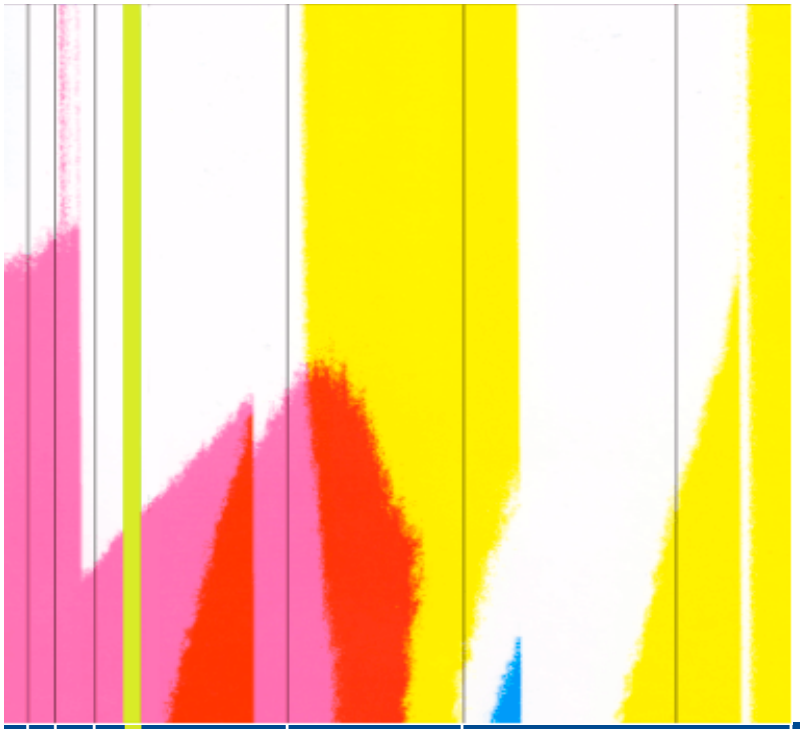


Digital genetics with *aevol*: a view on the (evolutionary) origin of biological complexity

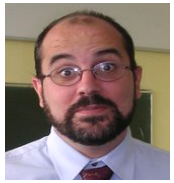
TAGp, Annecy - October 20, 2010

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
INRIA
centre de recherche GRENOBLE - RHÔNE-ALPES

Guillaume Beslon
COMBINING

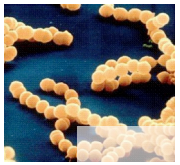
cnrs
Lyon 1
INSA LYON
LIRIS
IXXI Institut des Systèmes Complexes Complex Systems Institute Rhône-Alpes
INRIA

Molecular structure of biological systems



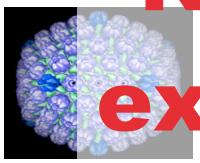
Homo sapiens

~3 billions bp
~25 000 genes



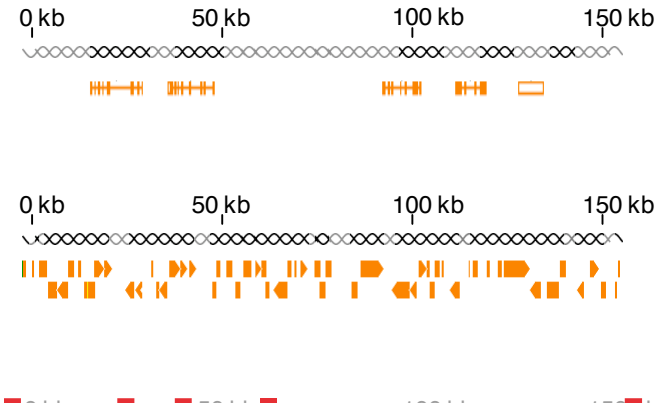
Neisseria meningitidis

~2 millions bp
~2 000 genes

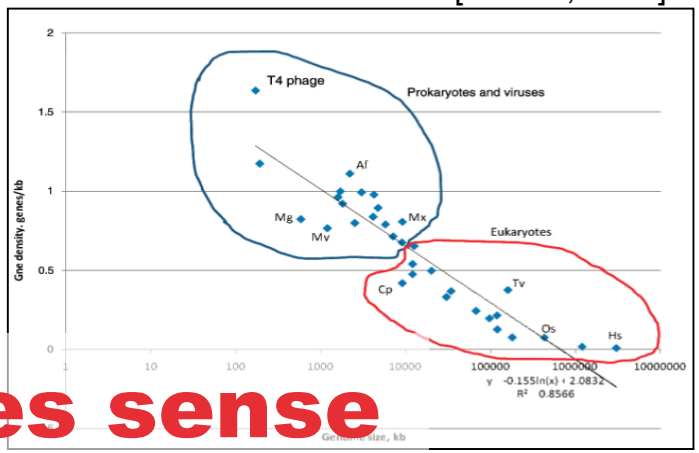


Herpes HSV-1

~150 000 bp
~100 genes

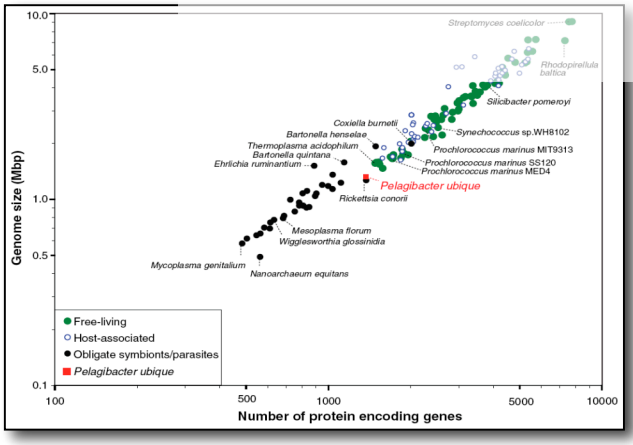


[Koonin, 2009]

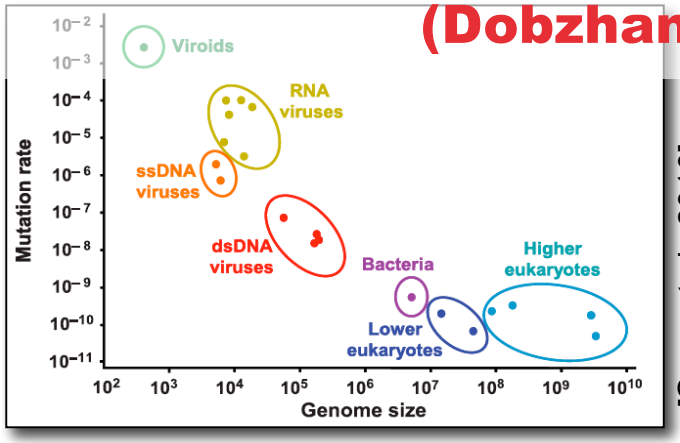


"Nothing in biology makes sense except in the light of evolution"

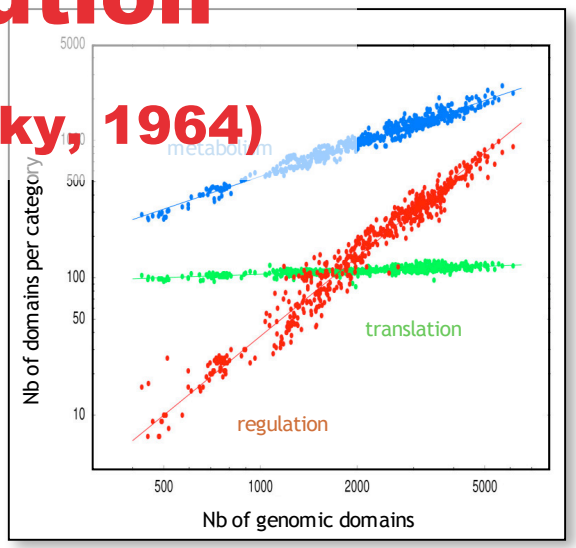
(Dobzhansky, 1964)



[Giovanonni et al., 2005]

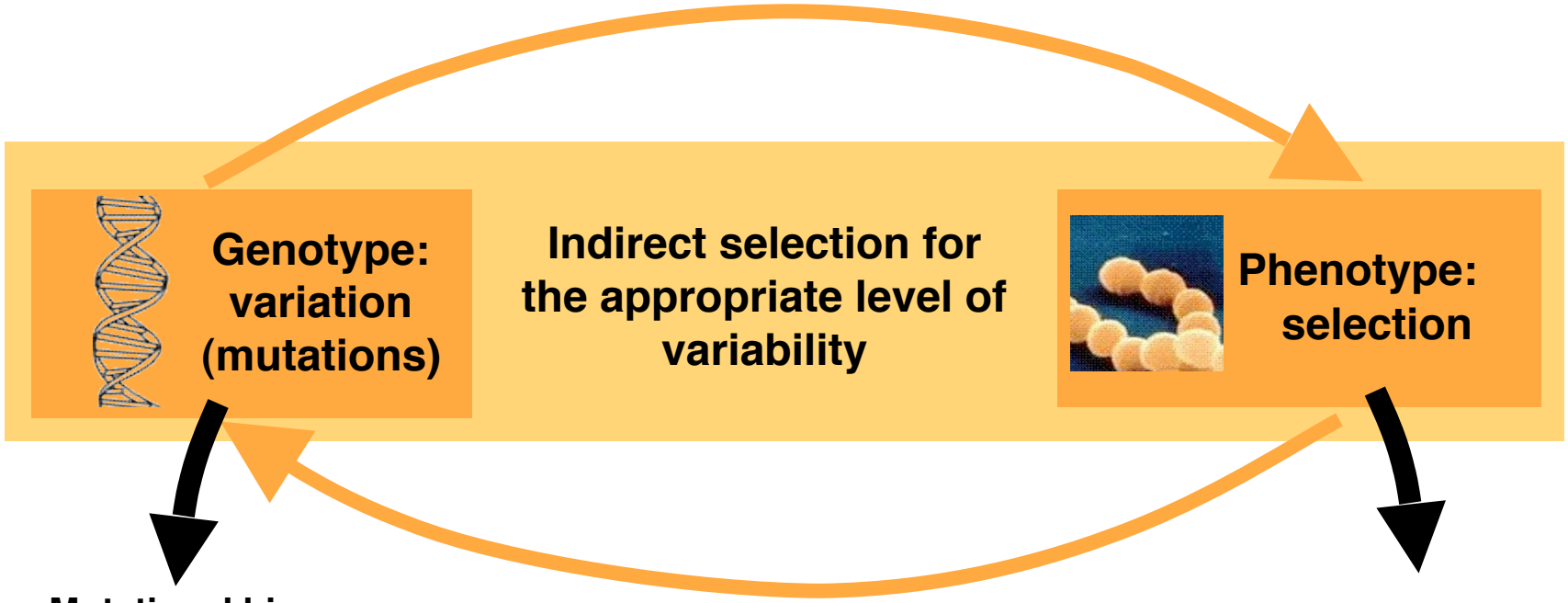


[Gago et al., 2010]



[Molina & Van Nimwegen, 2008]

Evolutionary origin of biological systems



Mutational biases:

