Who Cares?

Graph Products in Theoretical Biology

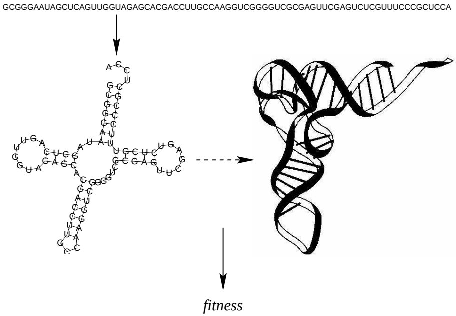
Peter F. Stadler

Universität Leipzig

Bled, February 2007

Outline

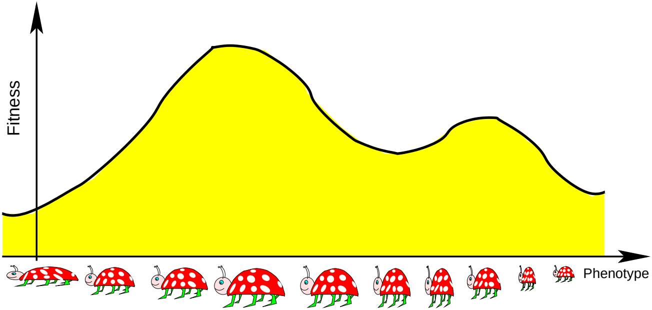
Evolutionary Biology: Genotype-Phenotype Maps



Given:

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

The “usual” view

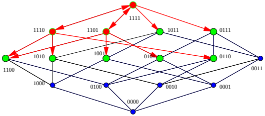


- Phenotypes are “somehow” numbers or *vectors*
- Accessible phenotypes are within a small (Euclidean) distance (in this vector space)
- Fitness is a (more or less) smooth function

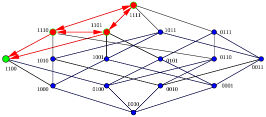
Population Genetics is perfectly happy . . .

So, what is wrong with this picture?

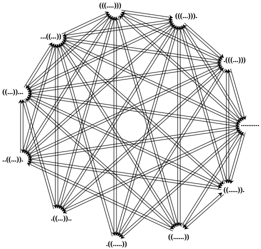
(1) Genotype space is **discrete**



(2) Its structure depends on the **genetic operator**



(3) Phenotype space inherits its structure from genotype space



Accessibility at genotypic levels **implies** accessibility at phenotypic level

Does Evolution really "live" on an Euclidean space????

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
$A \subseteq A'$.

This is the same as the C-operator for the chemical network!

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:

Generalized Closure Spaces

... instead of vector spaces ...

Set X, closure function cl : $\mathcal{P}(X) \to \mathcal{P}(X)$

Equivalent formulations:

$$\text{int}(A) = X \setminus \text{cl}(X \setminus A)$$

The interior is the dual of the closure function

A set N is a neigbhorhood of x if and only if $x \in \text{int}(N)$.

Let $\mathcal{N}(x)$ be the set of all neighborhood of x.

$\mathcal{N} : X \to \mathcal{P}(\mathcal{P}(X))$.

Closure, interior and neighborhood functions are equivalent.

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** For mutation in haploid populations: (K0), (K1), (K2), (K5) [and thus (K3)]: **additive pretopological space**

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

Cool!

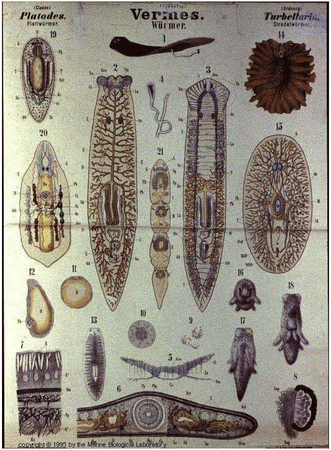
. . . so, real evolution, genetic algorithms, evolution strategies, multi-objective optimization heuristics, genetic programming, etc., etc., live on a neighborhood space.
. . . for mutation only, it is even a pretopology.
Thus:
Directed graphs and finite pretopological spaces are the same thing

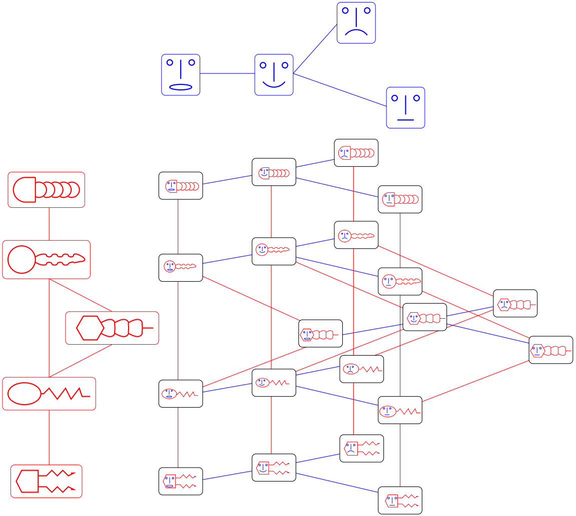
Should we care that our closure function is NOT idempotent? NO, Eduard Čech in the 1960s wrote a big, fat textbook on point set topology, where he showed that pretty much everything works in pretopologies — thus you can do topology without every talking about open or closed sets. (Just the proofs get a bit more tedious without this convenience.)

Product Spaces

Let (X_1, c_1) and (X_2, c_2) be two general closure spaces, with neighborhood systems $\mathcal{N}_1$ on X_1 and $\mathcal{N}_2$ on X_2.
Product space: Point set $X_1 \times X_2$.
The neighborhoods of the product space satisfy:
N is a neighborbood of (x_1, x_2) iff there are neighborhoods $N_1 \in \mathcal{N}_1(x_1)$ and $N_2 \in \mathcal{N}_2(x_2)$ such that $N_1 \times N_2 \subseteq N$.
The projections $\pi_i : (X_1 \times X_2, \mathcal{N}) \rightarrow (X_i, \mathcal{N}_i) : \ (x_1, x_2) \mapsto x_i$ are continuous functions for $i = 1, 2$.
(as in topological spaces)

What is a Phenotypic Character (Merkmal)?





Characters

Idea:

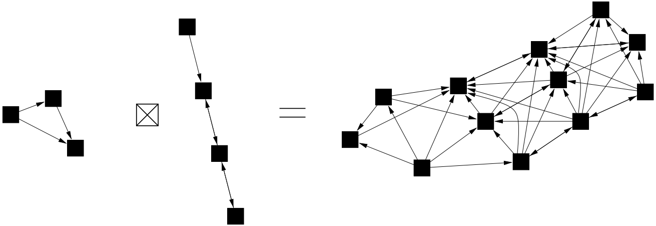
Characters **can** vary independently

⟺

Factors of phenotype space

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.

The Question

Can we actually compute such factorizations???
For real data, i.e., when the original graph is known only appproximately?
“Regionally”, i.e., when we want a coordinate system only is some part of the space?
$\Longrightarrow$ Wilfried’s Presentation