Grundlagen der Systembiologie und der Modellierung epigenetischer Prozesse

Sonja J. Prohaska

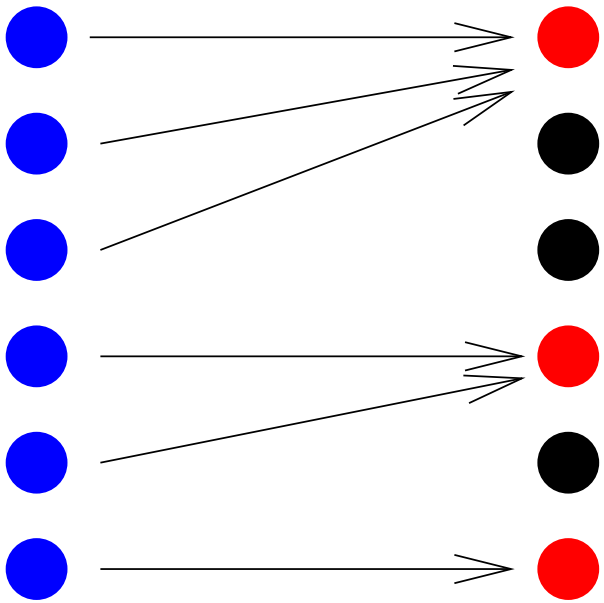
Bioinformatics Group
Institute of Computer Science
University of Leipzig

November 15, 2010

Genothype-Phenotype Map

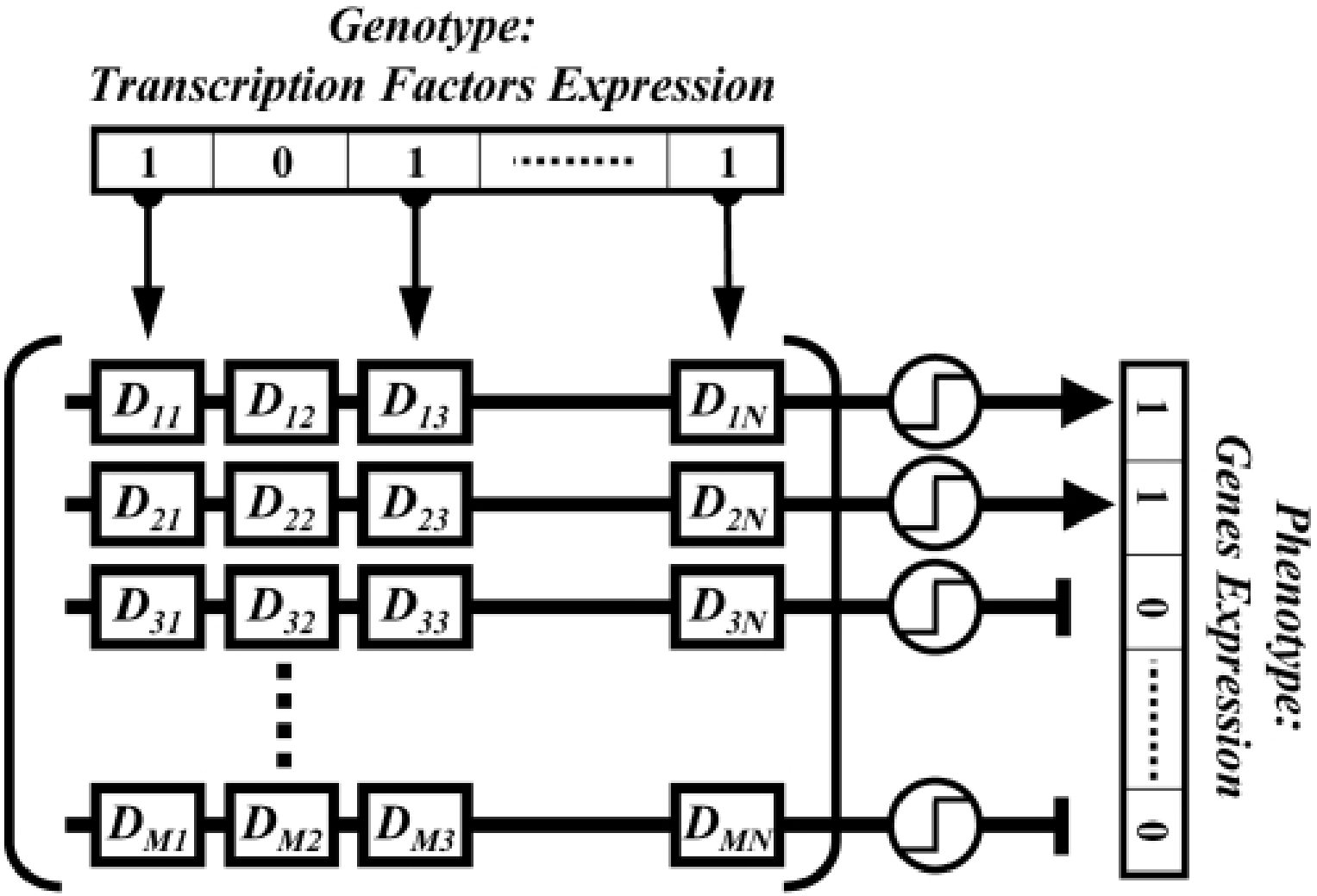
- morphological variation is extensive early in the history of multicellular life
- phenotypes are sparsely distributed in the space of "potential" phenotypes
- diversity is better predicted by variation in the structure of gene regulatory networks that variation in the presence and absence of structural genes
- the effects of one gene are modified by one or several other genes $\rightarrow$ epistasis
- a single gene influences multiple phenotypic traits $\rightarrow$ pleiotrophy

genotypes phenotypes



distinct visible phenotypes

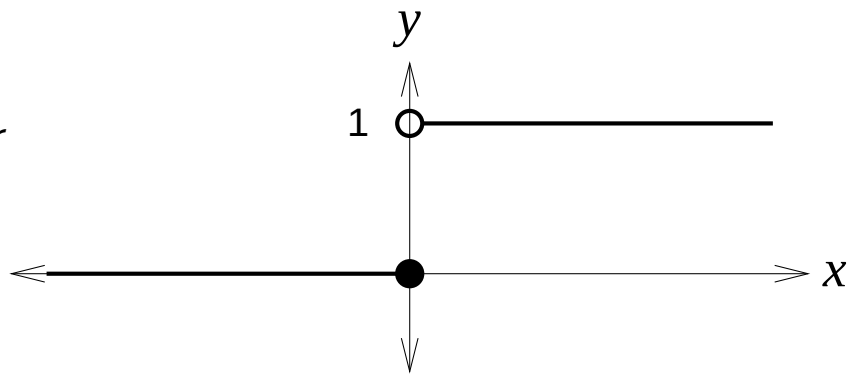
The Developmental Plan



Basic (Single-Layer) Model

$$\vec{p} = H(D\vec{g})$$

$\vec{p}$	...	phenotype vector, of length k
$\vec{g}$	...	genotype vector, of length r
D	...	developmental plan, matrix $k \times r$
D_{ij}	...	$\in \{-1, +1\}$
		-1 for repression of i by j
		$+1$ for activation of i by j
H	...	heaviside function (stepfunction)



if $r < k$ then 2^r phenotypes
if $r = k$ then $2^k = 2^r$ phenotypes
if $r > k$ then 2^k phenotypes (biological nonsense)

Multilayered Model

Allow multiple regulatory layers, where t denotes the layer number and $\vec{p}_0$ corresponds to the genotype $\vec{g}$. At least two layers are required to produce a universal Turing machine.

