Agent-Based Modeling of Evolution

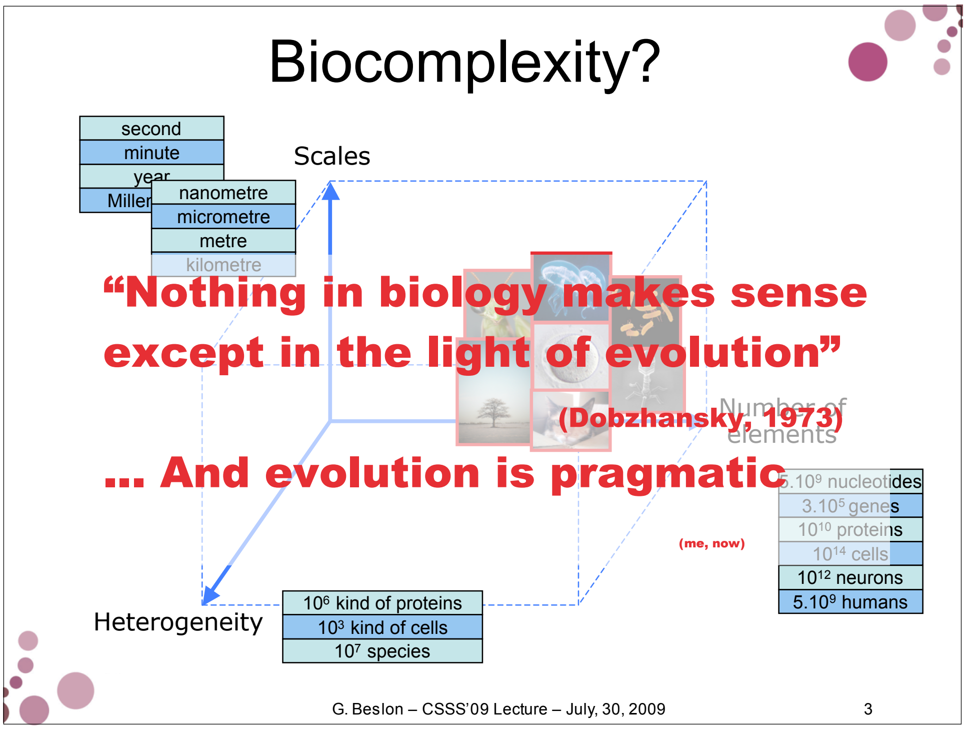
- introduction to digital genetics -

Guillaume Beslon
INSA – LIRIS – IXXI

Introduction

- Aim of the course:
  - Introduce main concepts of evolutionary biology
  - Introduce digital genetic modeling
  - Present may own work in digital genetic
- Who am I?
  - Guillaume BESLON (guillaume.beslon@liris.cnrs.fr)
  - Professor at the INSA-Lyon, LIRIS Lab. (Laboratoire d'Informatique en Image et Systèmes d'Information)
  - Assistant-director of IXXI (Rhône-Alpes Complex Systems Institute)
  - Research topics: Individual-based modeling of complex biological systems (mainly evolution)

Biocomplexity?



second
minute
year
Miller

nanometre
micrometre
metre
kilometre

Scales

“Nothing in biology makes sense except in the light of evolution”

(Dobzhansky, 1973)

Number of elements

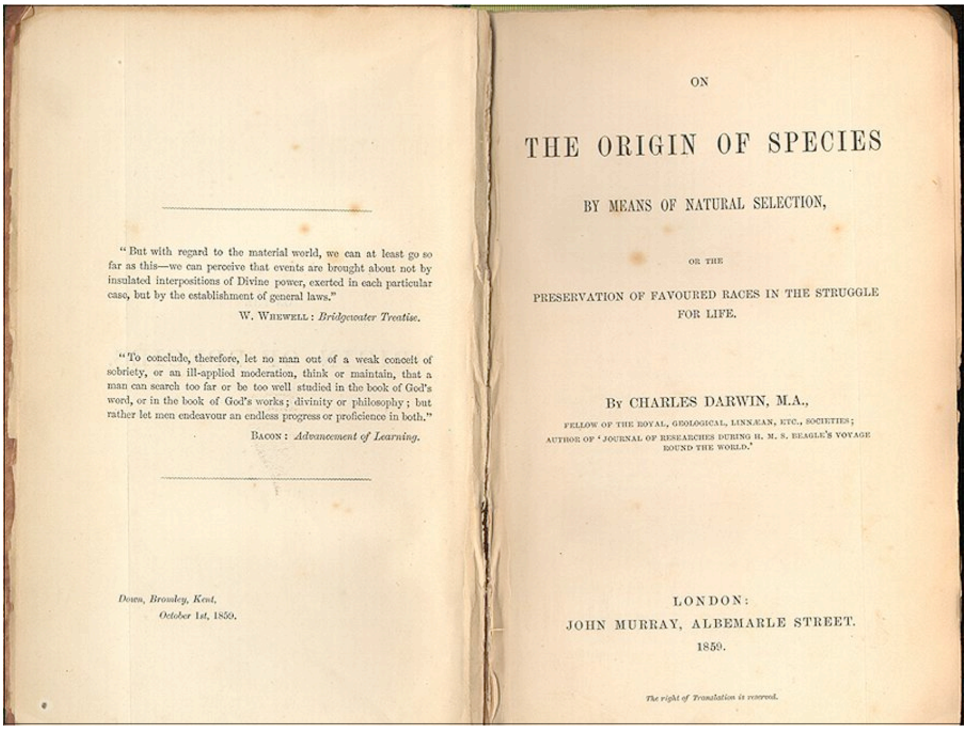
… And evolution is pragmatic

(me, now)

5.10^9 nucleotides
3.10^5 genes
10^{10} proteins
10^{14} cells
10^{12} neurons
5.10^9 humans

Heterogeneity

10^6 kind of proteins
10^3 kind of cells
10^7 species



“But with regard to the material world, we can at least go so far as this—we can perceive that events are brought about not by insulated interpositions of Divine power, exerted in each particular case, but by the establishment of general laws.”

W. Whewell: *Bridgewater Treatise.*

“To conclude, therefore, let no man out of a weak conceit of sobriety, or an ill-applied moderation, think or maintain, that a man can search too far or be too well studied in the book of God’s word, or in the book of God’s works; divinity or philosophy; but rather let men endeavour an endless progress or proficience in both.”

Bacon: *Advancement of Learning.*

Down, Bromley, Kent,
October 1st, 1859.

ON
THE ORIGIN OF SPECIES
BY MEANS OF NATURAL SELECTION,
OR THE
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

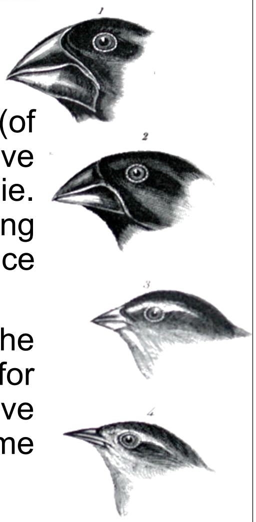
By CHARLES DARWIN, M.A.,
FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF ‘JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE’S VOYAGE
ROUND THE WORLD.’

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.

Principe général de l ’évolution darwinienne

- The context of evolution is a **population** (of organisms, objects, agents ...) that survive for a limited time (usually) and then die. Some produce **offspring** for succeeding generations, the **'fitter'** ones tend to produce more.
- Over many generations, the make-up of the population changes. Without the need for any individual to change, successive generations, the 'species' changes, in some sense (usually) adapts to the conditions.

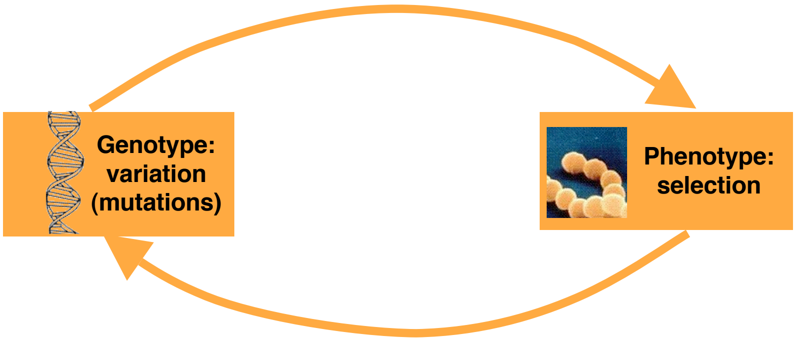


Galapagos Finches

Principe général de l ’évolution darwinienne

“Evolution will occur whenever and wherever three conditions are met: replication, variation (mutation), and differential fitness (competition).”

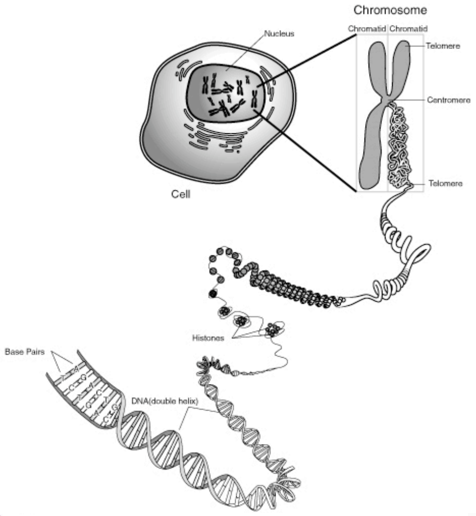
[Daniel Dennett]

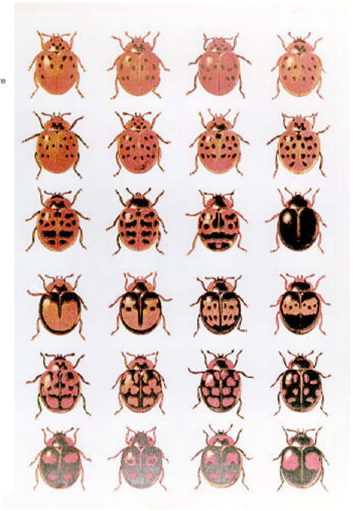


Genotype: variation (mutations)

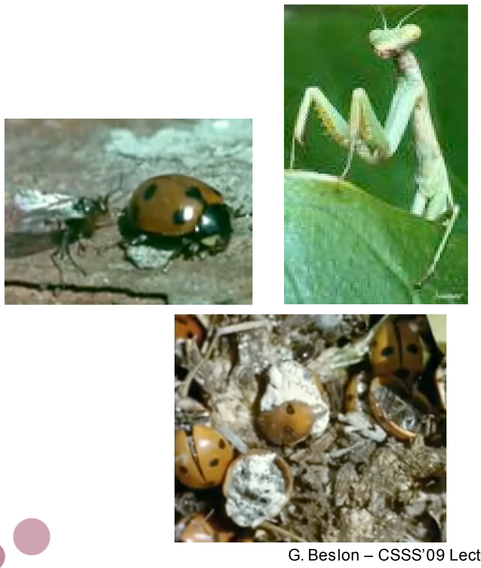
Phenotype: selection

Genetic variability

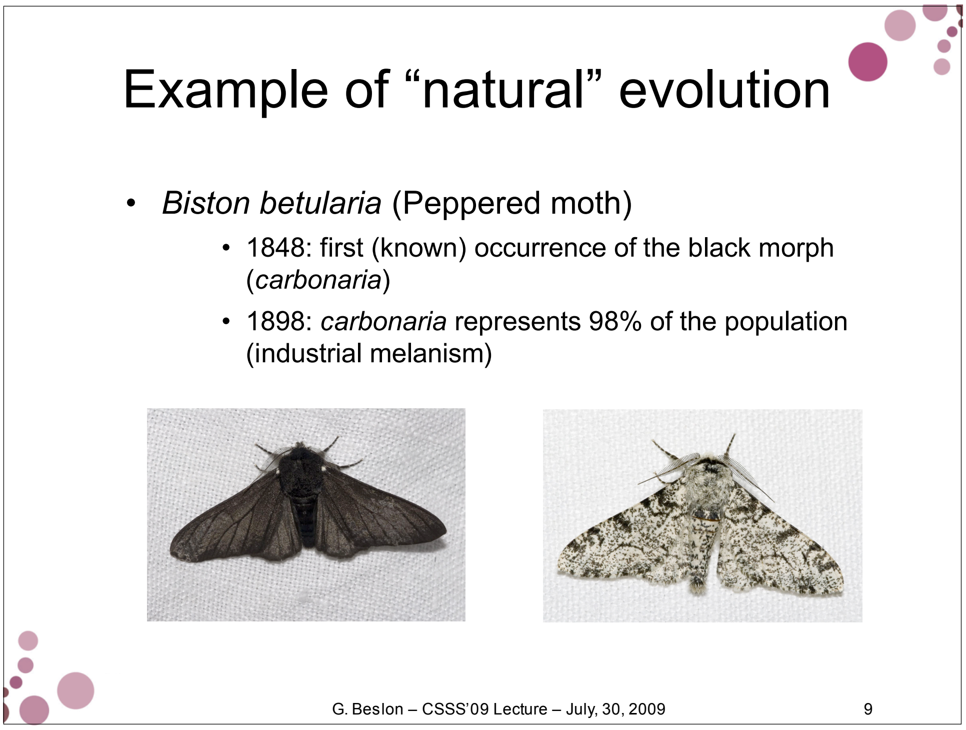




Natural selection

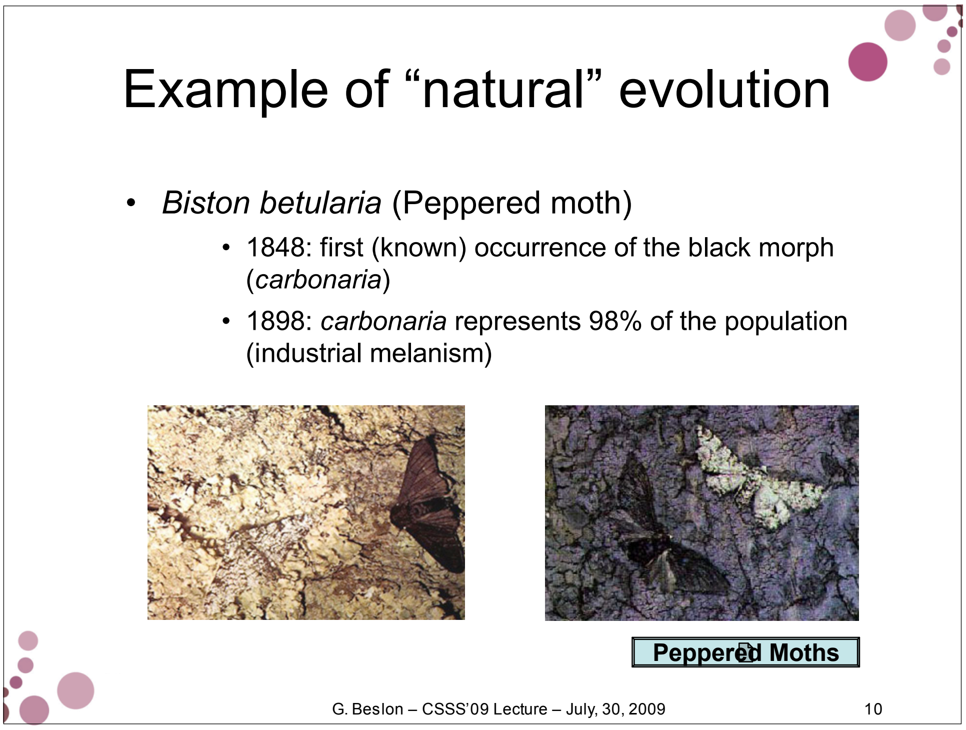


The fitness measures the probability of survival and reproduction



Example of “natural” evolution

- *Biston betularia* (Peppered moth)
  - 1848: first (known) occurrence of the black morph (*carbonaria*)
  - 1898: *carbonaria* represents 98% of the population (industrial melanism)