"Homo Sapiens genome spontaneously undergoes more insertions than deletions"

"Gene duplication spontaneously creates scale-free regulation networks"

Selective costs/advantages:

"In virus and bacteria genomes are physically constrained"

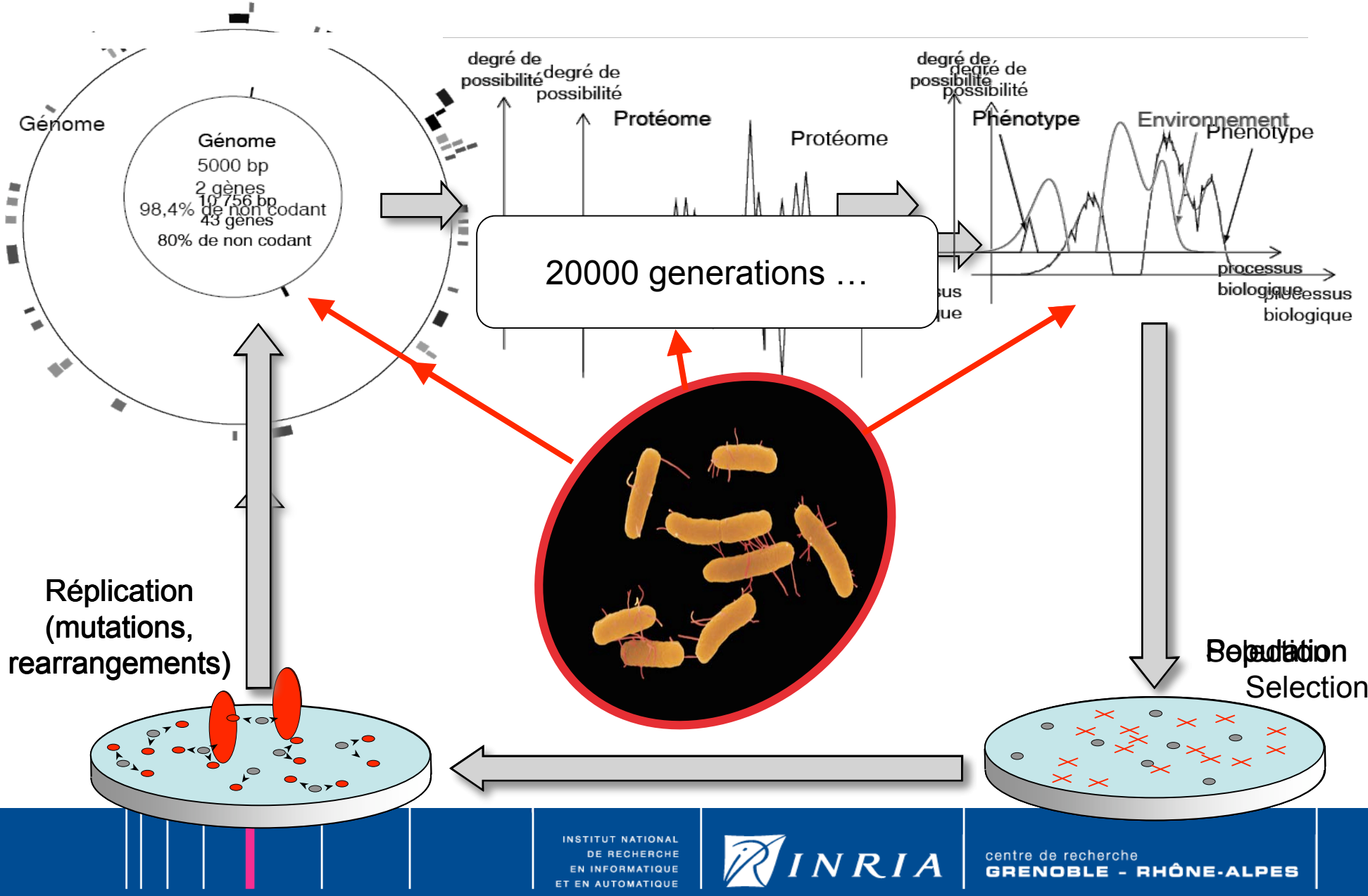
"Complex environments lead to complex regulation networks"

Biological question

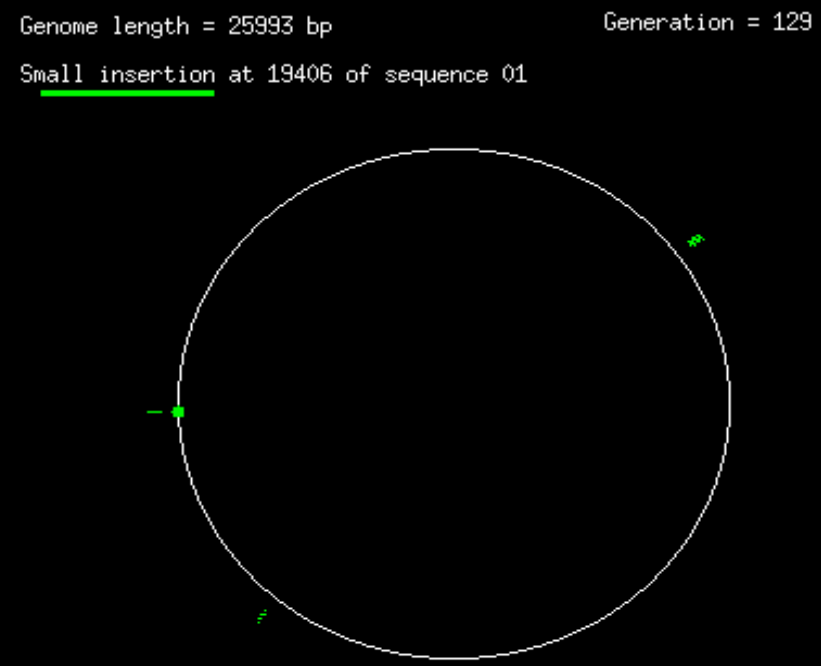
1. The scaling of most aspects of genomic architecture and gene structure with genome size fall on a continuum from viruses to prokaryotes to single-celled eukaryotes to multicellular eukaryotes. This suggests that general population-genetic mechanisms, transcending cellular and metabolic features, are the predominant drivers of interspecific divergence in genome architecture. (Lynch, Annu. Rev. Microbiol., 2006)

- Can indirect selection of an appropriate variability level explain the various genetic organizations observed in nature?
  - Mutational burden hypothesis
  - Is there a burden on non-coding region?
- Computational approach
  - Investigations using *in silico* experimental evolution
  - Digital genetics (“real evolution of false organisms”)
  - Individual-Based Modeling of (prokaryotes) evolution

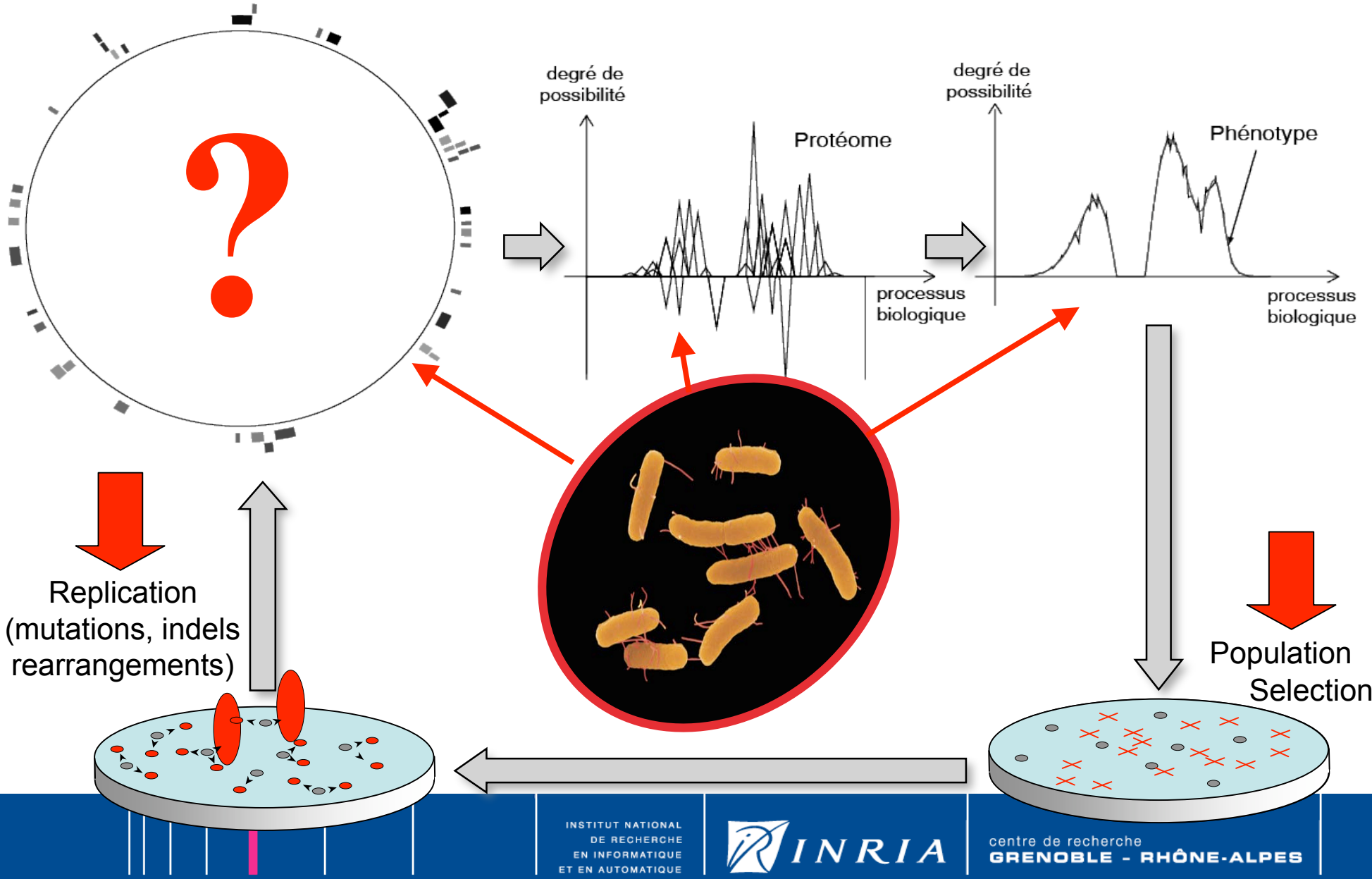
The aevol model (C. Knibbe, 2006)