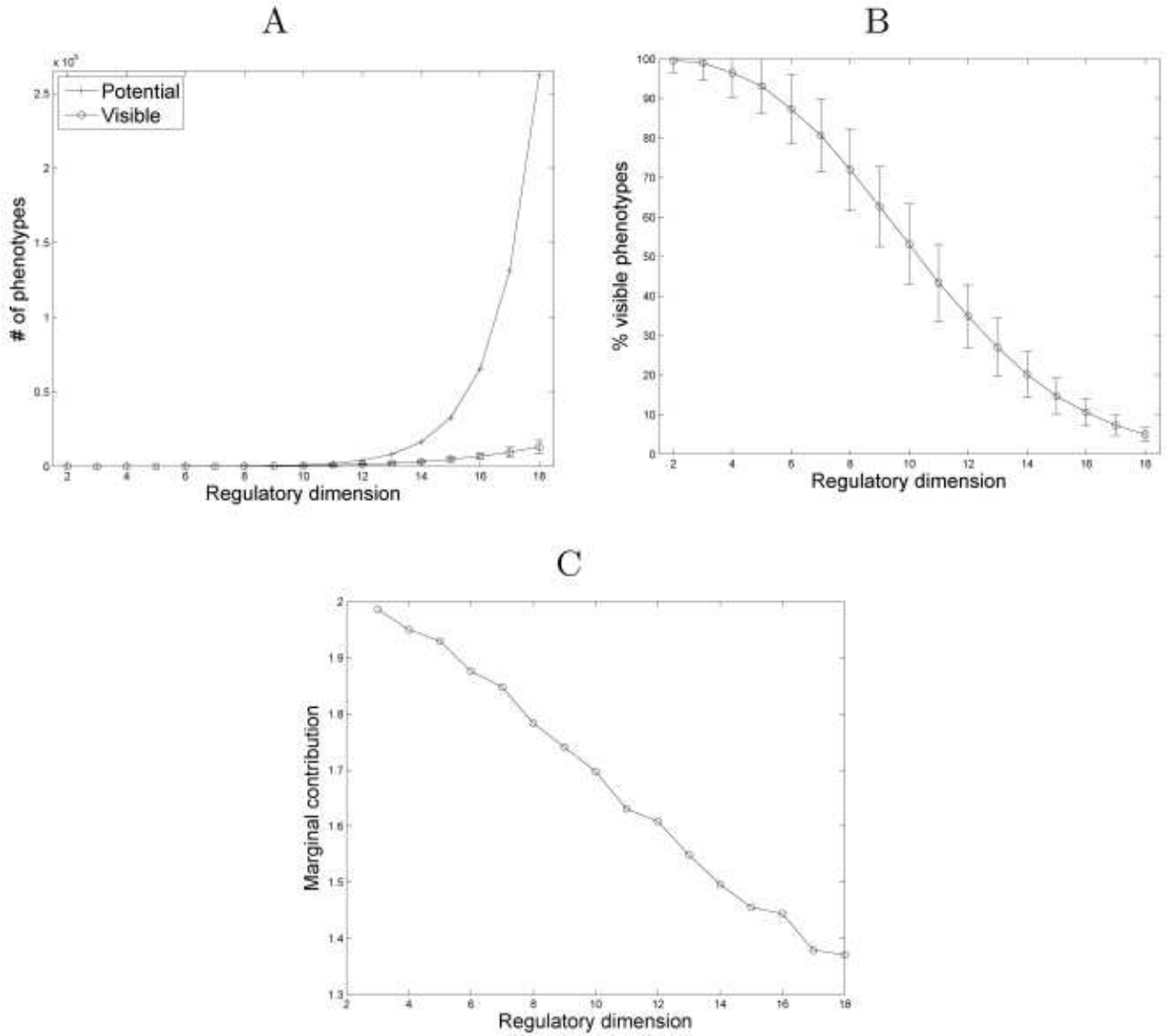
$$\vec{p}_t = H(D_t \vec{p}_{t-1})$$

- D is the same for every layer t
- independent D_t's for every layer t

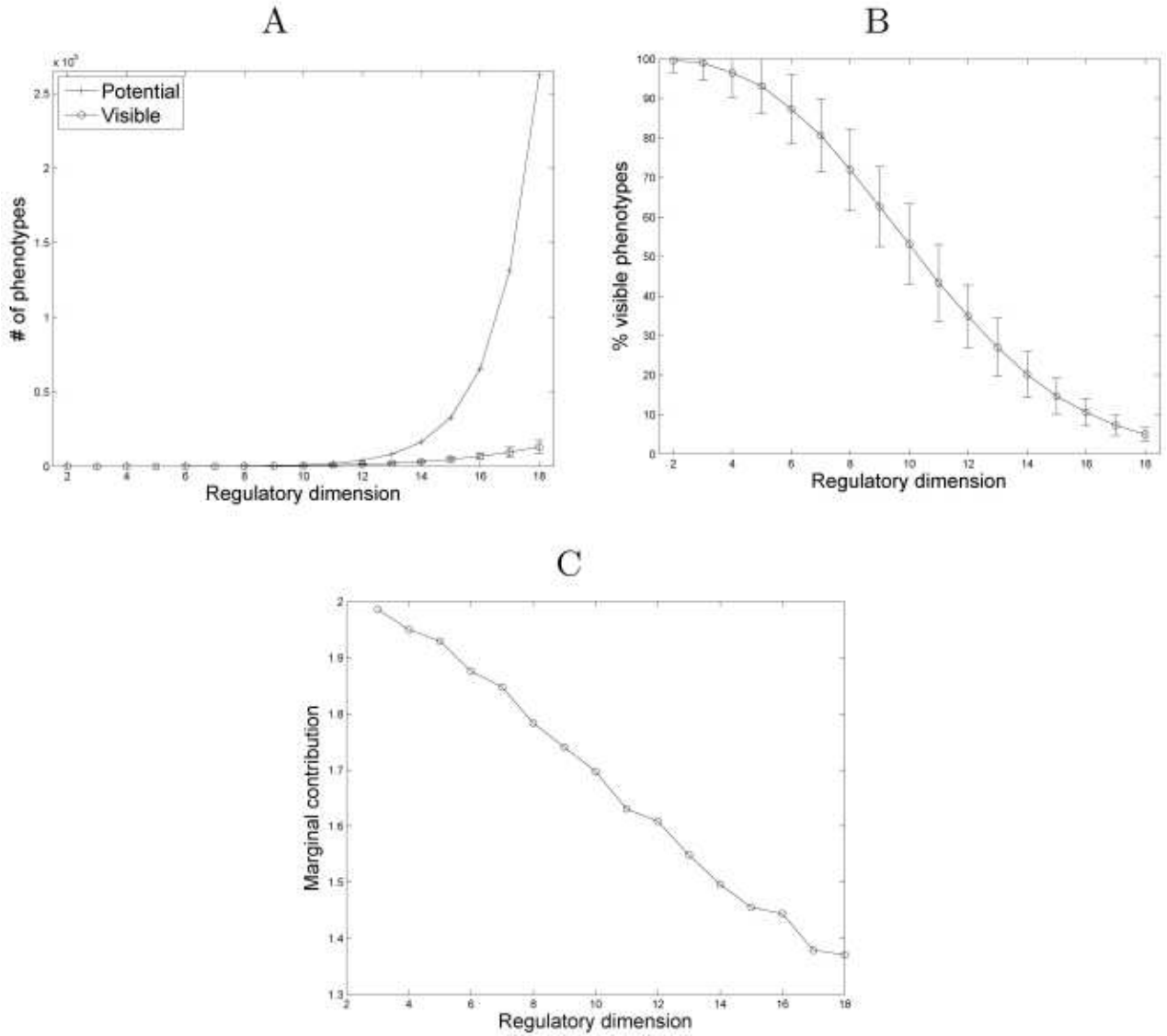
Partially Connected Model

c ... density parameter, the probability for each entry in the matrix to be attributed with a nonzero value $\rightarrow D_{ij} \in \{-1, 0, +1\}$

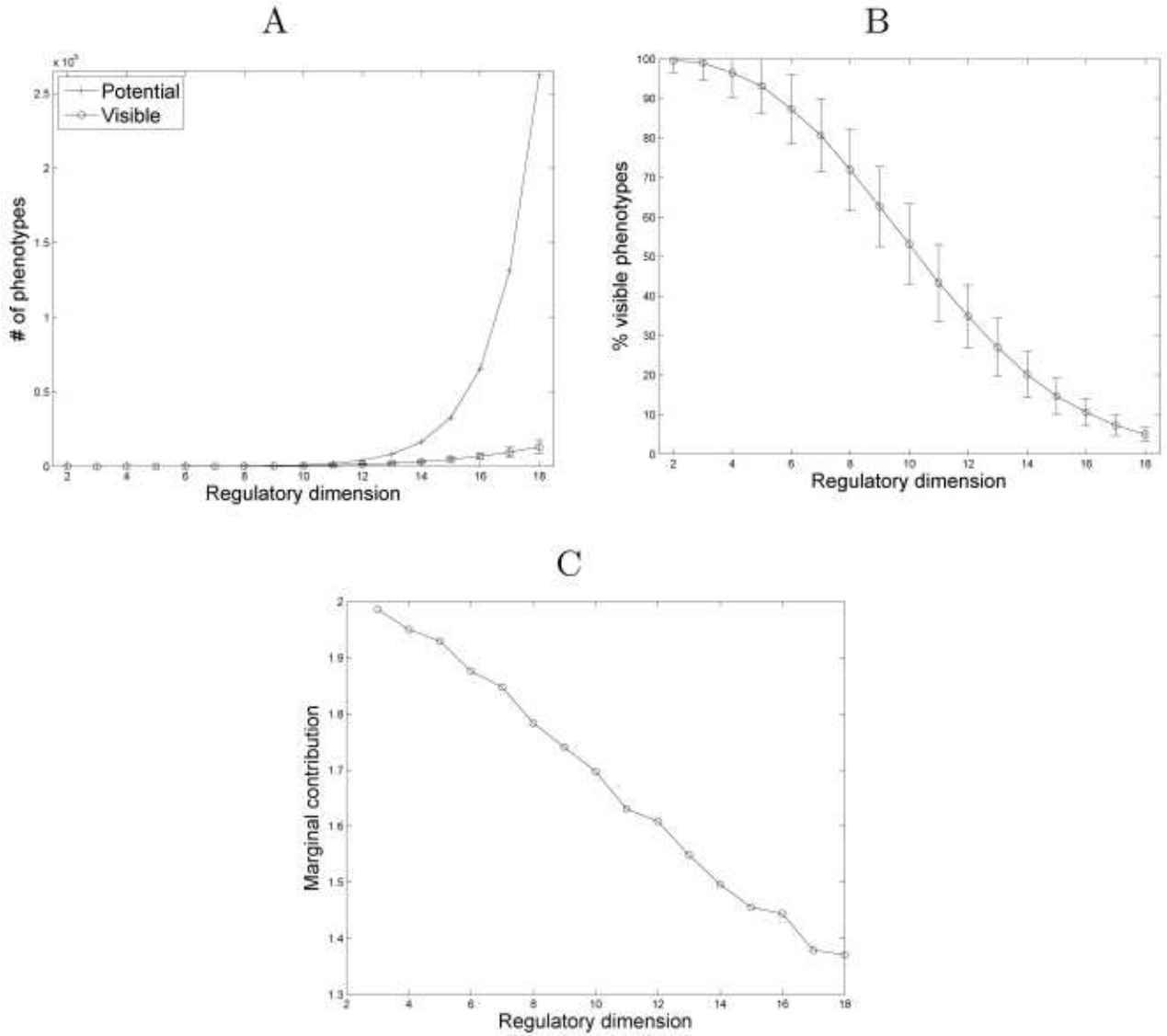
- "fully connected" ($c = 1$): all r regulators influence all k phenotypes (full epistasis and full pleiotrophy)
- "partially connected" ($c < 1$): a zero at D_{ij} indicates that regulator j has no influence on phenotype i. (a biologically more realistic model)