Example of “natural” evolution

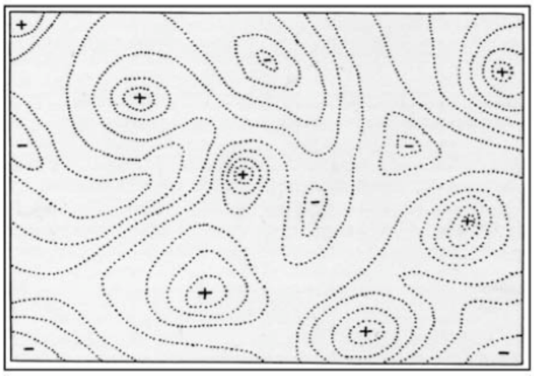
- *Biston betularia* (Peppered moth)
  - 1848: first (known) occurrence of the black morph (*carbonaria*)
  - 1898: *carbonaria* represents 98% of the population (industrial melanism)

Introduction

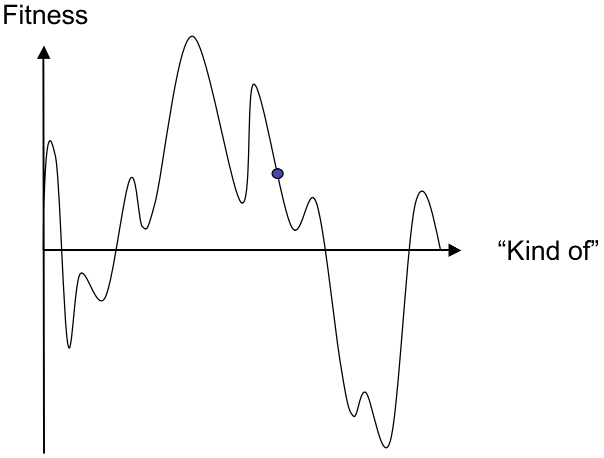
- Although it can be described in a few words, evolution give rise to many complex phenomenon that can be very difficult to understand
  - Evolution of cooperation, evolution of sex…
- Evolution is difficult to study
  - Well known snapshot (today)
  - Few fossil records
  - Difficult experiments
- Some evolutionary pressures are well-known but their relative contribution is almost impossible to assess
  - Modeling needed! **Bug-Hunt Speed**

The fitness landscape metaphore

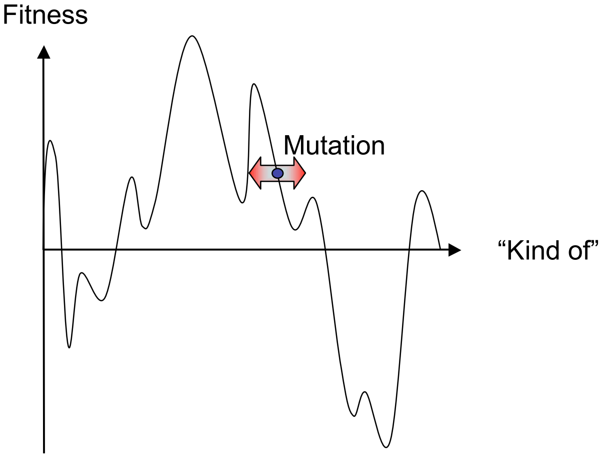
(Sewall Wright, 1932 - last publication: 1988!)

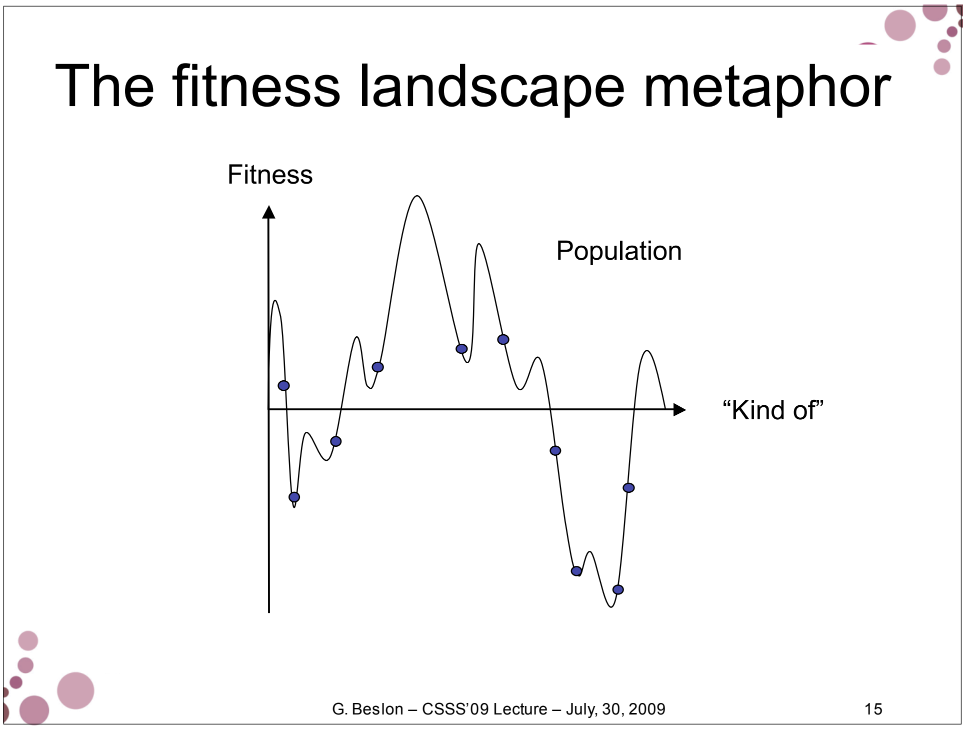


The fitness landscape metaphor

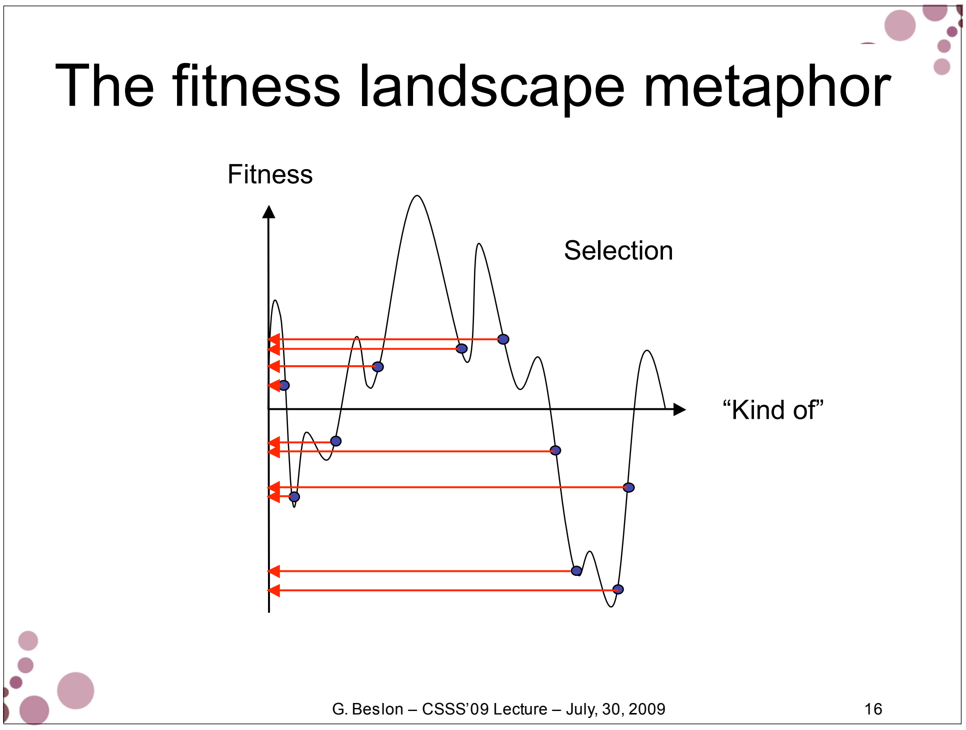


The fitness landscape metaphor

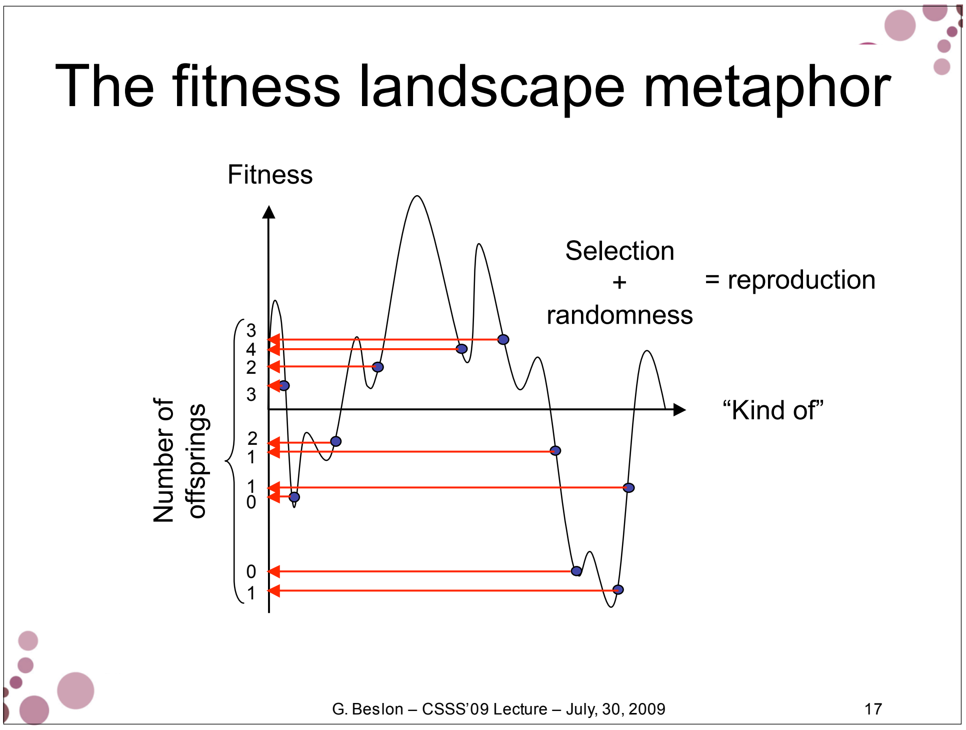




The fitness landscape metaphor

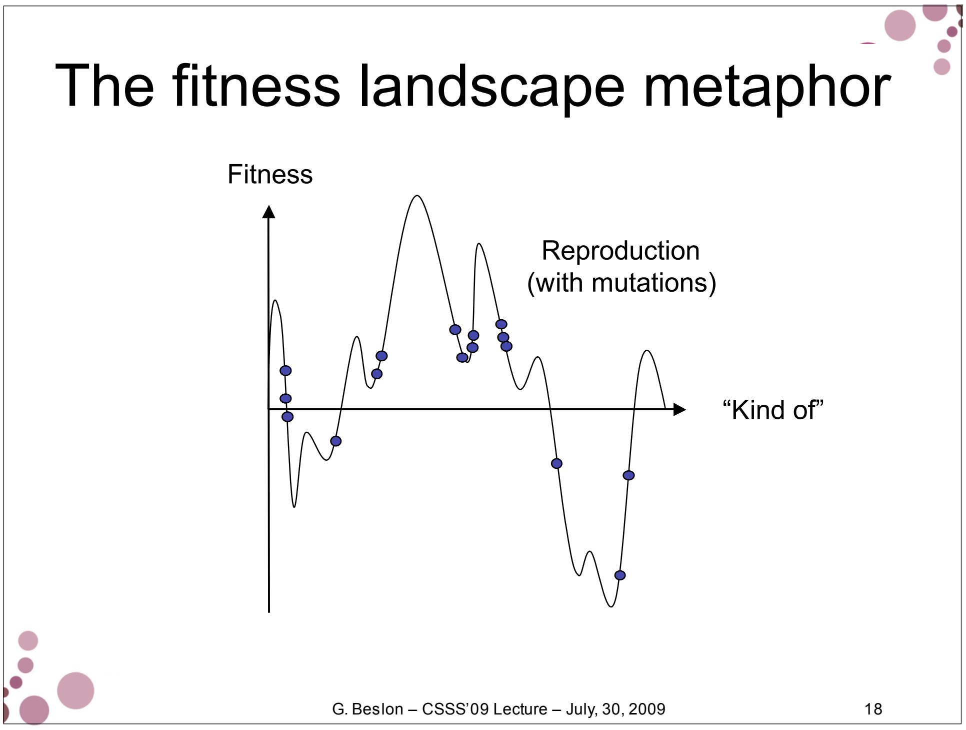


The fitness landscape metaphor



The fitness landscape metaphor

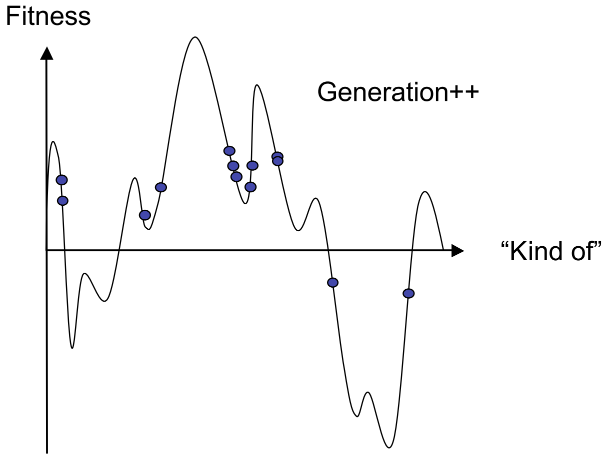
Selection
+
randomness
= reproduction



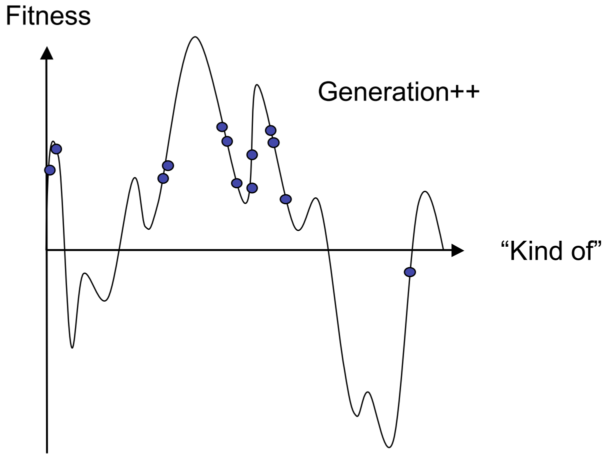
The fitness landscape metaphor

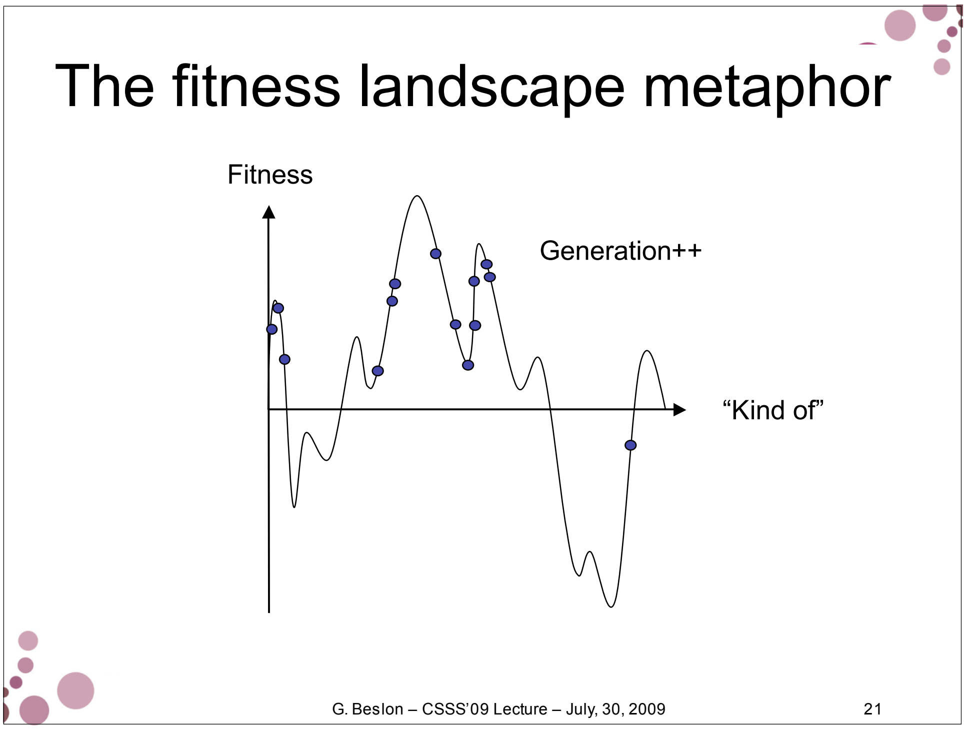
Reproduction
(with mutations)

