Generalized Topological Spaces

in

Evolutionary Theory

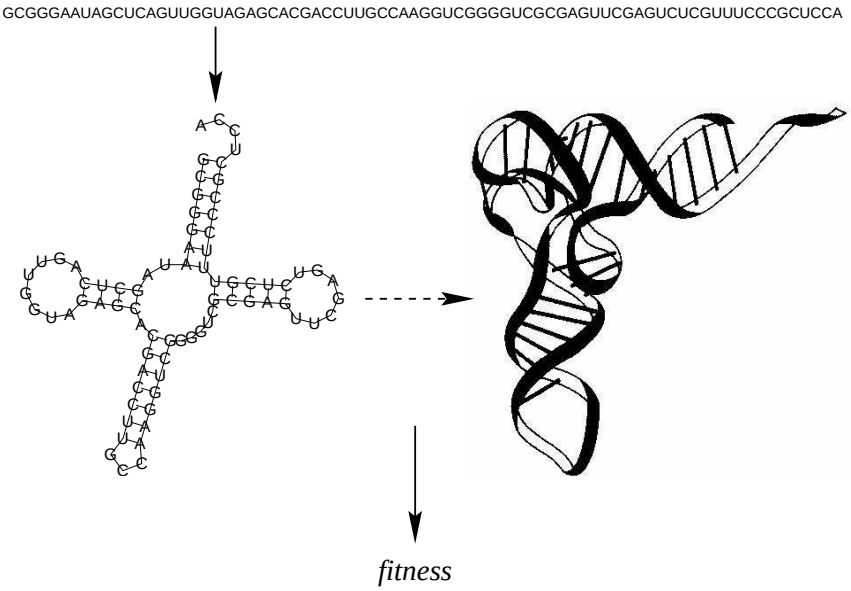
Bärbel STADLER

Institute f. Theoretical Chemistry & Structural Biology,
University of Vienna, Austria

http://www.tbi.univie.ac.at/~baer

Bled, Jan 2002

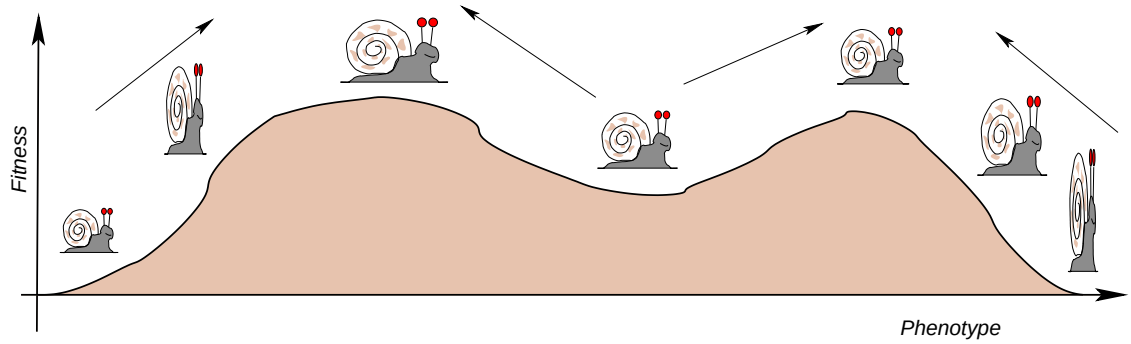
Genotype-Phenotype Maps



Given:

- A **set** X of genotypes (sequences)
- Genetic Operators
  (mutation and/or recombination ...)
- A **set** Y of (potential) phenotypes (structures)
- A **function** $f : X \rightarrow Y$ assigning a phenotype to each genotype.

Phenotype Space - for the happy population geneticist...



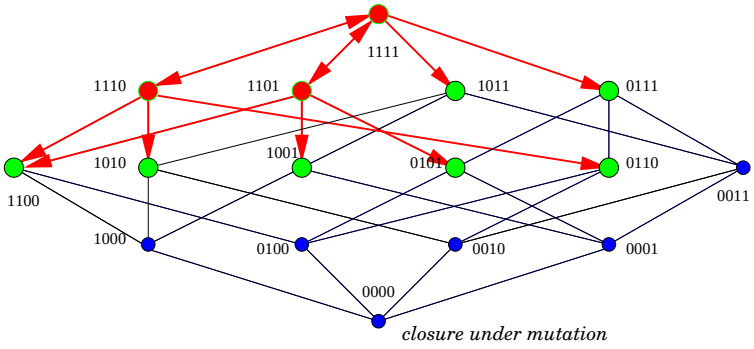
- Phenotypes are "somehow" numbers or *vectors*
- Neighboring phenotypes are within a small "neighborhood"
- Fitness is a (more or less) smooth function of one or more coordinates

BUT: There is a catch in this picture!