A – The number of (visible) phenotypes increases with r.



B – The fraction of visible phenotypes out of the number of potential phenotypes rapidly declines with r.



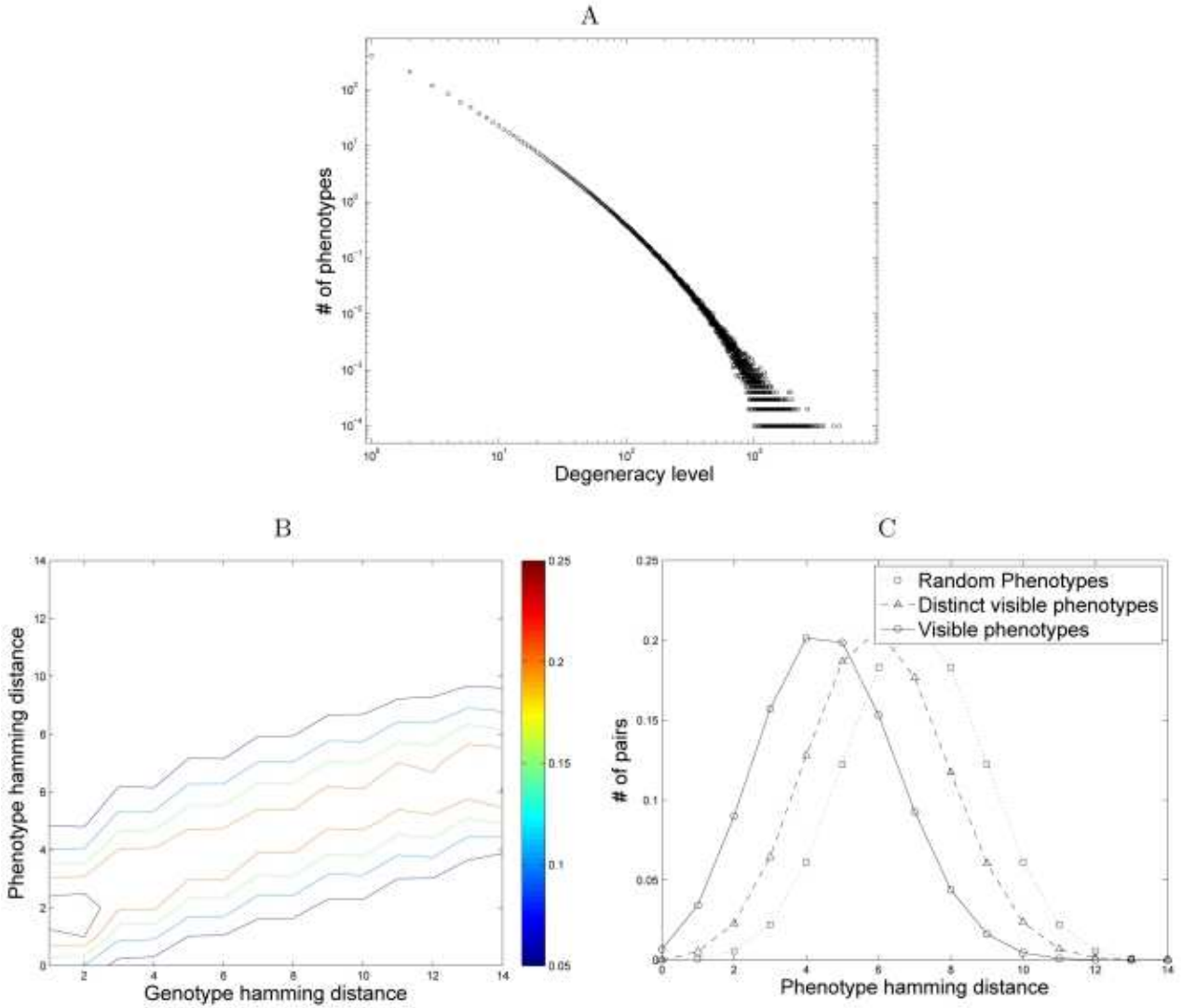
C – The (marginal) contribution of each genetic element to the visible phenotypic space decreases with r.

degeneracy level

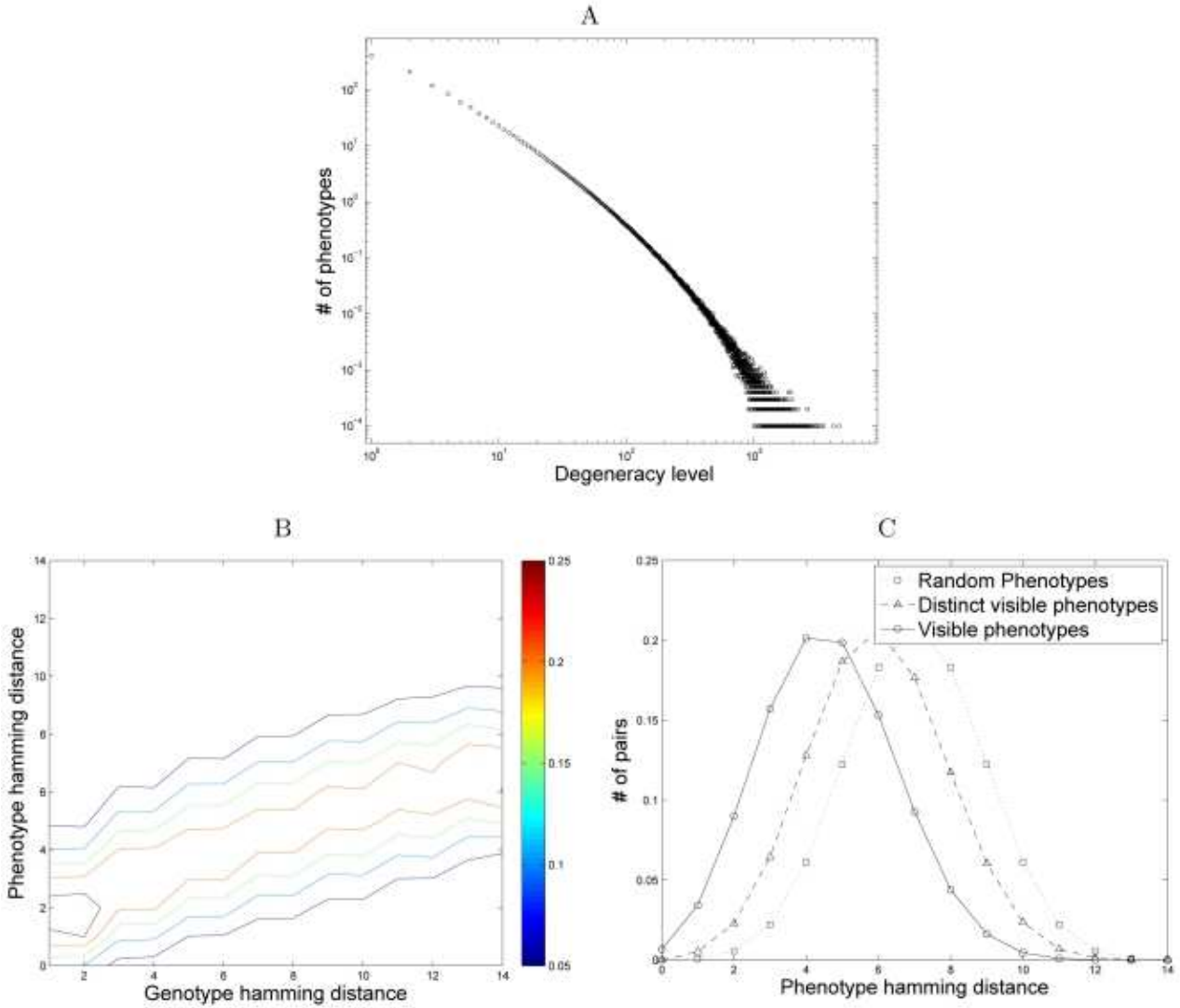
- fore each phenotype p we calculate the degeneracy level n_p denoting the number of genotypes that produce it
- for visible phenotypes $n_p > 0$