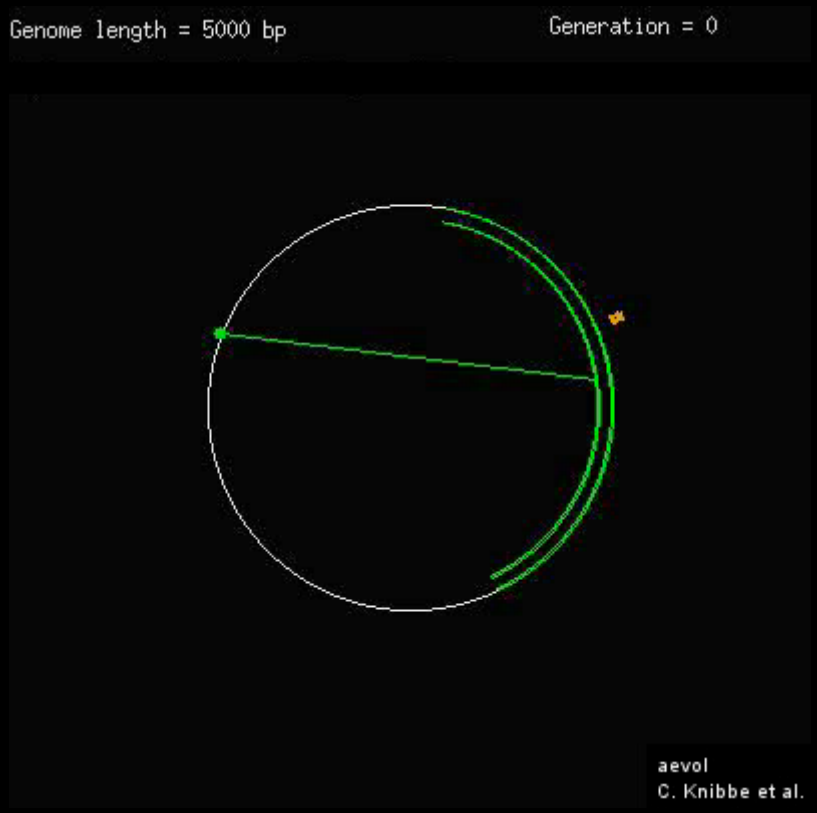
aevol: The movie (« winning » lineage)



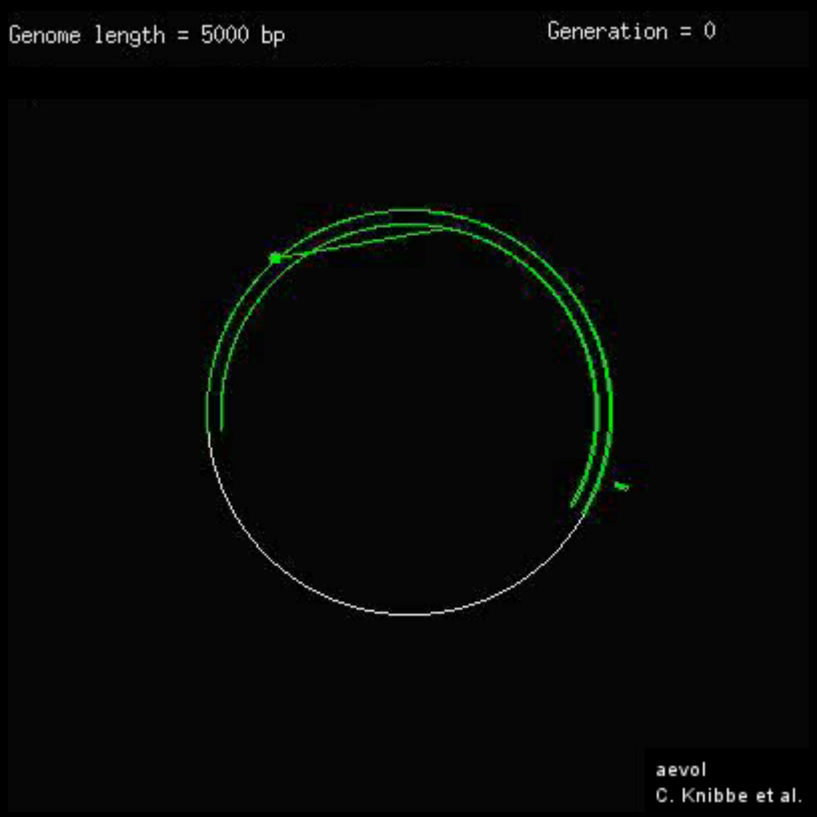
Experimental evolution with aevol



aevol: The movie (II) ...



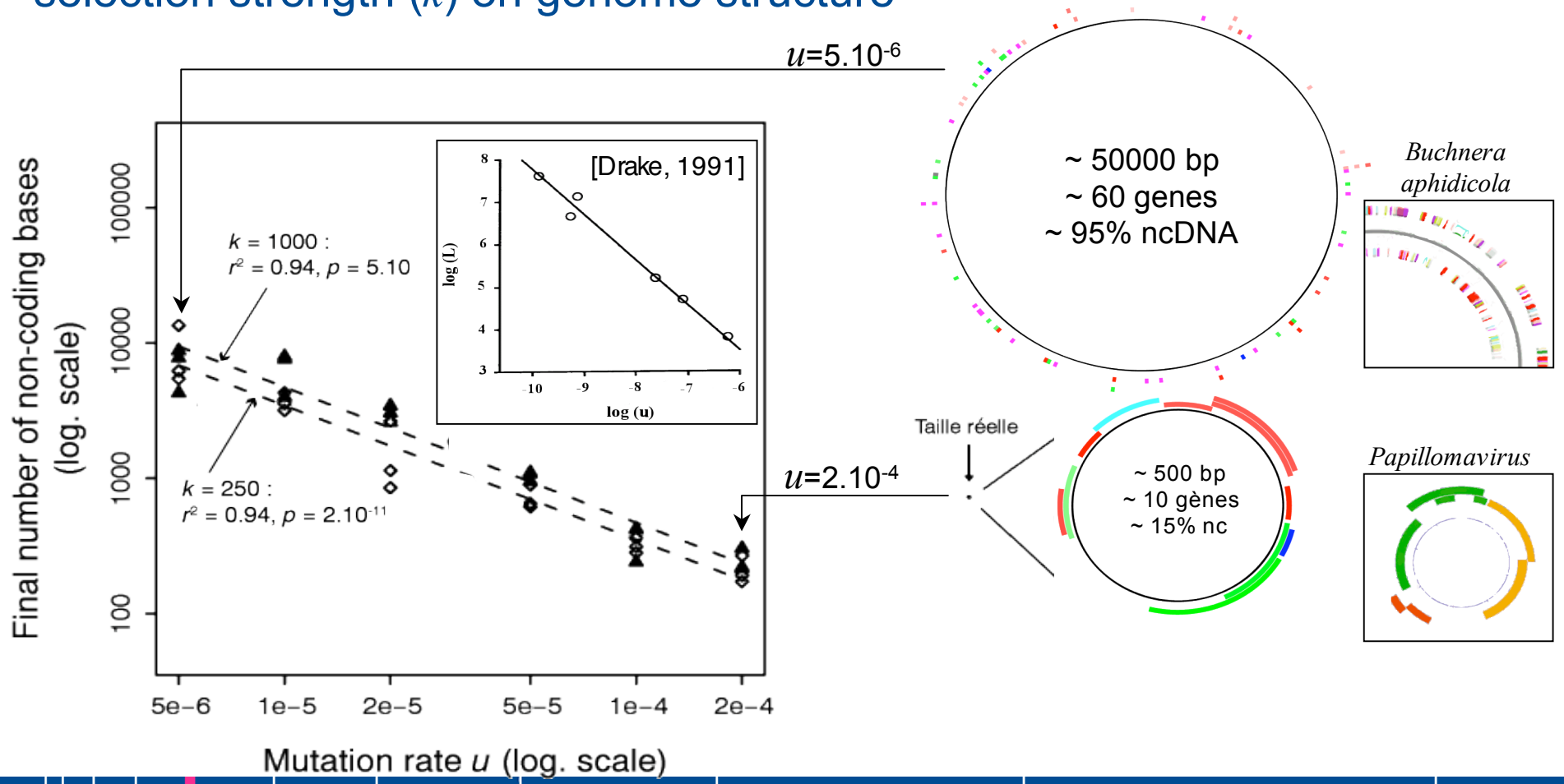
High mutation rates : 2.10^{-4} / pb



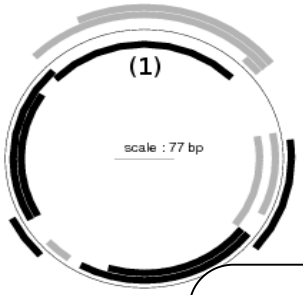
Low mutation rates : 5.10^{-6} / pb

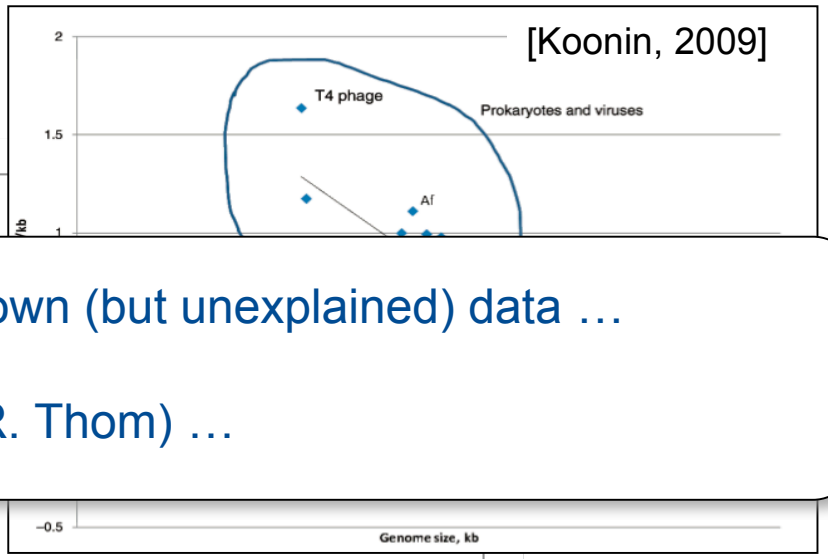
In silico experimental evolution

- Experimental study of the effect of mutation/rearrangement rates (u) and selection strength (k) on genome structure



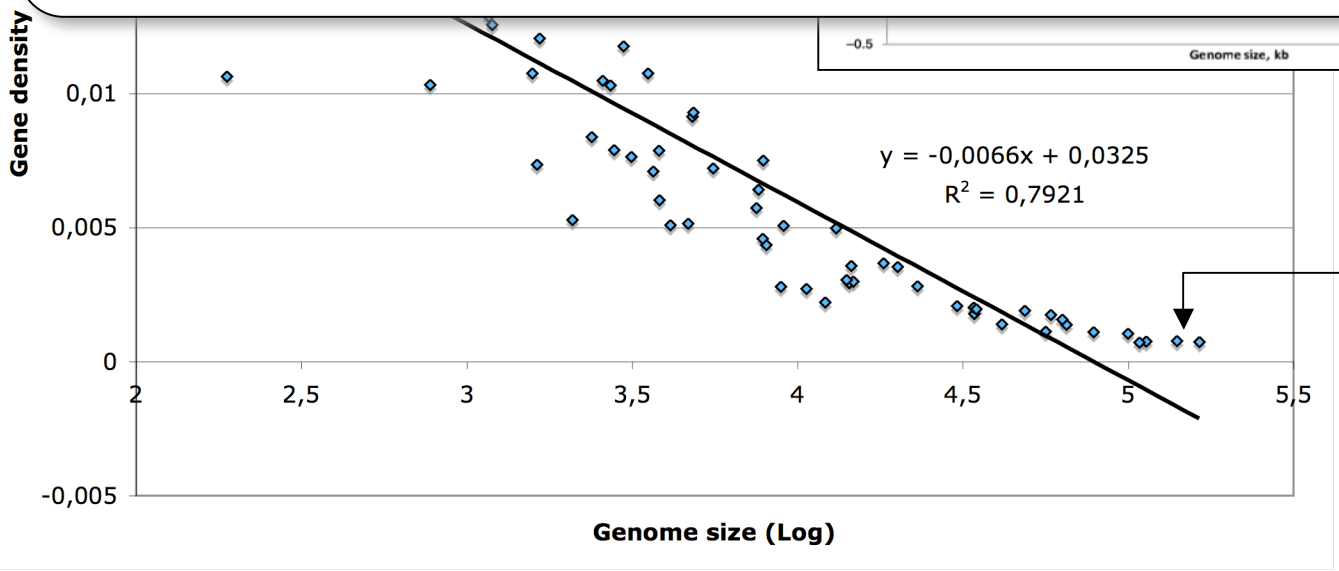
Yet another model explaining everything ;)