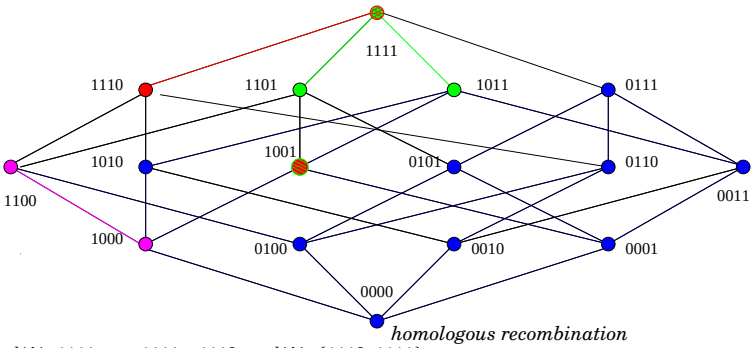
The RNA model

In RNA, genotype and phenotype are two features of one and the same molecule

(1) Genotype space is discrete



(2) Its structure depends on the genetic operator

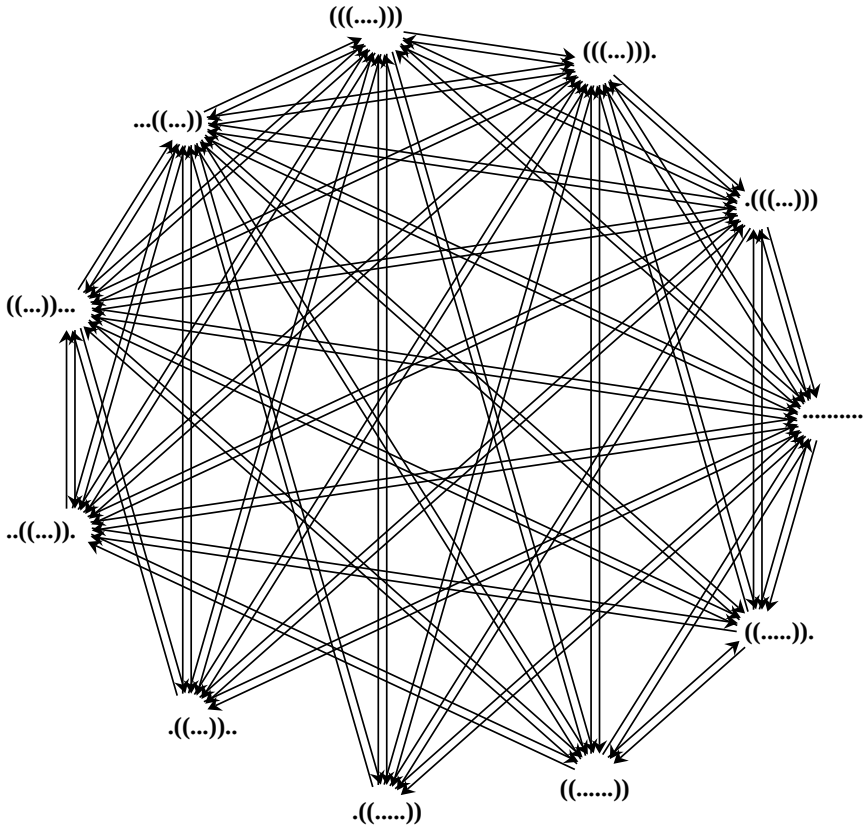


cl(A): 1111 ---> 1111 , 1110 cl(A)={1110, 1111}
1110 1110 1111

cl(B): 1101 ---> 1101 , 1011 cl(B)={1111, 1101, 1011}
1011 1011 1101

cl(C):1100 ---> 1100, 1000 cl(C)={1000, 1100} BUT: {1001} is in cl(A u B u C)
1000 1000 1100

(3) Phenotype space inherits its structure from genotype space

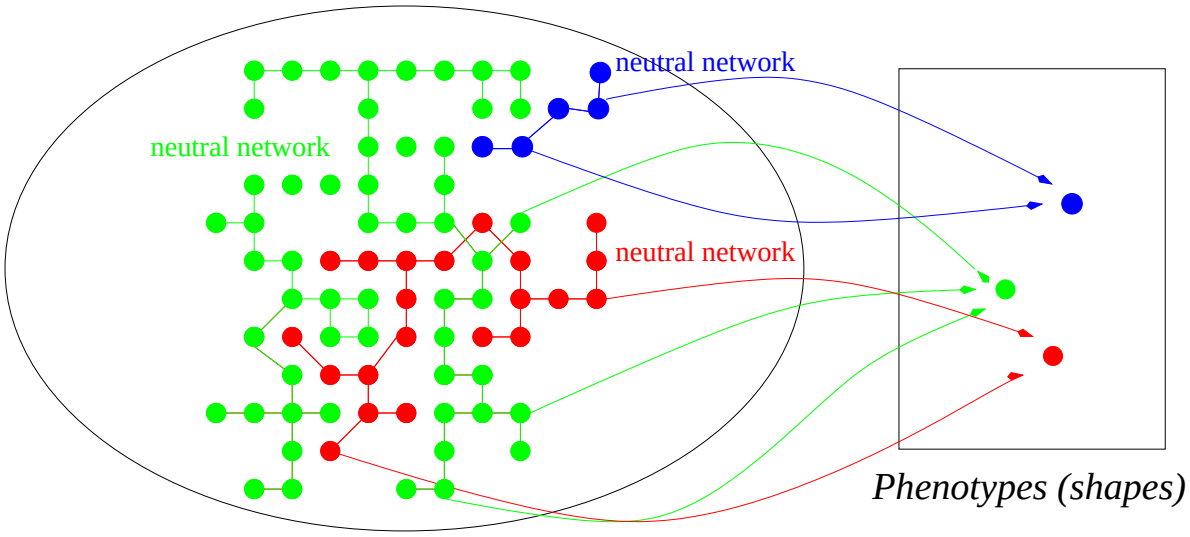


Accessibility at genotypic levels

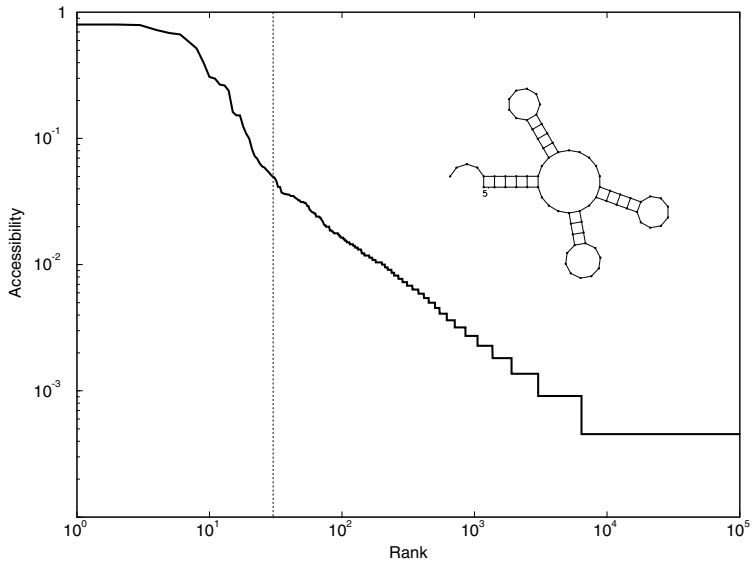
(i.e., the genetic operator)

implies accessibility at phenotypic level. There are many more sequences(=genotypes) than structures(=phenotypes).