The fitness landscape metaphor

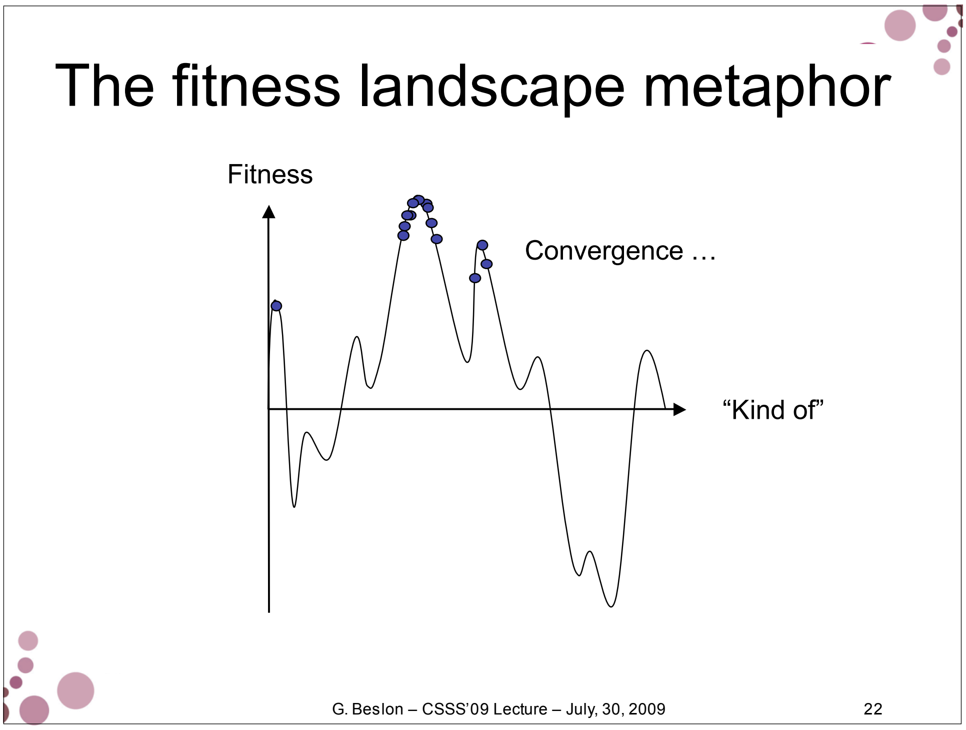


The fitness landscape metaphor



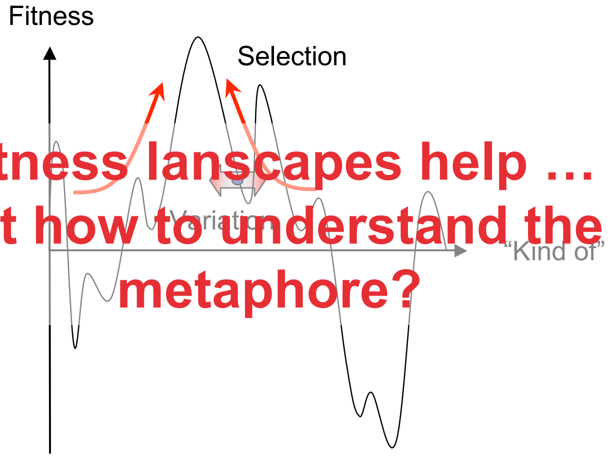


The fitness landscape metaphor



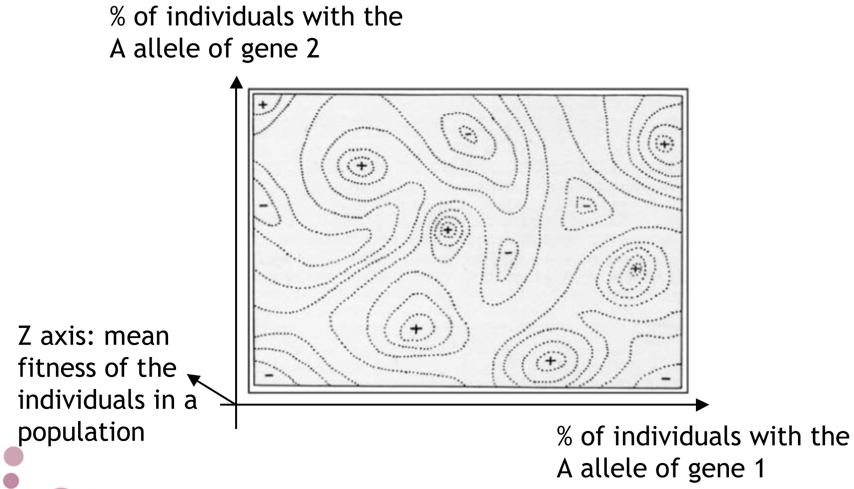
The fitness landscape metaphor

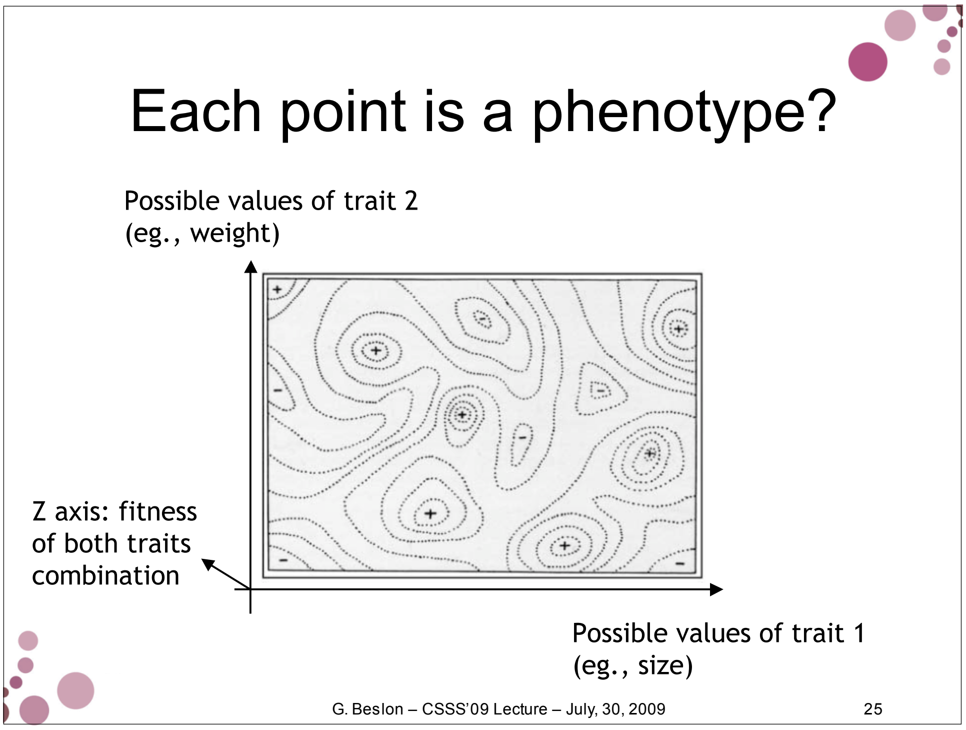
Two antagonist forces



**Fitness lanscapes help …
but how to understand the metaphore?**

Each point is a population?



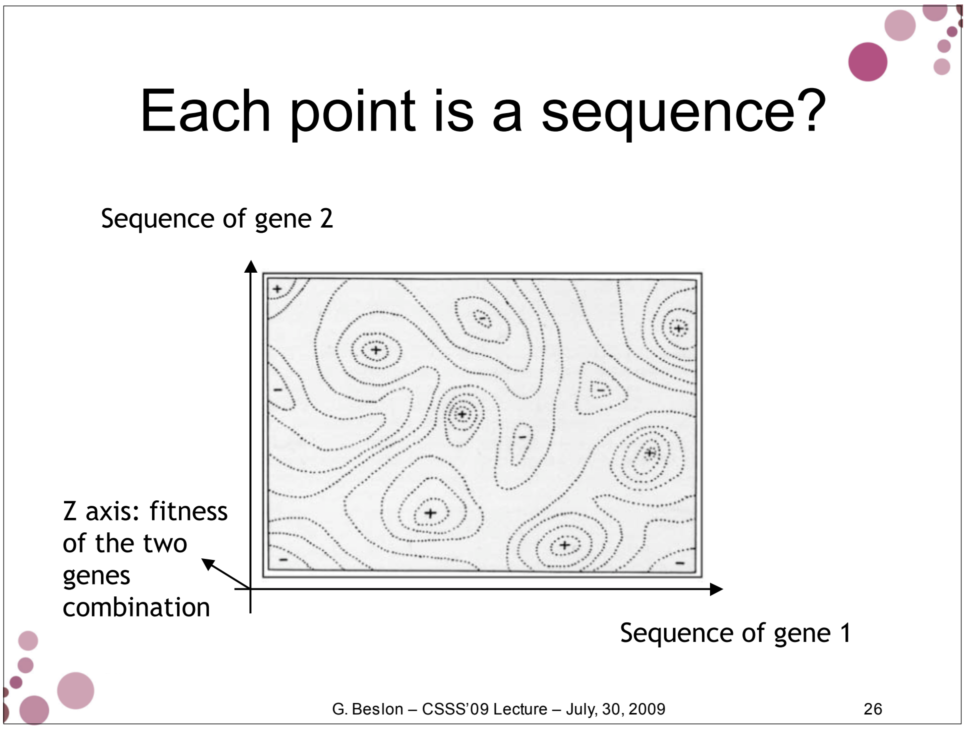


Each point is a phenotype?

Possible values of trait 2
(eg., weight)

Z axis: fitness
of both traits
combination

Possible values of trait 1
(eg., size)



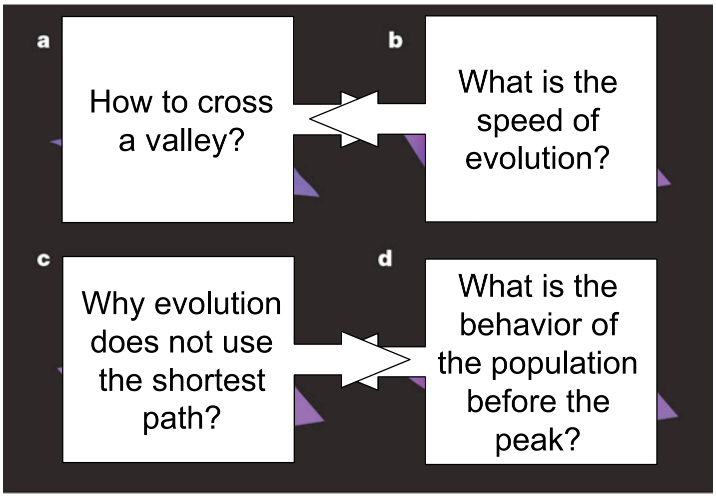
Each point is a sequence?

Sequence of gene 2

Z axis: fitness
of the two
genes
combination

Sequence of gene 1

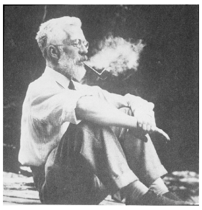
Fitness landscapes help thinking

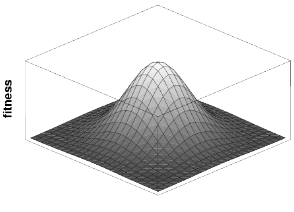


[Poelwijk et al. 2007]

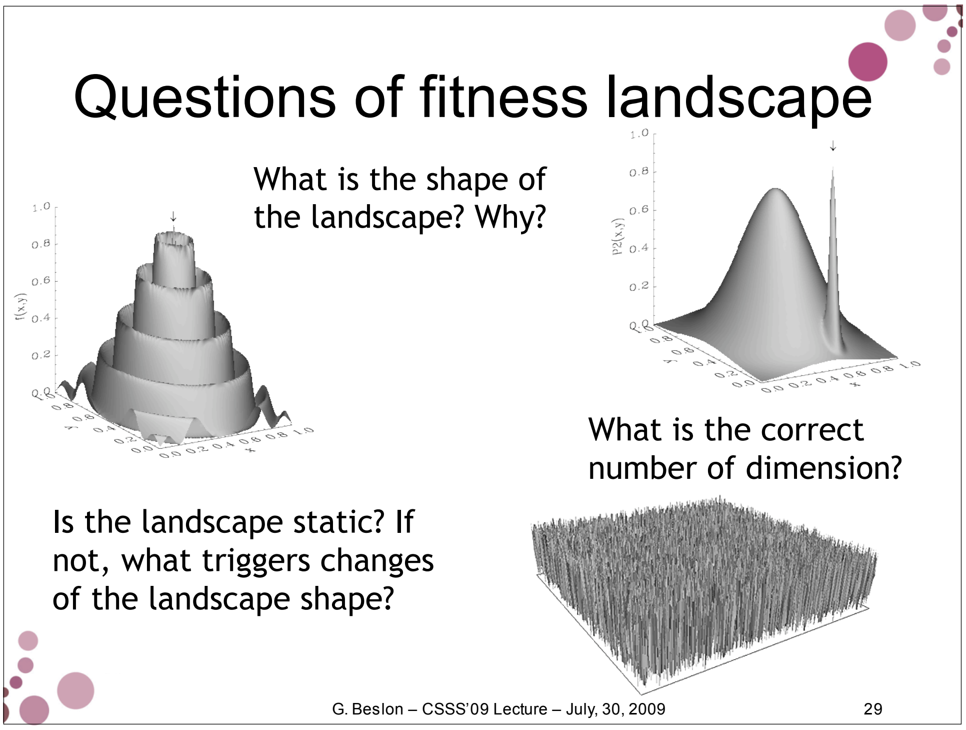
Quantitative interpretation

- Quantitative interpretation of fitness landscapes enables to built models in population genetics
  - Fisher model
  - Slow-down of evolution…





  - NK-Fitness landscapes (Stuart Kauffman)
    - NK fitness landscape enable to built landscapes with a controlled level of epistasis (links between genes' effects)
    - N is the size of the landscape, K is the level of epistatis; as K increases the ruggedness of the landscape increases



Questions of fitness landscape

What is the shape of the landscape? Why?

What is the correct number of dimension?

Is the landscape static? If not, what triggers changes of the landscape shape?

Need for experimental evolutionary studies

- Evolution if a general mechanism that relies on many random events
  - How can we distinguish between the effect of the mechanism and the random effects?
  - We only have a single "experiment" at our disposal!
- Many questions cannot be addressed without experiments (or only hardly addressed!)
  - Is there a trend in the evolution of biological complexity?
  - What if we start again?
  - Is evolution predictable?
  - Is evolution really universal? (Cf. Dennett)
  - What is true for E. coli is true for the elephant…

Experimental evolution

- Controlled experiments ARE possible for organisms which are
  - Cheep, small, abundant, controllable (organism and environment), fast (short generational time), measurable (sequence, fitness, …), freezable …
  - E.g., bacteria (*E. coli*, *salmonella*, …), viruses and phages, yeast, *C. elegans*, Drosophila, …
- Longest experiment in evolution
  - 12 strains of E. coli evolved during 40.000 generations in R. Lenski lab. at Michigan State University

http://myxo.css.msu.edu/index.html

Experimental evolution is not enough

- All known organisms share parts of their evolutionary history
  - We all come from LUCA (~3.5 billion years ago)
- Conditions are always changed by the experimental setup
  - What are the consequences on the evolutionary process?
- How can we analyze the results?
  - Real organisms are too complex for us!