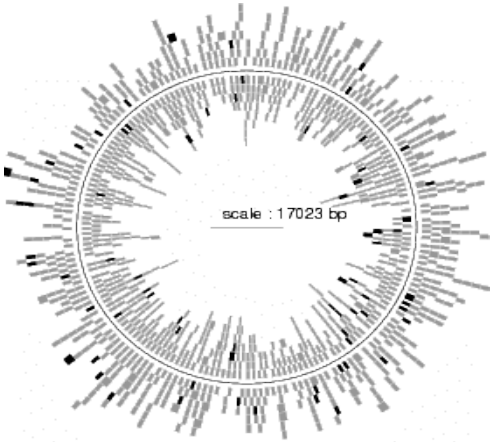




The model is able to reproduce known (but unexplained) data …

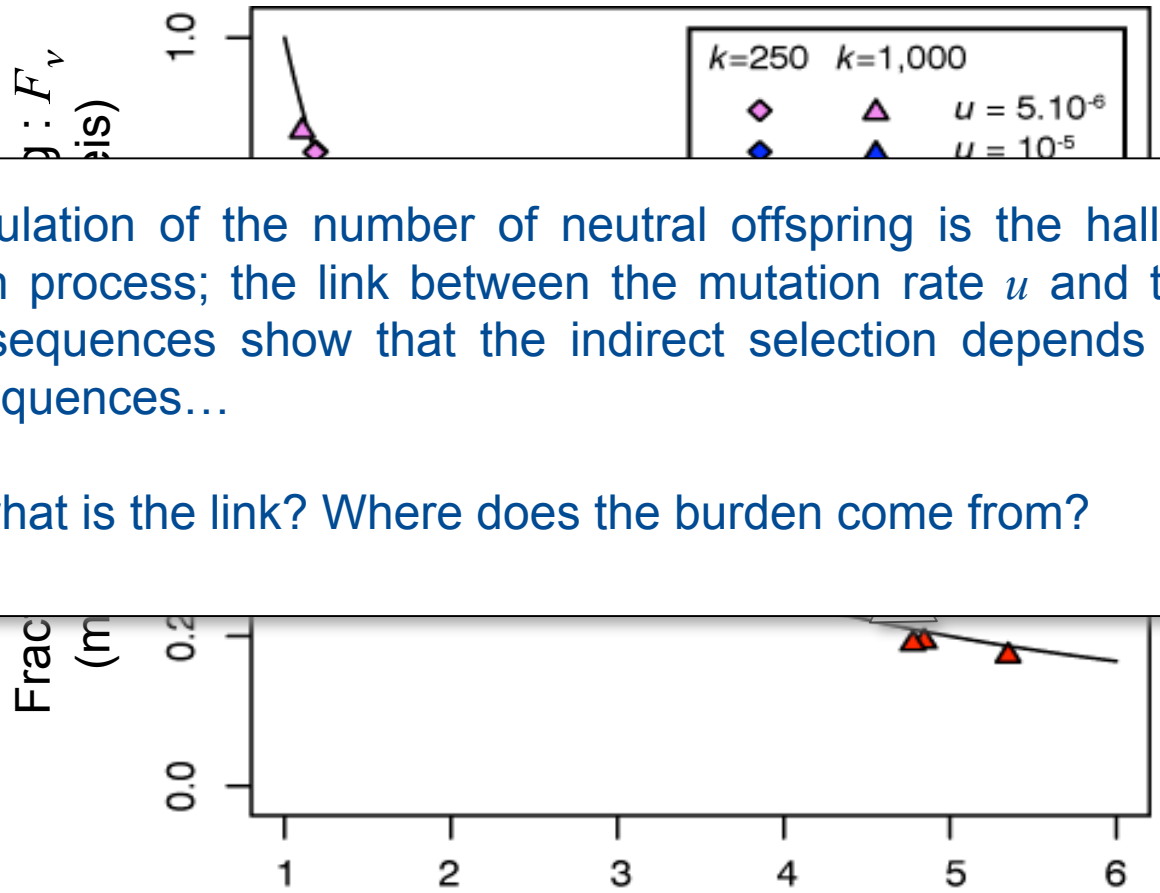
But "*Prédire n'est pas expliquer*" (R. Thom) …





[Parsons *et al.*, Alife, 2010]

Measure of best individual reproduction neutrality (at t = 20000)



The regulation of the number of neutral offspring is the hallmark of an indirect selection process; the link between the mutation rate u and the size of the non-coding sequences show that the indirect selection depends (at least partly) on these sequences…

… But what is the link? Where does the burden come from?

Number of reproductive trials : W
(depends on mean fitness)

Modeling the model …

- Mathematical model of reproduction
  - The math model is “true” for aevol *AND* for the “real world”…
- E : Probability of neutral reproduction as a function of genome size (L),

If: (i) genomes undergo large duplications and deletions, (ii) the number and the average size of these events increase with genome size, Then: the mutational variability of a lineage depends on the amount of non-coding DNA (it is mutagenic for the genes it surrounds).

Thus the indirect selection for an appropriate level of variability actually selects for a specific amount of non-coding DNA

[Knibbe et al., Mol. Biol. Evol., 2007]

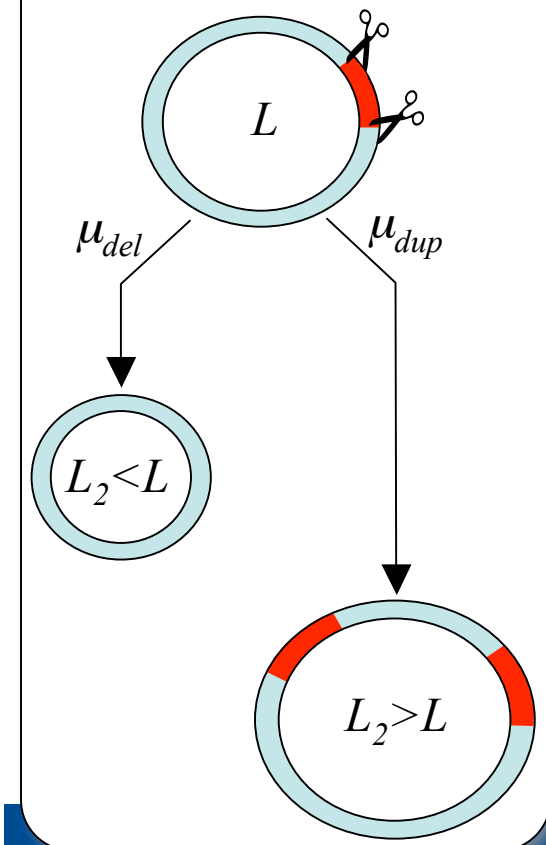
$$\begin{cases} \tilde{\nu}_{\text{ponct}} &= \tilde{\nu}_{\text{ins}} = \tilde{\nu}_{\text{del}} = 1 - \frac{l}{L} \\ \tilde{\nu}_{\text{inv}} &= \left(1 - \frac{l}{L}\right)^2 \\ \tilde{\nu}_{\text{transloc}} &= \left(1 - \frac{l}{L}\right)^3 \end{cases}$$

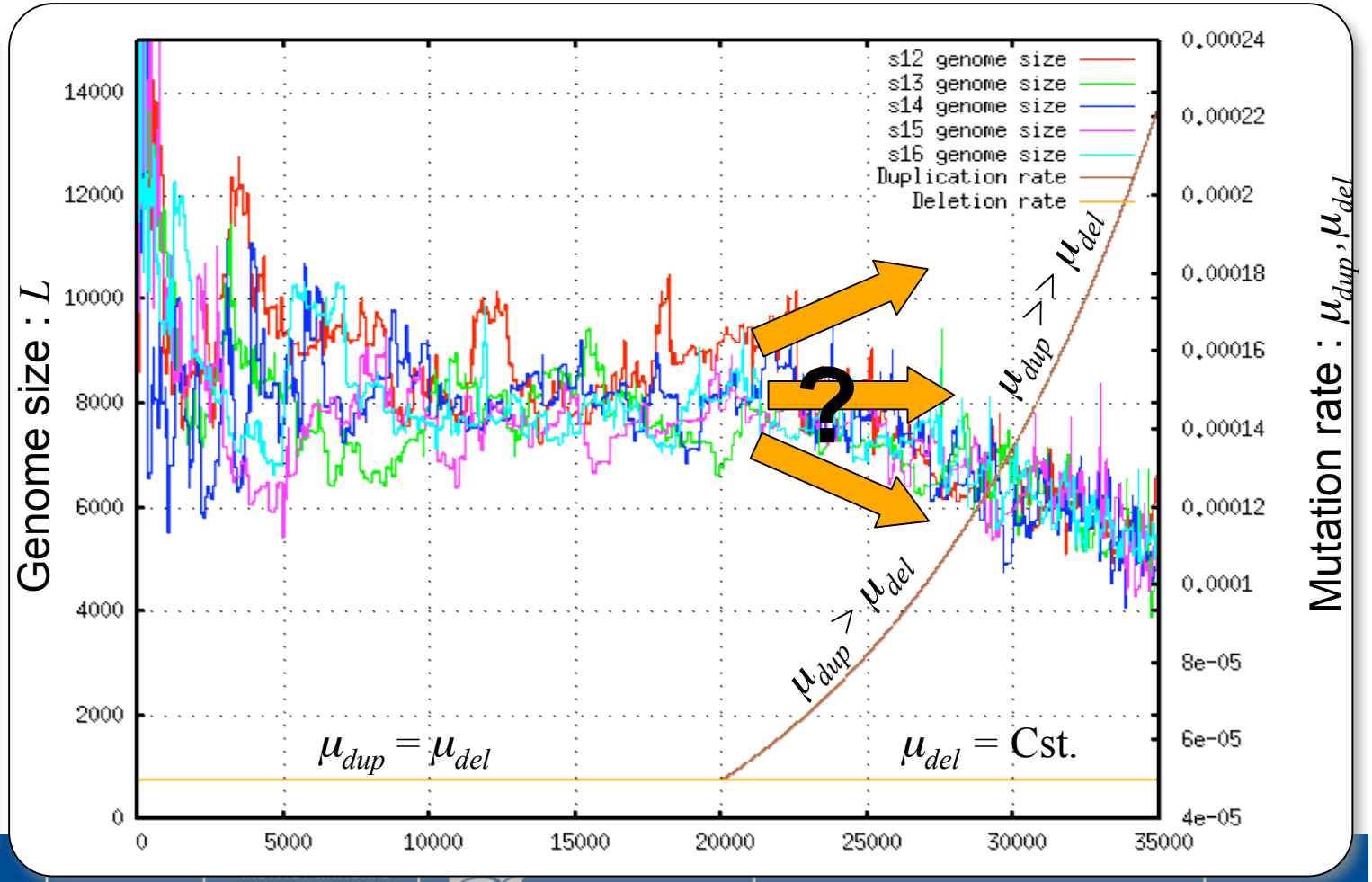
$$\begin{cases} \tilde{\nu}_{\text{gdel}} &= \frac{1}{2L^2} \sum_{j=1} \lambda_j \left(\lambda_j + 1\right) \\ \tilde{\nu}_{\text{dup}} &= \frac{1}{2L^2} \left(1 - \frac{l}{L}\right) \sum_{j=1}^{N_G} \lambda_j \left(\lambda_j + 1\right) \end{cases}$$