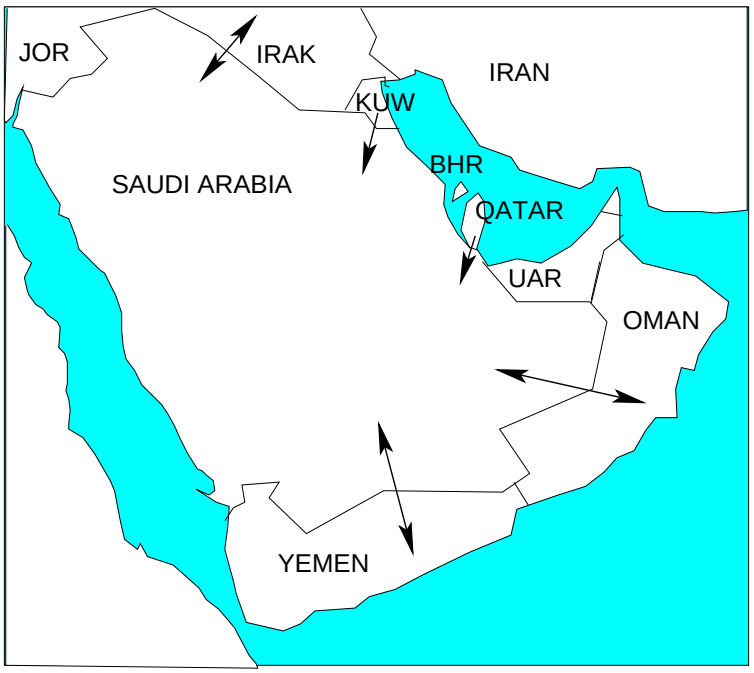
Neutral Networks



There are many more sequences(=genotypes) than structures(=phenotypes). Sequences folding into the same structure form a neutral net in genotype space.
Various degrees of accessibility of neighboring phenotypes:

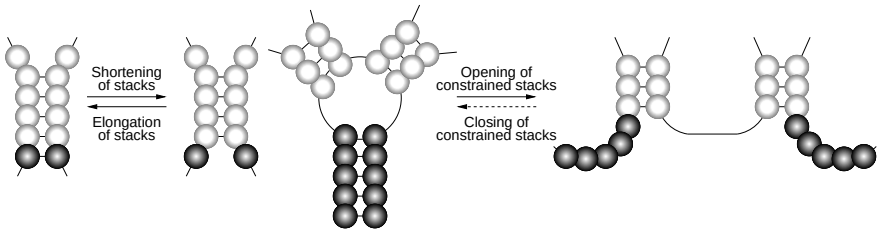


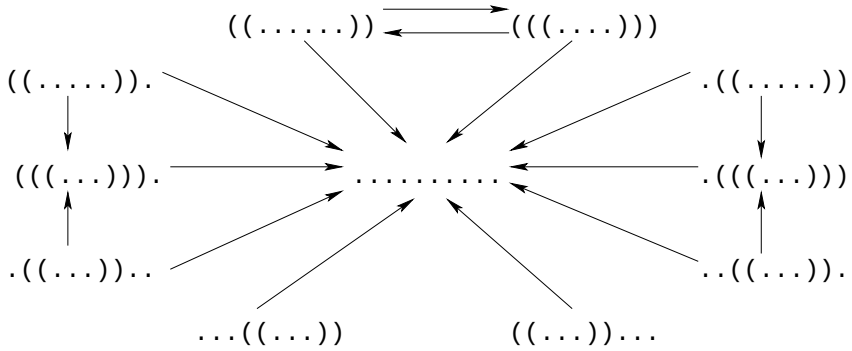
Accessibility is not symmetric



Sometimes

. . . there is a relation between **accessibility** and a **representation** of the phenotypes





(Fontana & Schuster, *J. Theor. Biol.* **194**, 491-515 (1998)

but not always !!!

Goal: A "Relative" Theory

We want a theory of phenotypes that can deal with concepts such as

- Continuity and Discontinuity
- Character
- Homology
- Innovation

WITHOUT recourse to a
specific representation of the phenotype

Genotype Spaces

Given:
a set X of possible genotypes
a set A of realized genotypes
a fixed collection of genetic operators
[such as mutation, recombination, gene-rearrangement]

define the set A' of genotypes accessible from A.

Properties

(i) No spontaneous creation, i.e, $\emptyset' = \emptyset$.

(ii) A more diverse population produces more diverse offsprings:
$A \subseteq B$ implies $A' \subseteq B'$

(iii) All parental genotypes are also accessible in the next time step
$A \subseteq A'$.

In the case of mutation as the only source of diversity:
haploid populations, no sex, no recombination, etc

(iv) Diversity of offsprings depends only on the parent:
$A' = \bigcup_{x \in A} \{x\}'$

Generalized Closure Spaces

	closure	neighborhood
K0	$\mathrm{cl}(\emptyset) = \emptyset$	$X \in \mathcal{N}(x)$
K1	$A \subseteq B \implies \mathrm{cl}(A) \subseteq \mathrm{cl}(B)$ $\mathrm{cl}(A \cap B) \subseteq \mathrm{cl}(A) \cap \mathrm{cl}(B)$ $\mathrm{cl}(A) \cup \mathrm{cl}(B) \subseteq \mathrm{cl}(A \cup B)$	$N \in \mathcal{N}(x),\ N \subseteq N'$ $\implies$ $N' \in \mathcal{N}(x)$
K2	$A \subseteq \mathrm{cl}(A)$	$N \in \mathcal{N}(x) \Rightarrow x \in N$
K3	$\mathrm{cl}(A \cup B) \subseteq \mathrm{cl}(A) \cup \mathrm{cl}(B)$	$N', N'' \in \mathcal{N}(x) \implies$ $N' \cap N'' \in \mathcal{N}(x)$
K4	$\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$	$N \in \mathcal{N}(x) \iff$ $\mathrm{int}(N) \in \mathcal{N}(x)$
K5	$\bigcup_{i \in I} \mathrm{cl}(A_i) = \mathrm{cl}\left(\bigcup_{i \in I} A_i\right)$	$\mathcal{N}(x) = \emptyset$ or $\exists N(x) : N(x) \subseteq N$ iff $N \in \mathcal{N}(x)$

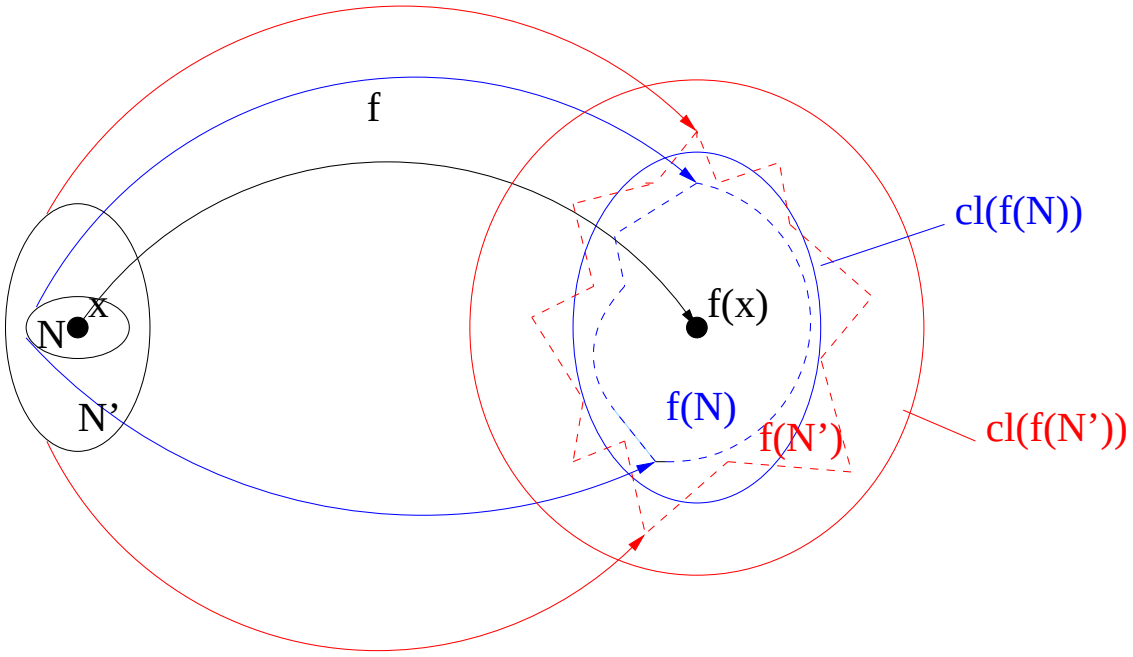
In general: only (K0), (K1), (K2) hold.
neighborhood space
e.g. recombination spaces - no graph representation

