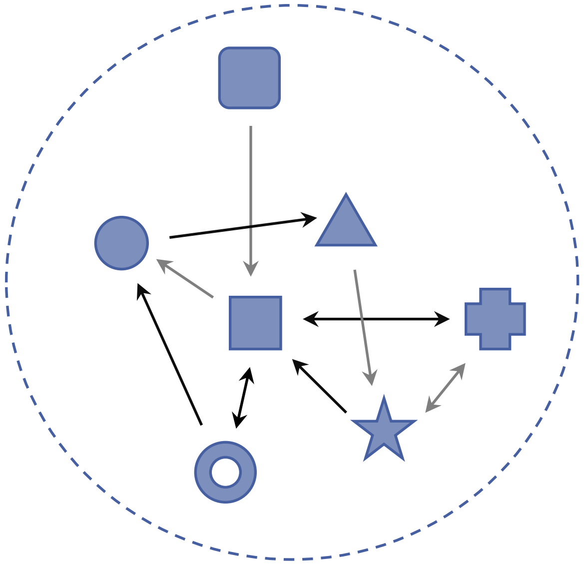
- Variation

  The observed phenotypic differences among individuals within a population or among taxa within a clade.

- Variability

  The potential or propensity of a phenotype to vary.



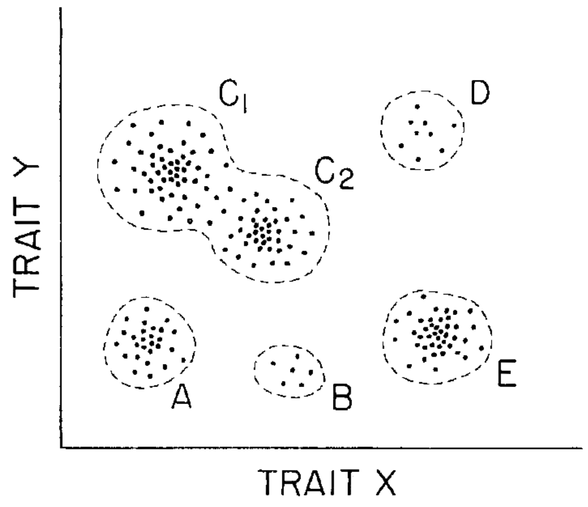


Approaches to phenotypic evolution

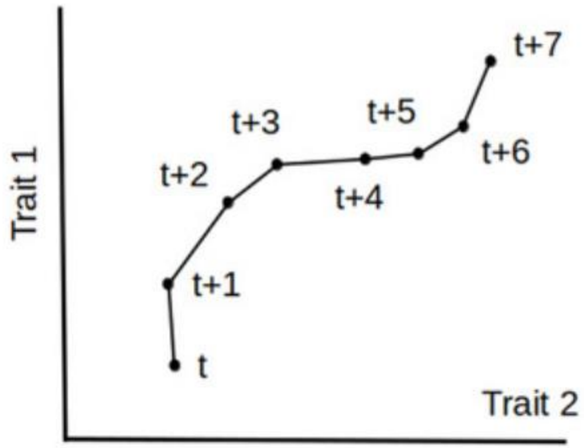
- Empirical studies document patterns of variation

- Patterns of variation are generally cast in terms of distribution (of variants) or displacement (path) in phenotype space.

- While the notion of constraint on variation is widely acknowledged as an important class of processes involved in phenotypic evolution, it does not seem to have permeated the methodological framework of phenotype space and morphospace analyses.



Alberch (1991)



Salazar-Ciudad & Cano-Fernández (2023)

Morphospaces

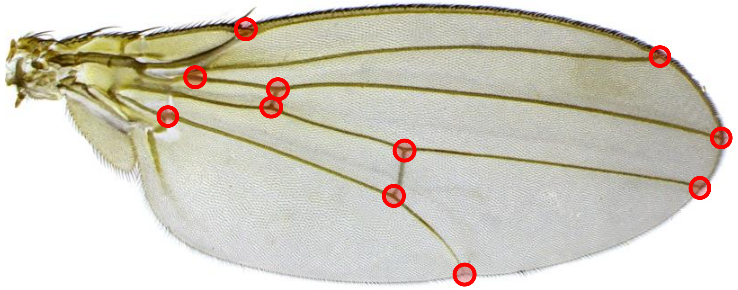
- Morphospace is an umbrella term for a variety of abstract or visual representations of morphological patterns.

- Morphospace:

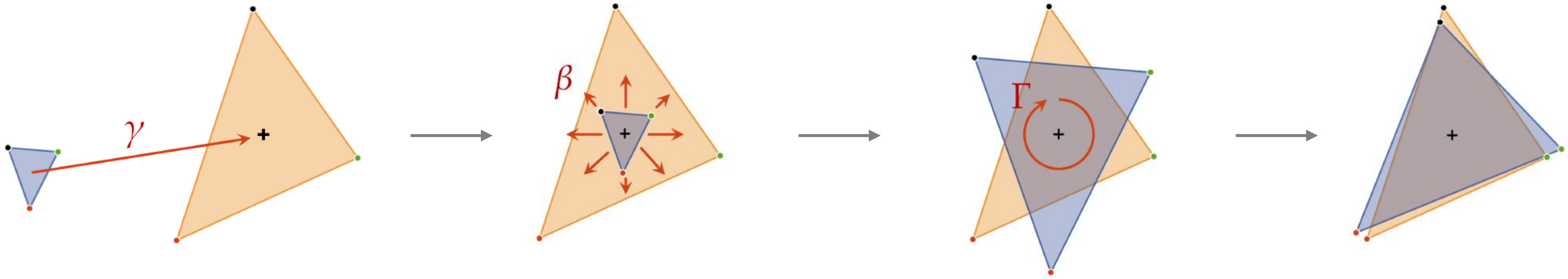
  A set of morphological variants equipped with some measure of distance reflecting their geometric, structural, or anatomical (dis)similarities.

Geometric morphometrics

- Procrustes analysis, Procrustes distance(s), and tangent projection



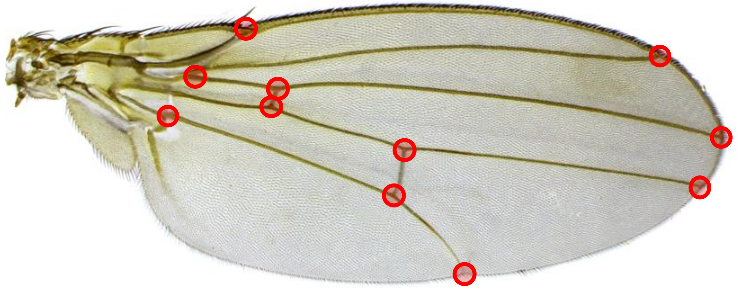
p landmarks in k dimensions



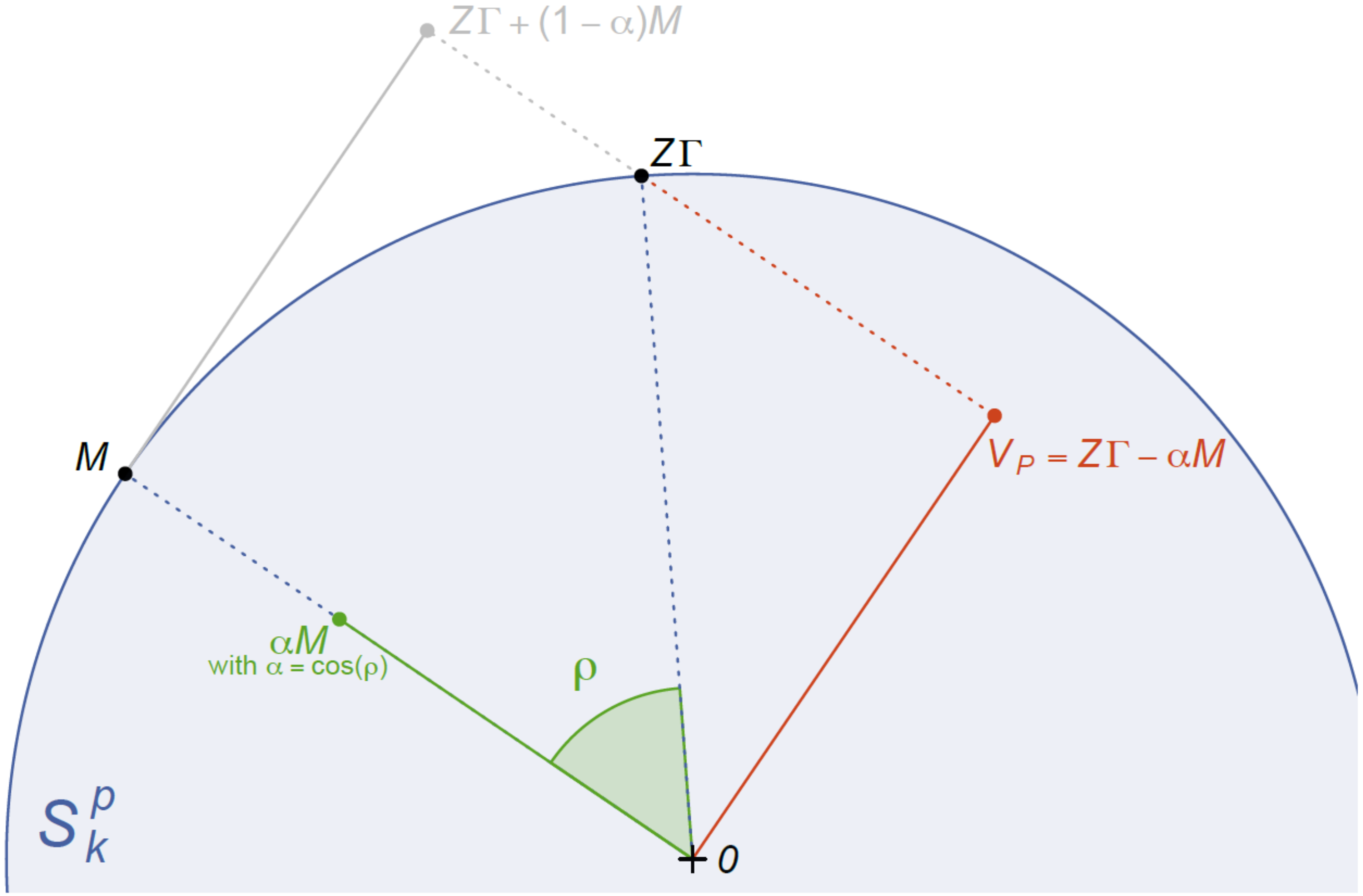
with $\gamma \in \mathbb{R}^k$, $\beta \in \mathbb{R}^+$, and $\Gamma \in SO(k)$

Geometric morphometrics

- Procrustes analysis, Procrustes distance(s), and tangent projection

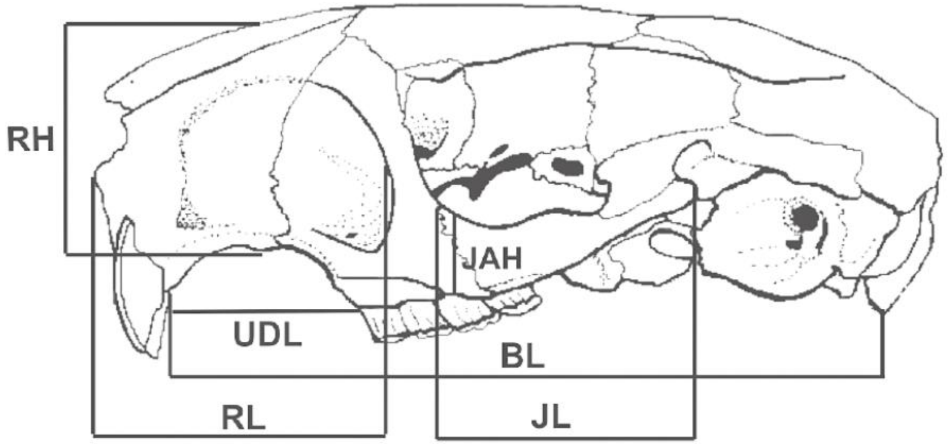


p landmarks in k dimensions

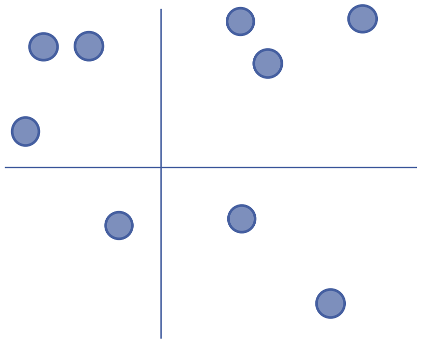


Traditional (multivariate) morphometrics

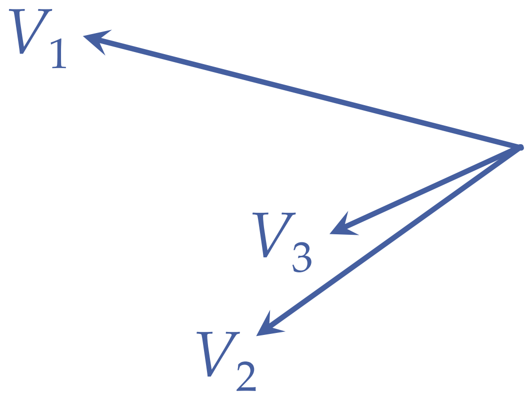
- Individuals described as vectors of k measurements.