Surprising consequences

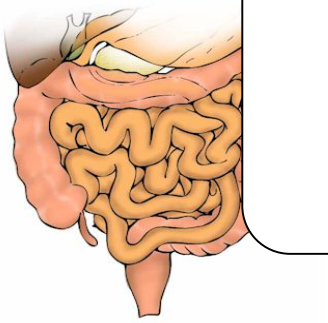
« It is simply a truism that the observed genome size is the result of a balance between the rate of DNA gain and loss » (Gregory, 2004)

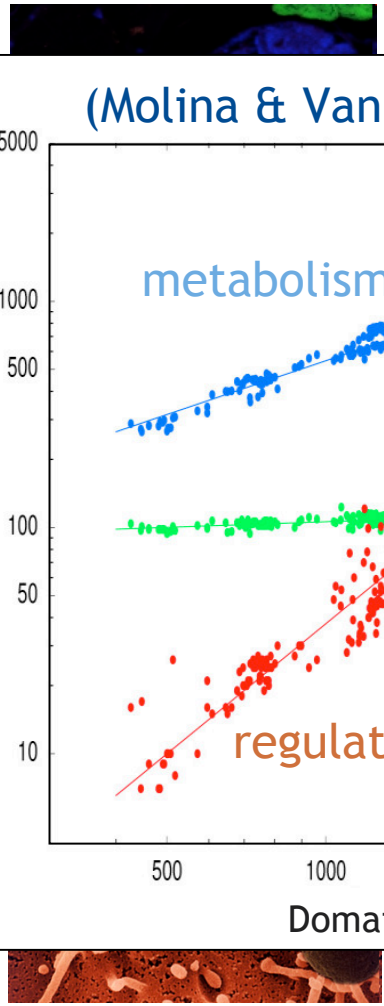
DNA gain: duplications
DNA loss: deletions

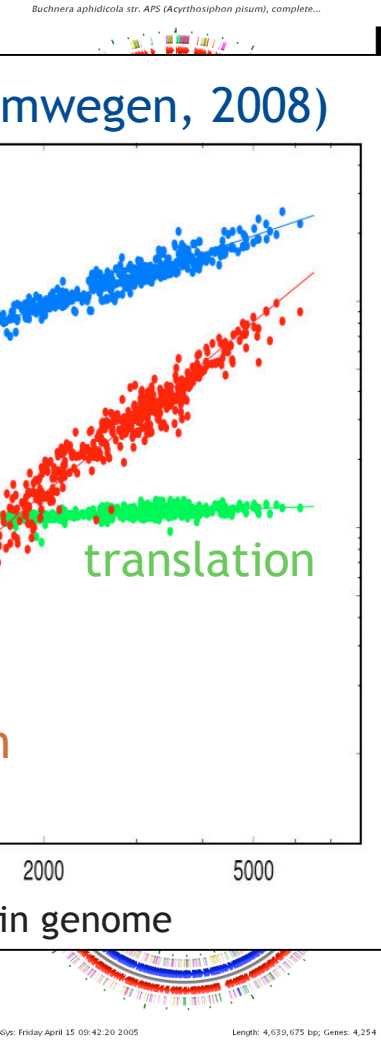




What about gene networks?







(Molina & Van Nimwegen, 2008)

Buchnera aphidicola

1 bp
nes
cription factors

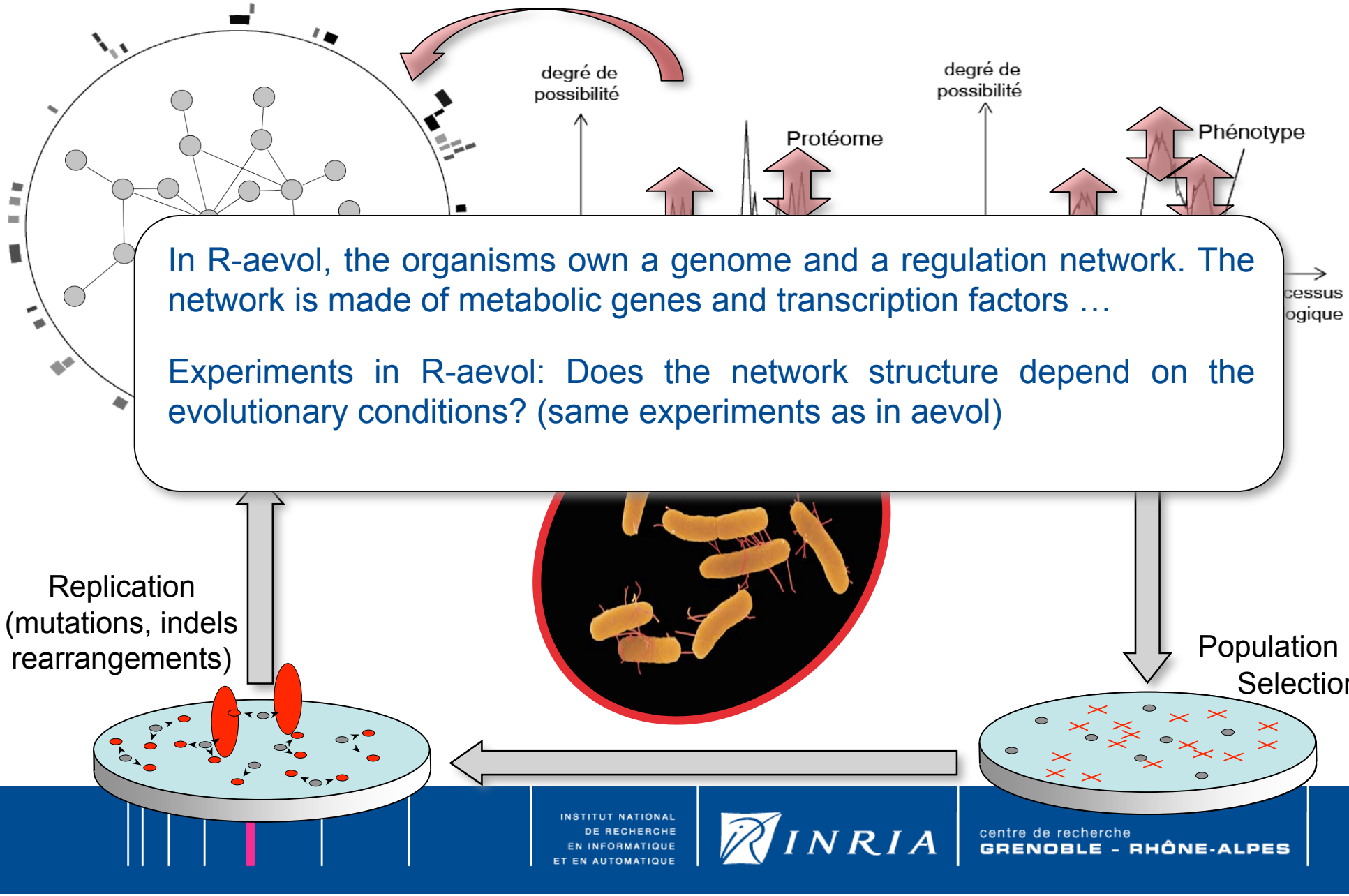
bium tepidum

946 bp
enes
scription factors

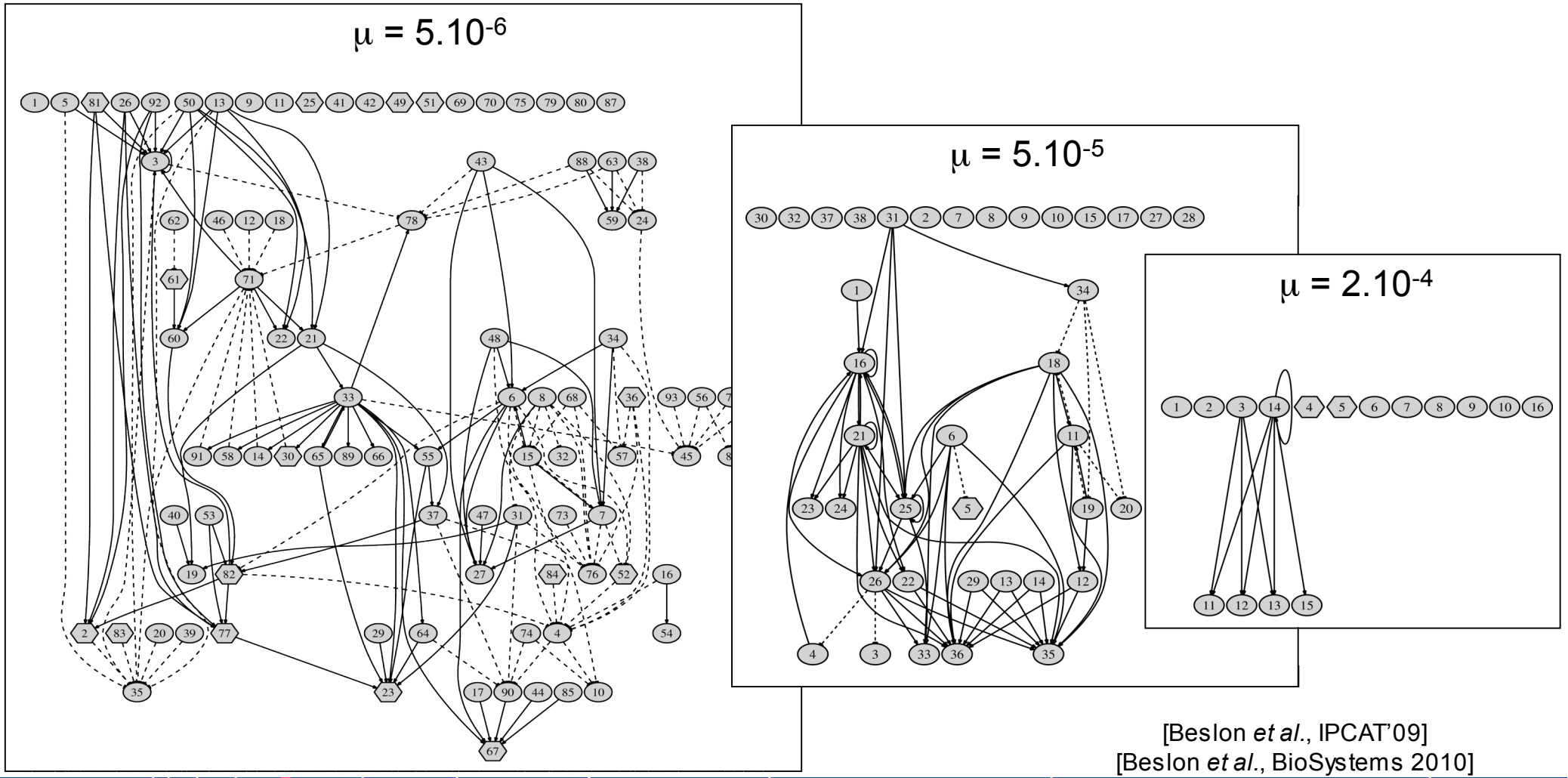
richia coli

675 bp
genes
275 transcription factors

R-aevol (Y. Sanchez-Dehesa, 2009)