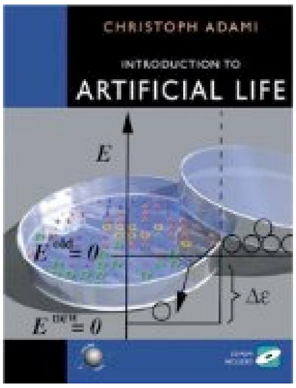
"So far, we have been able to study only one evolving system and we cannot wait for interstellar flight to provide us with a second. If we want to discover generalizations about evolving systems, we have to look at artificial ones."

[John Maynard Smith, 1992]

Artificial life

- Life inside a computer?
  - Free forms …

- Digital experiment on controlled organisms (artificial life)

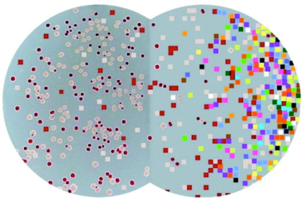


CHRISTOPH ADAMI

INTRODUCTION TO

ARTIFICIAL LIFE

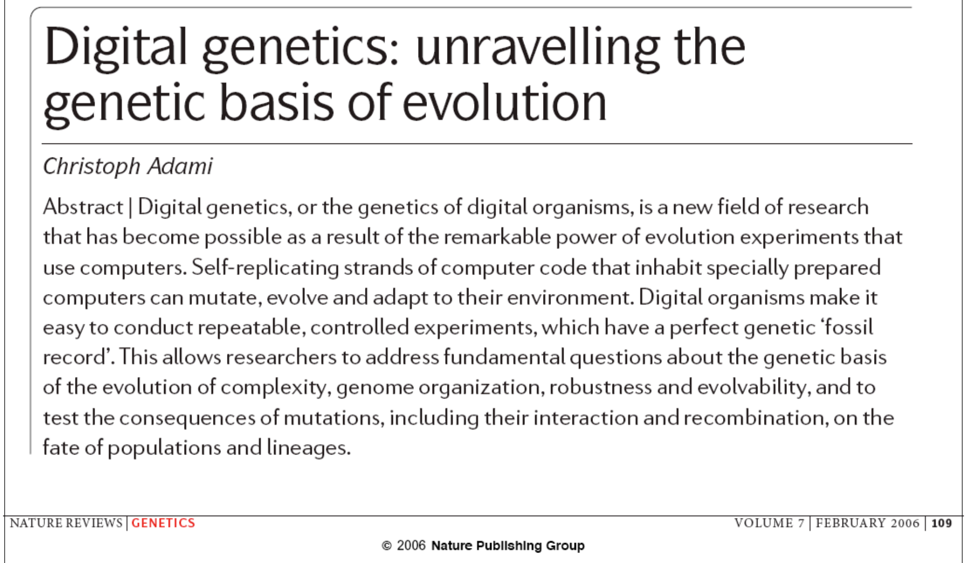
Artificial Life in few steps



- 1978 First attempts (C. Langton, LANL)
  - *"Life as it could be"*
- 1990 *Venus* simulator (S. Rasmussen, LANL)
- 1991 *Tierra* (T. Ray, U. of Delaware)
- 1992 *Creatures* (K. Sims, digital corp.)
- 1993 *Avida* (C. Adami., C.T. Brown, C. Ofria, Caltech)
  - Probably the most classical digital genetic software today
- 1996 *Amoeba* (A. Pargellis, Lucent)
- 2000 *Golem project* (H. Lipson, J. B. Pollack, Brandeis Univ.)
- 2005 *Aevol* (G. Beslon, C. Knibbe, INSA-Lyon)
- 2006 *Evolving robots* (D. Floreano, L. Keller, EPFL/UNIL)

- Note 1: Lots of researchers don't use the term but construct models close to these ones (e.g., Paulien Hogeweg at Utrecht)
- Note 2: Artificial life not only focuses on evolution but evolution is the heart of artificial life

---



Digital genetics: unravelling the genetic basis of evolution

Christoph Adami

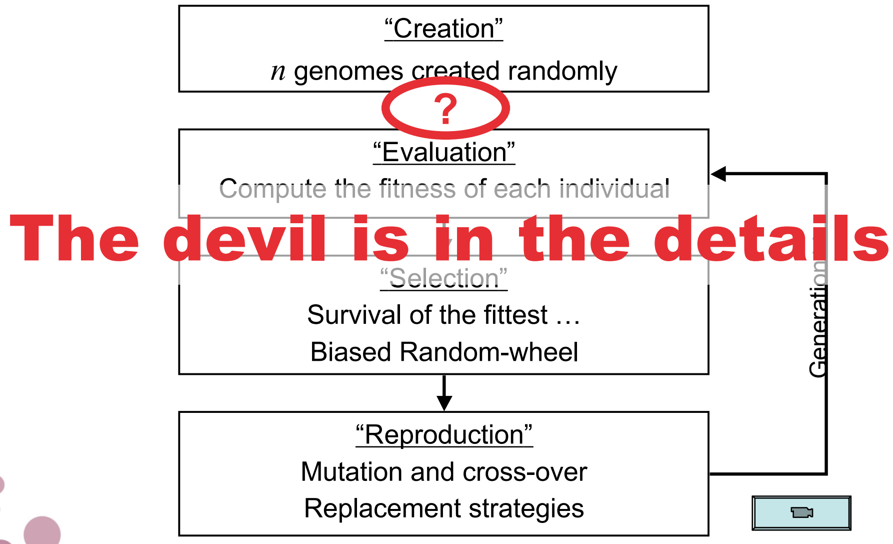
Abstract | Digital genetics, or the genetics of digital organisms, is a new field of research that has become possible as a result of the remarkable power of evolution experiments that use computers. Self-replicating strands of computer code that inhabit specially prepared computers can mutate, evolve and adapt to their environment. Digital organisms make it easy to conduct repeatable, controlled experiments, which have a perfect genetic 'fossil record'. This allows researchers to address fundamental questions about the genetic basis of the evolution of complexity, genome organization, robustness and evolvability, and to test the consequences of mutations, including their interaction and recombination, on the fate of populations and lineages.

NATURE REVIEWS | GENETICS VOLUME 7 | FEBRUARY 2006 | 109

Digital genetics

- Software that creates environment inside of a computer for populations of *self-replicating* elements, subject to *mutation* and *survival of the fittest*
  - “Real evolution of false organisms” (real Darwinism)
- This software can be used in a experimental setup
  - Modify some parameters of the simulation, look at the consequences on the organisms and/or on the ecosystem
  - Look for regularities…
- Experiments can be *repeated* many times for statistical accuracy.
  - All mutational events are known

Pseudo-code



“Creation”

n genomes created randomly

?

“Evaluation”

Compute the fitness of each individual

The devil is in the details

“Selection”

Survival of the fittest …

Biased Random-wheel

“Reproduction”

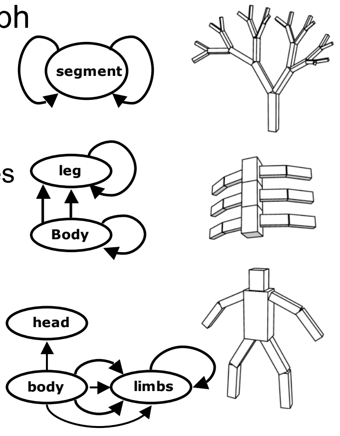
Mutation and cross-over

Replacement strategies

Evolved Virtual Creatures

(Karl Sims 1994)

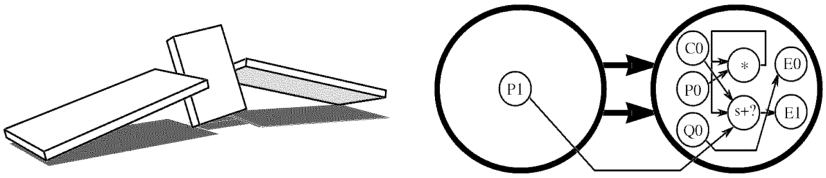
- Each creature is defined by a graph
  - One node = one body element
  - One link = one joint
  - Dual-links = multiple bodies
  - Recursive links = repeated structures
- Nodes and links are valued
  - Dimensions
  - Joint limits
  - Relative position
  - Recursion control
  - Joint control
  - ...



Evolved Virtual Creatures

(Karl Sims 1994)

- Each creature owns a distributed brain that receives stimuli and produces motor output at the joints …

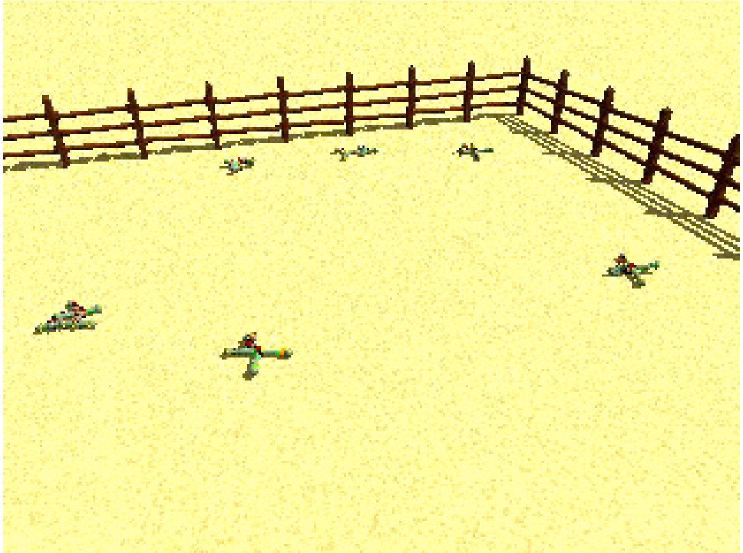


- Example:
  - P1: body light-sensor
  - C0, P0, Q0 : “wings” light-sensors
  - *, s+? : computation elements
  - E0, E1 : joint motor control

Evolved Virtual Creatures

(Karl Sims 1994)

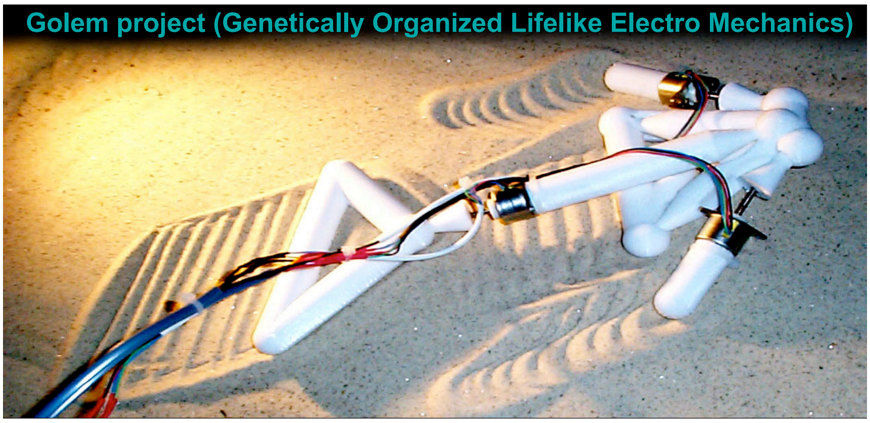
- Each creature “lives” in a precisely controlled world (viscosity, gravity, obstacles, light, …)
  - The emergent morphology and behavior is strongly dependent on the environment condition (although highly variable)
- The main difficulty is the computation of the fitness values (i.e. the simulation part!)
  - Each simulation error is rapidly detected and used by the creatures!
- Nice! What can we conclude?
  - Hmm … good question
  - It is almost impossible to disentangle the effect of evolution and environmental conditions from the effect of the (very complicated) genotype to phenotype mapping!
  - But Sims paved the way for many models…



[http://www.framsticks.com/]

The Golem Project

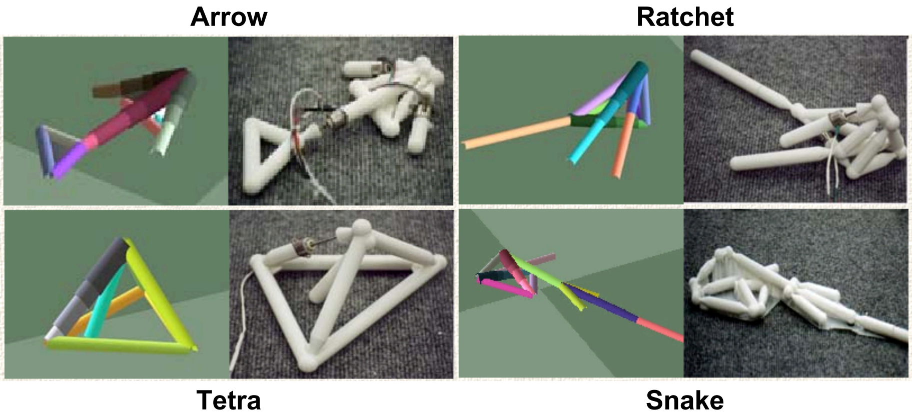
Automatic Design and Manufacture of Robotic Lifeforms
(Hod Lipson and Jordan Pollack)



First attempt to export artificially evolved creatures in the real world…

The Golem Project

Automatic Design and Manufacture of Robotic Lifeforms
(Hod Lipson and Jordan Pollack)

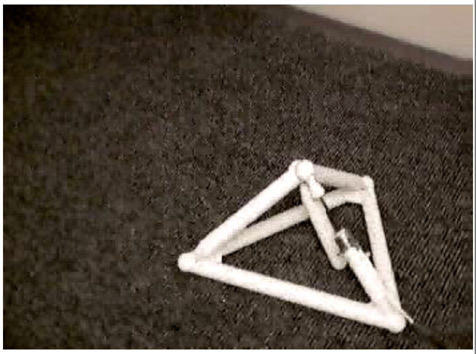


The Golem Project

Automatic Design and Manufacture of Robotic Lifeforms
(Hod Lipson and Jordan Pollack)

Simulated Tetra

Real Tetra



The Golem Project

Automatic Design and Manufacture of Robotic Lifeforms
(Hod Lipson and Jordan Pollack)

[http://www.demo.cs.brandeis.edu/golem/]

Artificial chemistries

- Creatures and the successive models aim at simulating real “high level” organisms like mammals, birds, worms or snakes
  - Too complex to comprehend?
  - Interesting for engineering and computer graphics
  - Actually “real results” in evolutionary biology
- We need a more simple genotype to phenotype mapping
  - Models based on artificial chemistries
- Artificial chemistries
  - Computer instructions or sequences interpreted by a virtual CPU to produce the behavior of the organism
  - Historically artificial chemistries come from “core-war” games
  - Various formalisms [Dittrich et al., 2001] …

Tierra: the ancestor

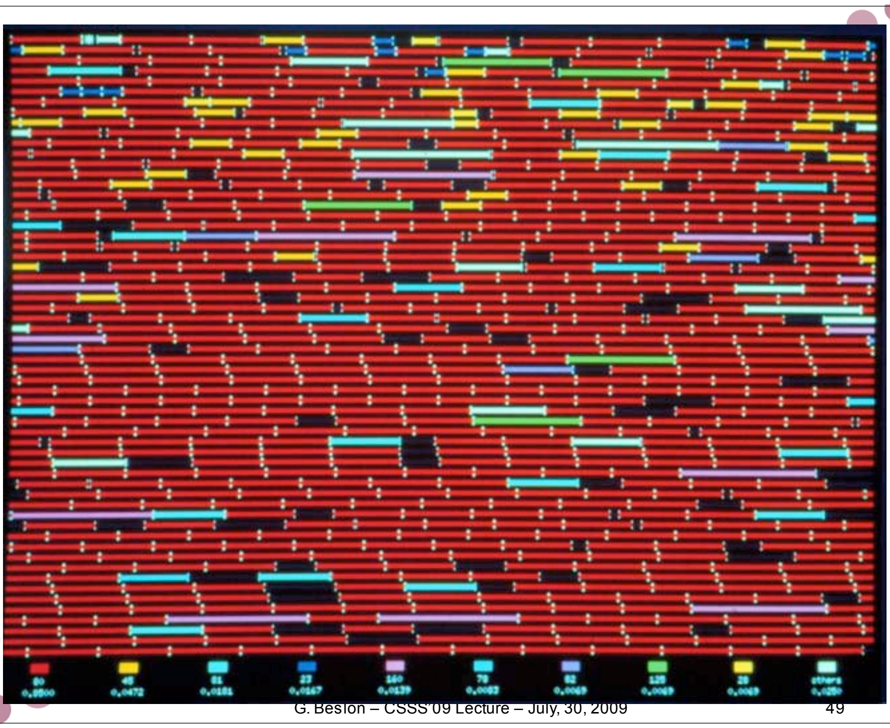
(Tom Ray, 1992)