Space of individuals
(a morphospace)



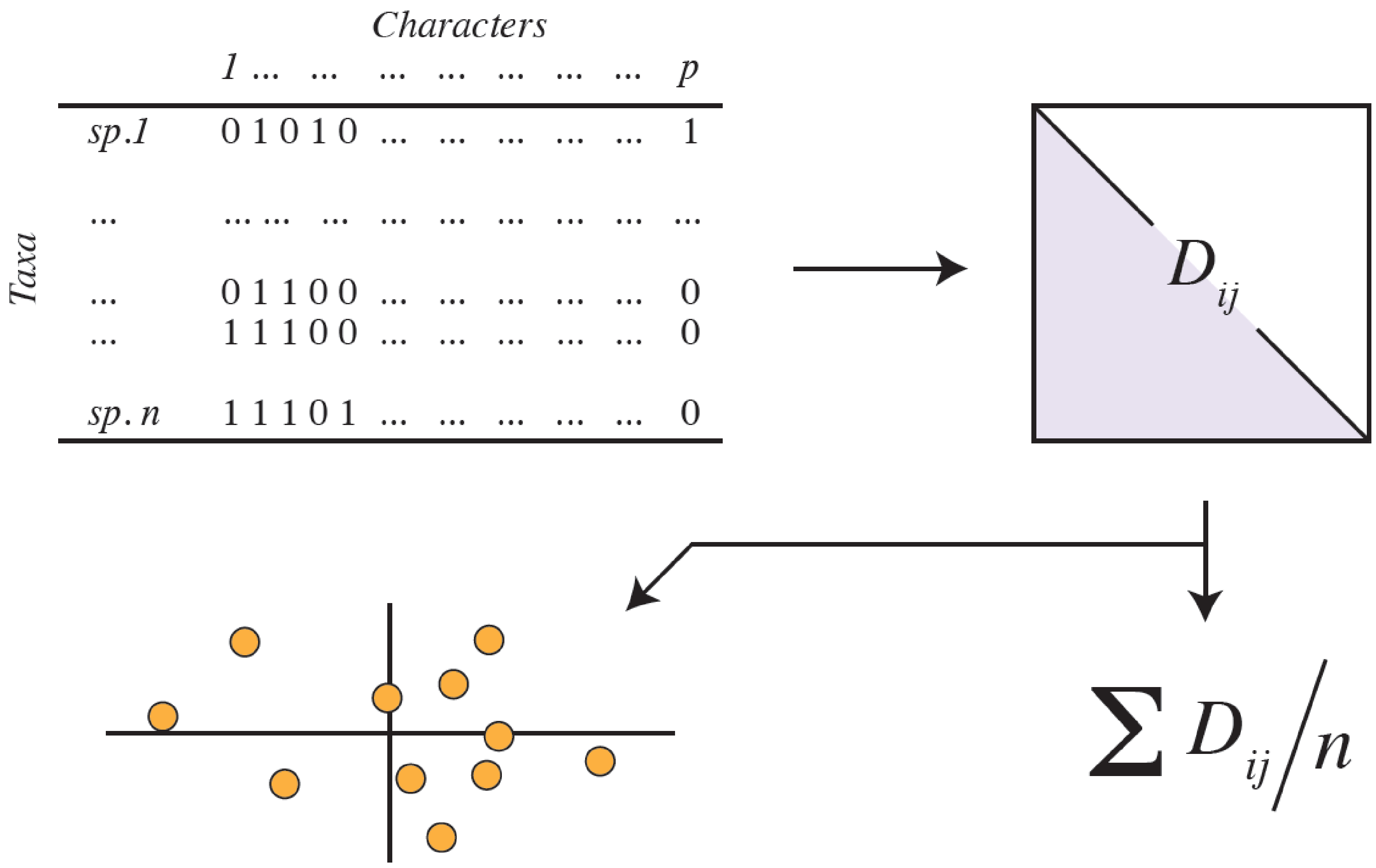
Space of variables



- Mosimann’s framework and requirements for the measurement function (Mosimann 1970, 1975).

- (a comment on affine spaces)

Discrete character schemes

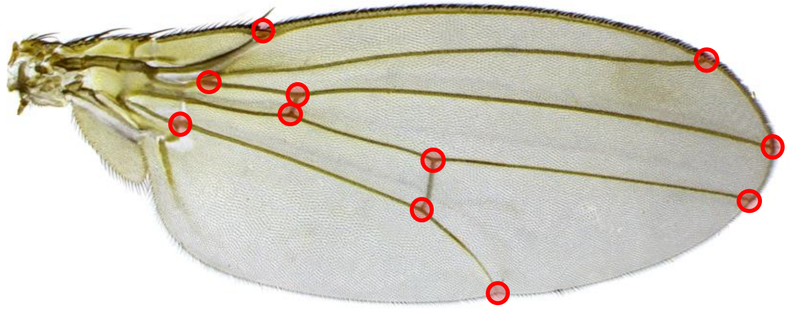


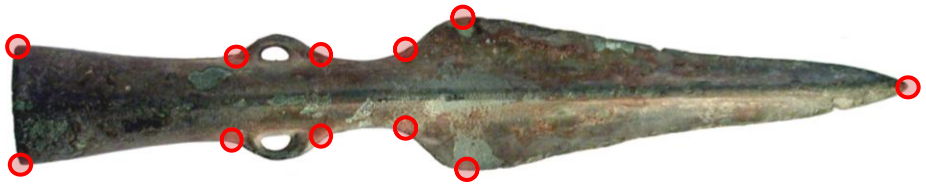
$$\sum D_{ij} / n$$

Limits of current approaches and shift in perspective

- Morphospaces are not ideally suited for the mechanistic interpretation of phenotypic patterns:

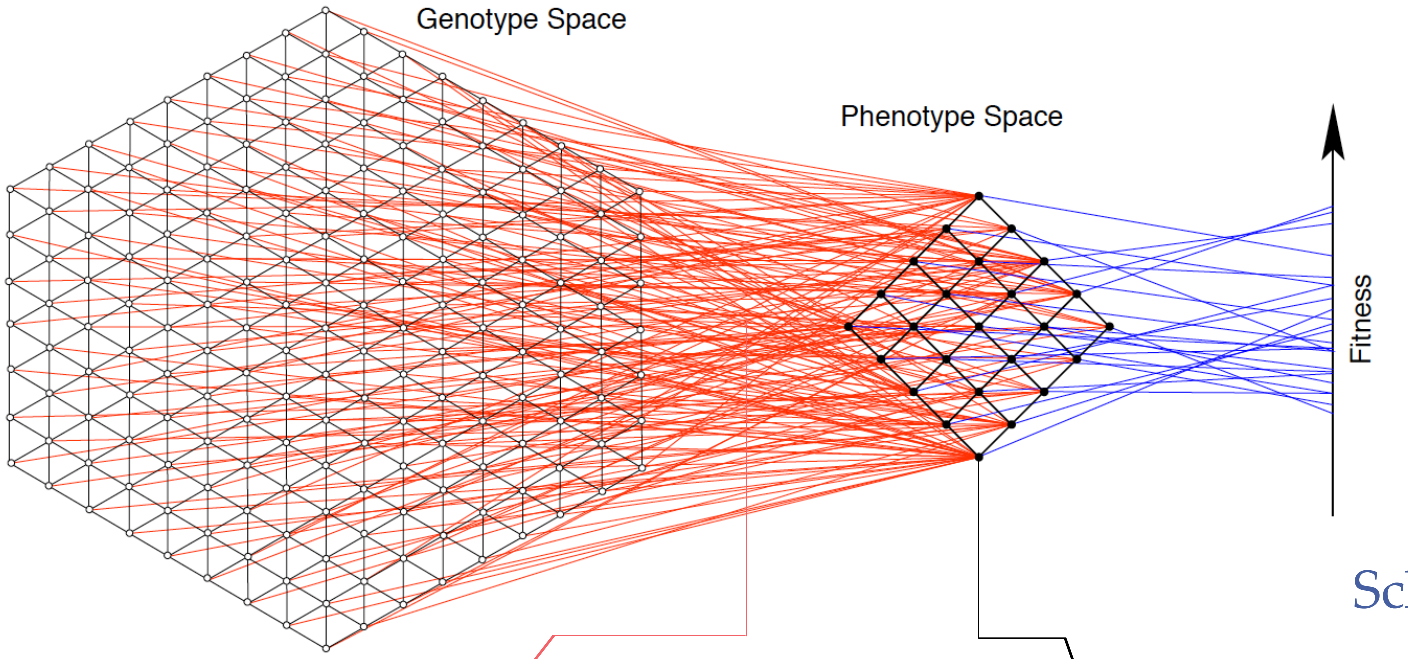
  - Morphospaces display patterns of variation.
  - Morphospaces are non-specific to the objects they ordinate





- One needs to consider the phenotype space topology naturally induced by the phenotype's variational properties.

Limits of current approaches and shift in perspective



Schuster (2002)

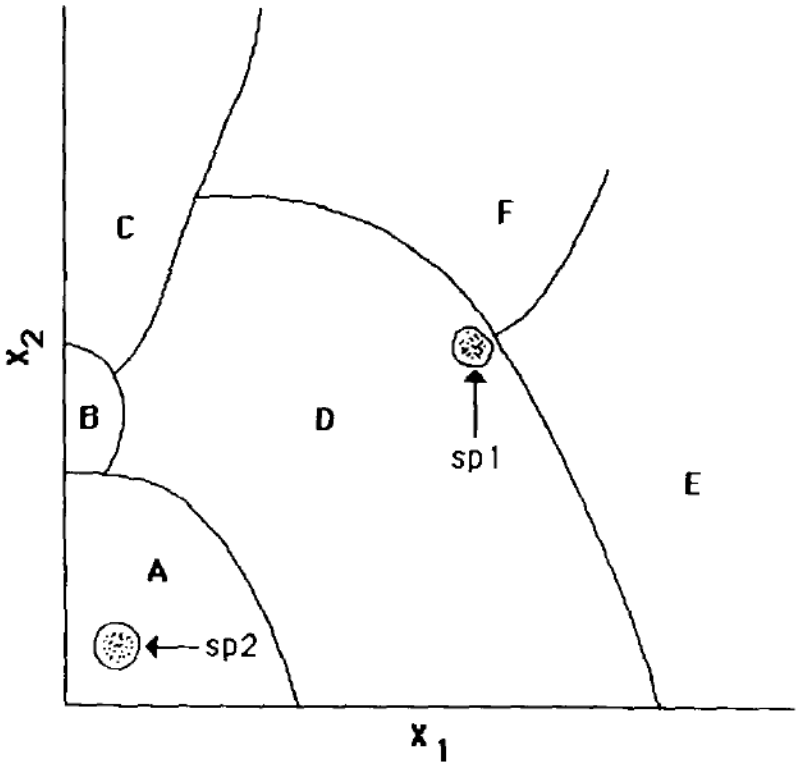
Evolutionary Developmental Biology (Evo-Devo)

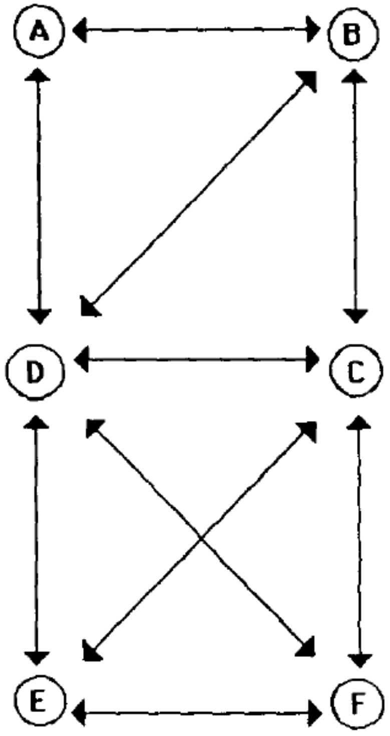
Morphometrics and Morphospaces

Previous work

"The true neighborliness of two developmental patterns is not established by raw comparison of the patterns, but implicitly refers to changes required in the developmental mechanism which generates them."

Kauffman (1983)



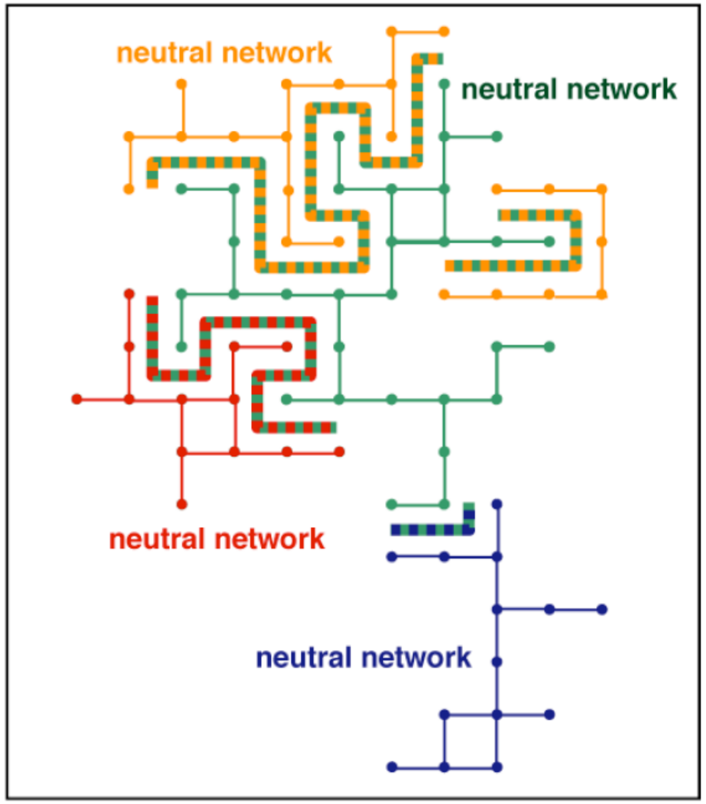


Alberch (1991)