For mutation in haploid populations:
(K0), (K1), (K2), (K5) [and thus (K3)]
additive pretopological space
e.g. RNA space under mutation - Hamming graph

For comparison: (K0), (K1), (K2), (K3), and (K4) are equivalent to the axioms of a **topology**.

Continuity

Genotype-Phenotype map: $(X, \mathrm{cl}) \to (Y, \mathrm{cl})$

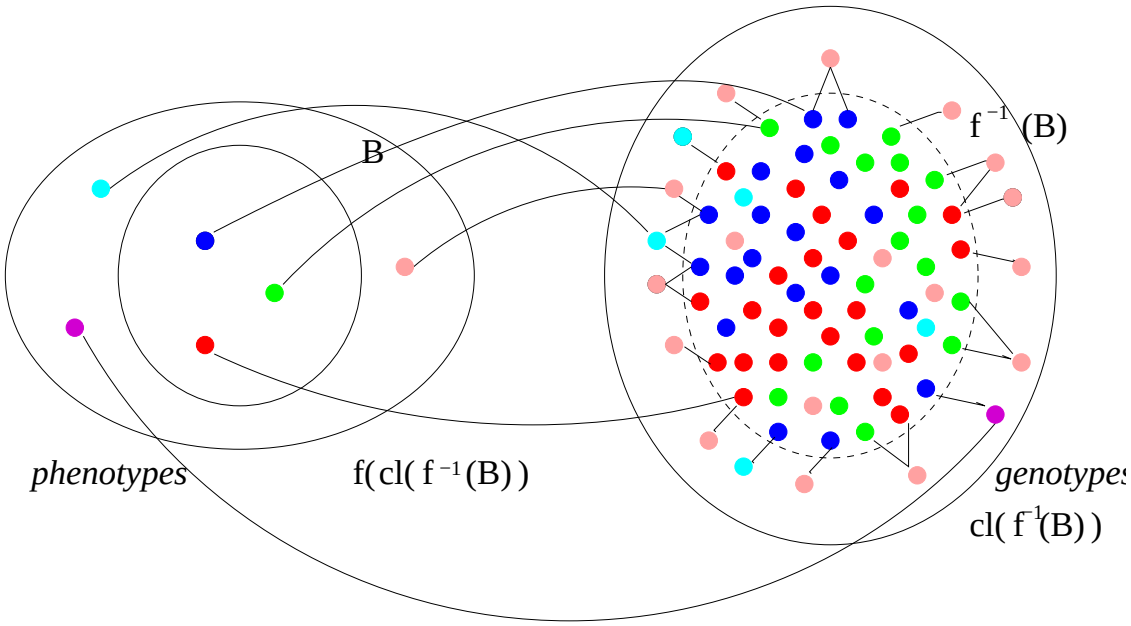


Equivalent in (K1)-spaces:
closure preservation: $f(\mathrm{cl}(A)) \subseteq \mathrm{cl}(f(A))$.

BUT: What is closure in phenotype space?

$\Phi \in \mathrm{cl}(B)$ ... Φ is "readily accessible" from B

i.e., there are "enough" genotypes that fold into members of B who can mutate or recombine into an offspring with phenotype Φ.

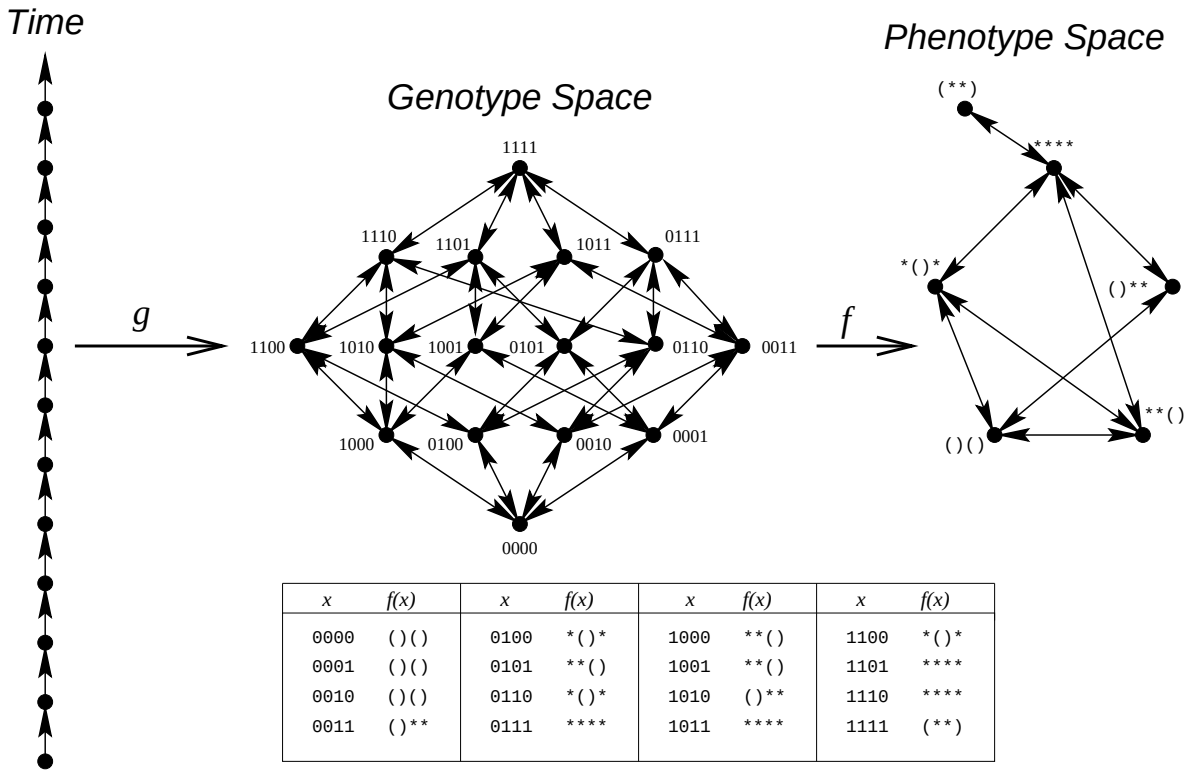


NOTE: closure in phenotype space (Y, cl) depends on closure in genotype space.