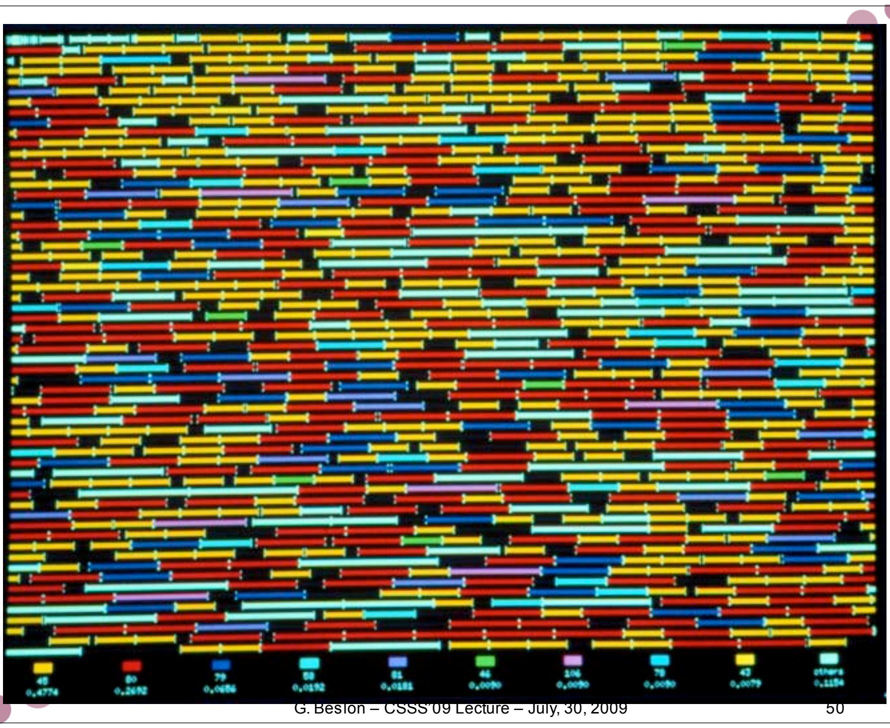
“*In Tierra, the self-replicating entities are executable machine code programs, which do nothing more than make copies of themselves in the RAM memory of the computer. Thus the machine code becomes an analogue of the nucleic acid based genetic code of organic life*”

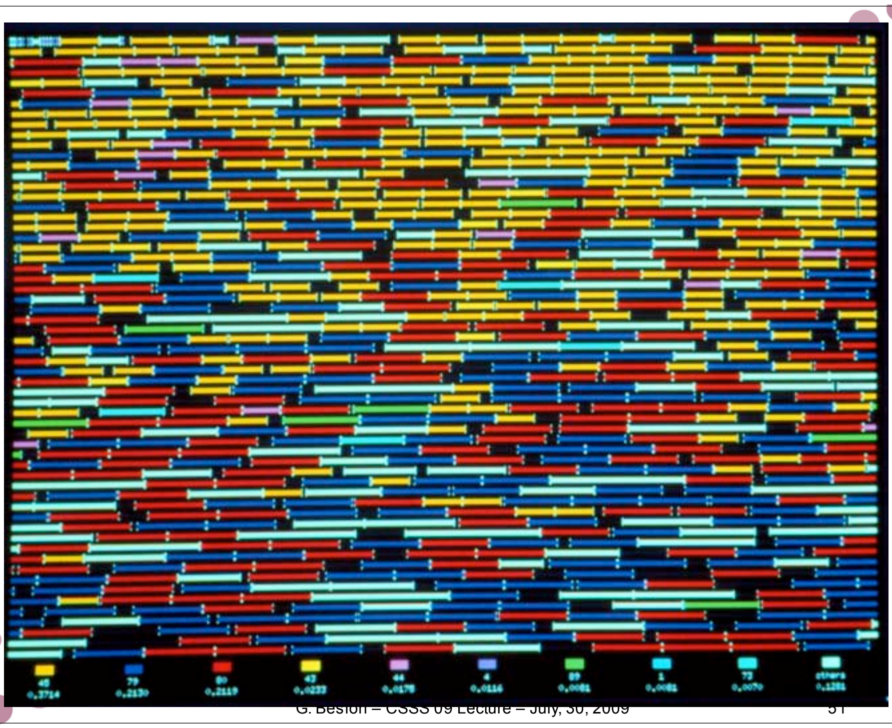
[T. Ray]

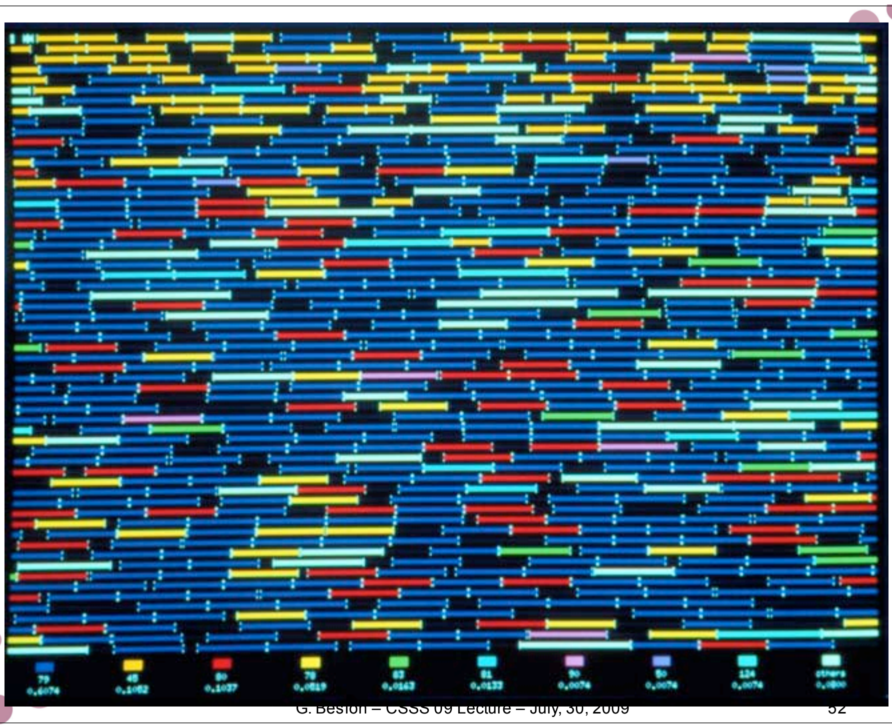
- Tierra enables to study the evolutionary behavior of evolving entities engaged in an “open-ended evolution”
  - No goal but (implicitely) survive and reproduce
  - Need to be sowed by some predefined code able to self-reproduce
- Tierra is an evolving ecological system

[http://life.ou.edu/pubs/fatm/fatm.html]


80
0,8500
45
0,0472
81
0,0181
27
0,0167
160
0,0139
78
0,0083
82
0,0069
125
0,0069
28
0,0069
others
0,0250


45
0,4774
80
0,2692
79
0,0656
58
0,0192
81
0,0181
46
0,0090
106
0,0090
78
0,0090
47
0,0079
others
0,1154


45
0,3714
79
0,2130
80
0,2119
43
0,0233
44
0,0178
4
0,0116
89
0,0081
1
0,0081
72
0,0070
others
0,1281

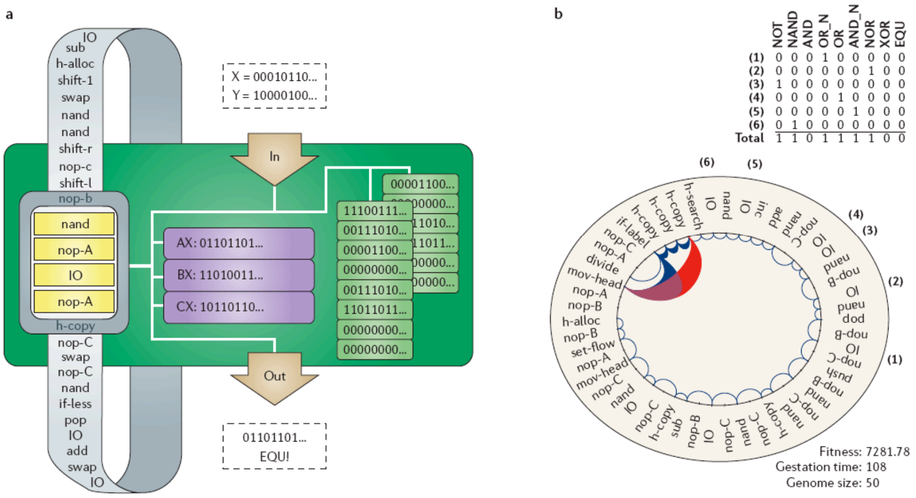

79
0,6074
45
0,1052
80
0,1037
78
0,0519
83
0,0163
81
0,0133
90
0,0074
50
0,0074
124
0,0074
others
0,0800

Avida: the maturity

(Chris Adami, 1997)

- Avida is not only a “better” model; it also starts with better questions
  - Chris Adami interacts with biologists on an almost daily basis …
  - Important collaboration with Richard Lenski
- Avida uses a simpler artificial chemistry than Tierra
  - Each “avidian” contains its own CPU (no interactions during code execution)
  - Avidians are immerged in a 2D space
  - The evolution is no more open-ended (the “fitness” don’t have the same meaning!) but the results are easier to analyze!
  - Better trade-of between simplicity and complexity of the model
- Many results in biology
  - See e.g., C. Adami, T. Collier, S. F. Elena, C. Ofria, C. Wilke, R. Lenski, D. Misevic…

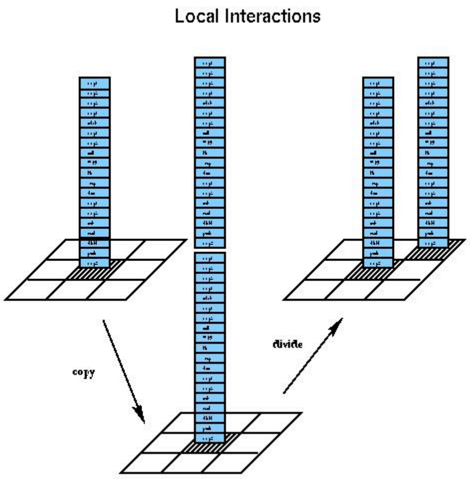
“Avidians”



	NOT	NAND	AND	OR_N	OR	AND_N	NOR	XOR	EQU
(1)	0	0	0	1	0	0	0	0	0
(2)	0	0	0	0	0	0	1	0	0
(3)	1	0	0	0	0	0	0	0	0
(4)	0	0	0	0	1	0	0	0	0
(5)	0	0	0	0	0	1	0	0	0
(6)	0	1	0	0	0	0	0	0	0
Total	1	1	0	1	1	1	1	0	0

[Adami, 2006]

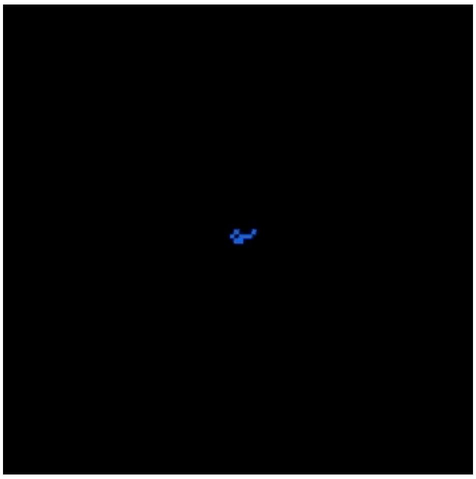
Replication

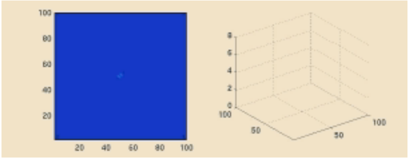


Avida: Metabolism

- Organisms have limited CPU time to execute their code
  - Each organism earns a minimal “salary”
  - Organisms can win extra CPU time if they perform computations on their inputs and write the result on their output
  - Avida rewards 78 different logical operations if the correct genotype appears
- Note that, in Avida, the genotype is confounded with the phenotype
  - Similar to an RNA-World
  - Evolution of active molecules without transcription-translation mechanisms

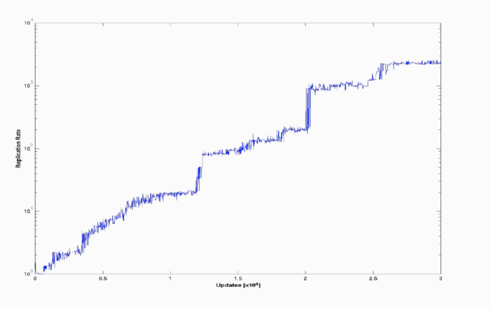
Evolutionary runs

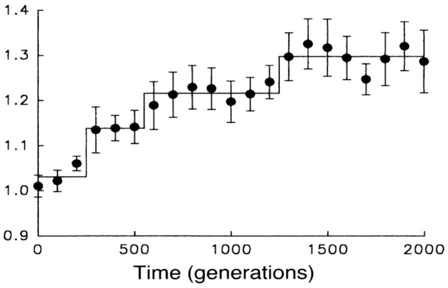




Try it!
http://

Typical evolutionary run in Avida and in *E. coli*





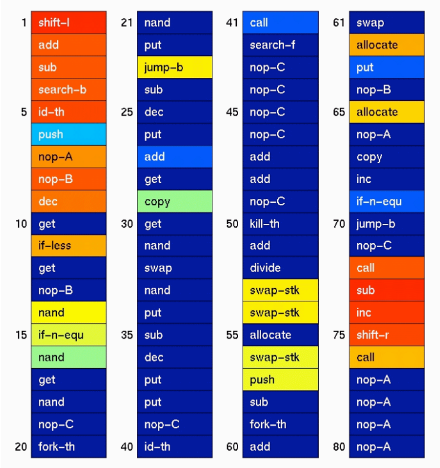
Avidians

(Chris Adami)

E. Coli

(Richard Lenski)

Final genome



1	shift–l	21	nand	41	call	61	swap
	add		put		search–f		allocate
	sub		jump–b		nop–C		put
	search–b		sub		nop–C		nop–B
5	id–th	25	dec	45	nop–C	65	allocate
	push		put		nop–C		nop–A
	nop–A		add		add		copy
	nop–B		get		add		inc
	dec		copy		nop–C		if–n–equ
10	get	30	get	50	kill–th	70	jump–b
	if–less		nand		add		nop–C
	get		swap		divide		call
	nop–B		nand		swap–stk		sub
	nand		put		swap–stk		inc
15	if–n–equ	35	sub	55	allocate	75	shift–r
	nand		dec		swap–stk		call
	get		put		push		nop–A
	nand		put		sub		nop–A
	nop–C		nop–C		fork–th		nop–A
20	fork–th	40	id–th	60	add	80	nop–A

- 80 “nucleotides”
- The color indicates the number of mutations undergone by each “nucleotide”
  - Blue: Stable nucleotide
  - Red: High polymorphism