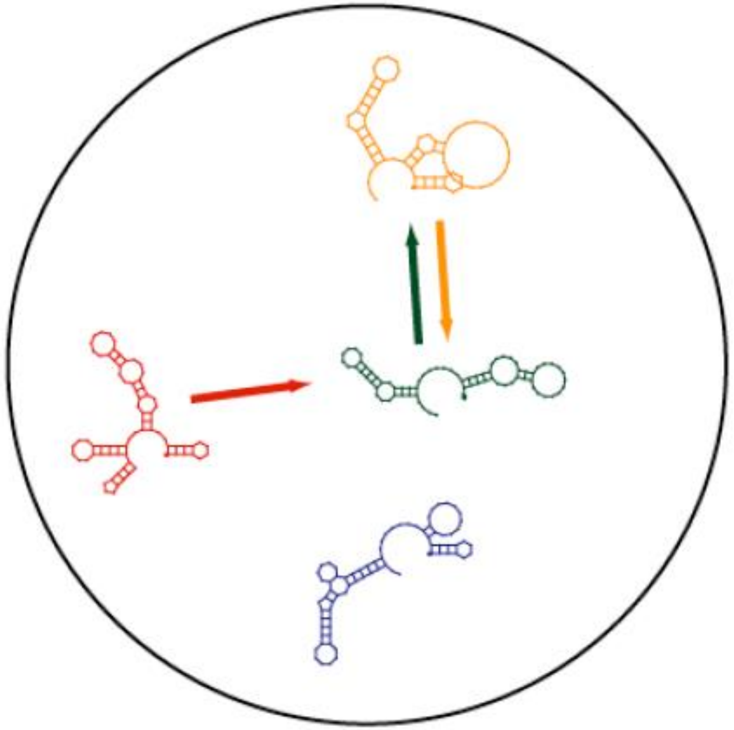
Previous work

The mapping from RNA sequences to RNA shapes as a model of GP-map

Genotype space

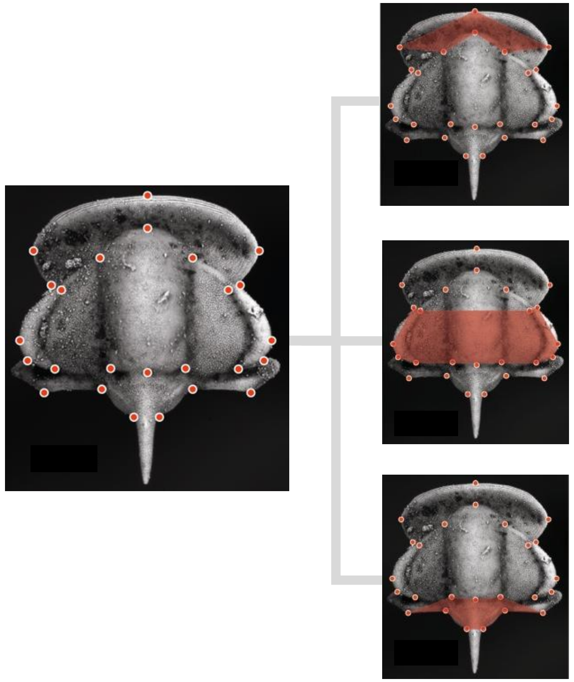


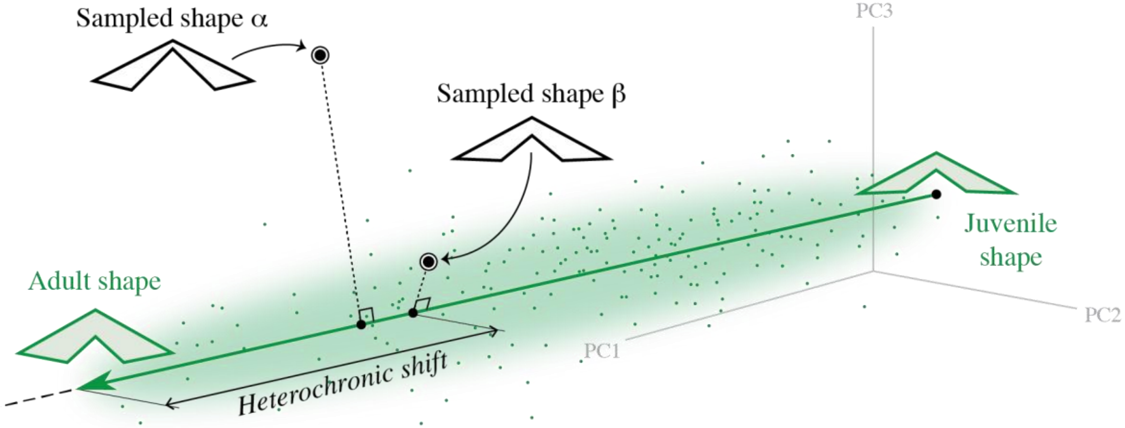
Phenotype space induced by the GP-map

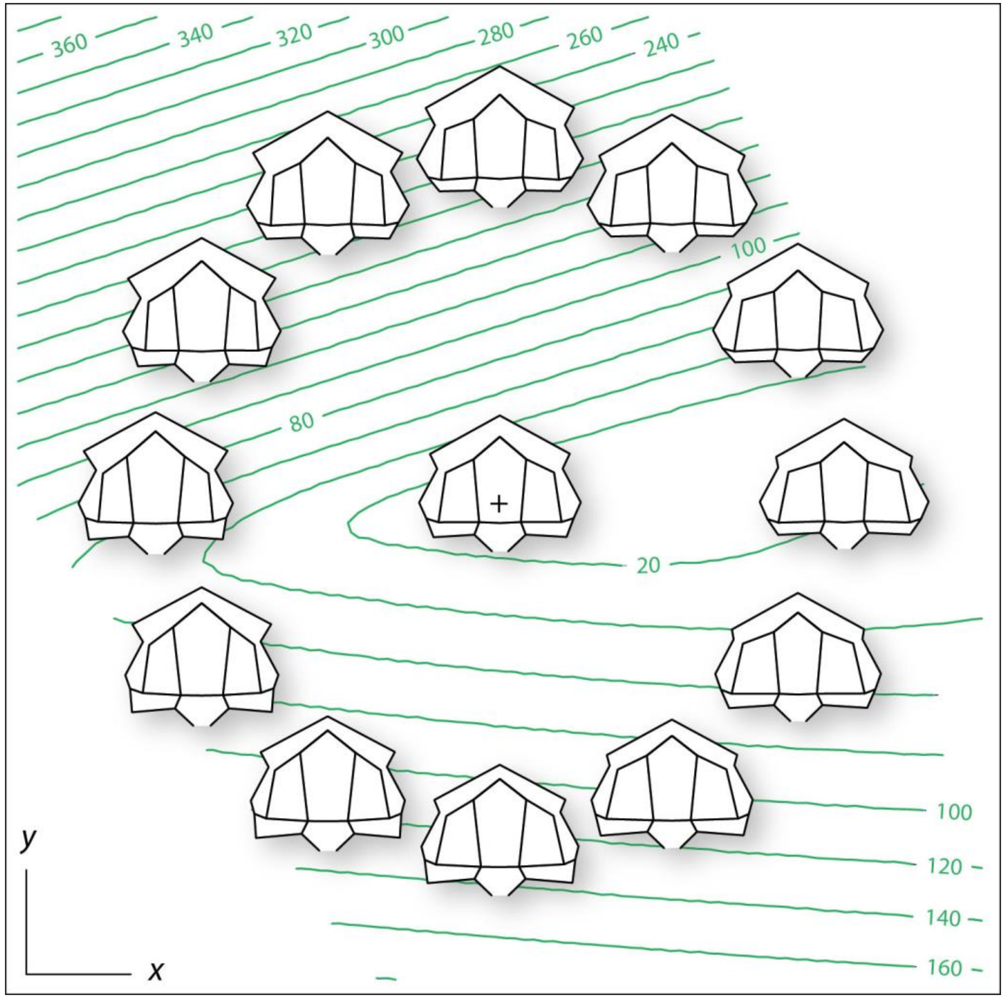


Fontana (2006)

Previous work



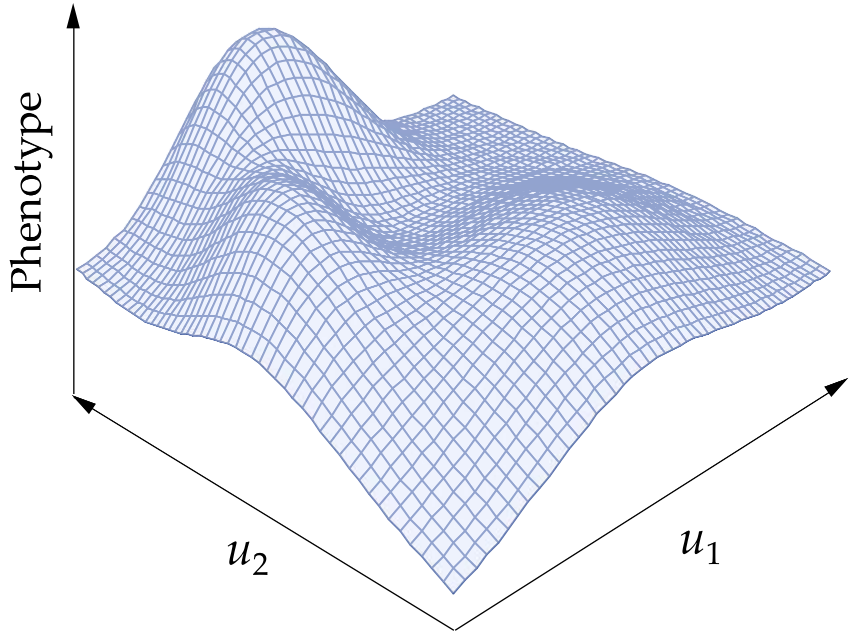


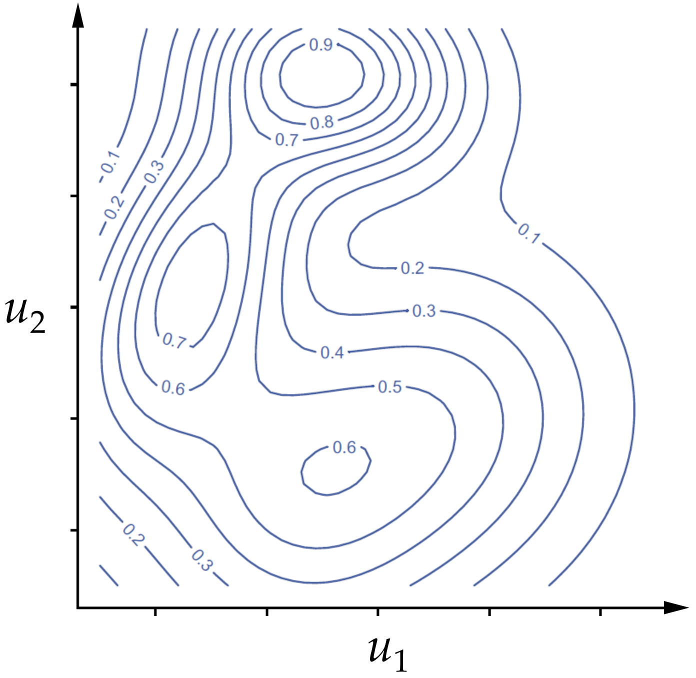


Gerber (2014)

Phenotype landscape

A phenotype landscape* is a representation of the possible values of a phenotype (phenotypic variants) as a function ϕ of some underlying factors $u_1, \ldots, u_n$.





* See Rice (1998, 2004)

Phenotype landscape

- The phenotype can be any measurable trait

- The underlying factors can be:

  Degree of expression of gene

  Degree of enzymatic activity

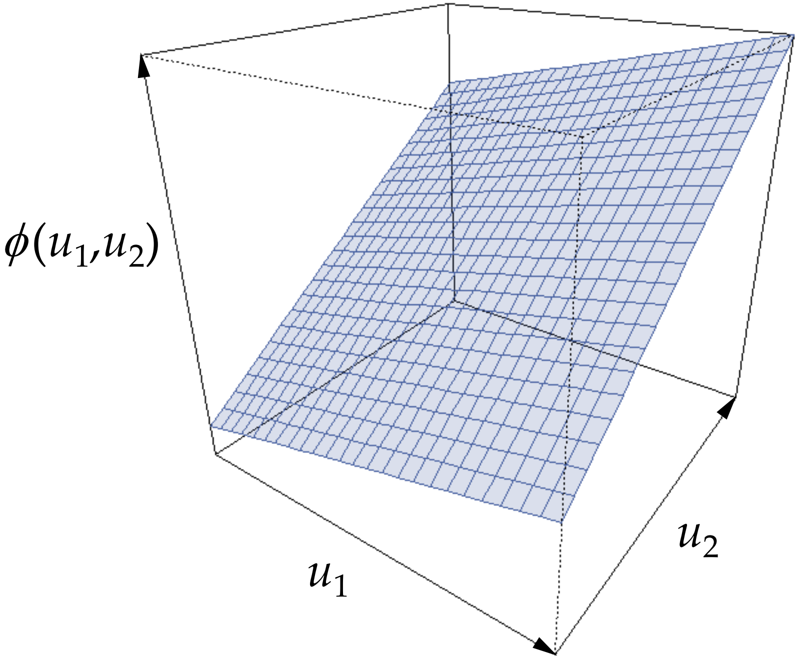
  Quantitative developmental factors

  (environmental, i.e., non-heritable, factors)

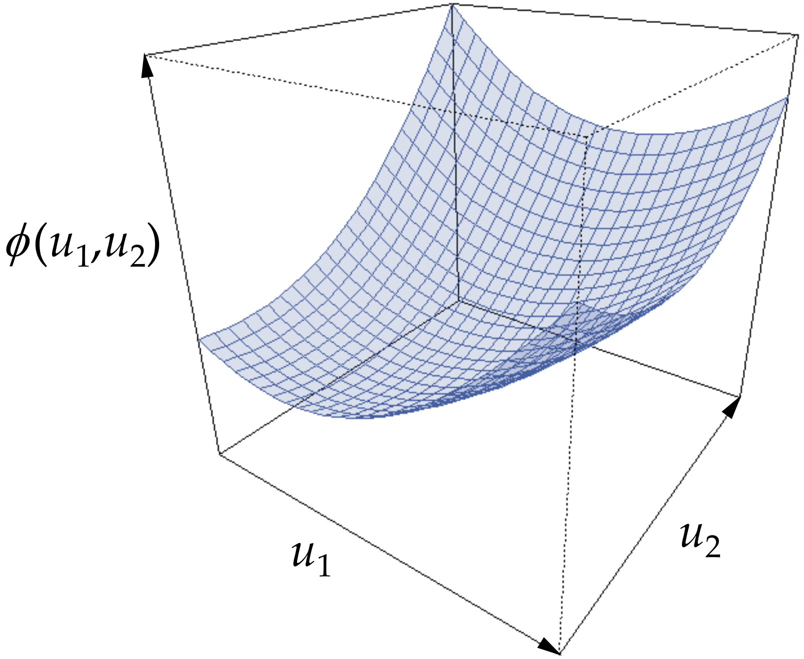
- Important aspect: the nature of the interactions among these factors

Phenotype landscape

The type of interaction determines the shape of the landscape



Additive interaction
uncurved landscape



Nonadditive interaction
curved landscape
(epistasis)