gain function

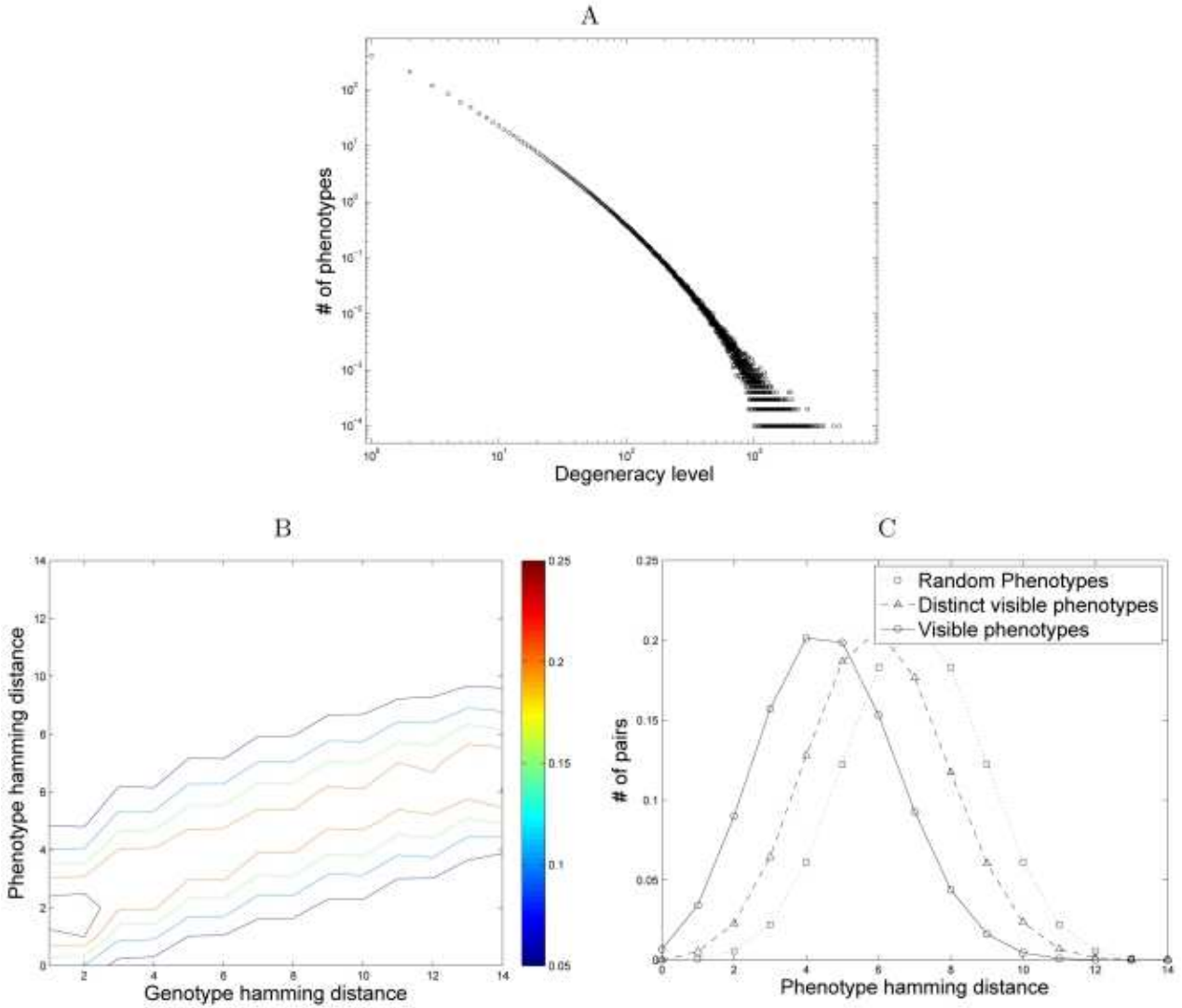
- describes the average pairwise Hamming distance among phenotypes whose corresponding genotypes are a certain Hamming distance apart
- it measures how the magnitude of a perturbation in the genotype space maps onto perturbations in the phenotype space



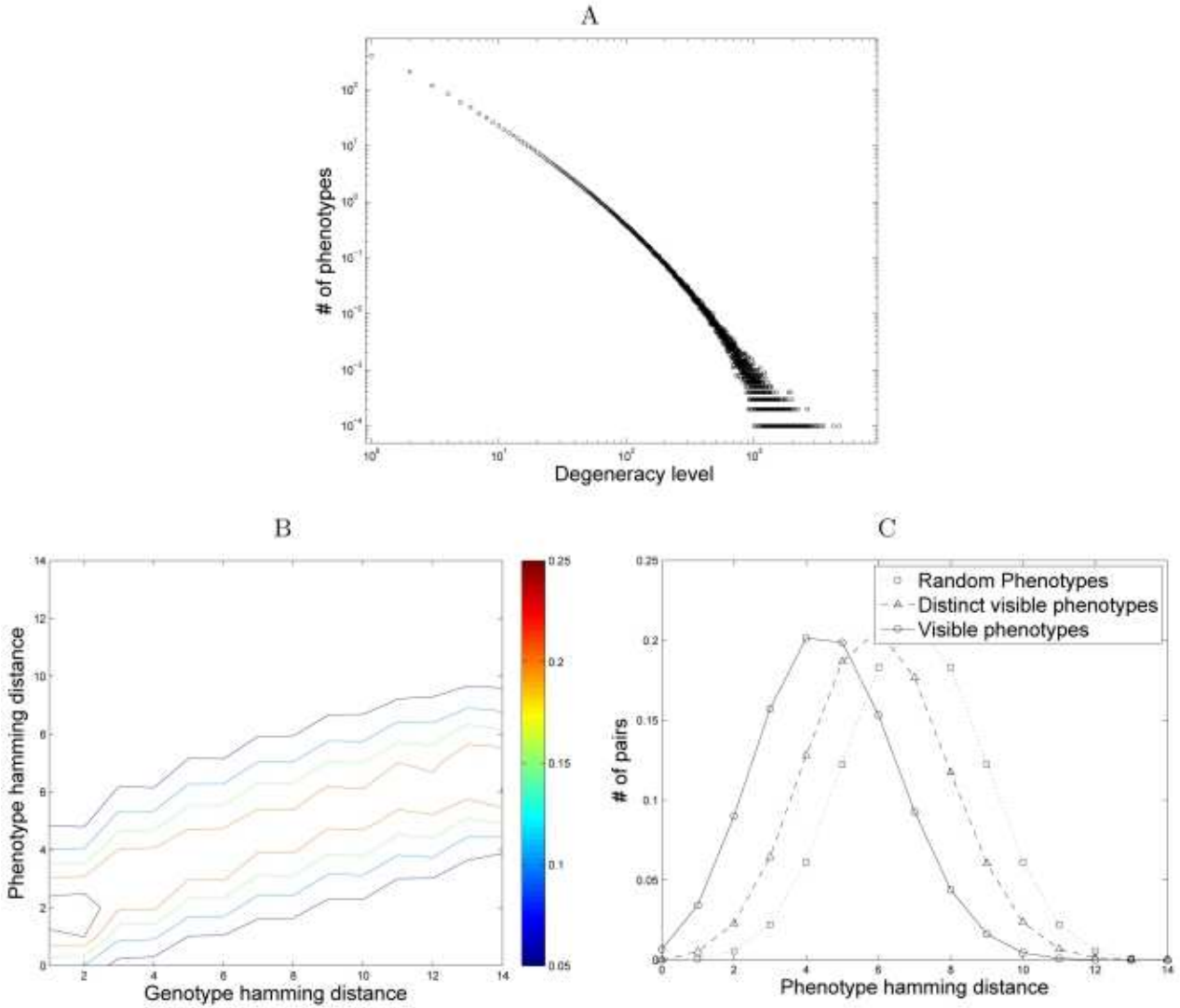
A – The distribution of degeneracy levels fits a generalized power-low distribution.