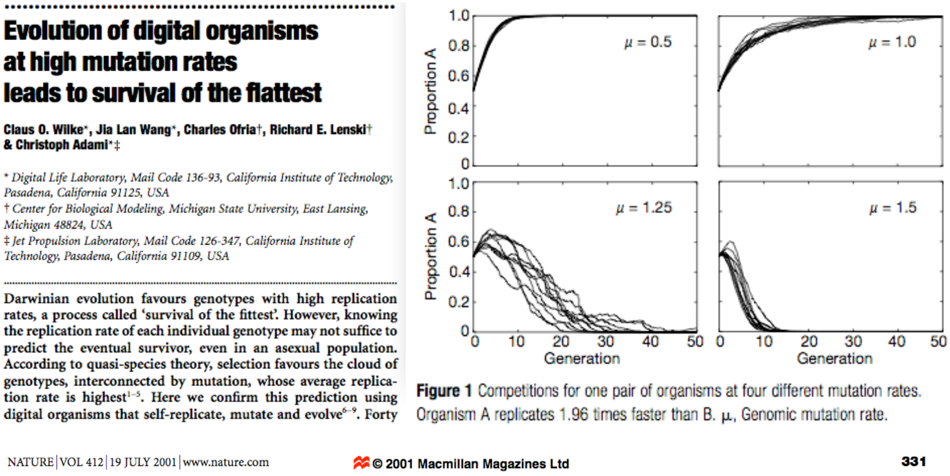
Experiment

- Two different organisms, same conditions
  - Yellow organism: good but not robust
  - Blue organism = not so good but robust

Mutation rate: 0.5

Mutation rate: 1.5

“Survival of the flattest”



Evolution of digital organisms at high mutation rates leads to survival of the flattest

Claus O. Wilke*, Jia Lan Wang*, Charles Ofria†, Richard E. Lenski† & Christoph Adami*‡

** Digital Life Laboratory, Mail Code 136-93, California Institute of Technology, Pasadena, California 91125, USA*
† Center for Biological Modeling, Michigan State University, East Lansing, Michigan 48824, USA
‡ Jet Propulsion Laboratory, Mail Code 126-347, California Institute of Technology, Pasadena, California 91109, USA

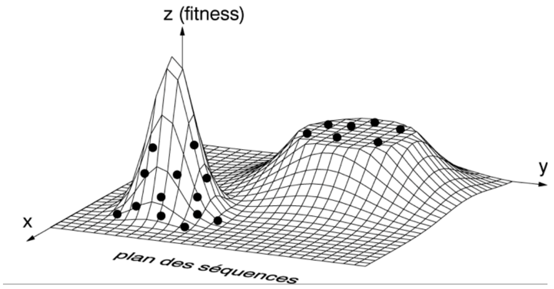
Darwinian evolution favours genotypes with high replication rates, a process called ‘survival of the fittest’. However, knowing the replication rate of each individual genotype may not suffice to predict the eventual survivor, even in an asexual population. According to quasi-species theory, selection favours the cloud of genotypes, interconnected by mutation, whose average replication rate is highest[1–5]. Here we confirm this prediction using digital organisms that self-replicate, mutate and evolve[6–9]. Forty

Figure 1 Competitions for one pair of organisms at four different mutation rates. Organism A replicates 1.96 times faster than B. μ, Genomic mutation rate.

NATURE | VOL 412 | 19 JULY 2001 | www.nature.com

Modeling the model

- “Survival of the flattest” [Wilke et al., Nature, 2001]
  - Under strong mutational pressure, sharp peaks are disadvantaged
  - When understood, the mechanism can be explained without the computational model (“*the model is no longer needed afterwards*”)
  - E.g., interpretation in terms of fitness landscape … the yellow is “high and thin”, the blue is “low but flat”



The Ævol model

An individual-based model of genome
evolution and genome structuring

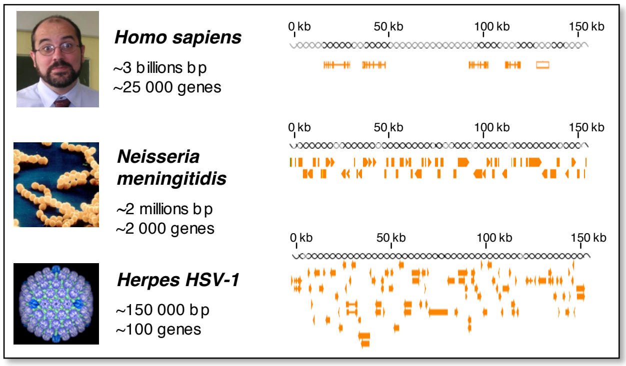
Knibbe, C. (2006) Structuration de génomes par sélection indirecte de la variabilité mutationnelle, une approche par modélisation et simulation, PhD Thesis, INSA-Lyon, décembre 2006, 174 p.

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

LIRIS INSA LYON Equipe BF2I BSMC Biologie des Systèmes et Modélisation Cellulaire IXXI Institut des Systèmes Complexes Complex Systems Institute Rhône-Alpes

---

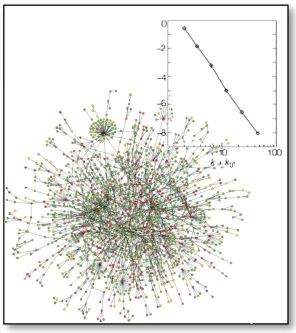
Origin of genomic structures?

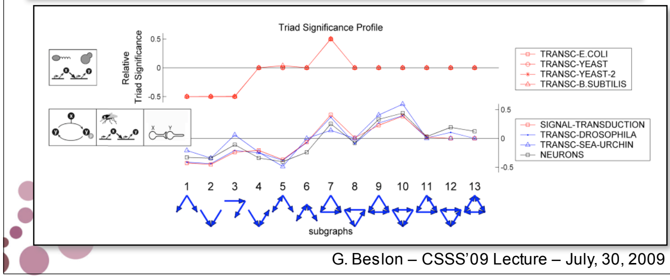


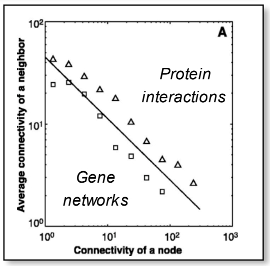
Homo sapiens
~3 billions bp
~25 000 genes

Neisseria meningitidis
~2 millions bp
~2 000 genes

Herpes HSV-1
~150 000 bp
~100 genes

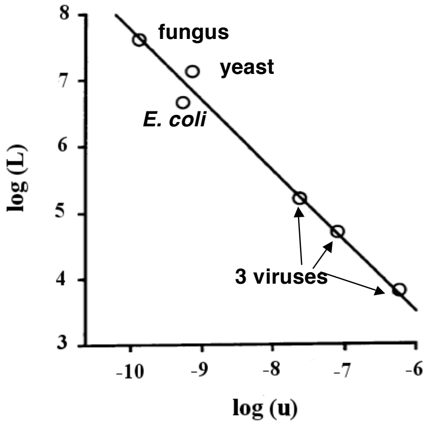




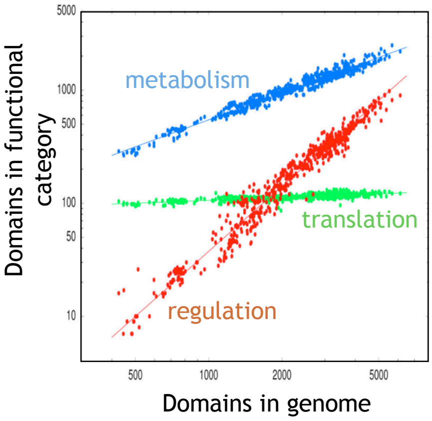


Scaling laws in real micro-organisms

Genome size
(Drake, 1991)



Domain content of genomes
(Molina & Van Nimwegen, 2008)



Origin of genomic structures?

Genotype: variation (mutations)

Indirect selection for the appropriate level of variability

Phenotype: selection

Mutational biases:

"Homo Sapiens genome spontaneously undergoes more insertions than deletions"

Selective costs:

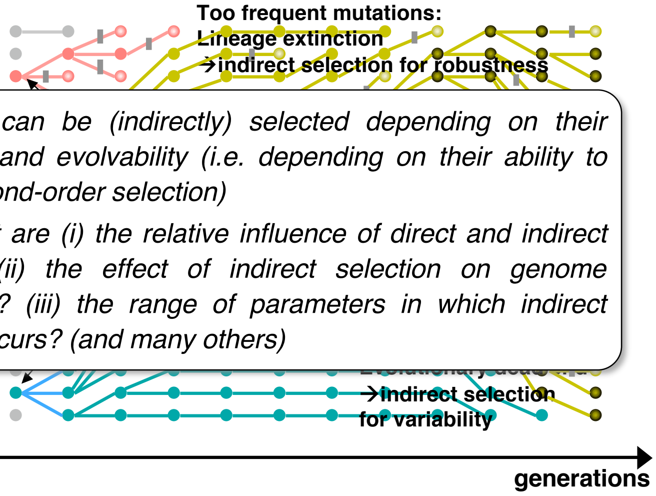
"A long genome can be disadvantageous for a bacteria or a virus"

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

Indirect selection, a thought experiment



High variability level
(Low probability to
reprodu
$F_v \approx 0$)

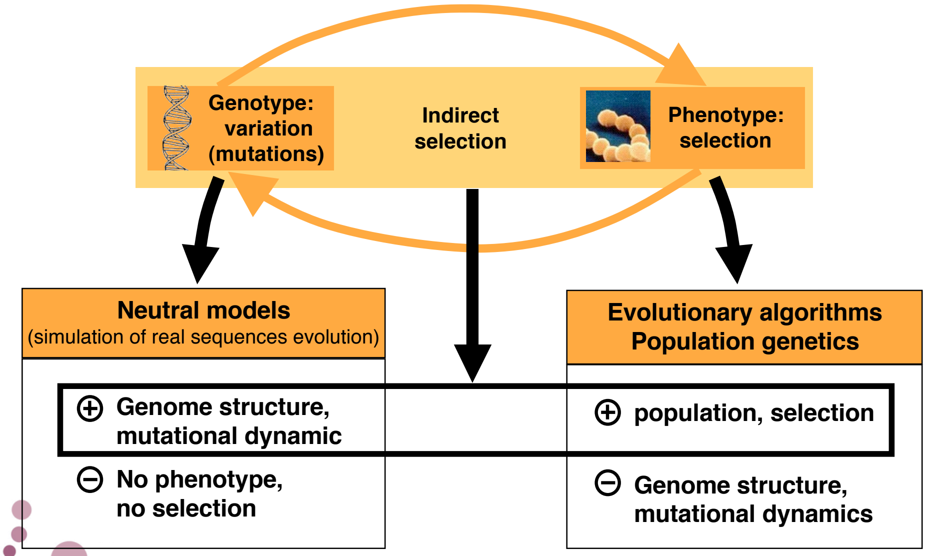
Mid var

Low var
(High probability to
reproduce neutrally:
$F_v >> 1$)

Organisms can be (indirectly) selected depending on their robustness and evolvability (i.e. depending on their ability to evolve; second-order selection)

… But what are (i) the relative influence of direct and indirect selection? (ii) the effect of indirect selection on genome architecture? (iii) the range of parameters in which indirect selection occurs? (and many others)

How to model indirect selection?



Genotype: variation (mutations)

Indirect selection

Phenotype: selection

Neutral models
(simulation of real sequences evolution)

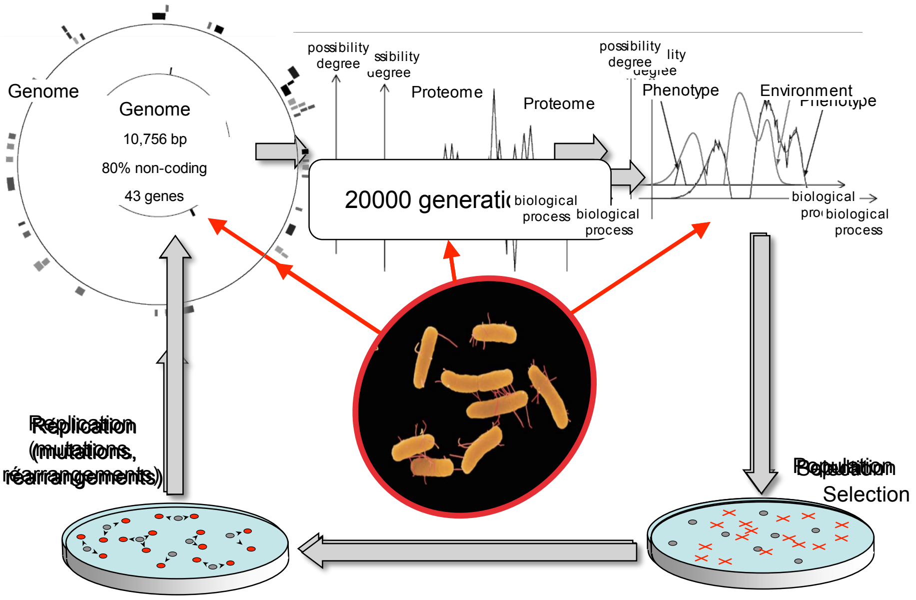
- ⊕ **Genome structure, mutational dynamic**
- ⊖ **No phenotype, no selection**

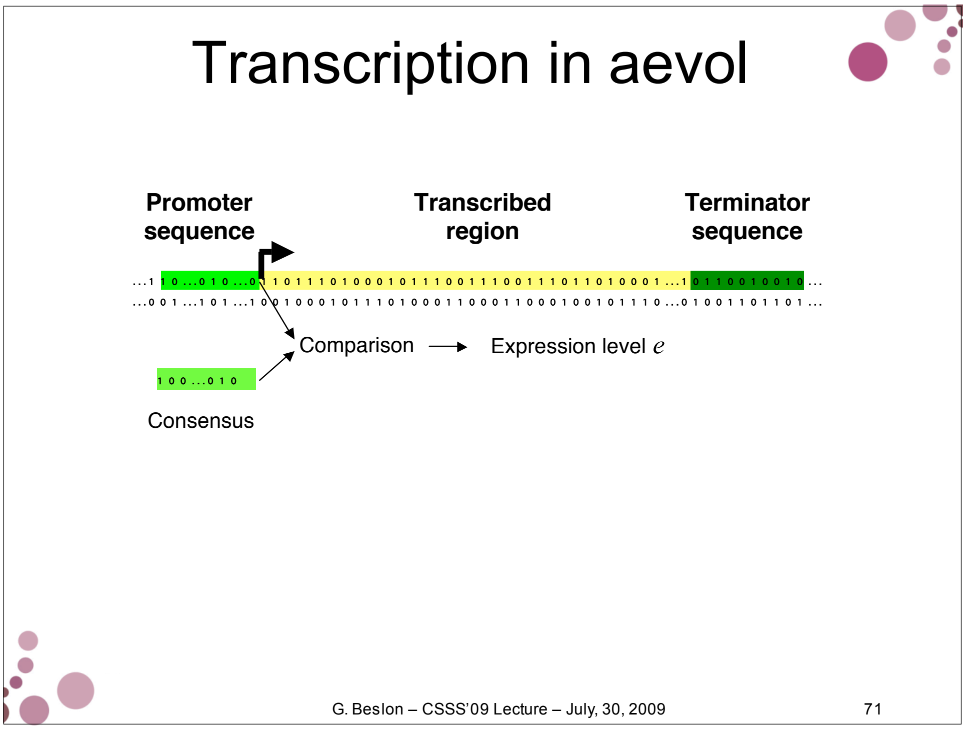
Evolutionary algorithms Population genetics