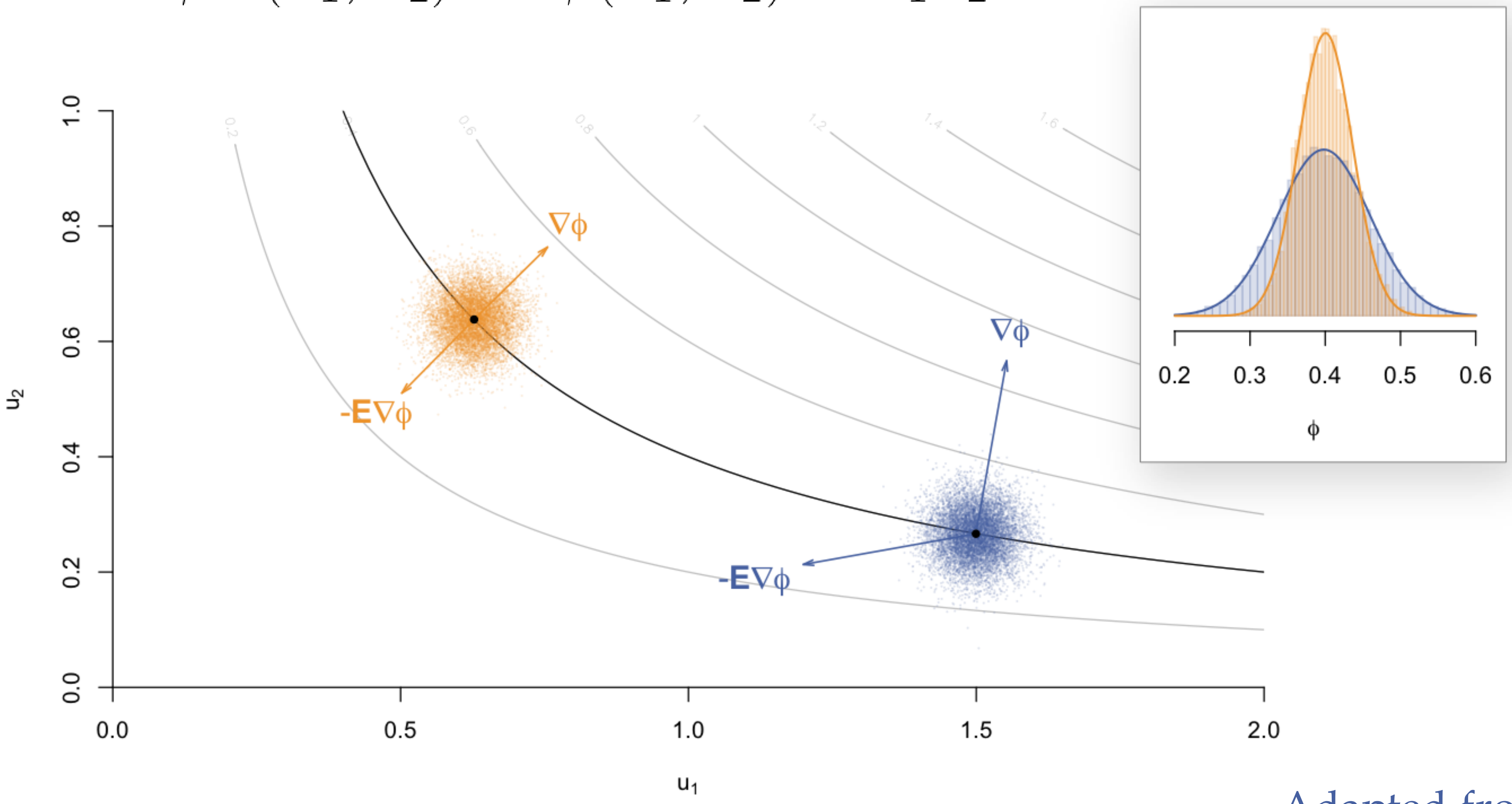
Phenotype landscape

A population only “experiences” the local geometry of the landscape

- The gradient vector $$\nabla\phi = \begin{pmatrix} \frac{\partial\phi}{\partial u_1} \\ \vdots \\ \frac{\partial\phi}{\partial u_n} \end{pmatrix}$$
- The epistasis matrix $$\mathbf{E} = \begin{pmatrix} \frac{\partial^2\phi}{\partial u_1^2} & \cdots & \frac{\partial^2\phi}{\partial u_1 \partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\phi}{\partial u_n \partial u_1} & \cdots & \frac{\partial^2\phi}{\partial u_n^2} \end{pmatrix}$$
- The Laplacian operator $$\nabla^2\phi = \sum_{i=1}^{n} \frac{\partial^2\phi}{\partial u_i^2}$$

Phenotype landscape

$$\phi : (u_1, u_2) \mapsto \phi(u_1, u_2) = u_1 u_2$$



Adapted from Rice (1998)

Phenotype representation

Notion of level set:

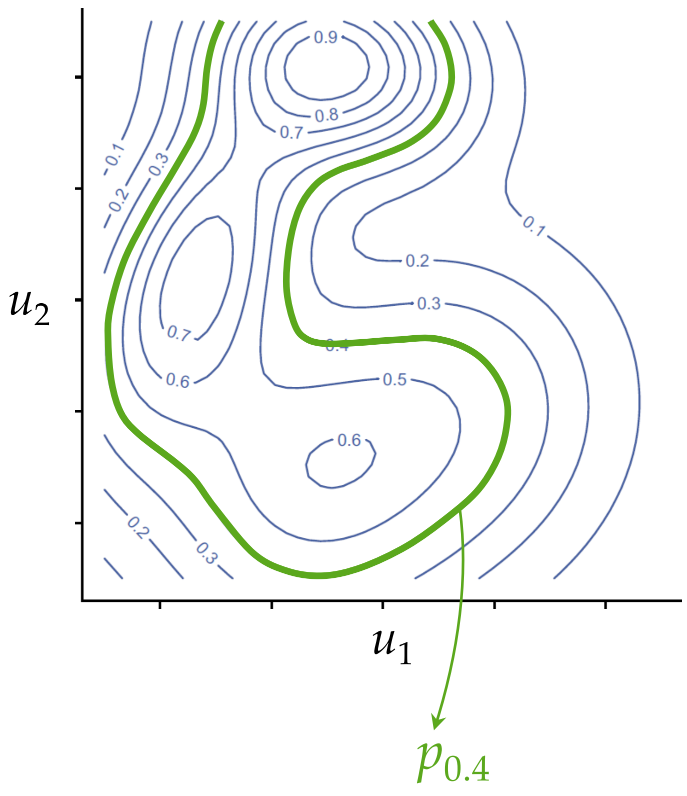
$$S_c(\phi) = \{(\alpha_1, \dots, \alpha_n) \in U : \phi(\alpha_1, \dots, \alpha_n) = c\}$$

The developmental function defines an equivalence relation $\sim_\phi$ on U, with the level sets as equivalence classes of "developmental n-tuples".

$$(\alpha_1, \dots, \alpha_n) \sim_\phi (\beta_1, \dots, \beta_n) \Leftrightarrow \phi(\alpha_1, \dots, \alpha_n) = \phi(\beta_1, \dots, \beta_n)$$

The phenotype set P is the quotient set:

$$P = U / \sim_\phi$$



Phenotype representation

Notion of level set:

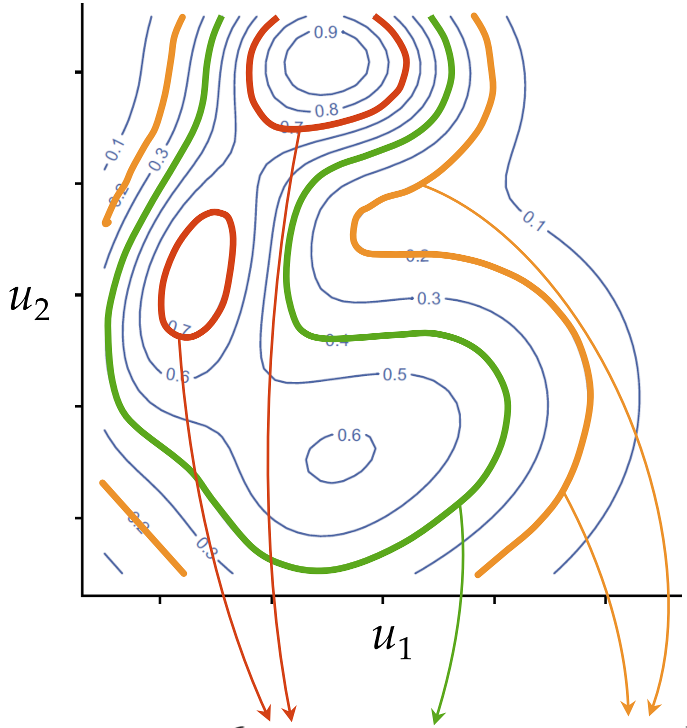
$$S_c(\phi) = \{(\alpha_1, \dots, \alpha_n) \in U : \phi(\alpha_1, \dots, \alpha_n) = c\}$$

The developmental function defines an equivalence relation $\sim_\phi$ on U, with the level sets as equivalence classes of "developmental n-tuples".

$$(\alpha_1, \dots, \alpha_n) \sim_\phi (\beta_1, \dots, \beta_n) \Leftrightarrow \phi(\alpha_1, \dots, \alpha_n) = \phi(\beta_1, \dots, \beta_n)$$

The phenotype set P is the quotient set:

$$P = U / \sim_\phi$$



$$P = \{ p_{0.7}, \dots, p_{0.4}, \dots, p_{0.2} \}$$

Phenotype representation

Notion of level set:

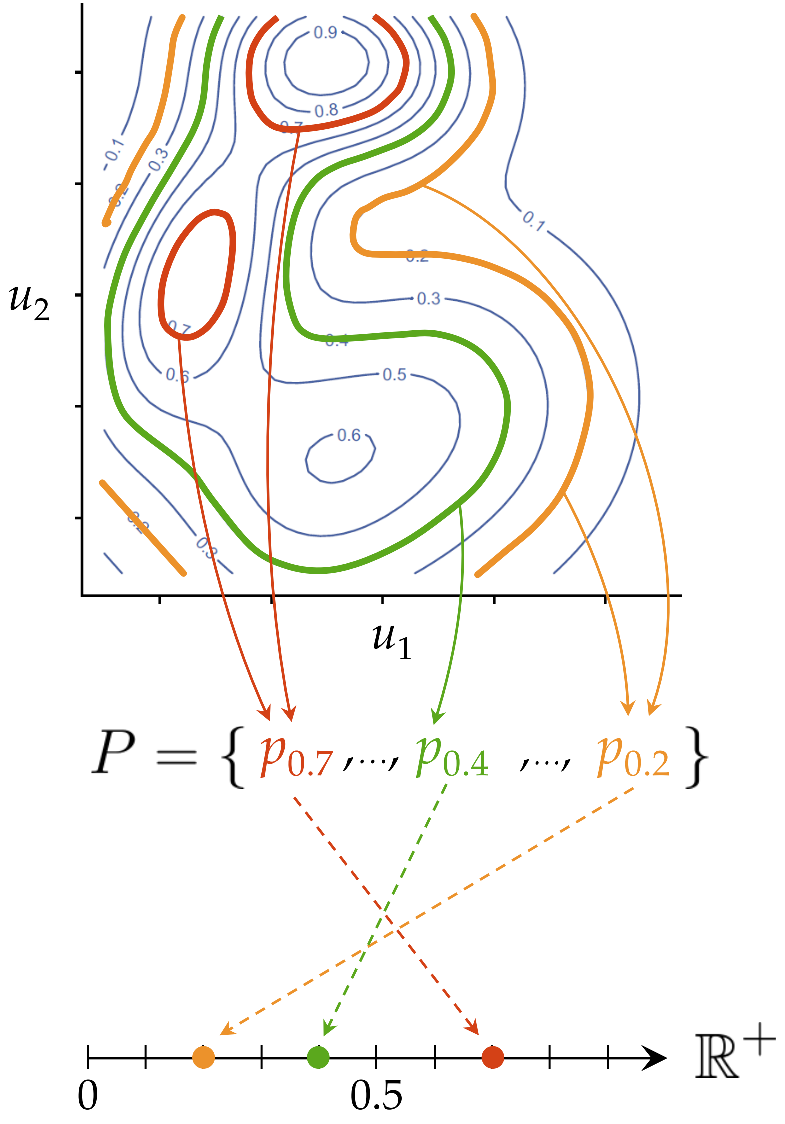
$$S_c(\phi) = \{(\alpha_1, ..., \alpha_n) \in U : \phi(\alpha_1, ..., \alpha_n) = c\}$$

The developmental function defines an equivalence relation $\sim_\phi$ on U, with the level sets as equivalence classes of "developmental n-tuples".

$$(\alpha_1, ..., \alpha_n) \sim_\phi (\beta_1, ..., \beta_n) \Leftrightarrow \phi(\alpha_1, ..., \alpha_n) = \phi(\beta_1, ..., \beta_n)$$

The phenotype set P is the quotient set:

$$P = U/\sim_\phi$$



Defining accessibility

Notion of superlevel set

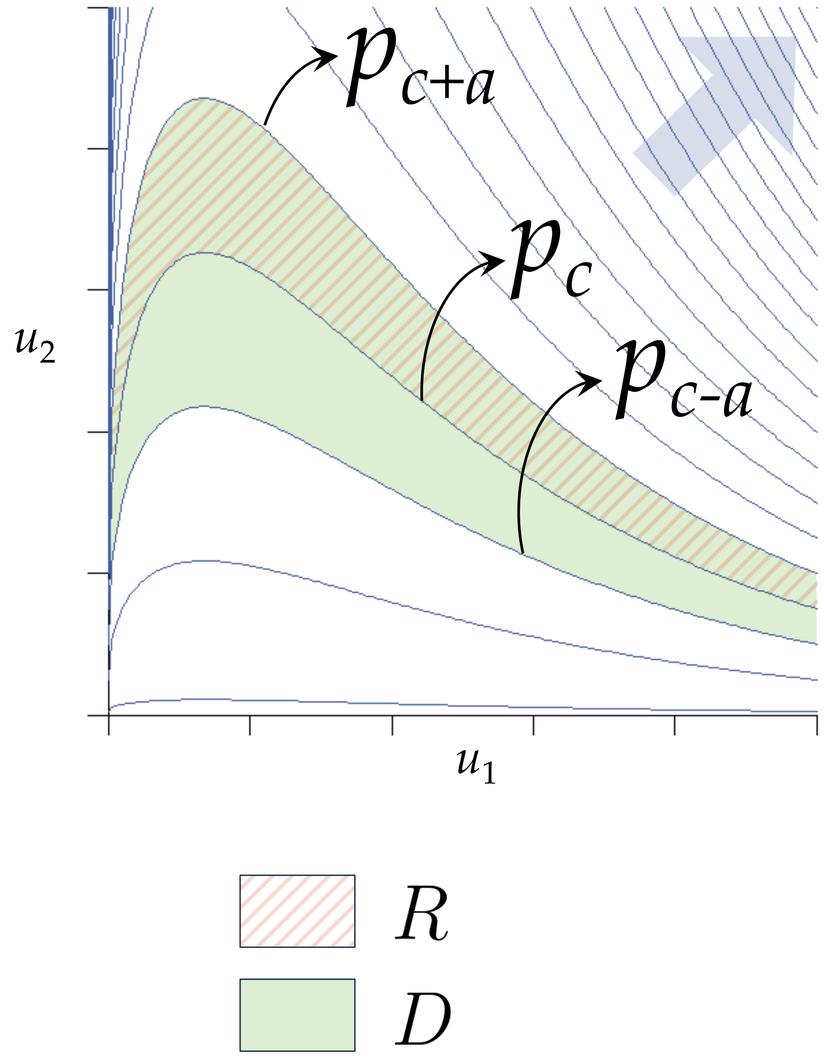
$$S_c^+(\phi) = \{(\alpha_1, ..., \alpha_n) \in U : \phi(\alpha_1, ..., \alpha_n) \geq c\}$$

Let p_c and p_{c+a} be two possible states of a univariate phenotype with values c and $c+a$.