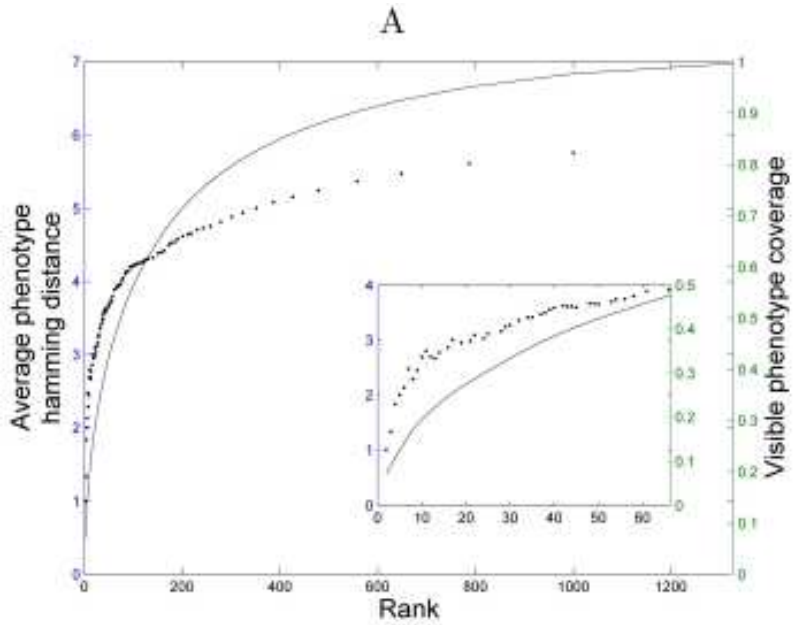
B – A large perturbation in the genotypic space produces on average, a significantly smaller perturbation in the phenotype $\rightarrow$ canalization.

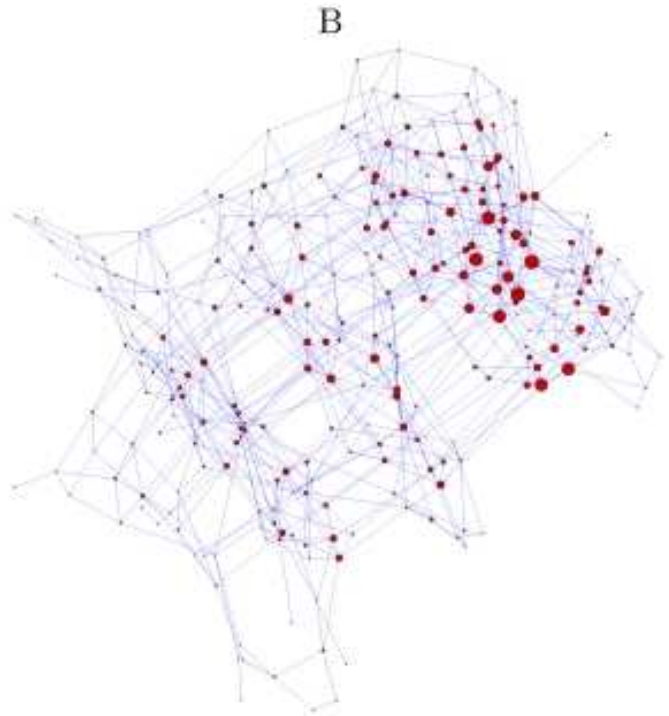


C – Visible phenotypes tend to be more similar than expected by chance.

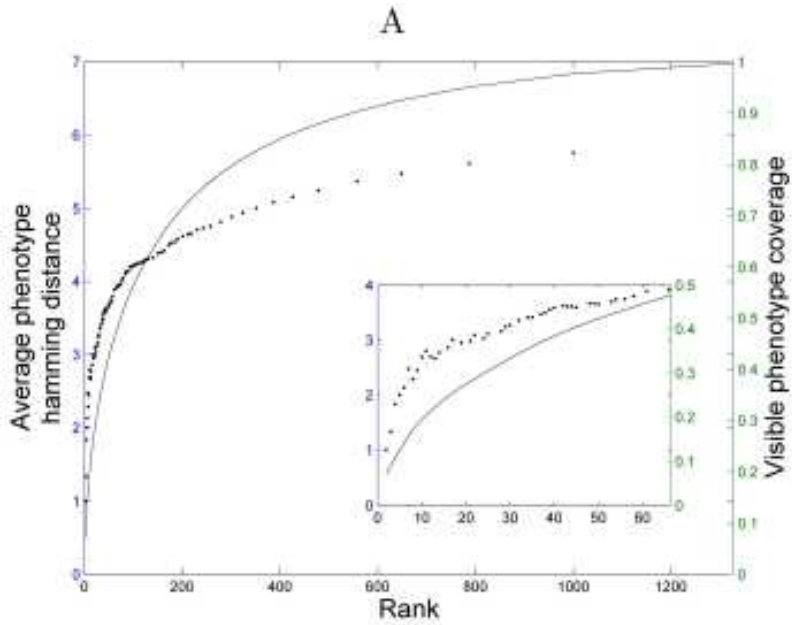


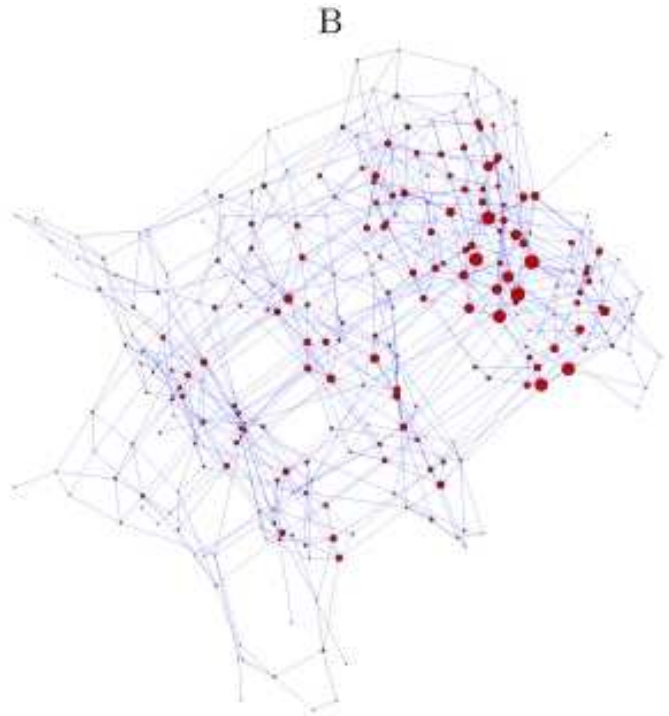
The most frequent phenotypes span a smaller subspace than the total visible phenotypes, and are located towards the center of the visible phenotype set.



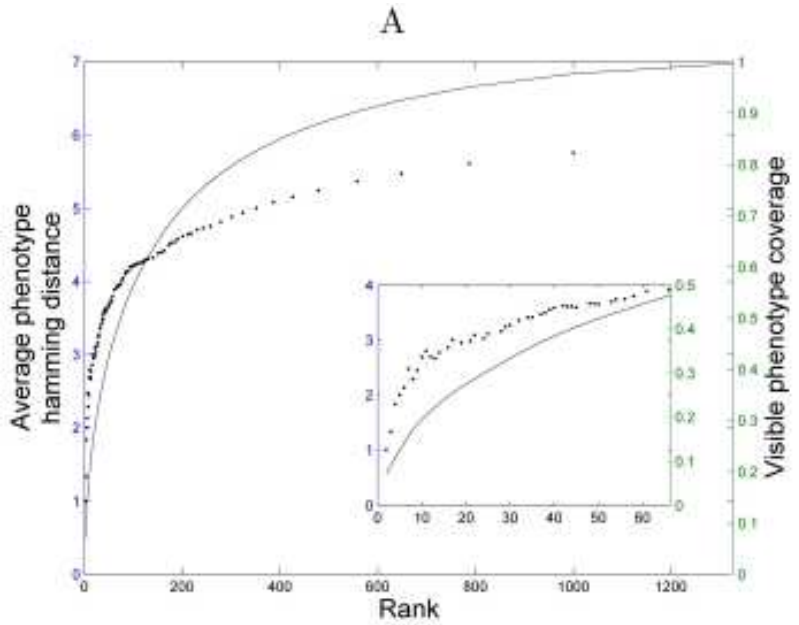


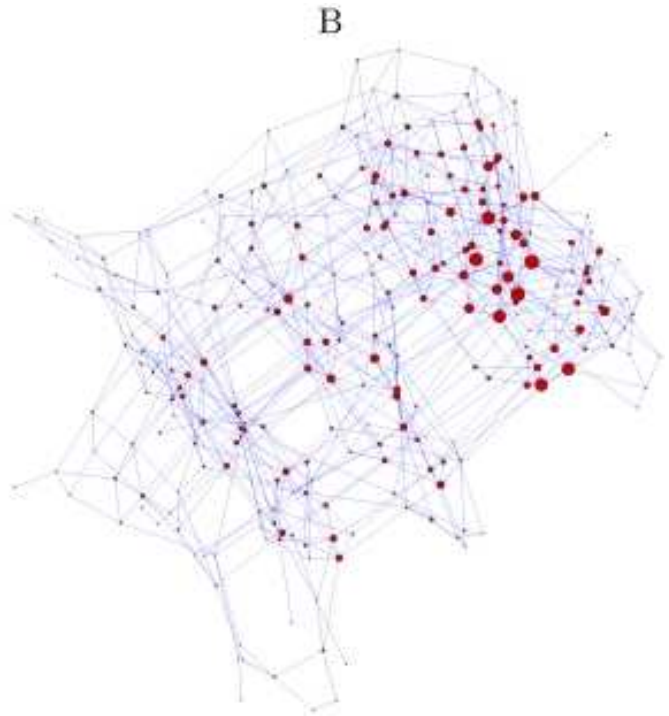
A – The distance between the most frequent phenotypes is significantly smaller than the average distance.