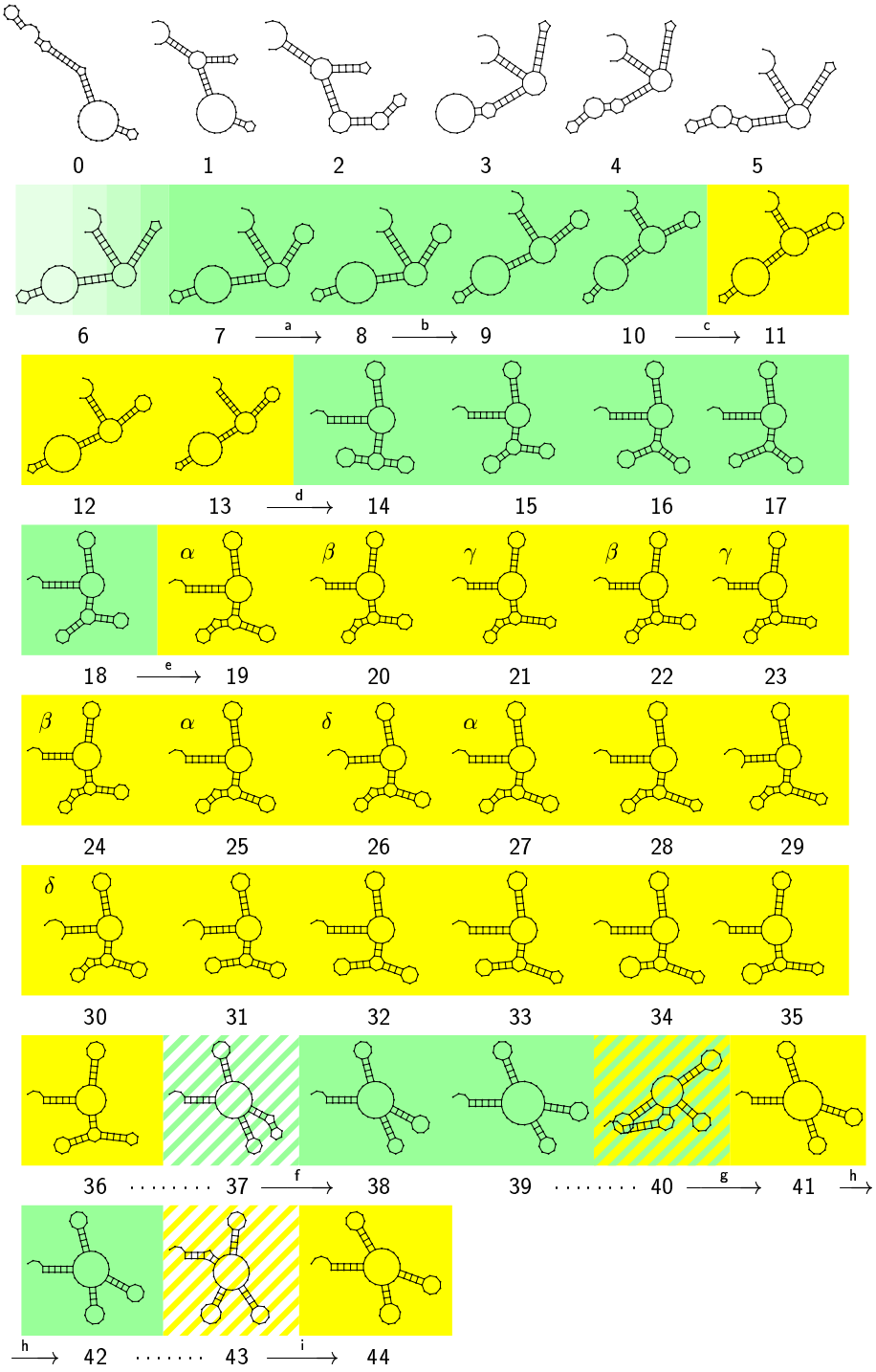
A useful closure function on Y is thus always finer than the **induced closure** $\phi(B) = f(\mathsf{cl}\big(f^{-1}(B)\big))$.

Of course $f : (X, \mathsf{cl}) \to (Y, \phi)$ is continuous.

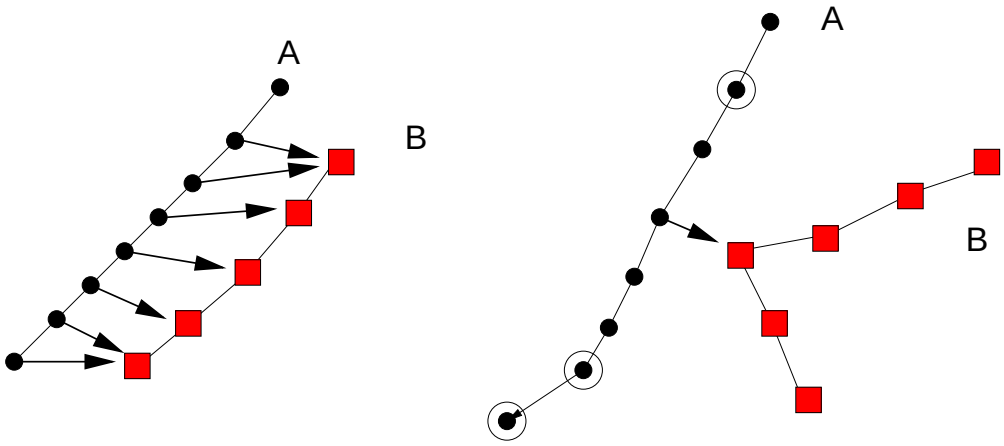
Evolutionary Trajectories



x	f(x)	x	f(x)	x	f(x)	x	f(x)
0000	()()	0100	*()*	1000	**()	1100	*()*
0001	()()	0101	**()	1001	**()	1101	****
0010	()()	0110	*()*	1010	()**	1110	****
0011	()**	0111	****	1011	****	1111	(**)