We delineate two subsets of U with the use of superlevel sets:

$$R = S_c^+(\phi) \setminus S_{c+a}^+(\phi)$$

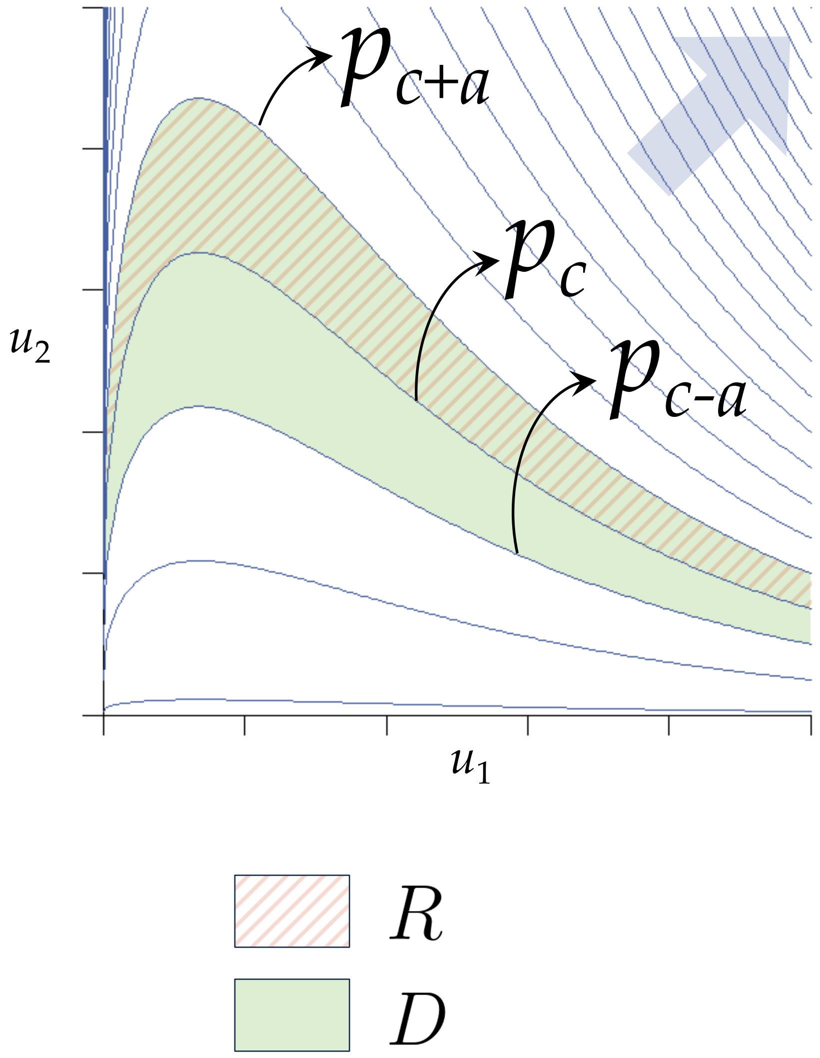
$$D = S_{c-a}^+(\phi) \setminus S_{c+a}^+(\phi)$$



Defining accessibility

The accessibility of p_{c+a} from p_c is measured as:

$$A(p_{c+a} \curvearrowleft p_c) = \frac{V_n(R)}{V_n(D)}$$



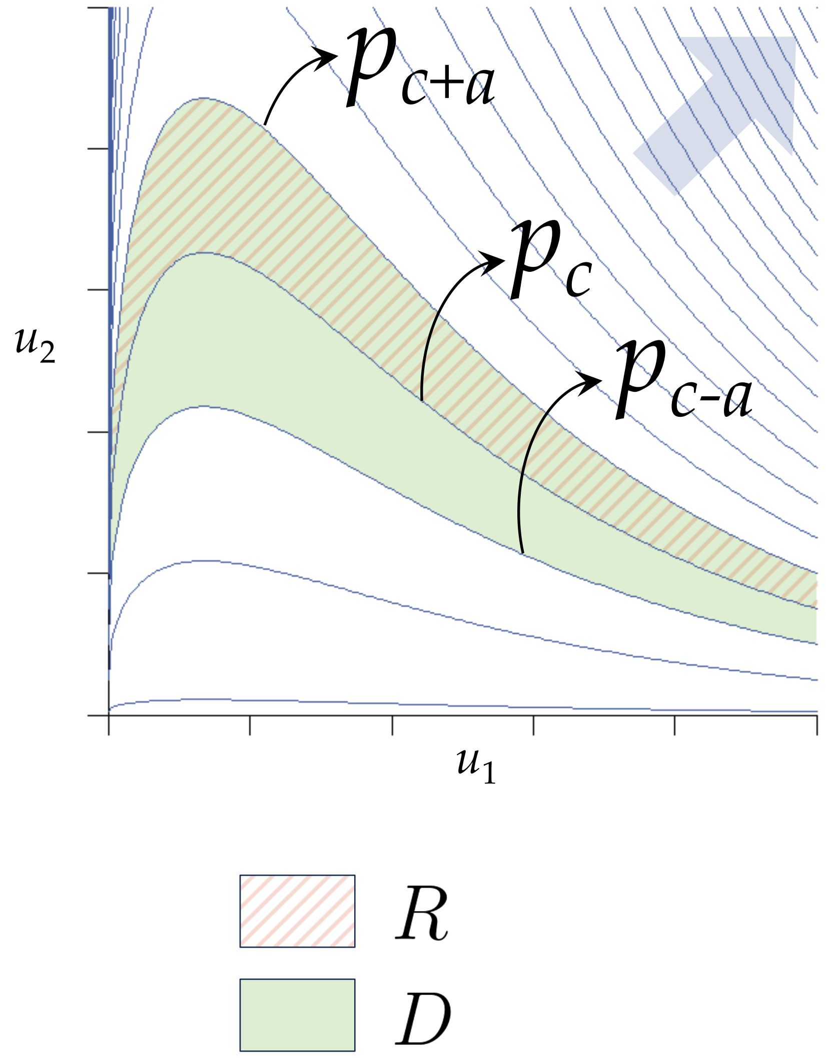
Defining accessibility

The accessibility of p_{c+a} from p_c is measured as:

$$A(p_{c+a} \curvearrowleft p_c) = \frac{V_n(R)}{V_n(D)}$$

and the inverse transition is

$$A(p_c \curvearrowleft p_{c+a}) = \frac{V_n(D)}{V_n\left(S_c^+(\phi) \setminus S_{c+2a}^+(\phi)\right)} A(p_{c+a} \curvearrowleft p_c)$$



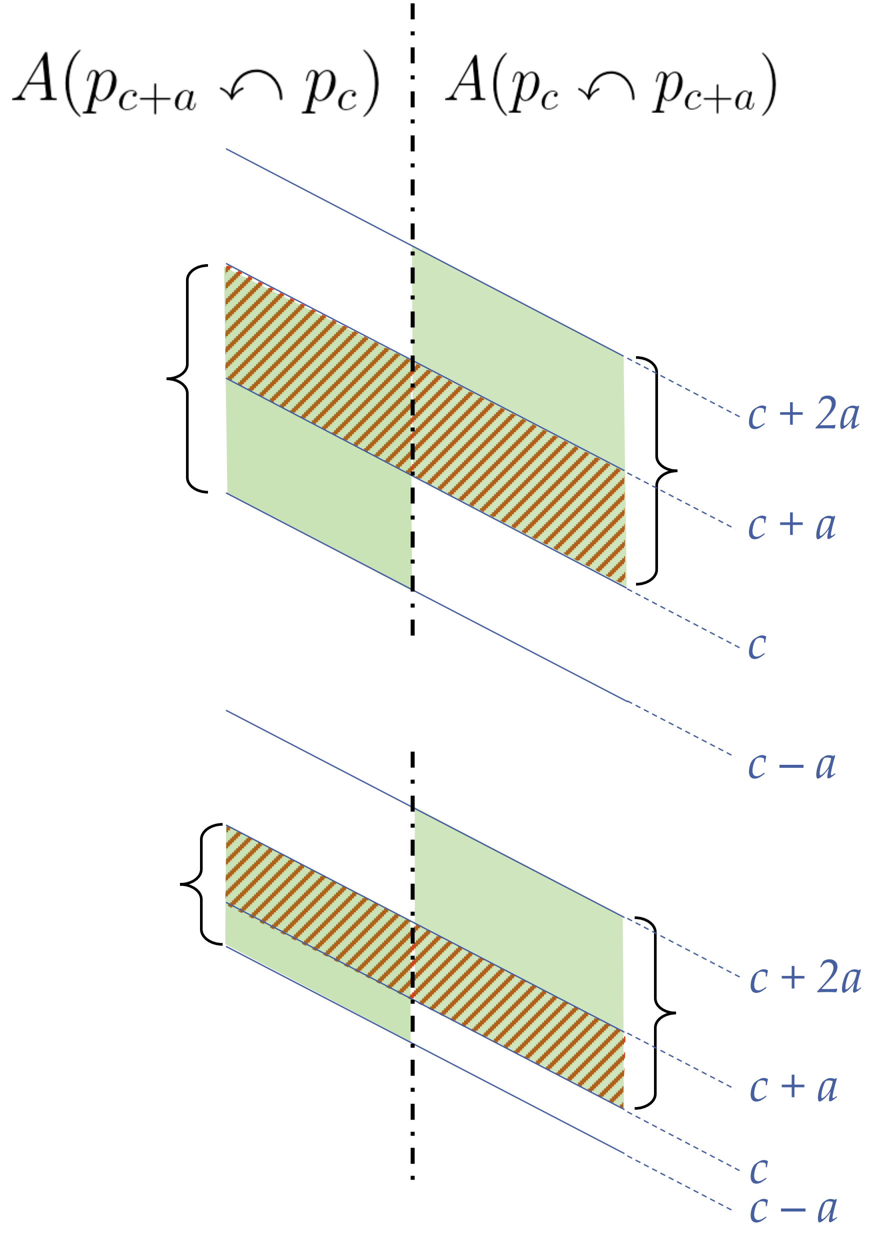
Defining accessibility

The accessibility of p_{c+a} from p_c is measured as:

$$A(p_{c+a} \curvearrowleft p_c) = \frac{V_n(R)}{V_n(D)}$$

and the inverse transition is

$$A(p_c \curvearrowleft p_{c+a}) = \frac{V_n(D)}{V_n\left(S_c^+(\phi) \setminus S_{c+2a}^+(\phi)\right)} A(p_{c+a} \curvearrowleft p_c)$$



Defining accessibility

The accessibility of p_{c+a} from p_c is measured as:

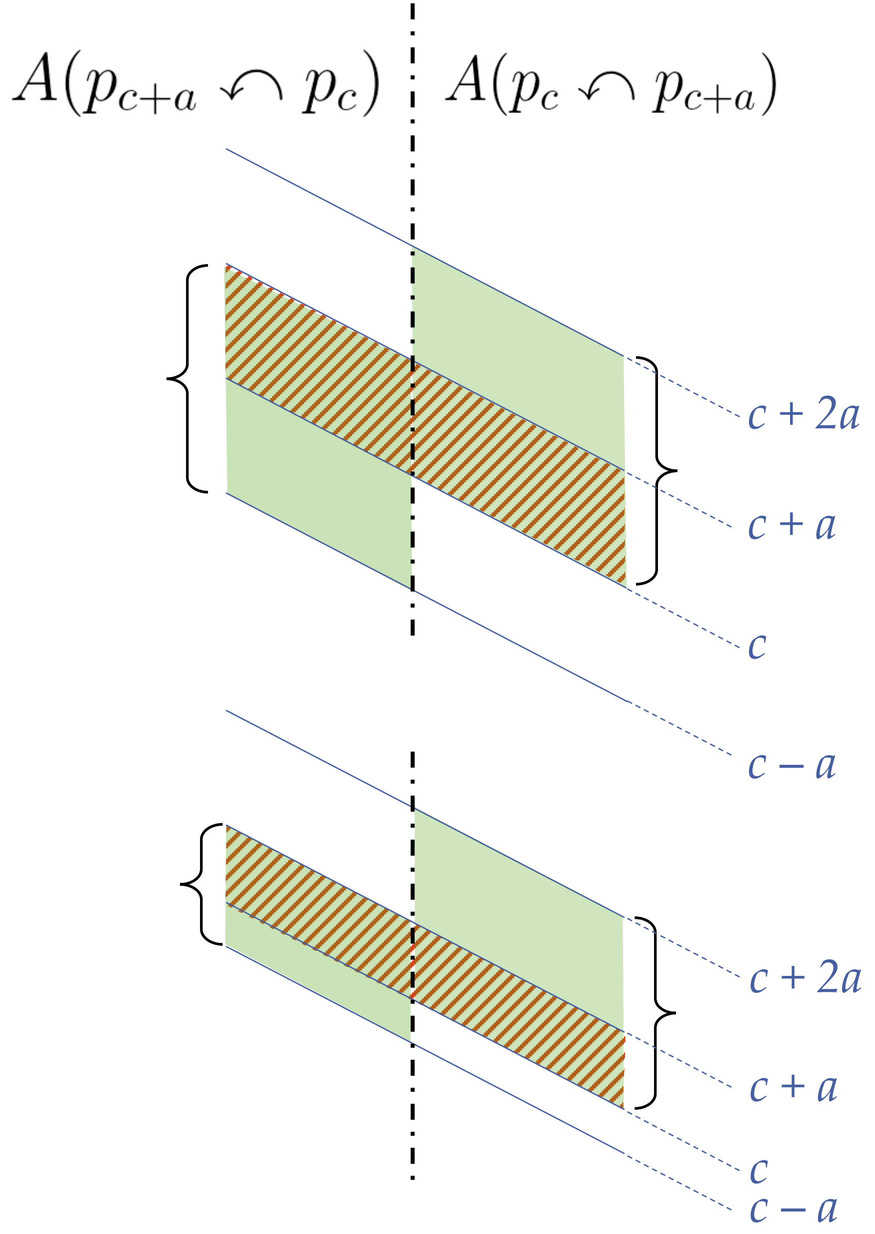
$$A(p_{c+a} \curvearrowleft p_c) = \frac{V_n(R)}{V_n(D)}$$

and the inverse transition is

$$A(p_c \curvearrowleft p_{c+a}) = \frac{V_n(D)}{V_n\left(S_c^+(\phi) \setminus S_{c+2a}^+(\phi)\right)} A(p_{c+a} \curvearrowleft p_c)$$