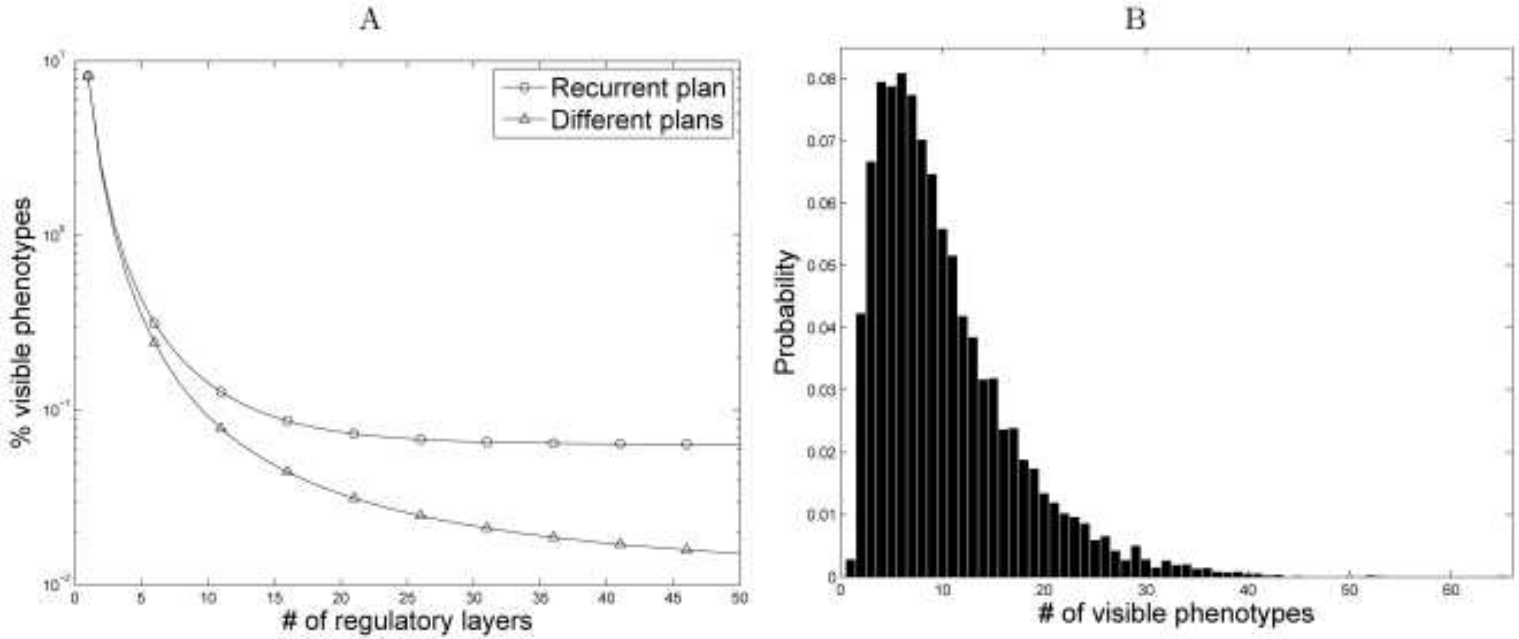


A inset – The top 5% most frequent phenotypes are very similar yet cover approximately 50% of all visible phenotypes.

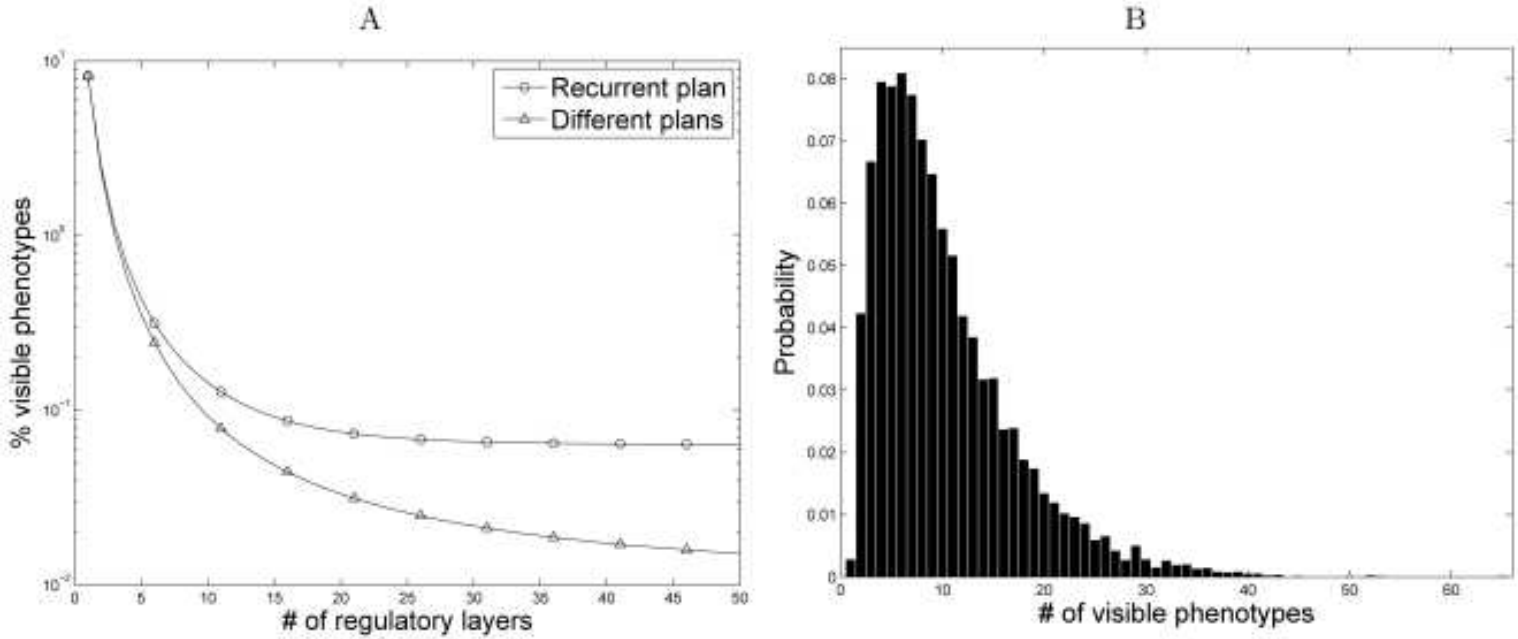




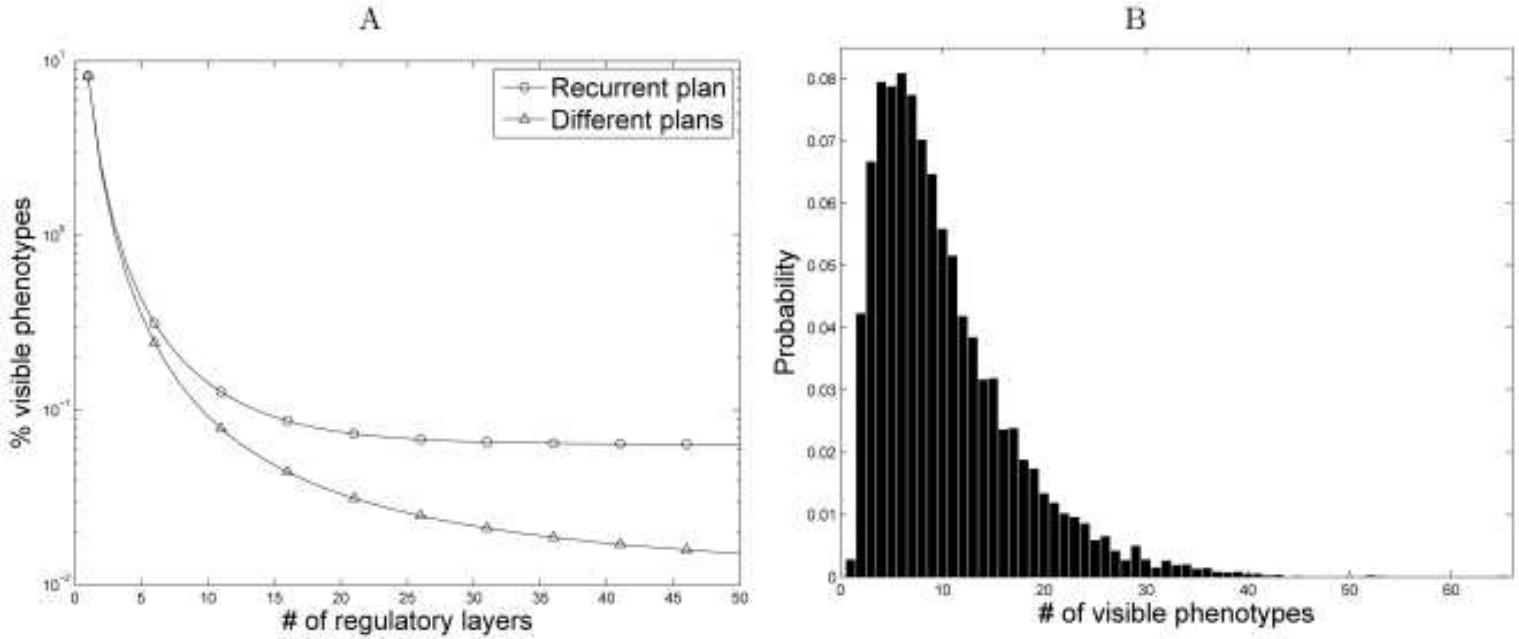
B – The network of visible phenotypes.