0
1
2
3
4
5
6
7
a
8
b
9
10
c
11
12
13
d
14
15
16
17
α
β
γ
18
e
19
20
21
22
23
24
25
26
27
28
29
δ
30
31
32
33
34
35
36
37
f
38
39
40
g
41
h
42
43
i
44

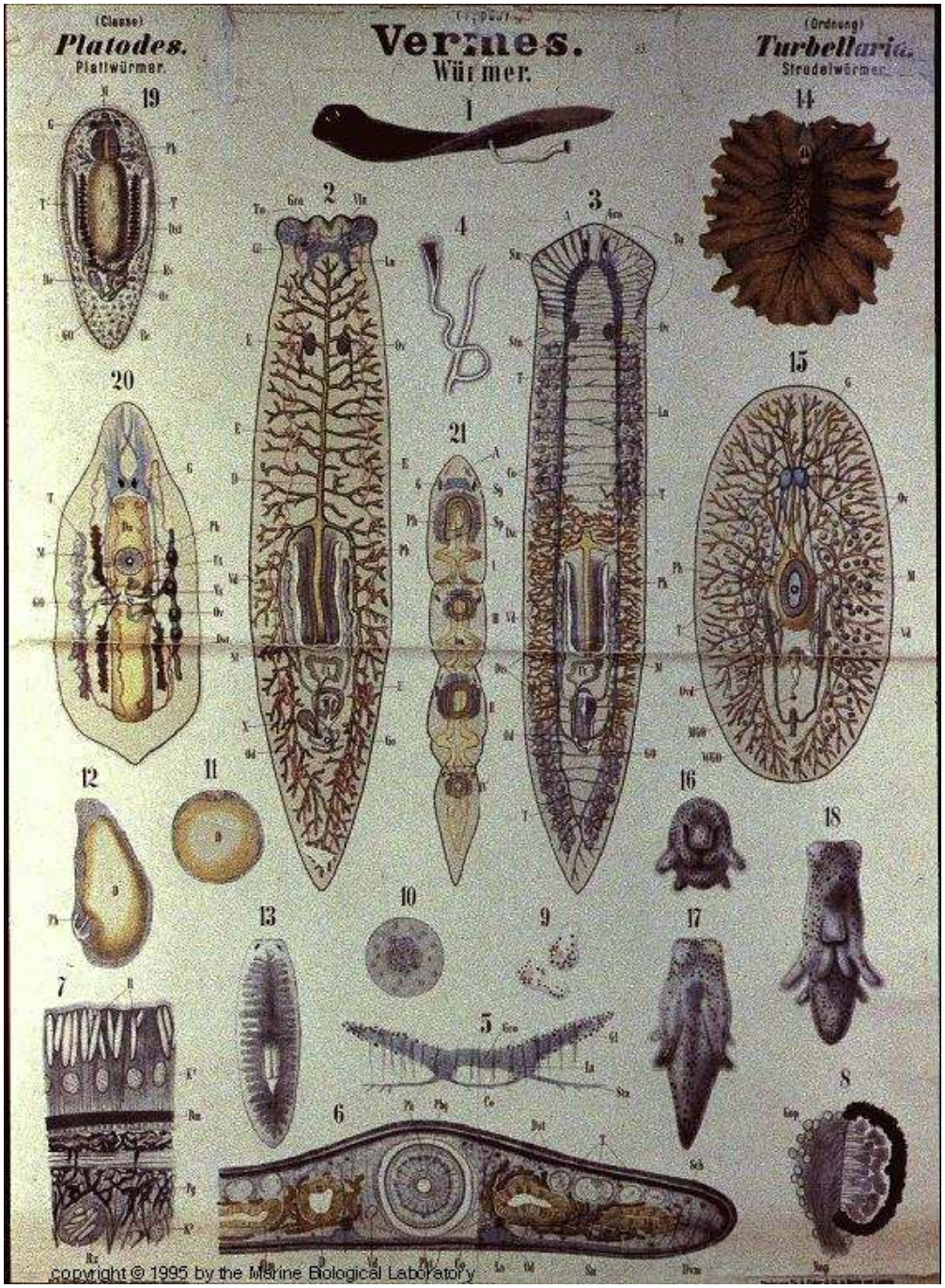
A possibly testable consequence:



On the r.h.s. phenotype B is not accessible from any of the extant species (with circles). The genetic changes that lead from ancestor A to phenotype B cannot be reproduced in any of the current species with A-phenotypes.

(Wagner, JEZ(MDE) to appear)

What is a Character?