By definition:

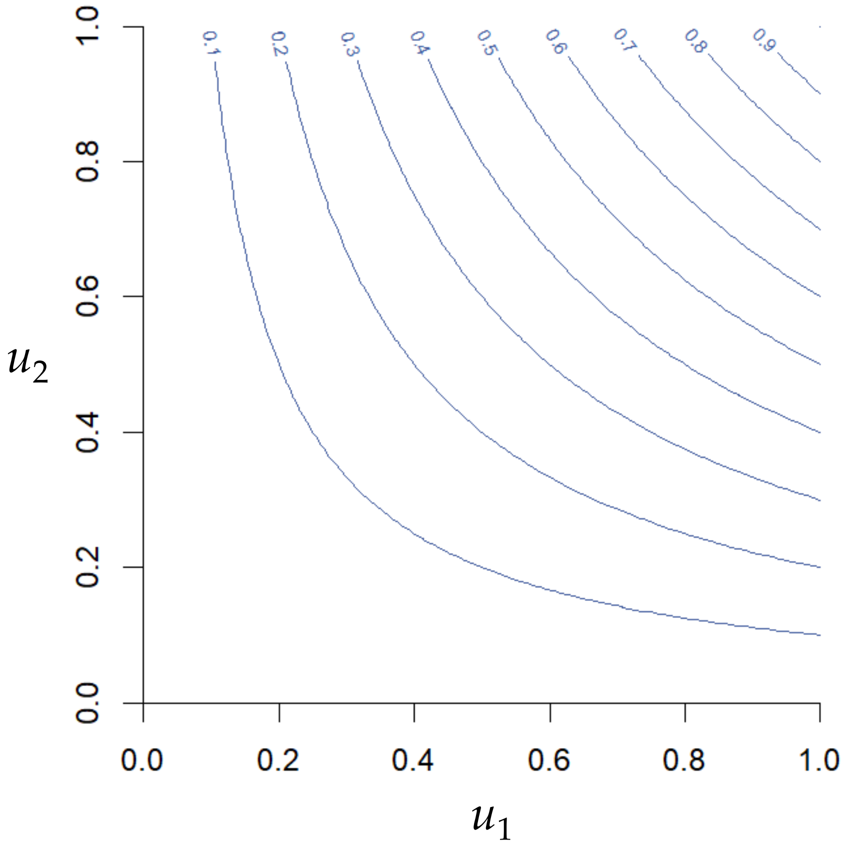
$$A(p_{c+a} \curvearrowleft p_c) = 1 - A(p_{c-a} \curvearrowleft p_c)$$



Empirical applications

We consider a univariate, quantitative phenotype controlled by two underlying factors u_1 and u_2, and with a developmental map ϕ from $U = [0,1] \times [0,1]$ to $\mathbb{R}^+$defined as:

$$(u_1, u_2) \mapsto \phi(u_1, u_2) = u_1 u_2$$



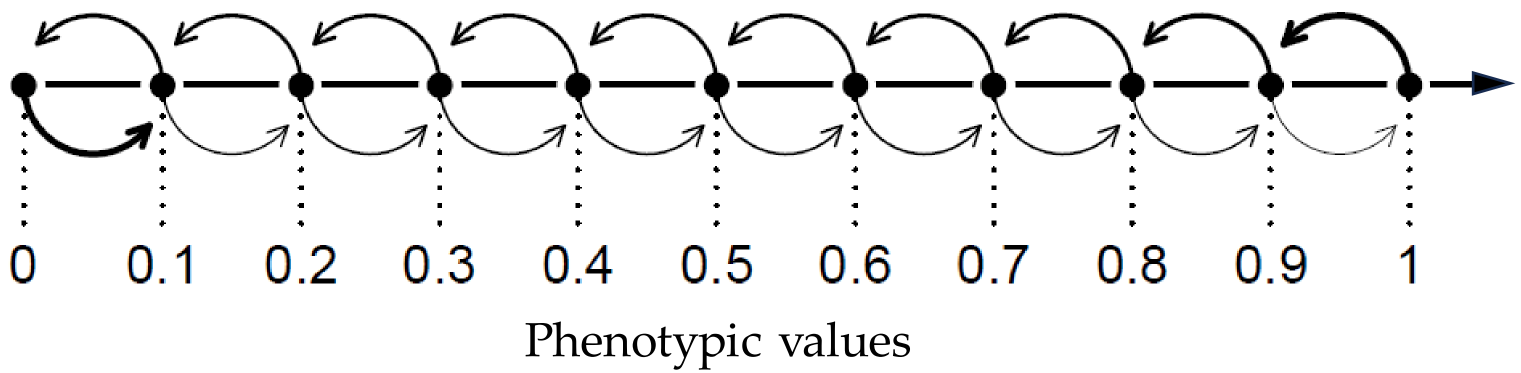
$$A(p_{c+a} \curvearrowleft p_c) = \frac{\iint_R 1 \mathrm{d}u_1 \mathrm{d}u_2}{\iint_D 1 \mathrm{d}u_1 \mathrm{d}u_2}$$

$$= \frac{\oint_{\partial R} u_1 \mathrm{d}u_2}{\oint_{\partial D} u_1 \mathrm{d}u_2} = \frac{c \log c - (c+a)\log(c+a) + a}{(c-a)\log(c-a) - (c+a)\log(c+a) + 2a}$$

Empirical applications

The accessibility structure is equivalent to a directed graph with adjacency matrix

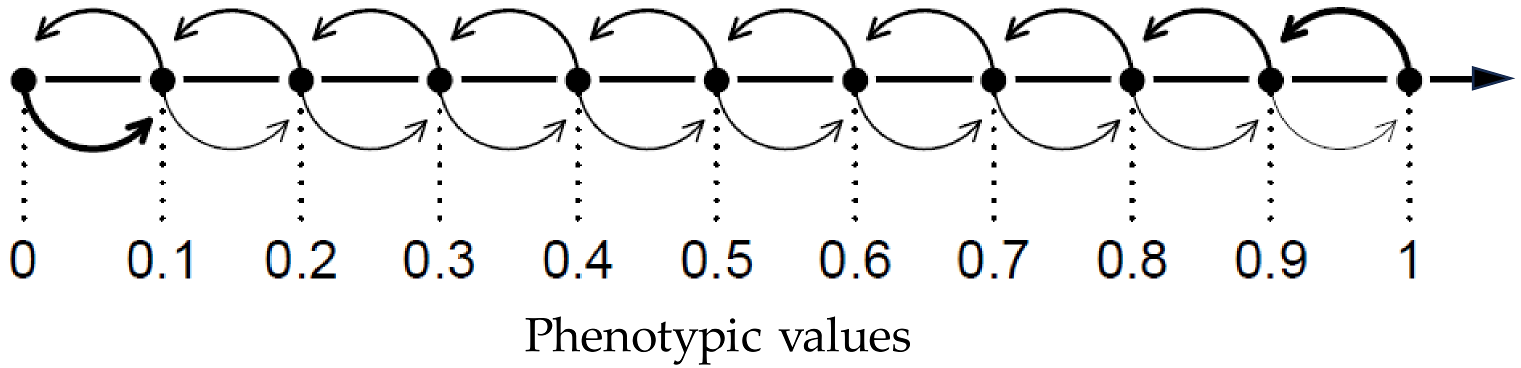
$$\mathbf{A} = (a_{ij}), \text{with } a_{ij} = A(p_j \curvearrowleft p_i)$$



Empirical applications

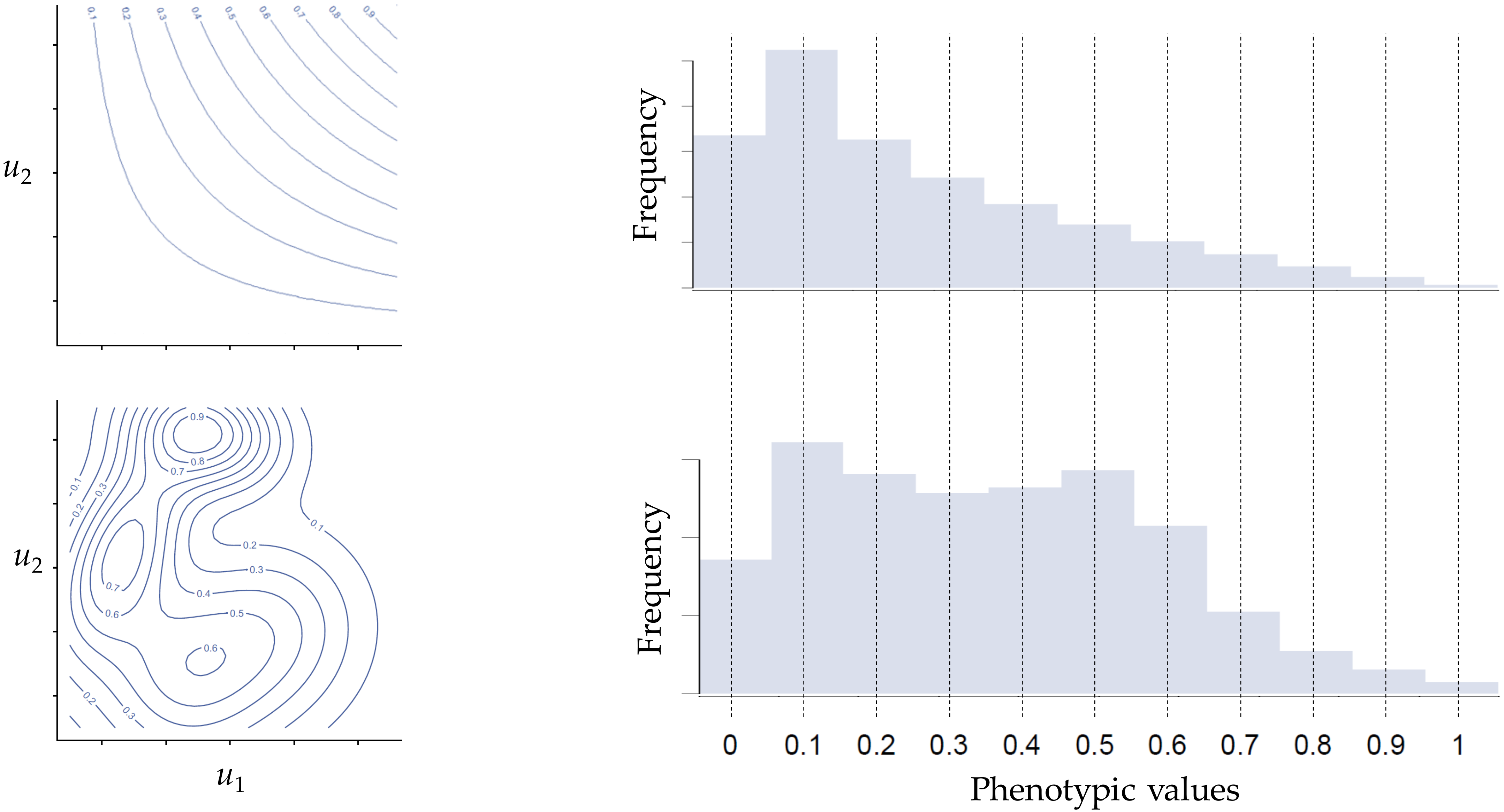
The accessibility structure is equivalent to a directed graph with adjacency matrix

$$\mathbf{A} = (a_{ij}), \text{with } a_{ij} = A(p_j \curvearrowleft p_i)$$

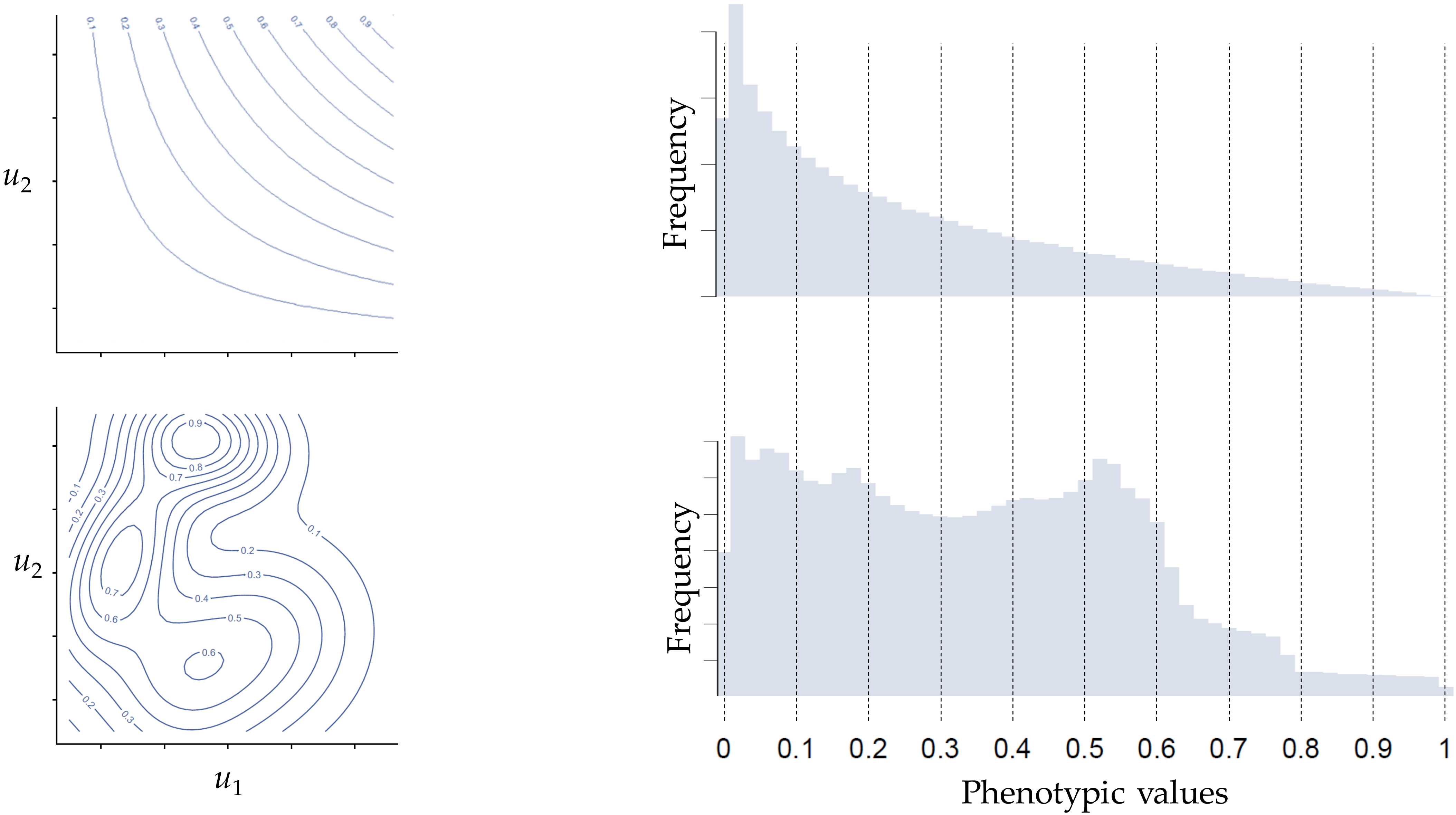


$\mathbf{A}$ is a stochastic matrix and describes a Markov chain over the phenotype space (as a finite state-space). Its stationary distribution π satisfies $\pi\mathbf{A} = \pi$.