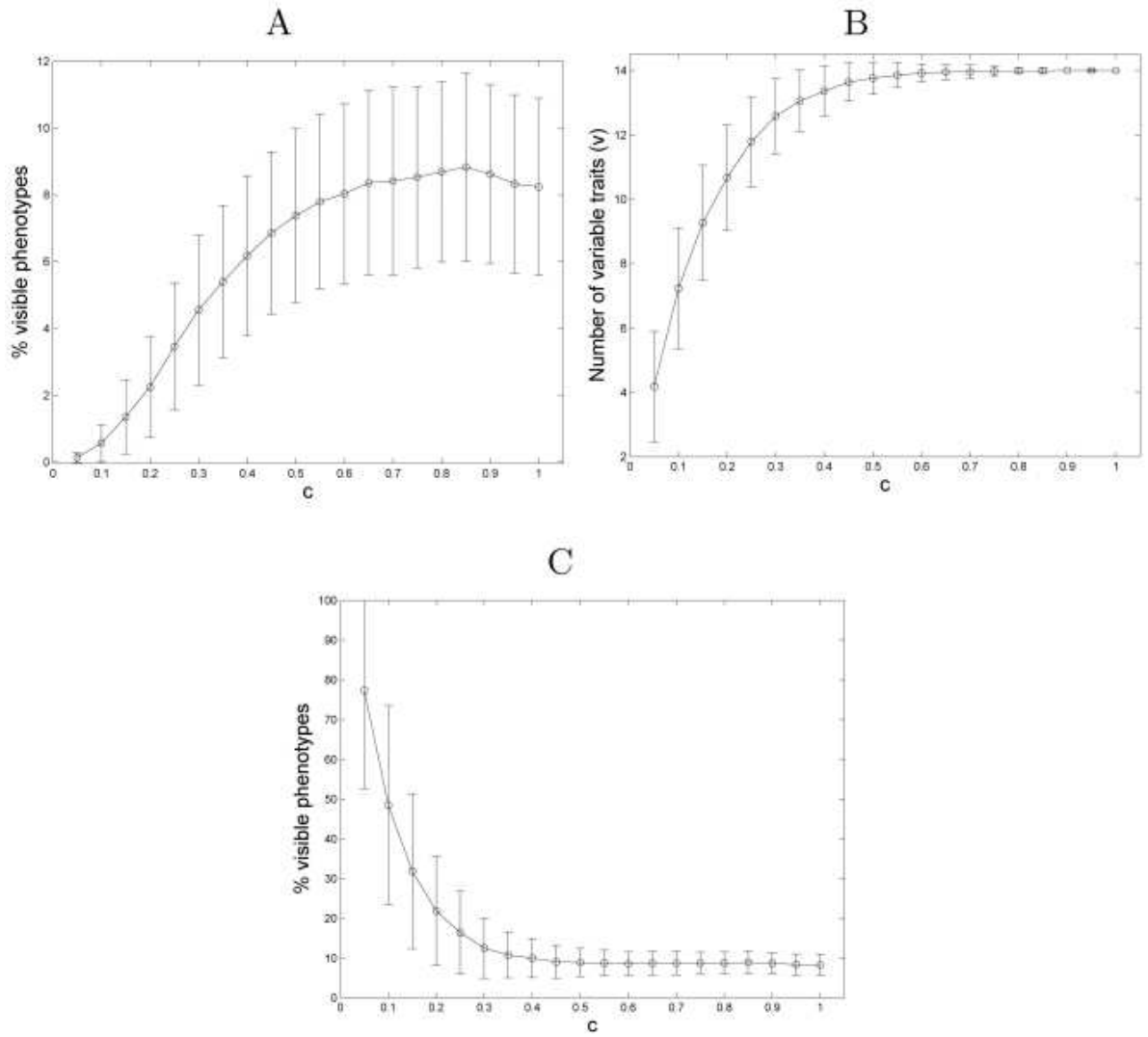
A – The Multilayered model shows with each additional layer a reduction in the number of visible phenotypes and an increase in canalization.



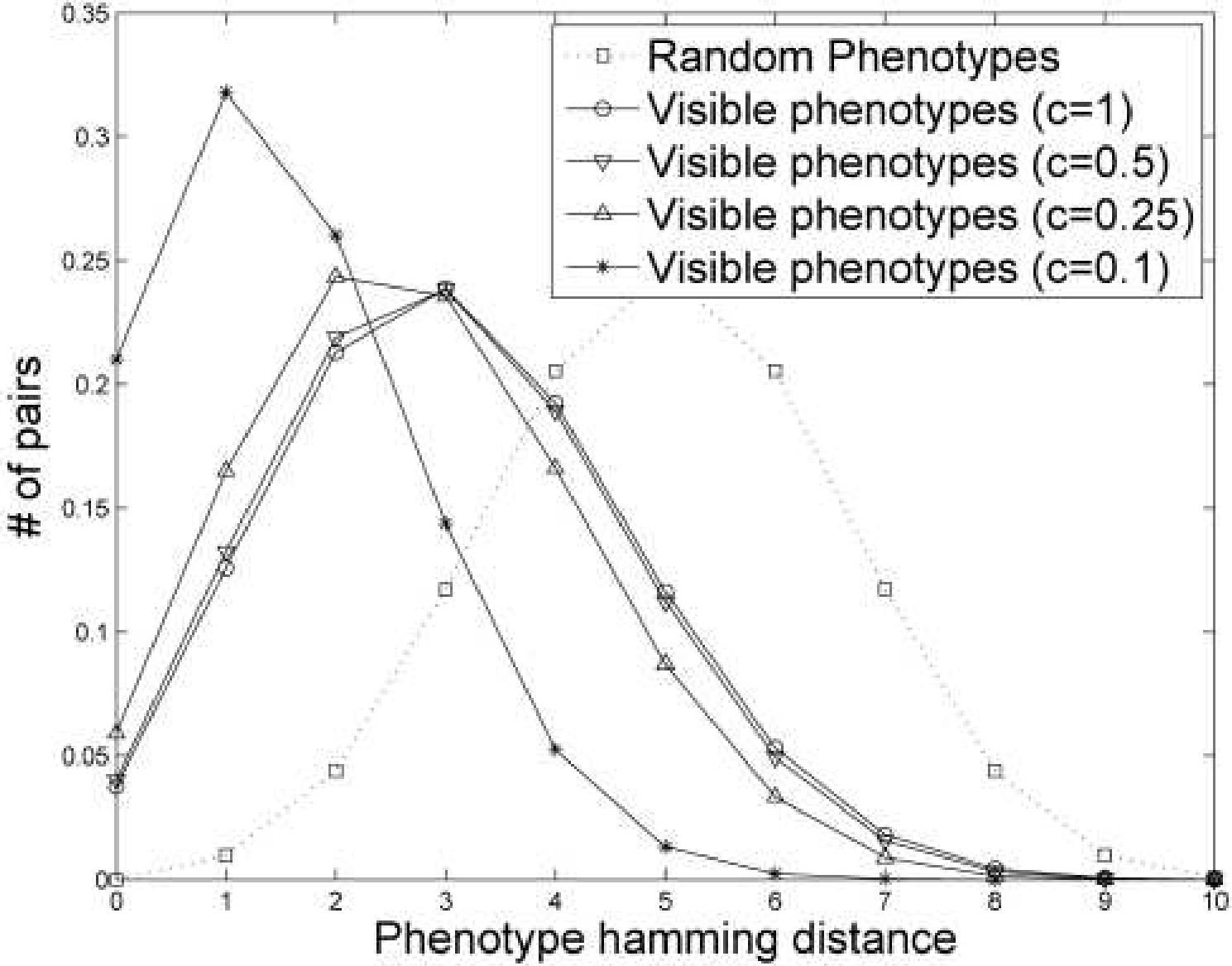
B – The number of visible phenotypes at steady state is only 10.3 ± 7 and this steady state is reached after 17.7 ± 7 layers.