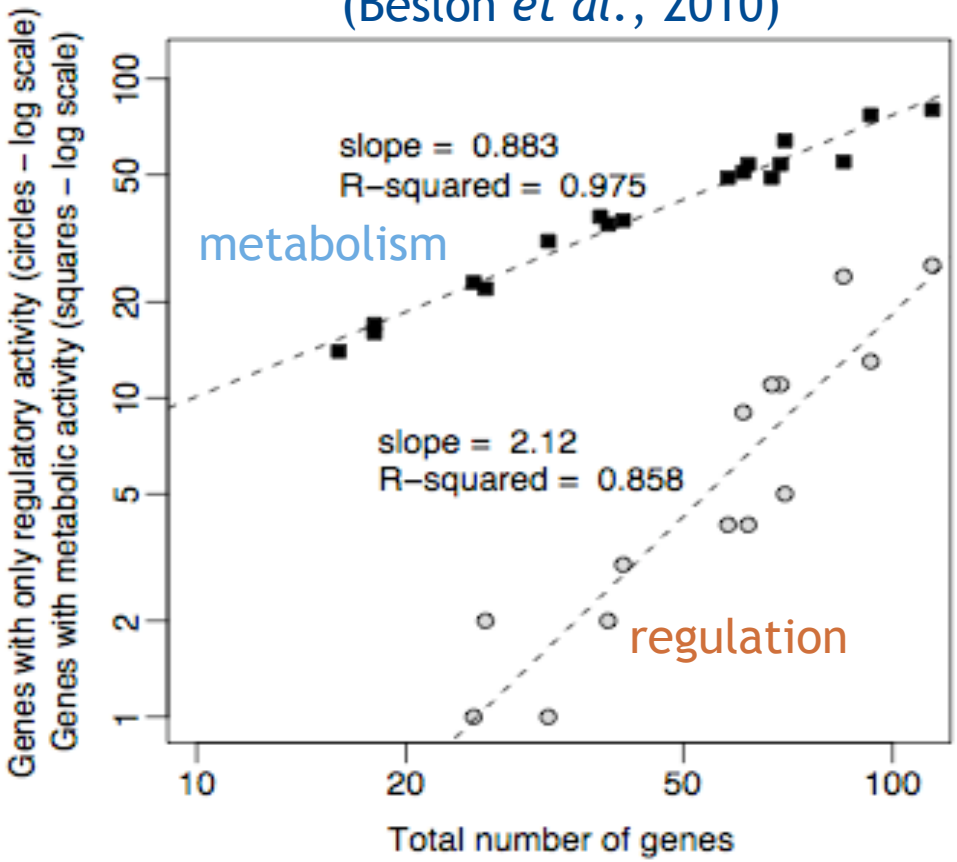
Impact of mutation rates on transcriptomic structures



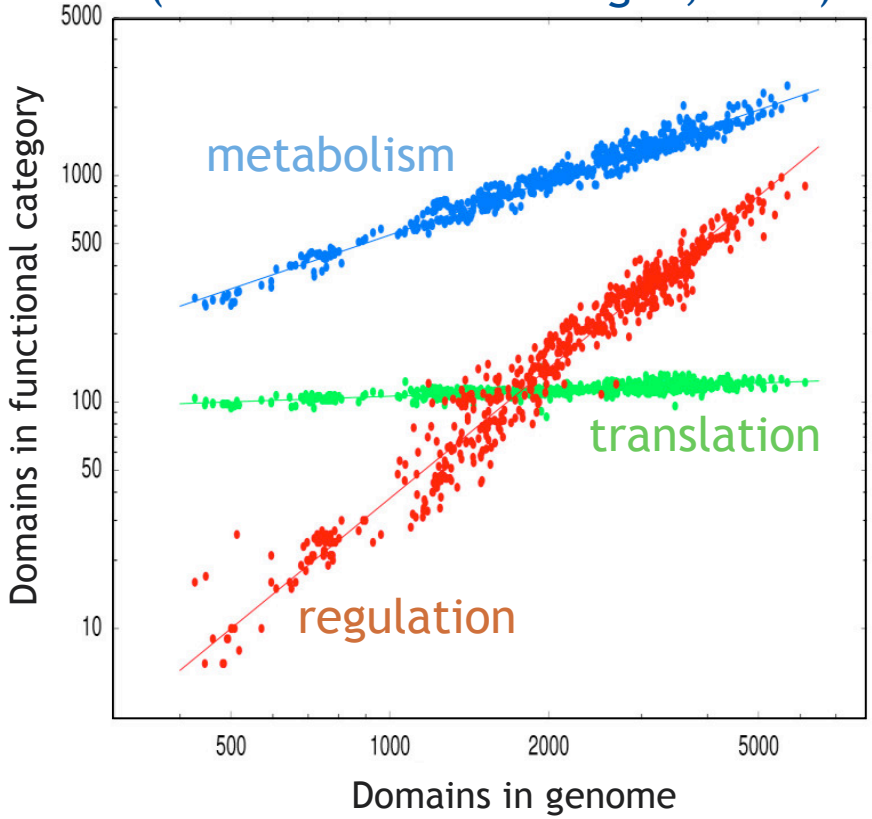
[Beslon *et al.*, IPCAT'09]
[Beslon *et al.*, BioSystems 2010]

R-aevol: emergence of scaling laws

R-aevol
(Beslon *et al.*, 2010)

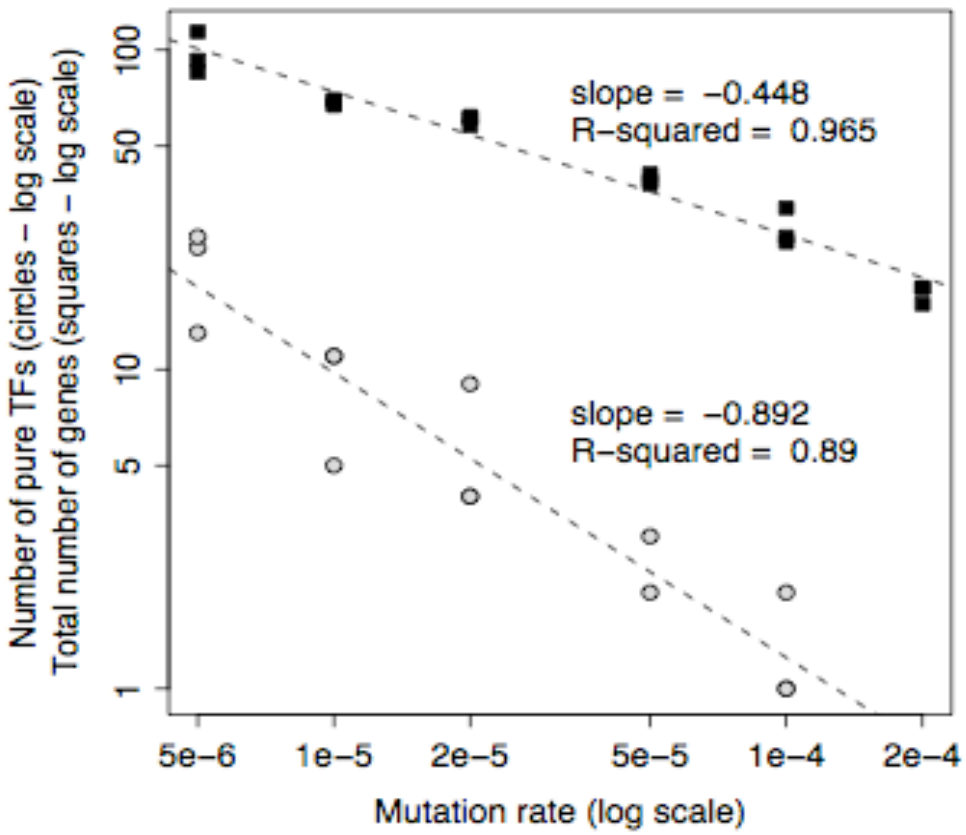


Biological data
(Molina & Van Nimwegen, 2008)

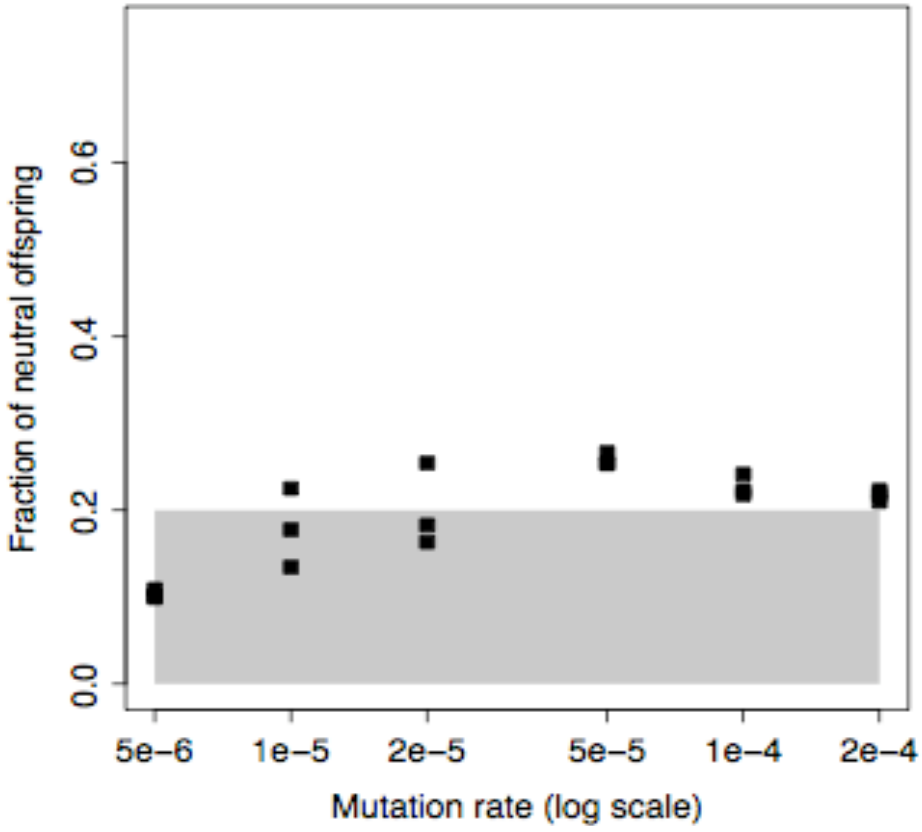


Side effect of the selection for robustness?