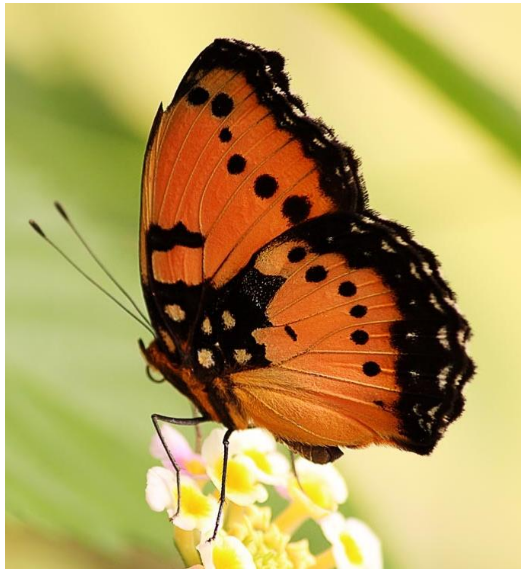
Empirical applications



Empirical applications



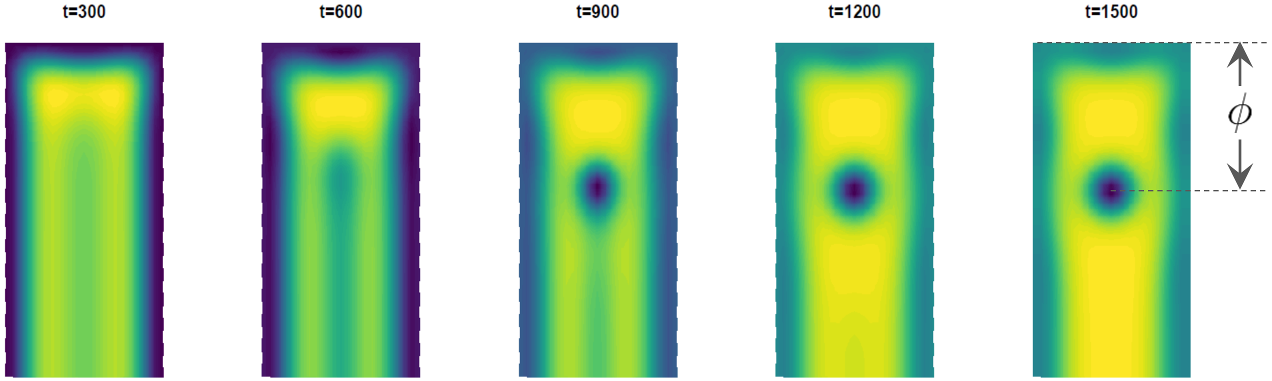
Empirical applications



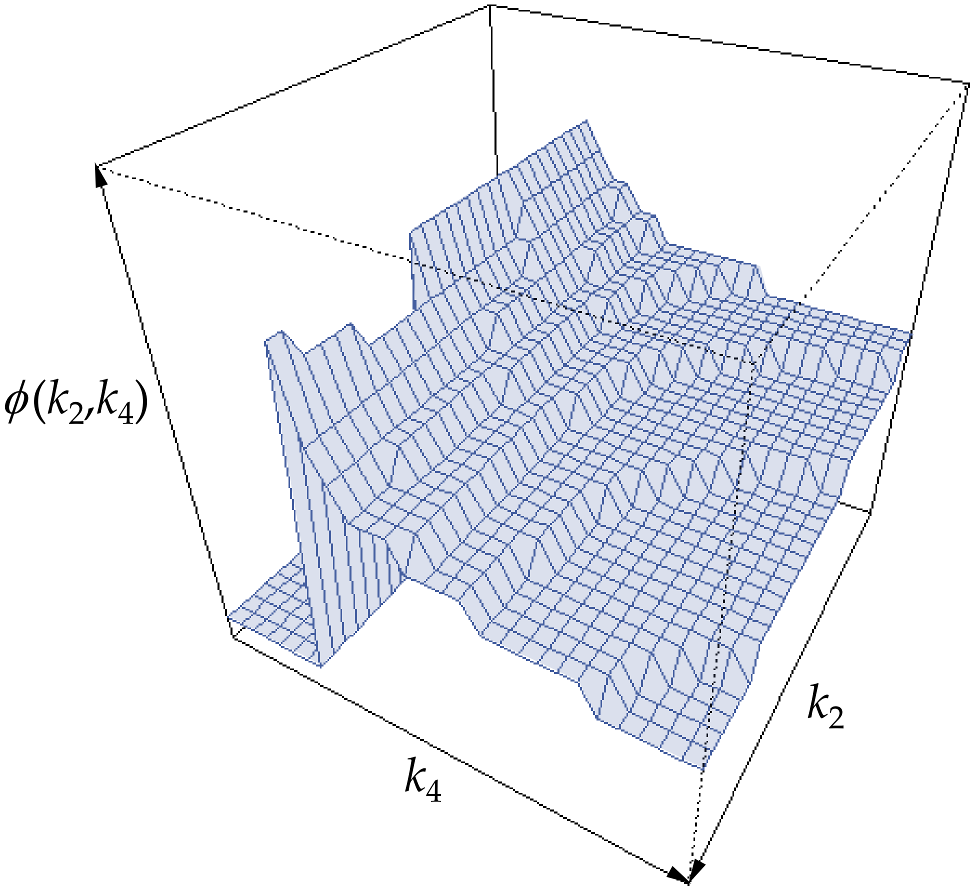
Nijhout (1990, 1991)

$$\frac{\partial A}{\partial t} = F(A, B) + D_A \nabla^2 A$$

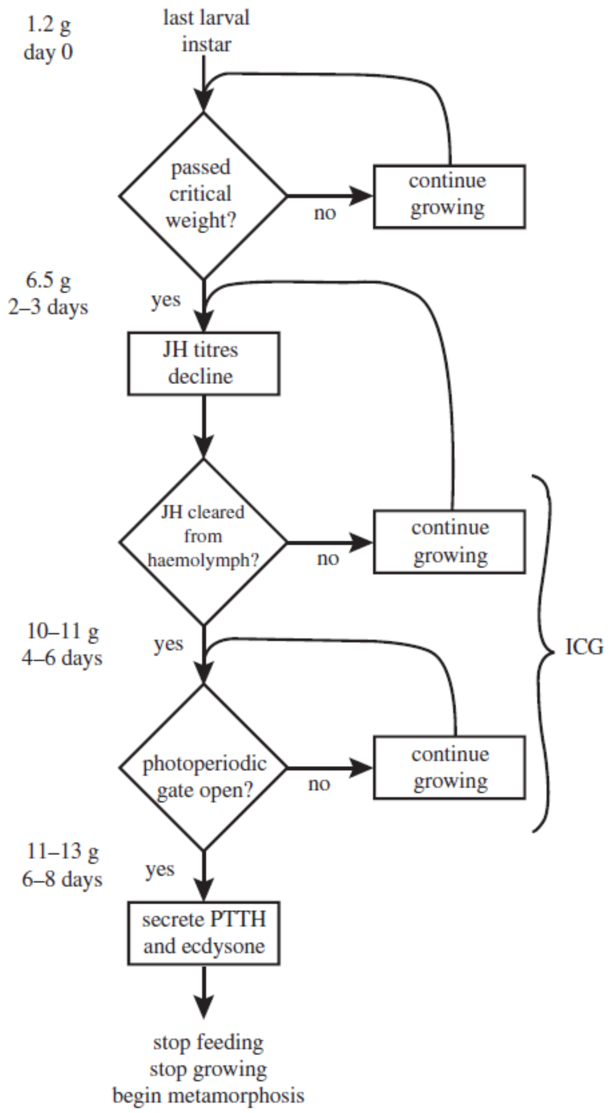
$$\frac{\partial B}{\partial t} = G(A, B) + D_B \nabla^2 B$$

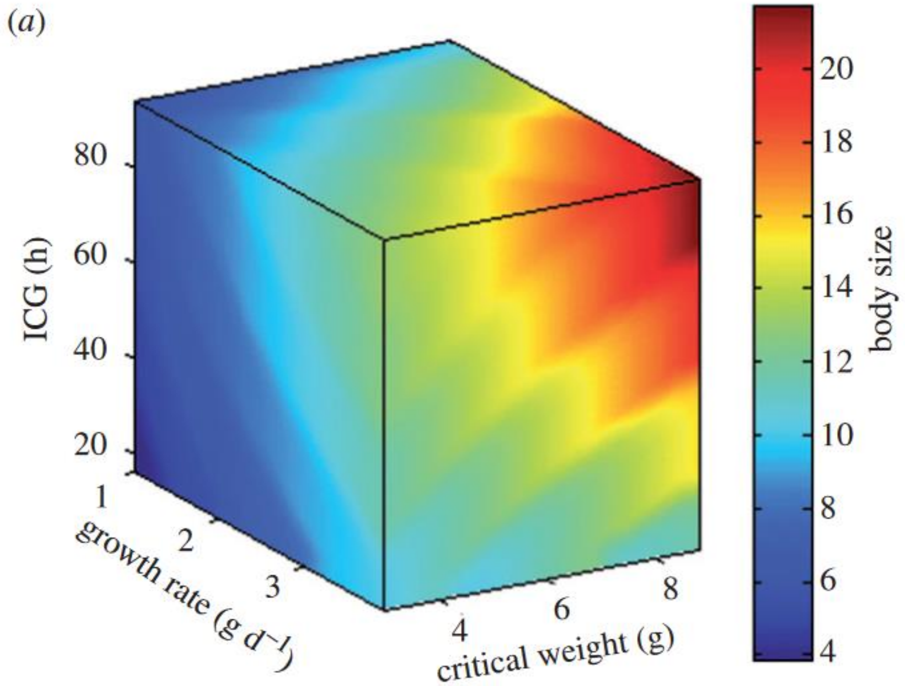


Developmental time



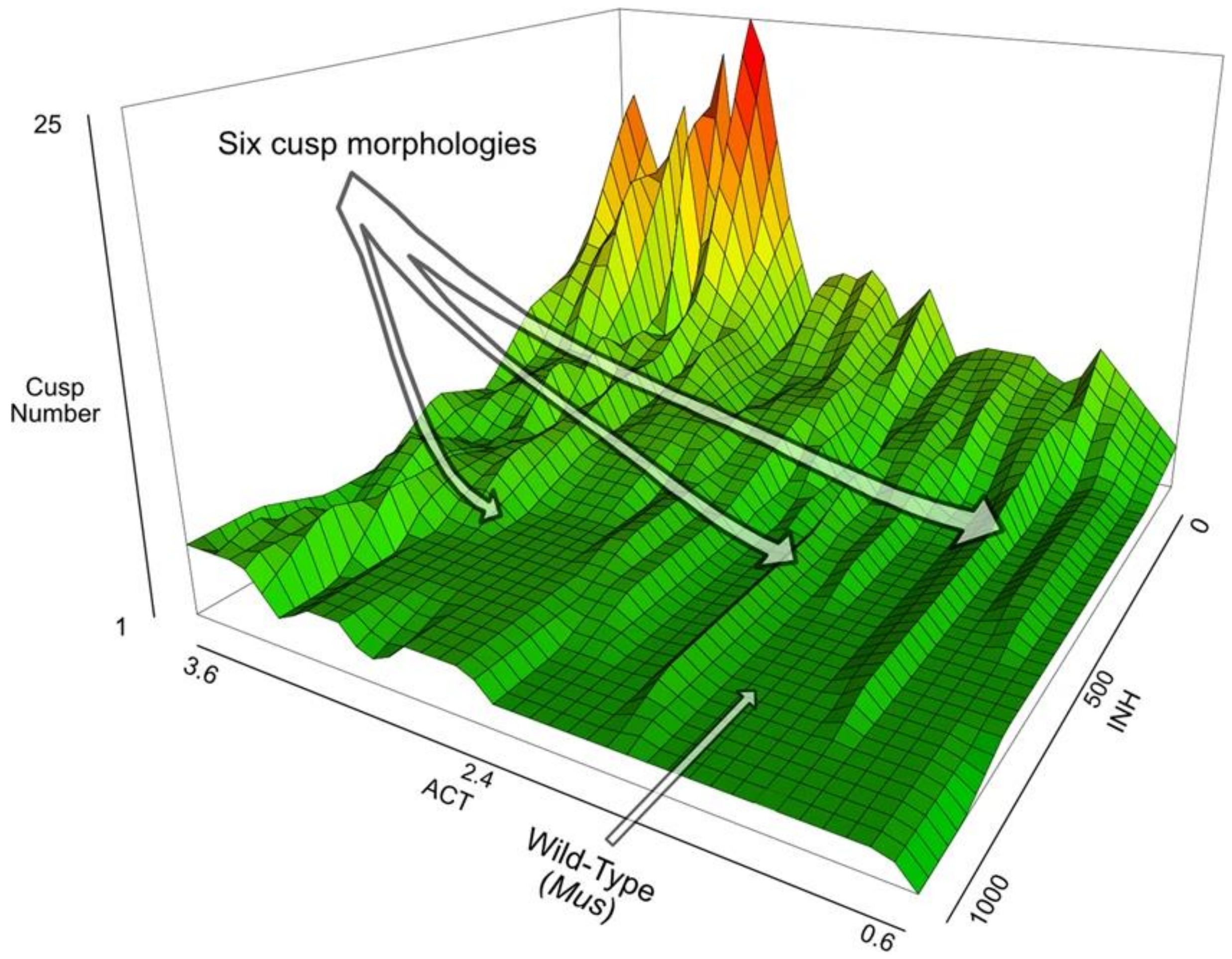
Empirical applications

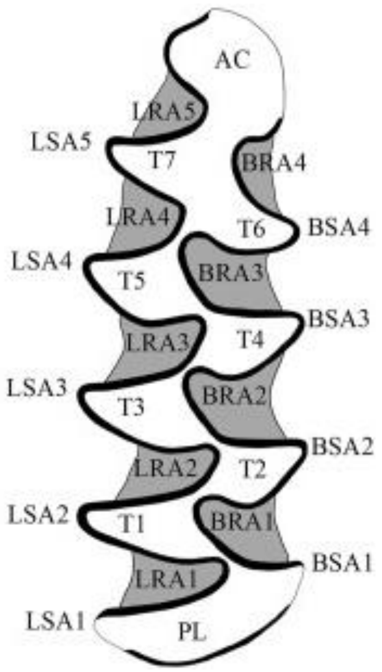




Nijhout et al. (2010)