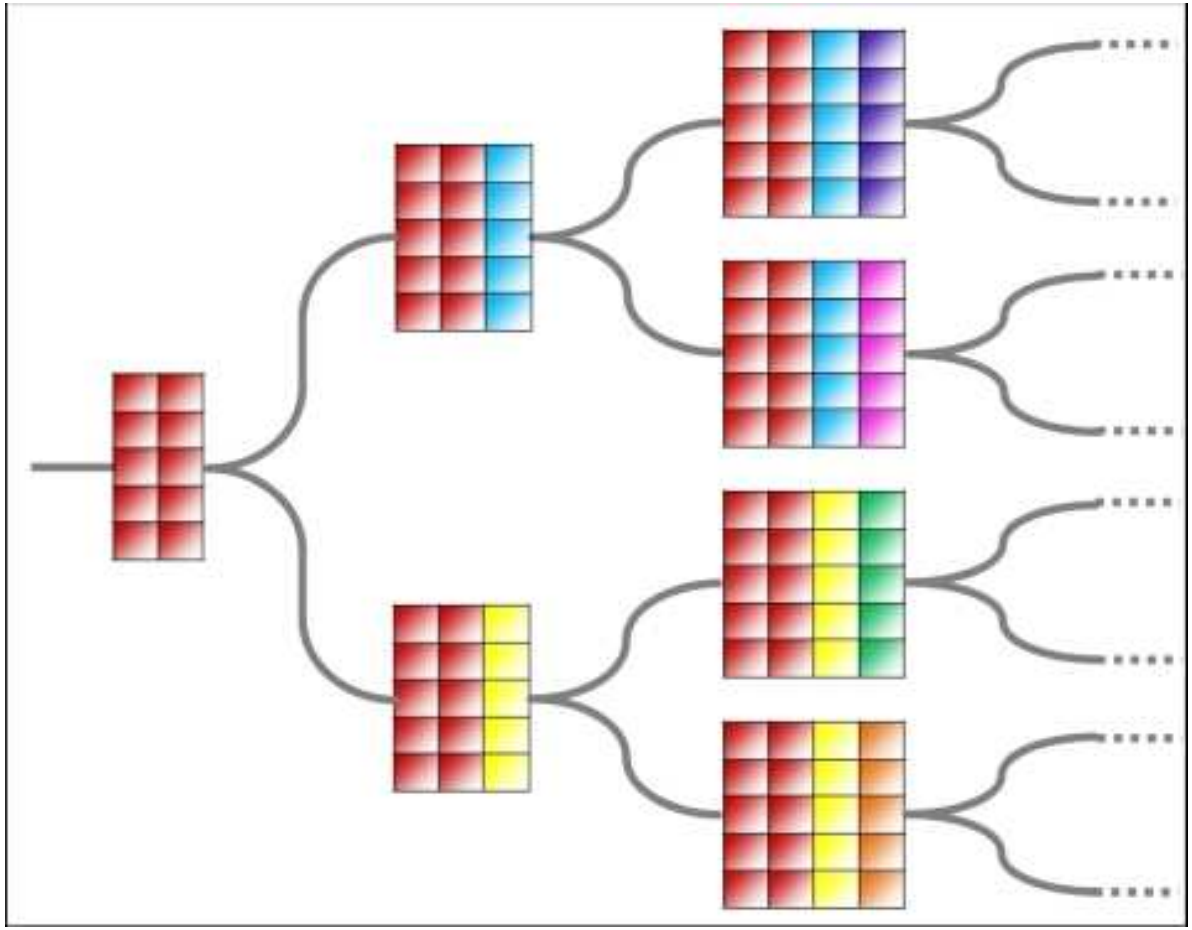
Different developmental plans at each layer t yield a more complete reduction in the number of visible phenotypes.



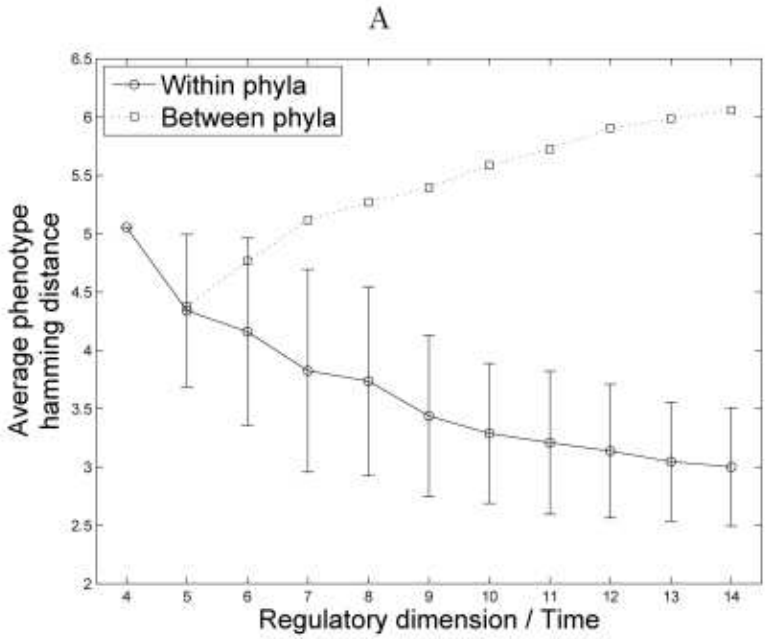
Entries in the phenotypic vector are never activated $\rightarrow$ reduce the number of visible phenotypes. Each phenotypic element is influenced by only a few genes, increasing their effect on the phenotype $\rightarrow$ increase the number of visible phenotypes.

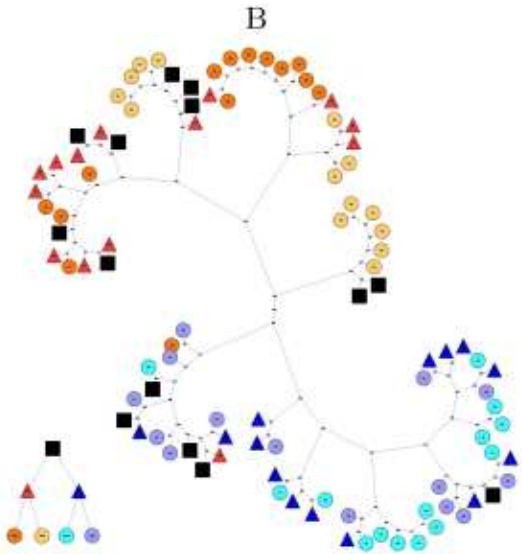

Random Phenotypes
Visible phenotypes (c=1)
Visible phenotypes (c=0.5)
Visible phenotypes (c=0.25)
Visible phenotypes (c=0.1)
# of pairs
Phenotype hamming distance
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10

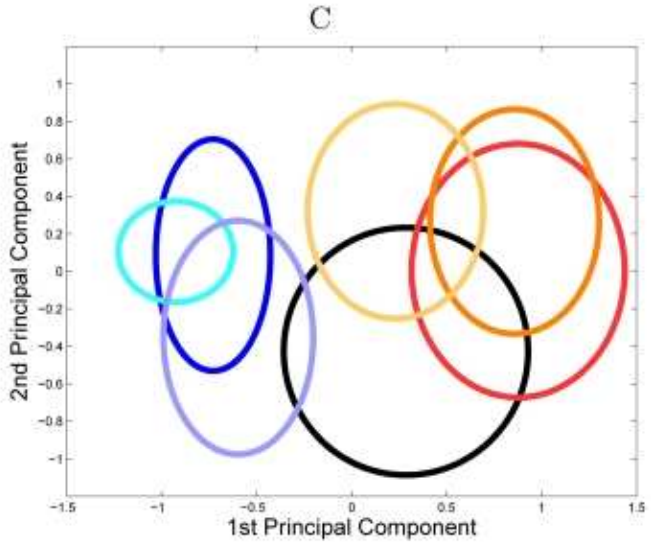
Evolution of Developmental Plans – A Model



Developmental plans evolve incrementally by addition of new genetic regulatory elements into existing regulatory networks.


A
Within phyla
Between phyla
Average phenotype hamming distance
Regulatory dimension / Time
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
4
5
6
7
8
9
10
11
12
13
14


B


C
2nd Principal Component
1st Principal Component
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.5
-1
-0.5
0
0.5
1
1.5

We observe that the phenotypes comprising a single developmental plan, become more similar throughout the evolutionary process, whereas disparity among members of different plans increases.
Intermediate ans ancestral groups span the same phenotypic space as their descendants.
The strong dependency of the phenotypic element on the number of $+1$ elements in the corresponding developmental matrix row is the source for the nonuniform distribution of degeneracy levels.
In the unit matrix, every change in the genotype induces a change in the phenotype. This would decrease the level of degeneracy of the map and produce more visible phenotypes.

References

[Borenstein, 2008] Borenstein E. and Krakauer DC. *An End to Endless Forms: Epistasis, Phenotype Distribution Bias, and Nonuniform Evolution*. PLoS Comp. Biol. (2008), 4(10):p1-13