R-aevol



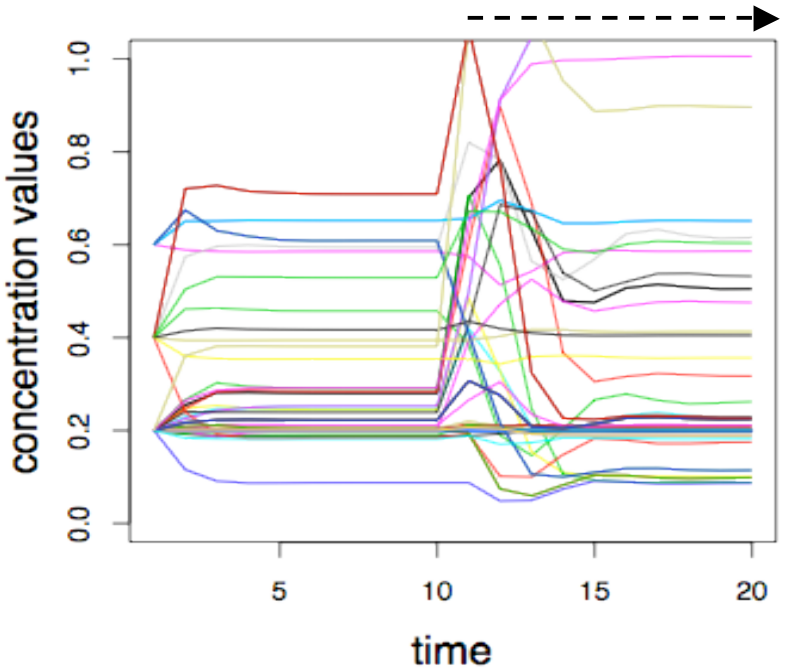
F_ν ~ constant



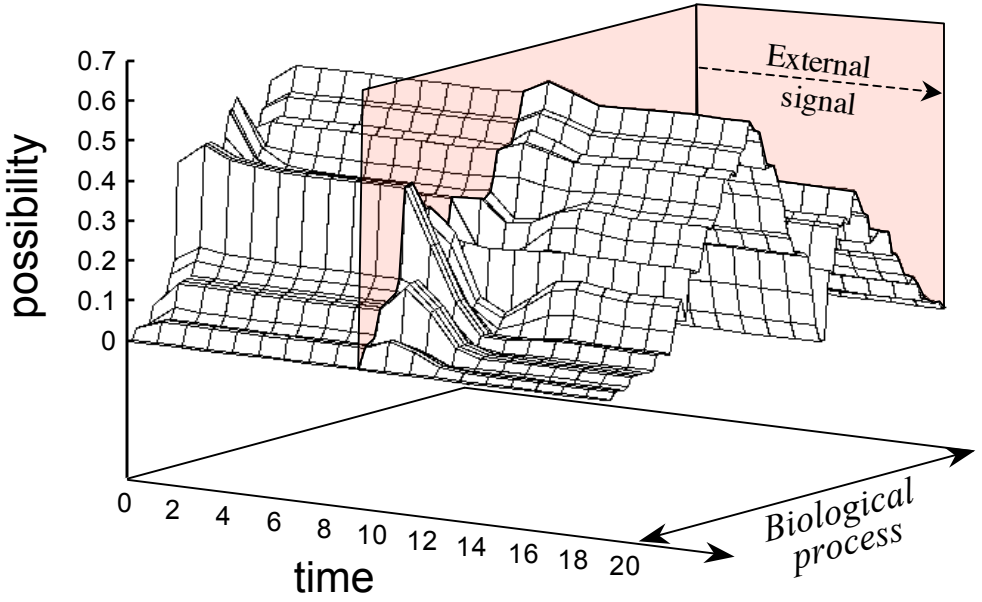
What if the organisms have “something to do”

- Organisms live for 20 time steps ; at t = 10 a signal is sent to the “cells” that must react by changing their phenotypes…