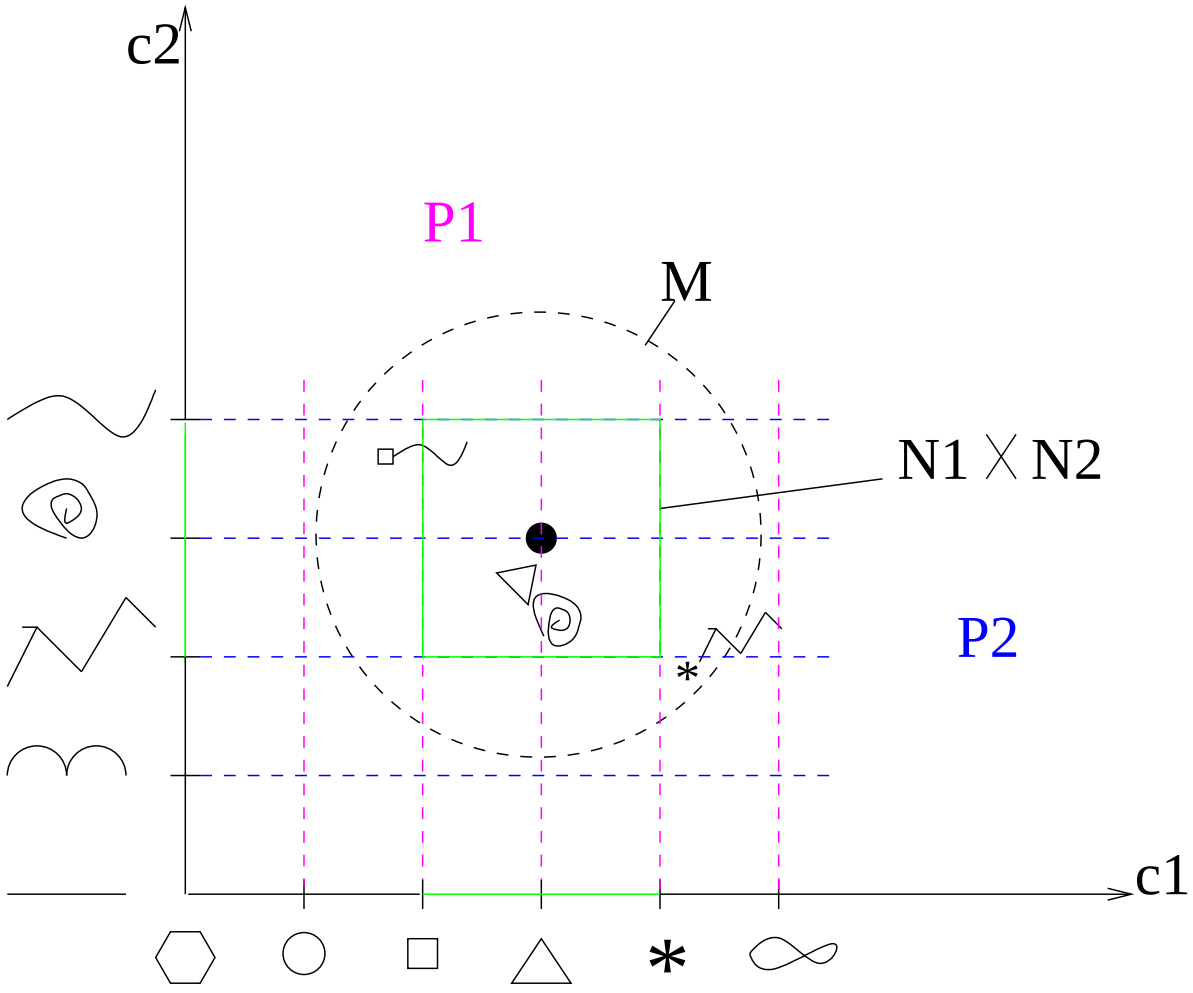


Characters

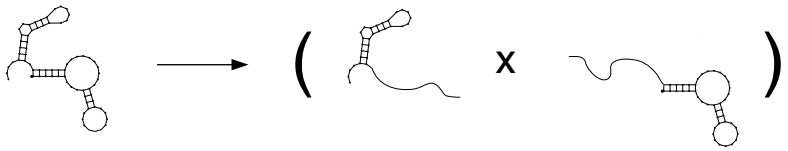
Idea: Characters **can** vary independently

$\Longleftrightarrow$

Factors of phenotype space

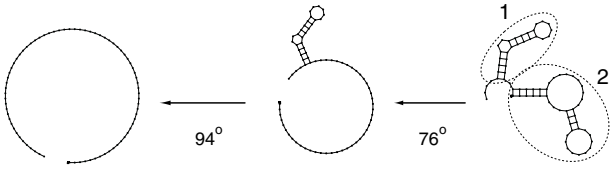


A

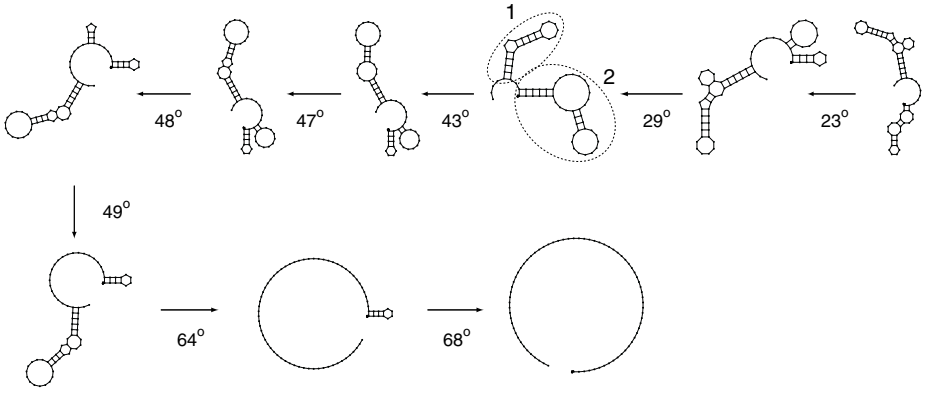


B

1 and 2 are modular characters

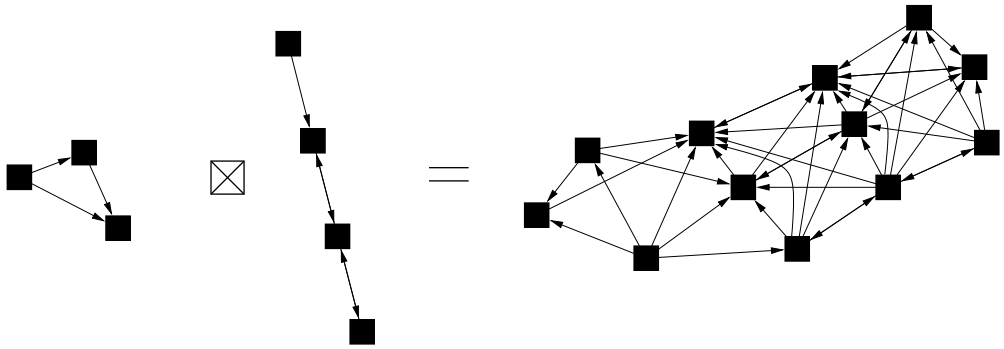


1 and 2 are not modular characters



Mutation only:Directed Graphs

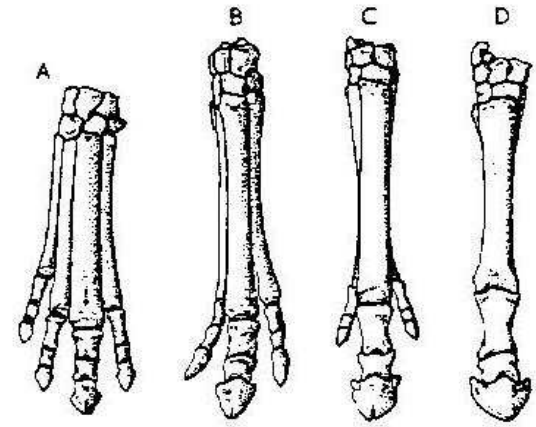
topological product $\Longleftrightarrow$ strong product of graphs



Unique prime factor decomposition of connected graphs and digraphs.

Allows identification of global characters.