- ⊕ **population, selection**
- ⊖ **Genome structure, mutational dynamics**

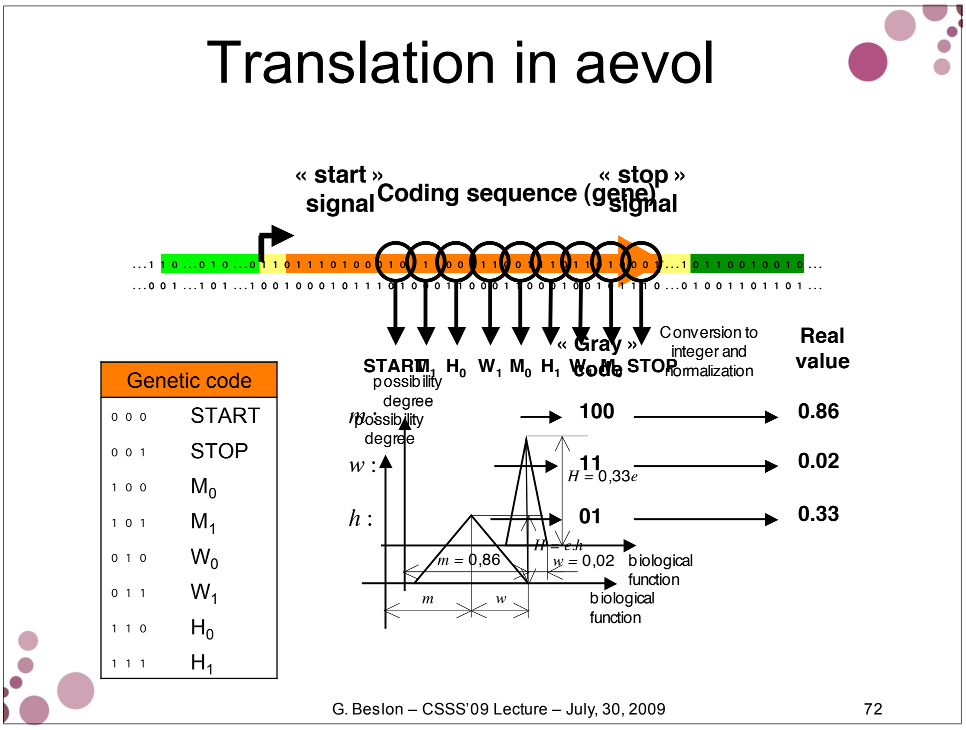
G. Beslon – CSSS'09 Lecture – July, 30, 2009

The aevol model



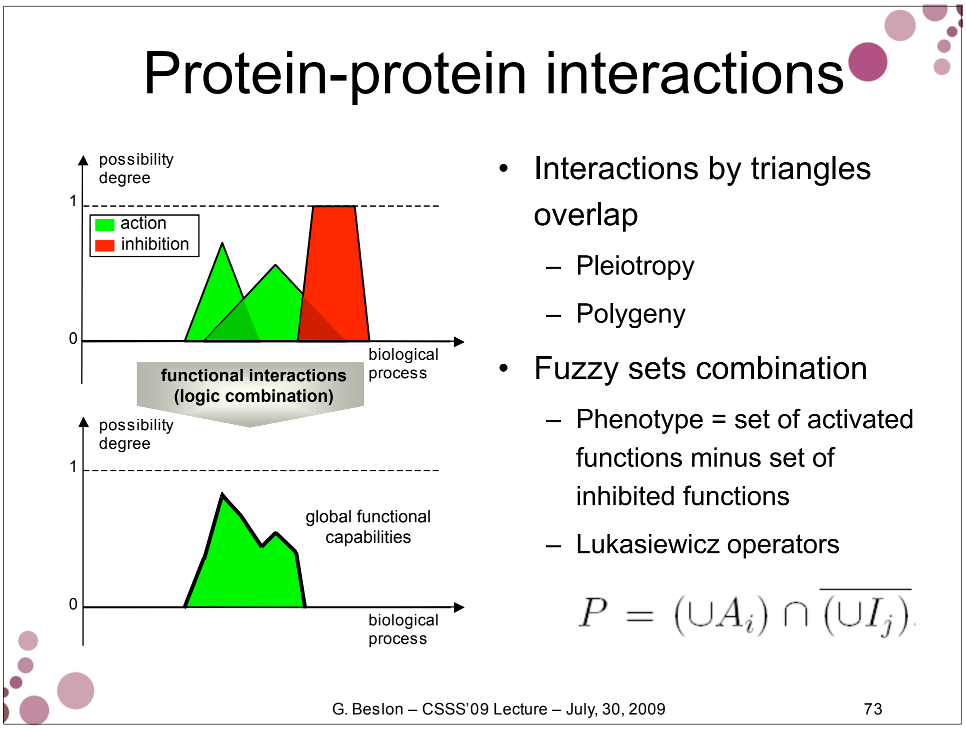


Transcription in aevol



Translation in aevol

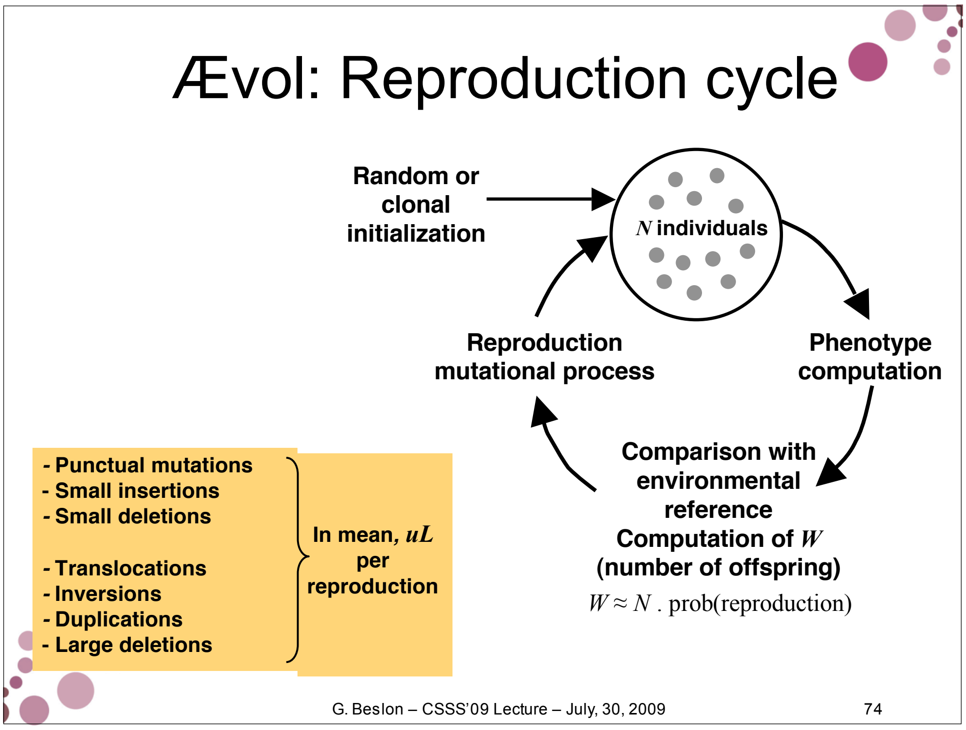
Genetic code	
0 0 0	START
0 0 1	STOP
1 0 0	M_0
1 0 1	M_1
0 1 0	W_0
0 1 1	W_1
1 1 0	H_0
1 1 1	H_1



Protein-protein interactions

- Interactions by triangles overlap
  - Pleiotropy
  - Polygeny
- Fuzzy sets combination
  - Phenotype = set of activated functions minus set of inhibited functions
  - Lukasiewicz operators

$$P = (\cup A_i) \cap \overline{(\cup I_j)}$$



Ævol: Reproduction cycle

Random or clonal initialization

***N* individuals**

Phenotype computation

Comparison with environmental reference
Computation of *W* (number of offspring)

$W \approx N$. prob(reproduction)

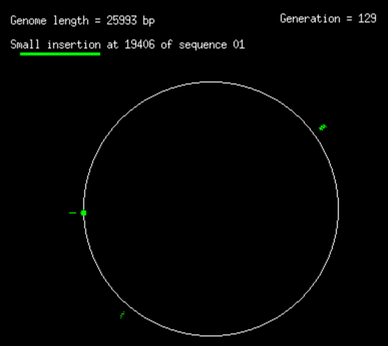
Reproduction mutational process

- Punctual mutations
- Small insertions
- Small deletions

- Translocations
- Inversions
- Duplications
- Large deletions

In mean, *uL* per reproduction

Ævol: The movie (« winning » lineage)

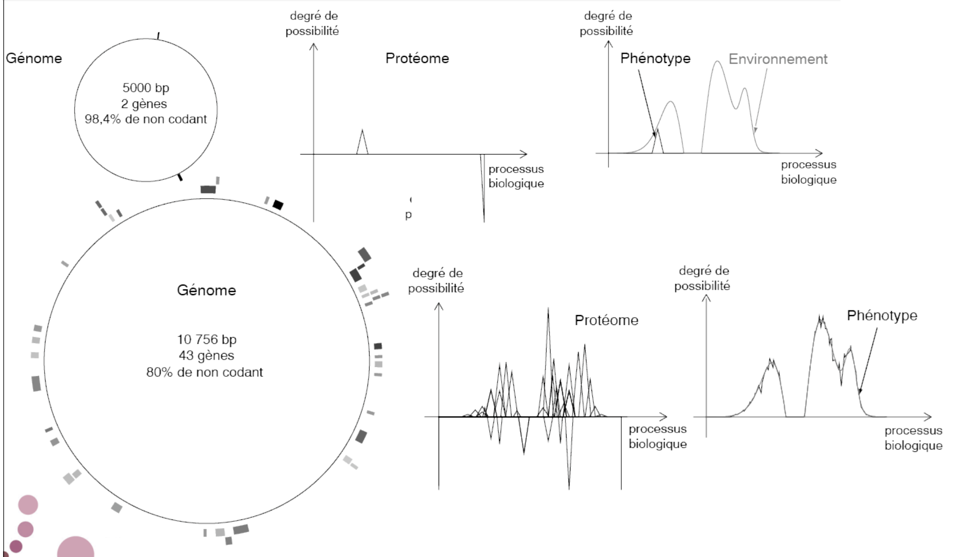


Genome length = 25993 bp

Generation = 129

Small insertion at 19406 of sequence 01

Organisms at $t = 0$ and $t = 20000$



In-silico experimental evolution

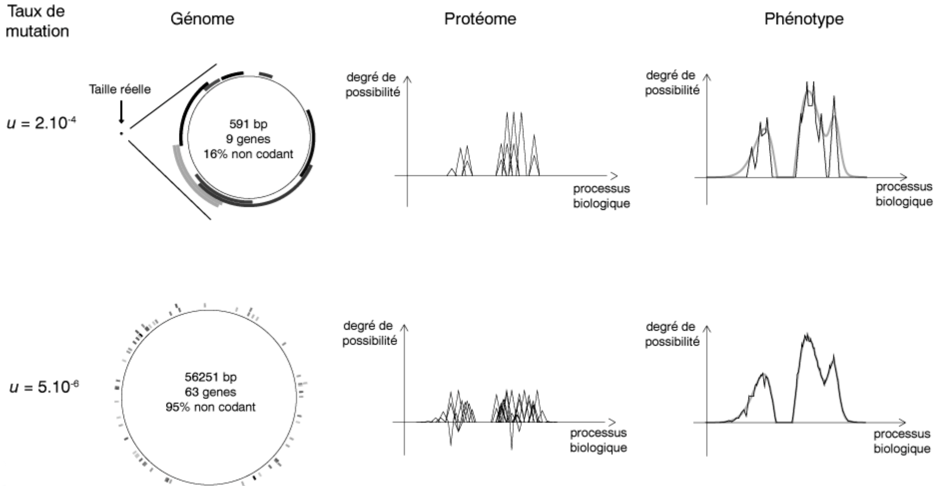
- Mutation rate u:
  - Six mutation rates from $u = 5.10^{-6}$ to $u = 2.10^{-4}$ per bp
  - Same mutation rates for point mutations and rearrangements
- Selection:
  - Two selection modes (fitness proportional or rank-based)
  - Different selection strength (here $k = 250$ or $k = 1000$)
- Experimental evolution during 20000 generations
  - Populations: 1000 individuals
  - Steady environment
- Three repetitions per couple (u,k)
  - More than 100 simulations
  - It's *really* an experimental approach …

Ævol: The movie (II) …

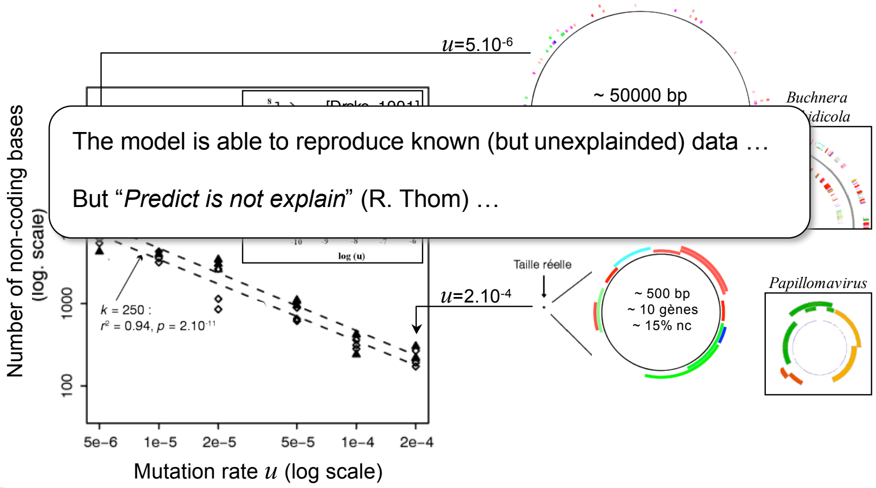
High mutation rates : 2.10^{-4} / pb

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

Best organisms for $u = 2.10^{-4}$ and $u = 5.10^{-6}$



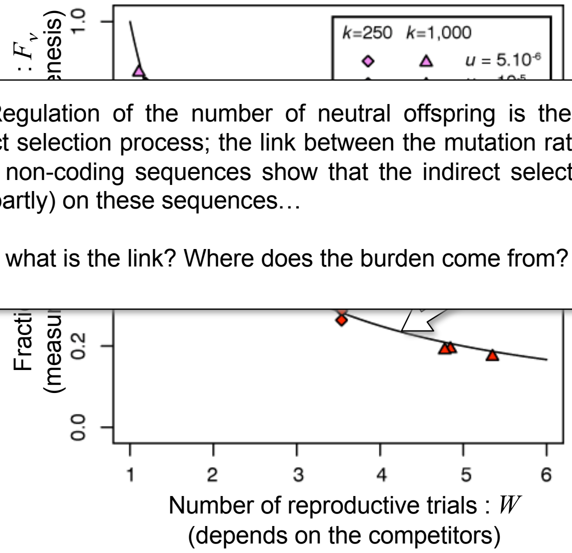
Scaling laws emerge *in silico*



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

But "*Predict is not explain*" (R. Thom) …

Experiments in the model



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?