Empirical applications



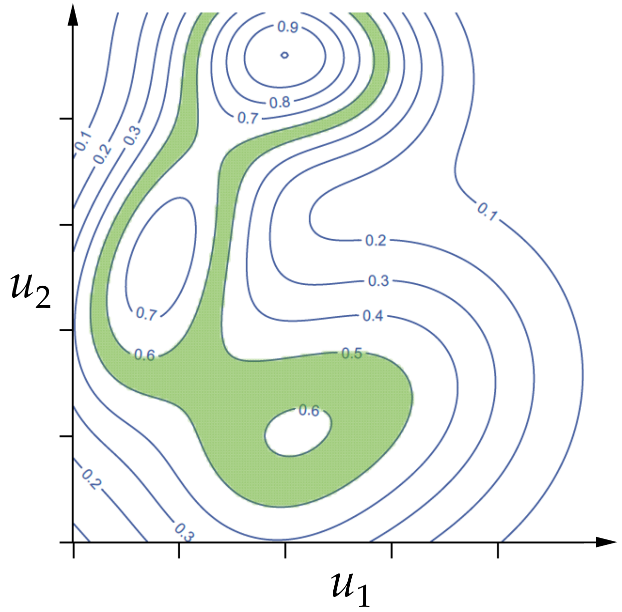


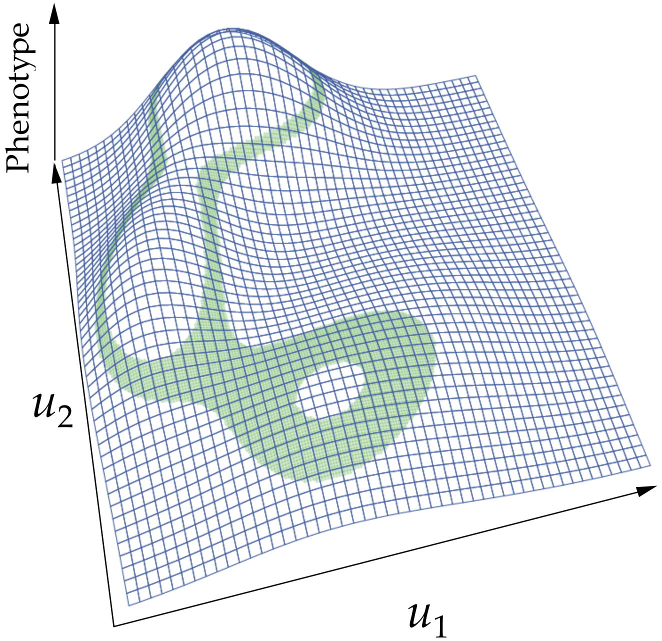
Burroughs (2021)

Conclusions

- General case

$$A(p_{c+a} \curvearrowleft p_c) = \frac{\sum_i \int \cdots \int_{R_i} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}{\sum_j \int \cdots \int_{D_j} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}$$

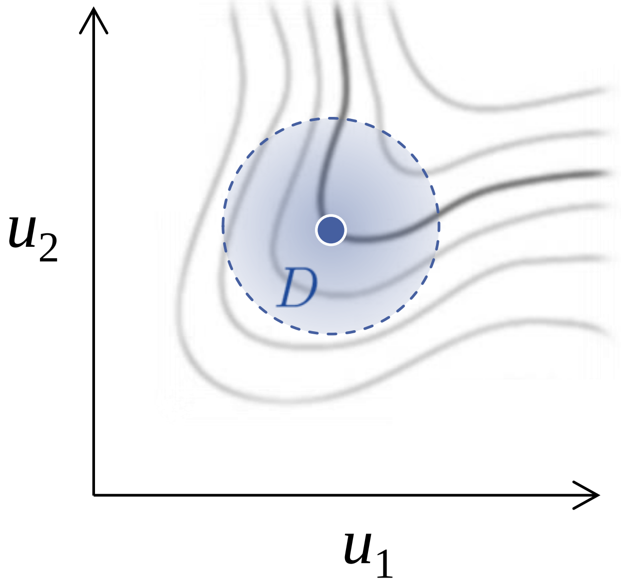




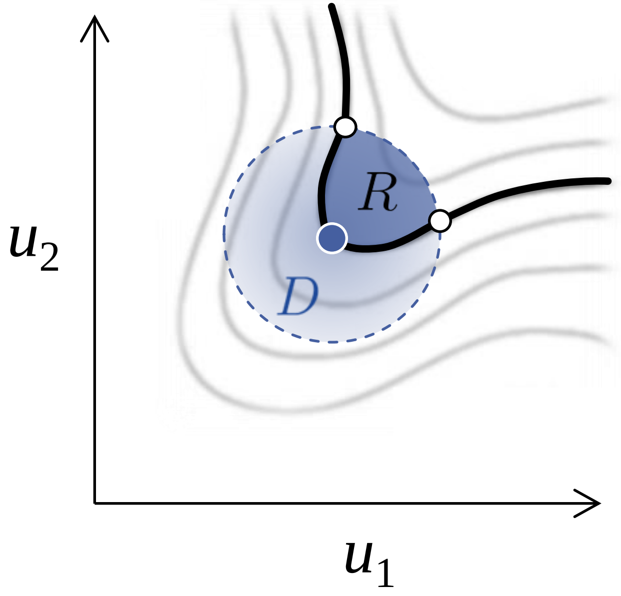
Conclusions

- General case

$$A(p_{c+a} \curvearrowleft p_c) = \frac{\sum_i \int \cdots \int_{R_i} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}{\sum_j \int \cdots \int_{D_j} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}$$



- "Population accessibility"



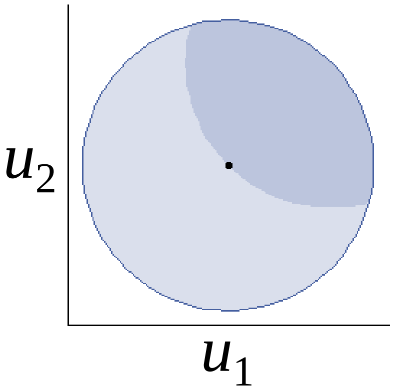
Conclusions

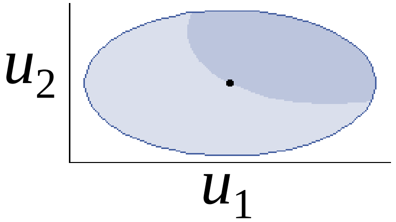
- General case

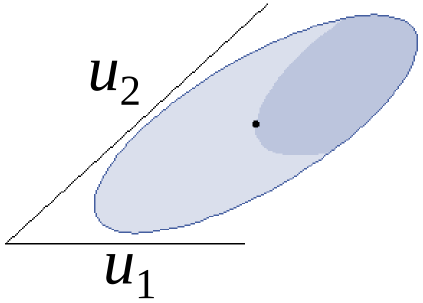
$$A(p_{c+a} \curvearrowleft p_c) = \frac{\sum_i \int \cdots \int_{R_i} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}{\sum_j \int \cdots \int_{D_j} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}$$

- “Population accessibility”

- Affine invariance



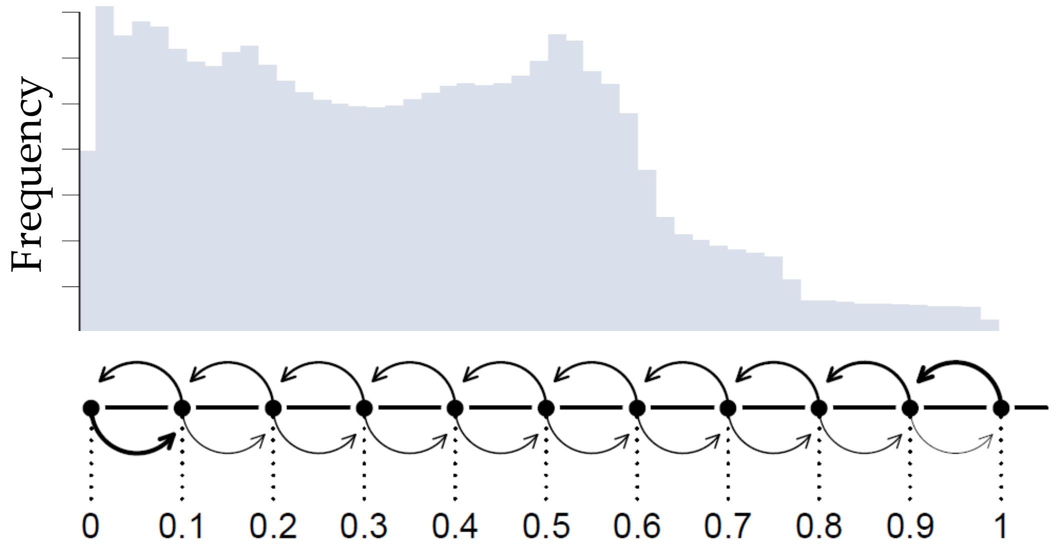




Conclusions

- General case

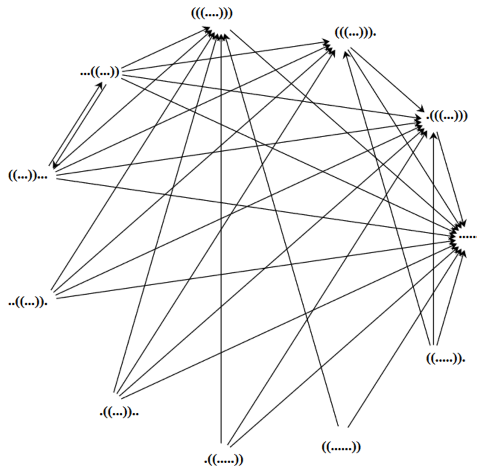
$$A(p_{c+a} \curvearrowleft p_c) = \frac{\sum_i \int \cdots \int_{R_i} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}{\sum_j \int \cdots \int_{D_j} 1 \mathrm{d}u_1 \ldots \mathrm{d}u_n}$$



- “Population accessibility”

- Affine invariance

- Null model of phenotype space occupation at a given life stage, prior to the action of external selection



Conclusions

- General case

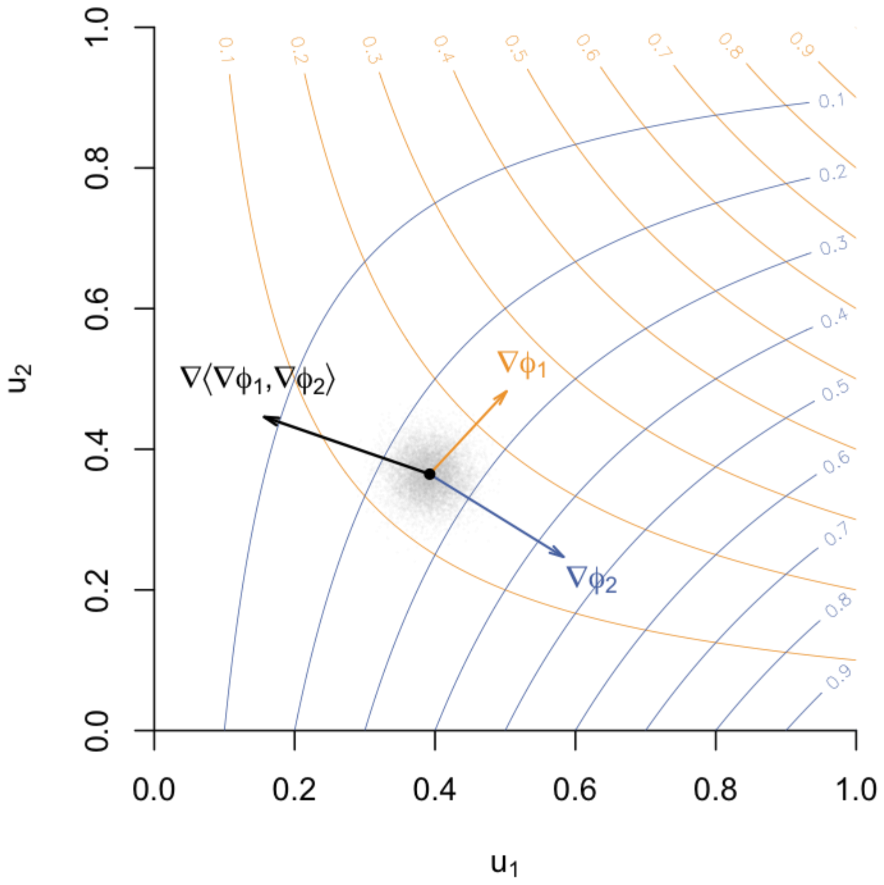
$$A(p_{c+a} \curvearrowleft p_c) = \frac{\sum_i \int \cdots \int_{R_i} 1 du_1 \ldots du_n}{\sum_j \int \cdots \int_{D_j} 1 du_1 \ldots du_n}$$

- "Population accessibility"

- Affine invariance

- Null model of phenotype space occupation at a given life stage, prior to the action of external selection

- Multiple traits and pleiotropy



Acknowledgments

W. Fontana, H. F. Nijhout, C. Venkataraman, T. Yamaguchi, T. Ubukata