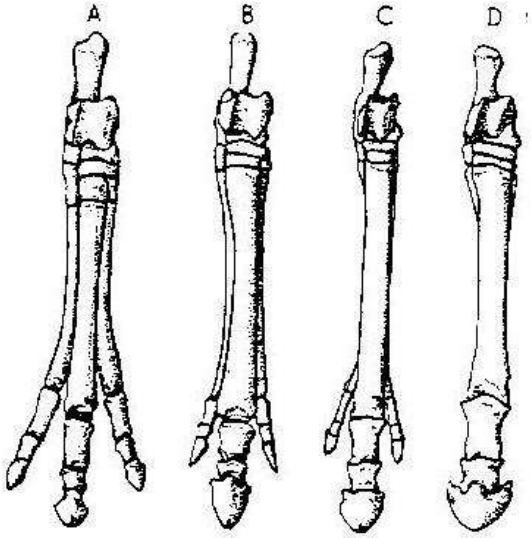
Homologous Characters



Horse Leg Evolution

A Hyracotherium
B Miohippus
C Merychippus
D Equus

Forefeet



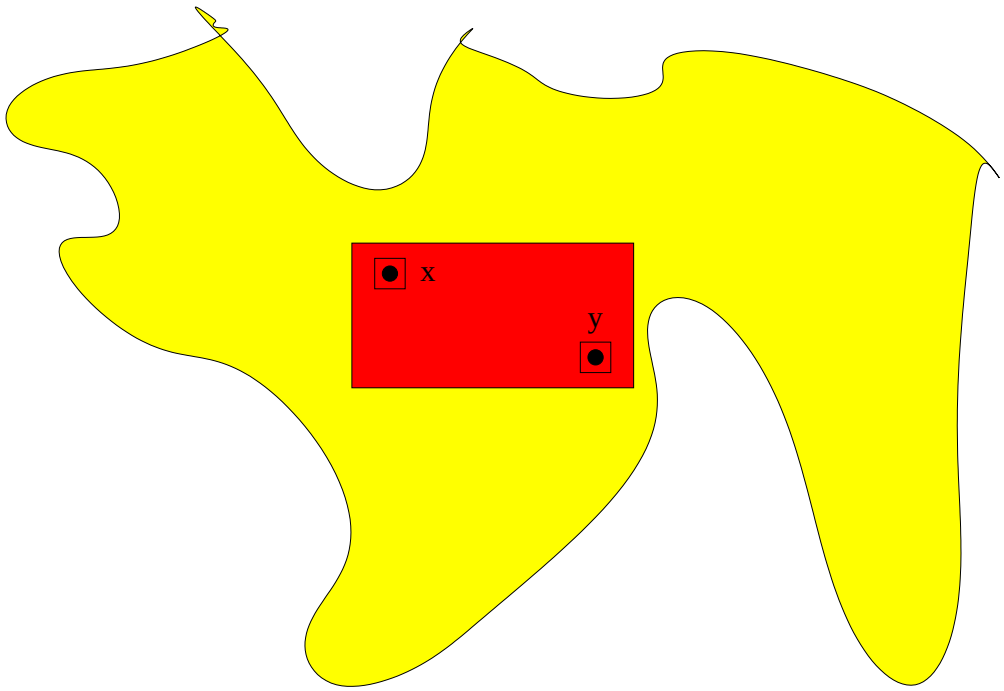
Hindfeet

Homology and Innovation

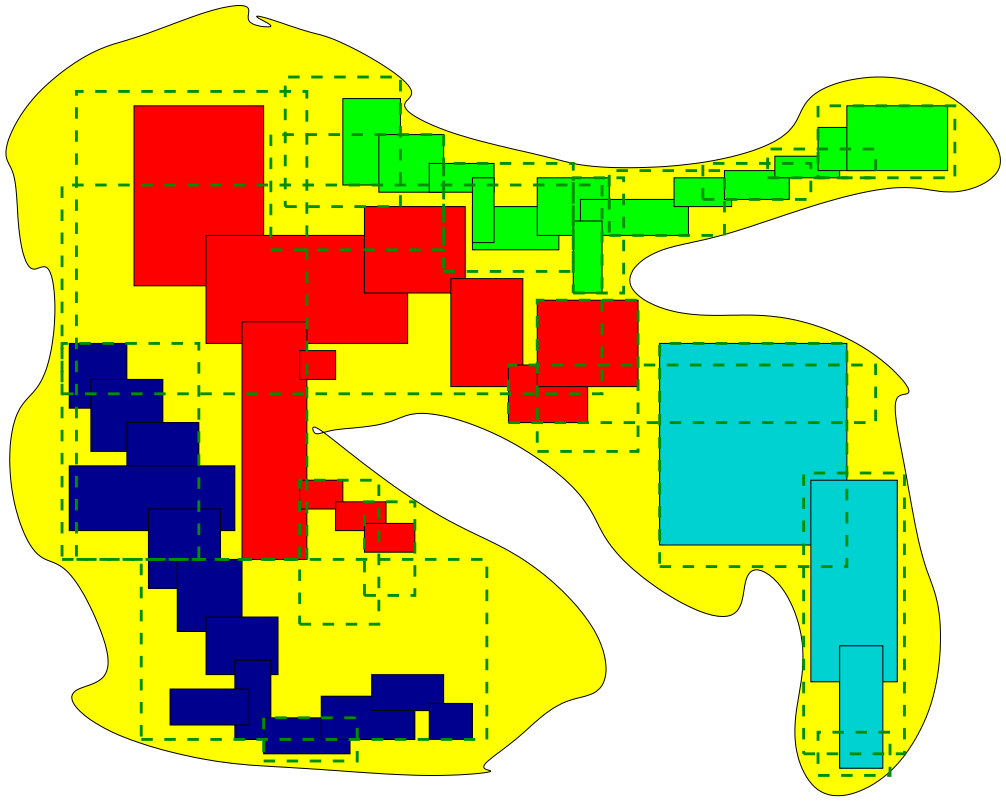
Innovation: Any major transition in evolution, e.g. origin of multicellularity, or significant modification of body plan.

Mathematically, it is a process transforming a $\{z_1, z_2, \ldots, z_N\}$ coordinate system into a $\{z_1, z_2, \ldots, z_N, z_{N+1}\}$ state space or vice versa. Characters are factors of phenotype space.

Problem: Which factors are "the same" in different places of phenotype space?



Idea: Characters are local factors. We can find a common coordinate system for two phenotypes if there is a region that factorizes. Hence these two phenotypes have the same local factors. The respective characters are homologous.



The identity of characters can be extended wherever the colored rectangles overlap.

We have developed here a **framework** (or a **language**) for formalizing evolution at large scales that can deal with:

- Continuous and discontinuous evolutionary transitions
- the concept of a **character**
- the concept of **homology**
- different notions of **innovation**
- & suggests (at least some) testable hypotheses

Of course, it is only a first step . . .