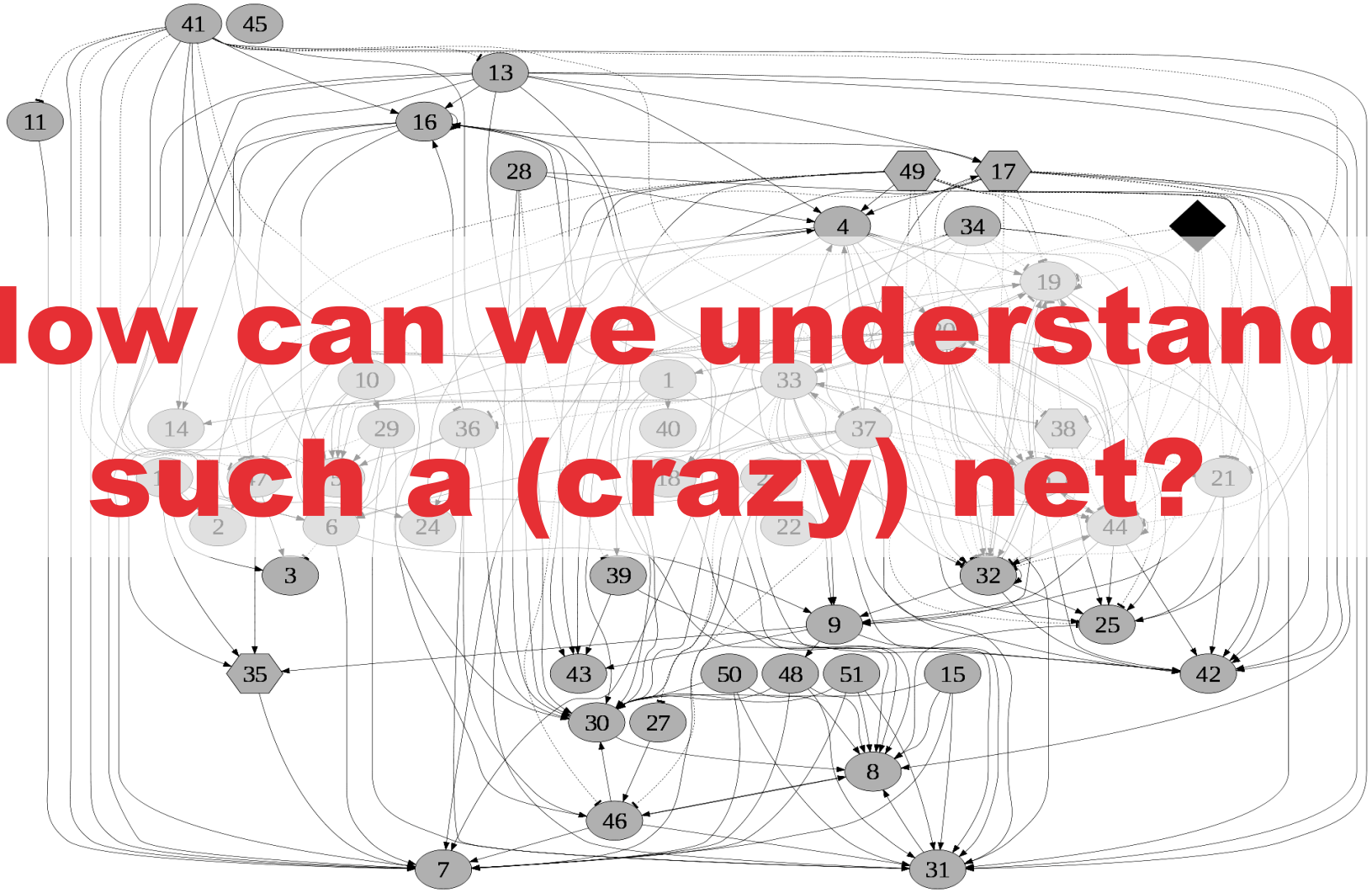
Protein concentration over time



Phenotype over time

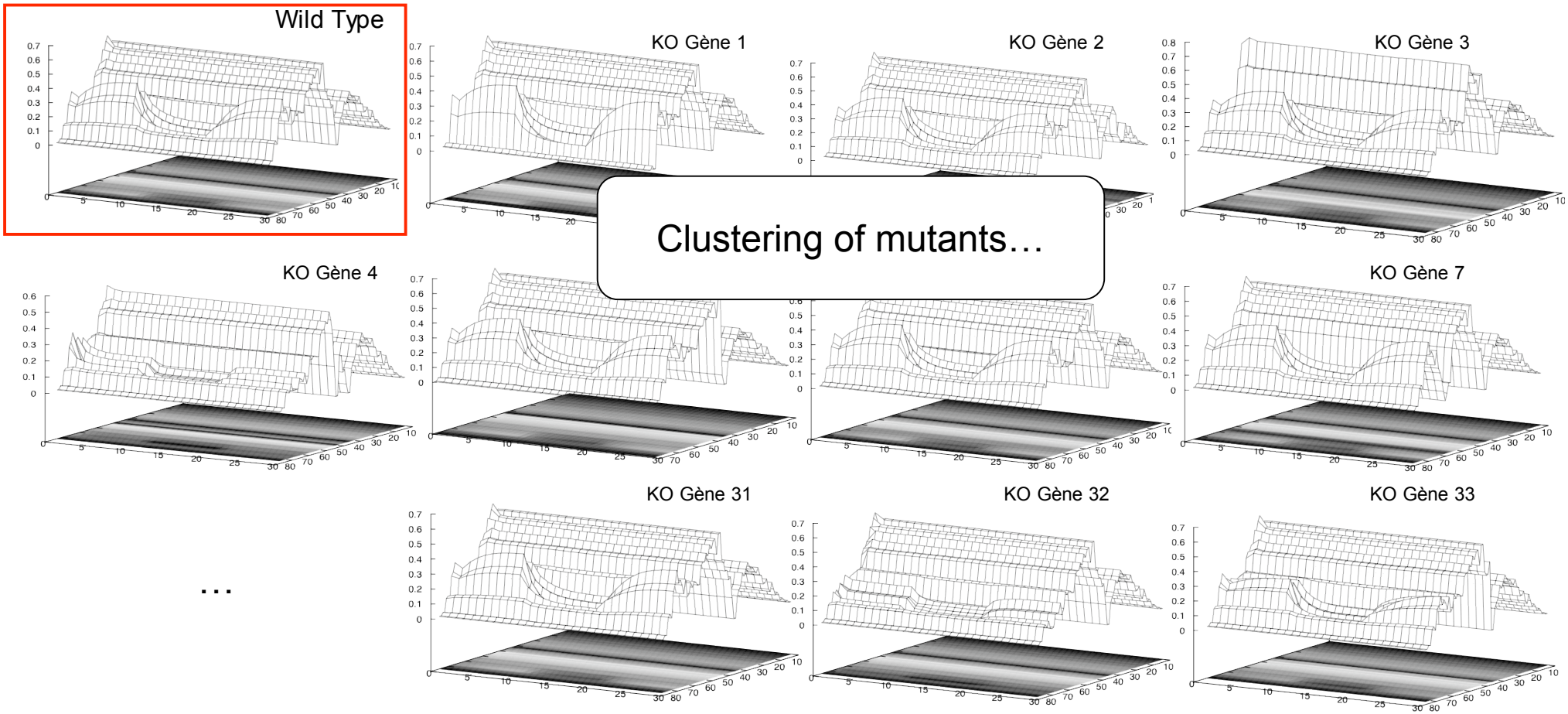


Evolved network after 15000 generations



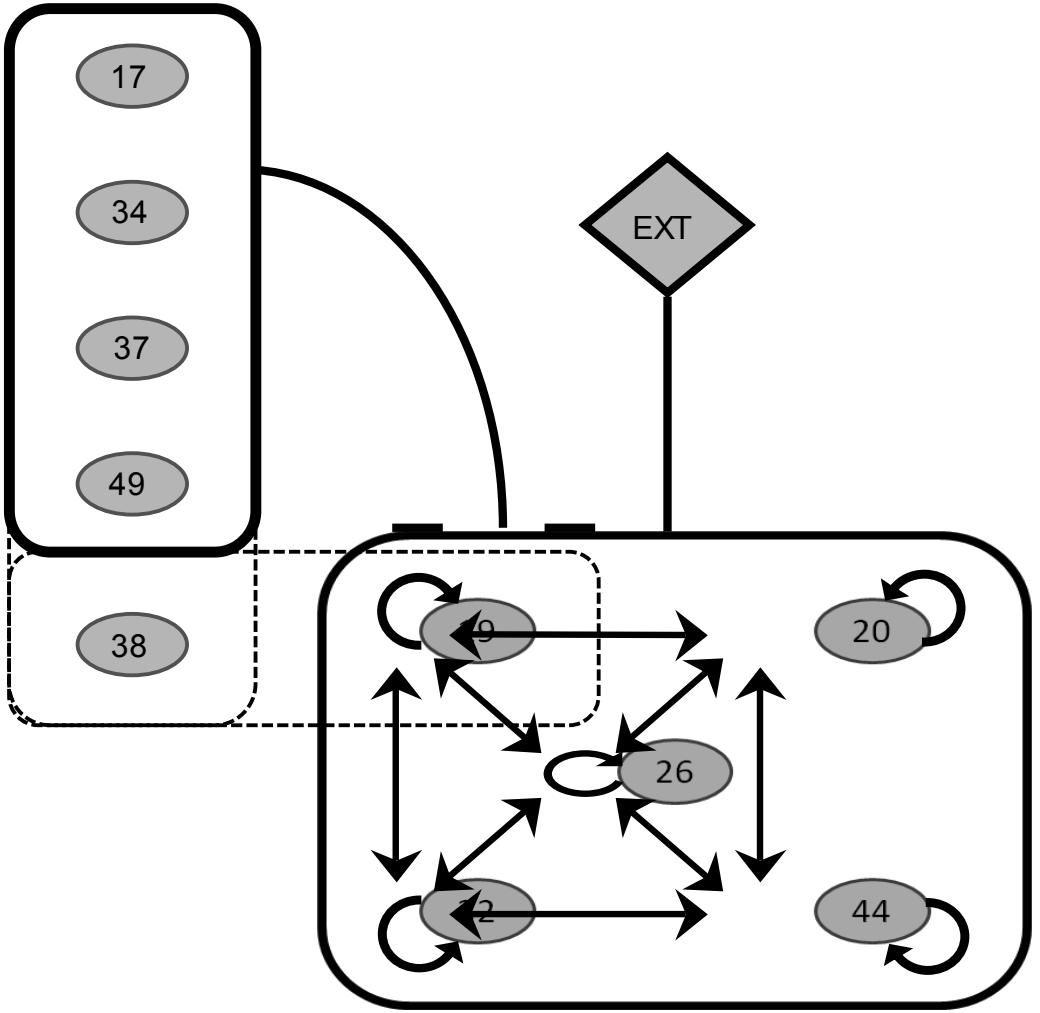
How can we understand such a (crazy) net?

Systematic Knock-Out experiments



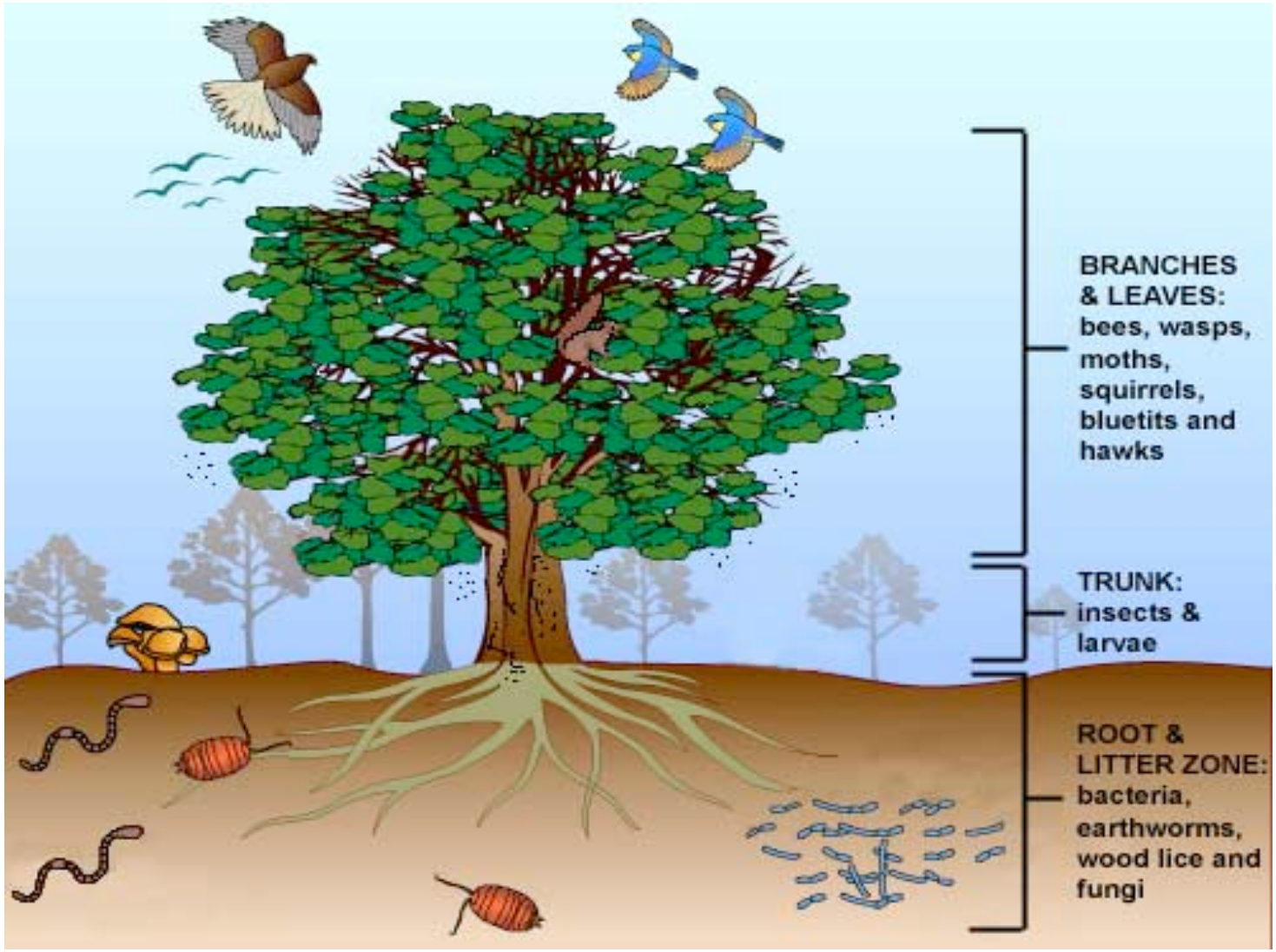
[Beslon et al., IDAj, 2010]

Network sketch with two modules



Origin of transcriptomics complexity?

- Where does the network complexity come from?
  - [In less stable, more changing environments, transcription factors are over-represented] … *This suggests that in ever-changing, highly competitive environments, there is a strong selective pressure towards regulated and coordinated gene expression, compared with very stable environments.* (Cases *et al.*, 2003)
- According to this view, the origin of (transcriptomic) complexity is another complexity (environmental)!
  - But in our experiments, the complex network emerges in a simple environment (one stable state) as well as in two-states environments
- Thus complexity emerges “for free” (at least in the model)
  - Environmental complexity is NOT a necessary condition
  - A new analysis paradigm for genetic networks understanding?


BRANCHES & LEAVES: bees, wasps, moths, squirrels, bluetits and hawks
TRUNK: insects & larvae
ROOT & LITTER ZONE: bacteria, earthworms, wood lice and fungi

References

G. Beslon, D. P. Parsons, Y. Sanchez-Dehesa, J.-M. Peña, C. Knibbe (2010) Scaling Laws in Bacterial Genomes: A Side-Effect of Selection of Mutational Robustness? *BioSystems,* 102(1):32-40.

G. Beslon, D. P. Parsons, J.-M. Peña, C. Rigotti, Y. Sanchez-Dehesa (2010) From Digital Genetics to Knowledge Discovery: Perspectives in Genetic Network Understanding. *Intelligent Data Analysis journal (IDAj),* 14(2):173-191.

D. P. Parsons, C. Knibbe, G. Beslon, Importance of the Rearrangement Rates on the Organization of Genome Transcription. In: Proc. Artificial Life 12, Odense, Denmark. pp. 479-486.

G. Beslon, Y. Sanchez-Dehesa, D.P. Parsons, J.M. Peña, C. Knibbe (2009) Scaling Laws in Digital Organisms In: Proc. Information Processing in Cells and Tissues IPCAT'09, Ascona, Switzerland. pp. 111-114.

C. Knibbe, J.M. Fayard, G. Beslon (2008) The Topology of the Protein Network Influences the Dynamics of Gene Order : From Systems Biology to a Systemic Understanding of Evolution. *Artificial Life* 14(1):149-156.

C. Knibbe, A Coulon, O. Mazet, J.M. Fayard, G. Beslon (2007) A Long-Term Evolutionary Pressure on the Amount of Noncoding DNA. *Molecular Biology and Evolution* 24(10):2344-2353.

C. Knibbe, O. Mazet, F. Chaudier, J.M. Fayard, G. Beslon (2007) Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences. *Journal of Theoretical Biology* 244(4):621-630.

W. Banzhaf, G. Beslon, S. Christensen, J. Foster, F. Képès, V. Lefort, J.F. Miller, M. Radman, J. Ramsden (2006) From artificial evolution to computational evolution: a research agenda. *Nature Review*

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
GRENOBLE - RHÔNE-ALPES