Modeling the model

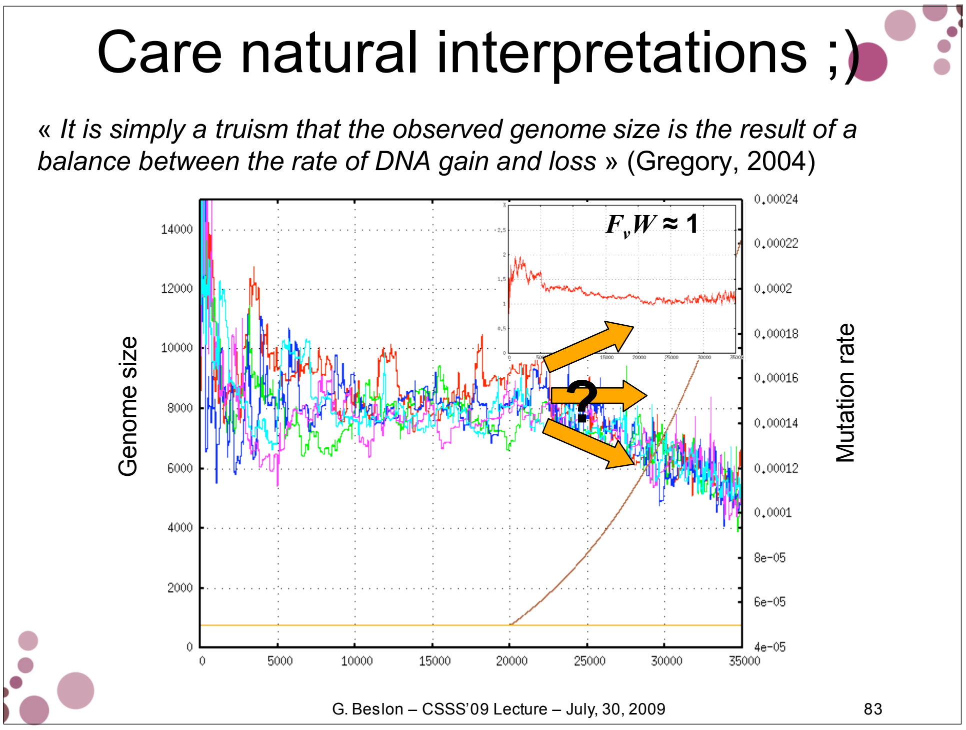
- Mathematical model of reproduction
  - The math model represents Aevol *AND* the "real world"…
- $\tilde{F}$: Probability of neutral reproduction as a function of

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

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

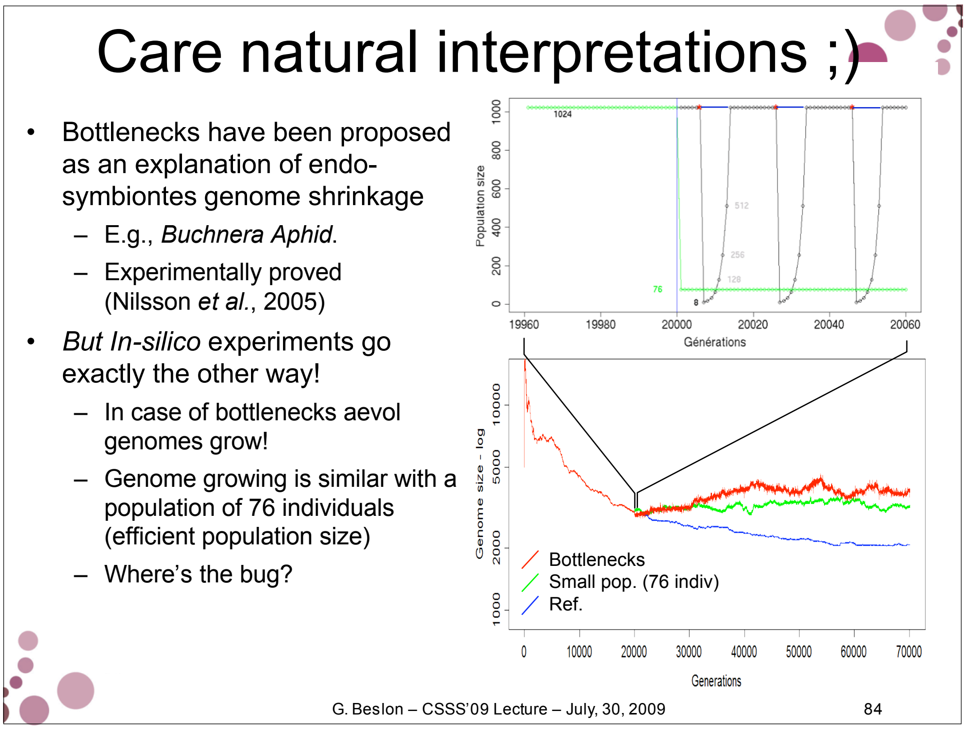
$$\left\{\begin{array}{lcl}\tilde{\nu}_{\text{ponct}} & = & \tilde{\nu}_{\text{ins}} = \tilde{\nu}_{\text{del}} = 1 - \frac{l}{L} \\ \tilde{\nu}_{\text{inv}} & = & \left(1 - \frac{l}{L}\right)^2 \\ \tilde{\nu}_{\text{transloc}} & = & \left(1 - \frac{l}{L}\right)^3 \end{array}\right.$$

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



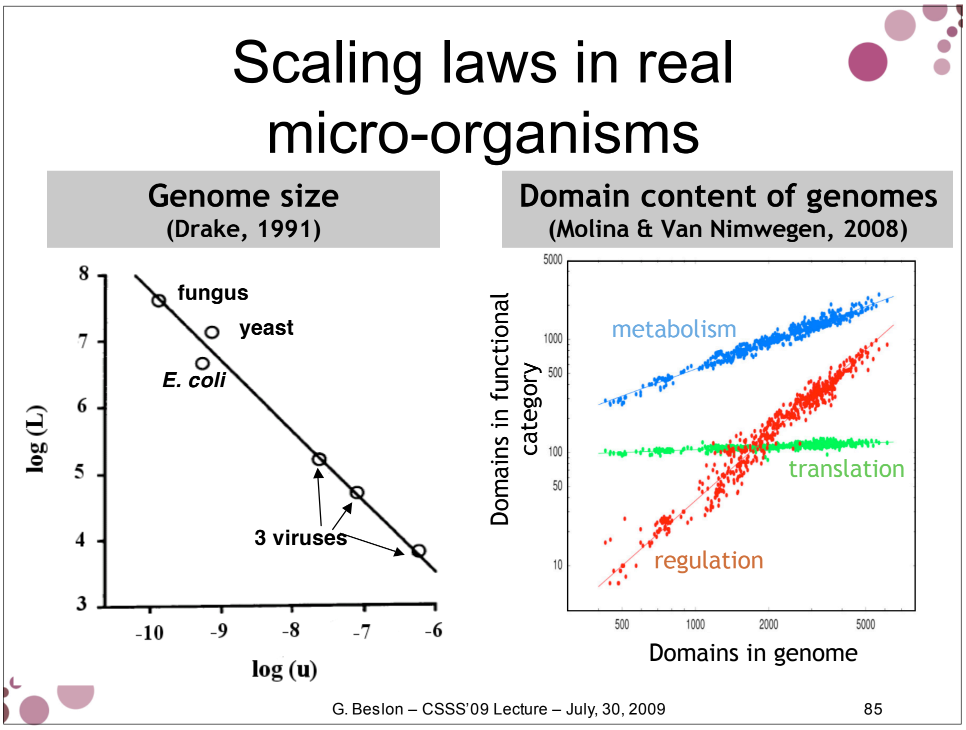
Care natural interpretations ;)

« *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)



Care natural interpretations ;)

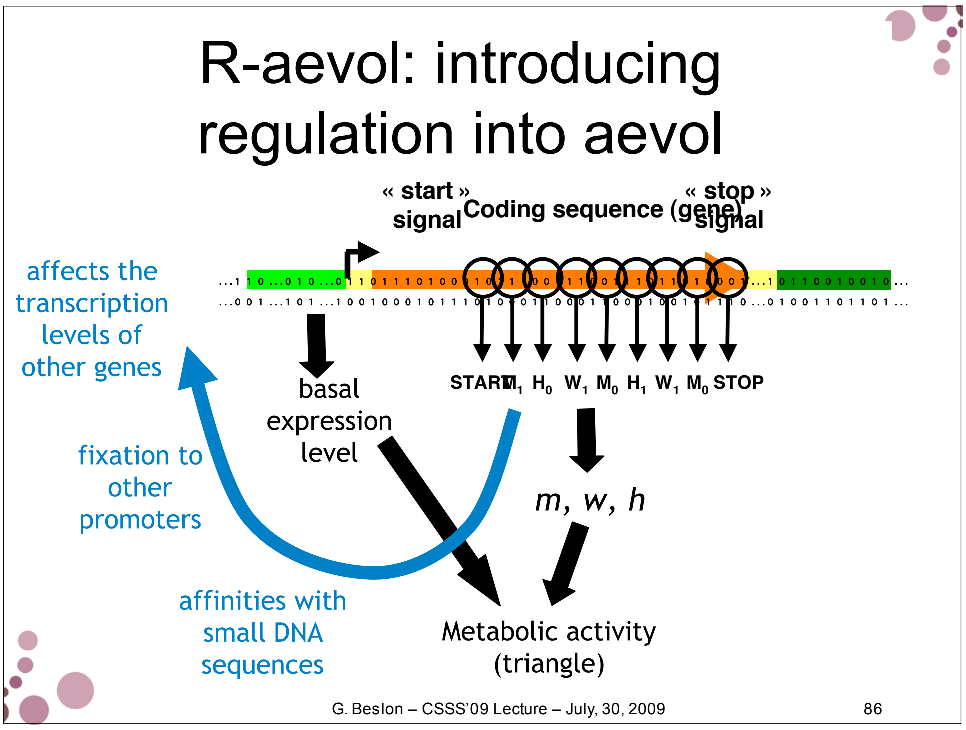
- Bottlenecks have been proposed as an explanation of endo-symbiontes genome shrinkage
  - E.g., *Buchnera Aphid*.
  - Experimentally proved (Nilsson *et al.*, 2005)
- *But In-silico* experiments go exactly the other way!
  - In case of bottlenecks aevol genomes grow!
  - Genome growing is similar with a population of 76 individuals (efficient population size)
  - Where's the bug?



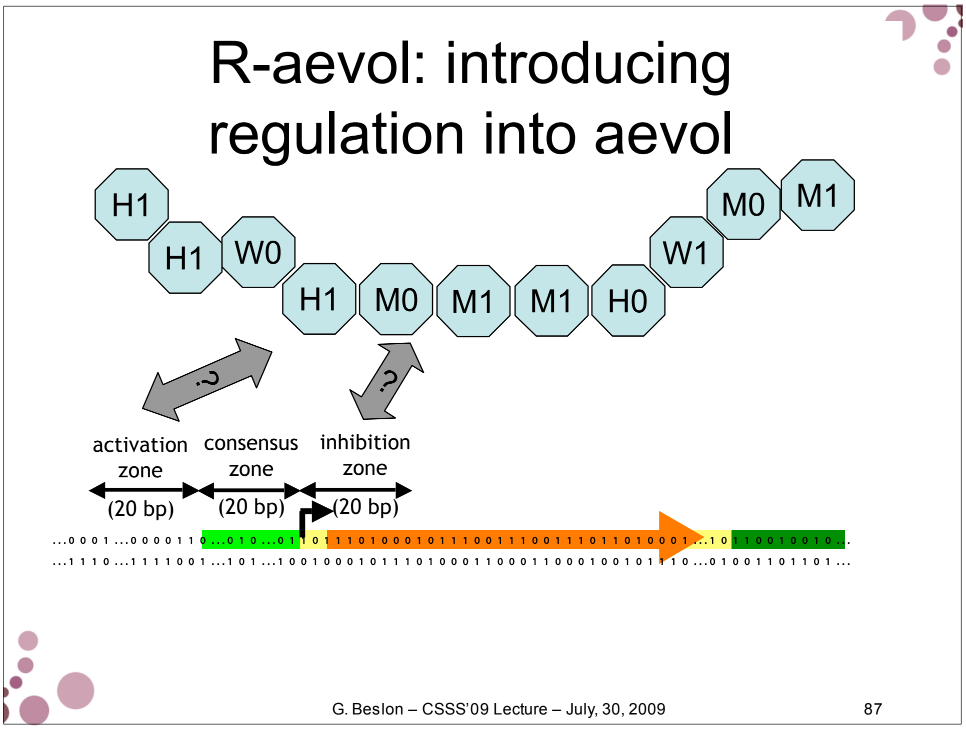
Scaling laws in real micro-organisms

Genome size
(Drake, 1991)

Domain content of genomes
(Molina & Van Nimwegen, 2008)



R-aevol: introducing regulation into aevol



R-aevol: introducing regulation into aevol

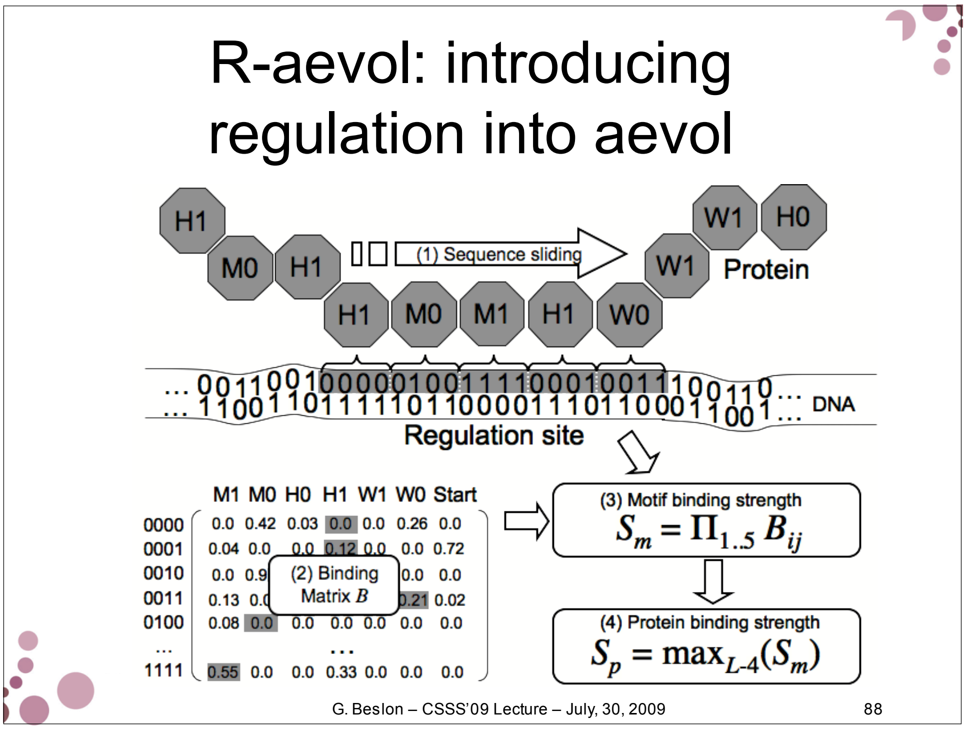
activation zone (20 bp)

consensus zone (20 bp)

inhibition zone (20 bp)

...0 0 0 1 ...0 0 0 0 1 1 0 ...0 1 0 ...0 1 1 0 1 1 1 0 1 0 0 0 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 1 1 0 1 0 0 0 1 ...1 0 1 1 0 0 1 0 0 1 0 ...

...1 1 1 0 ...1 1 1 1 0 0 1 ...1 0 1 ...1 0 0 1 0 0 0 1 0 1 1 1 0 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 0 0 1 0 1 1 1 0 ...0 1 0 0 1 1 0 1 1 0 1 ...



R-aevol: introducing regulation into aevol

(1) Sequence sliding

Protein

… 0011001000001001111000100111100110 …

… 1100110111110110000111011000011001 …

DNA

Regulation site

	M1	M0	H0	H1	W1	W0	Start
0000	0.0	0.42	0.03	0.0	0.0	0.26	0.0
0001	0.04	0.0	0.0	0.12	0.0	0.0	0.72
0010	0.0	0.9				0.0	0.0
0011	0.13	0.0				0.21	0.02
0100	0.08	0.0	0.0	0.0	0.0	0.0	0.0
…				…			
1111	0.55	0.0	0.0	0.33	0.0	0.0	0.0

(2) Binding Matrix B

(3) Motif binding strength

$$S_m = \Pi_{1..5} B_{ij}$$

(4) Protein binding strength

$$S_p = \max_{L-4}(S_m)$$

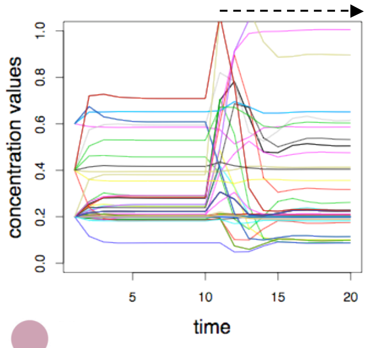
R-aevol: introducing regulation into aevol

- The expression level of a protein now depends on :
  - The similarity with the consensus
  - The affinities of other proteins for the activation zone
  - The affinities of other proteins for the inhibition zone
  - The expression levels of the activating proteins
  - The expression levels of the inhibiting proteins
- The expression level of a protein modulates :
  - The height of its triangle
  - The strength of its regulatory activity
- The organism's phenotype becomes a function of t
  - Organisms have a "life"

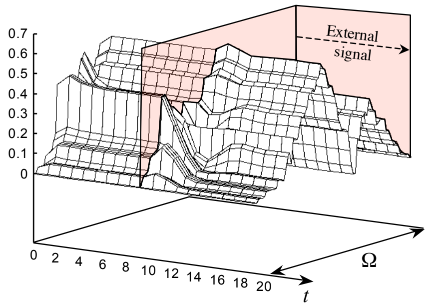
R-aevol: introducing regulation into aevol

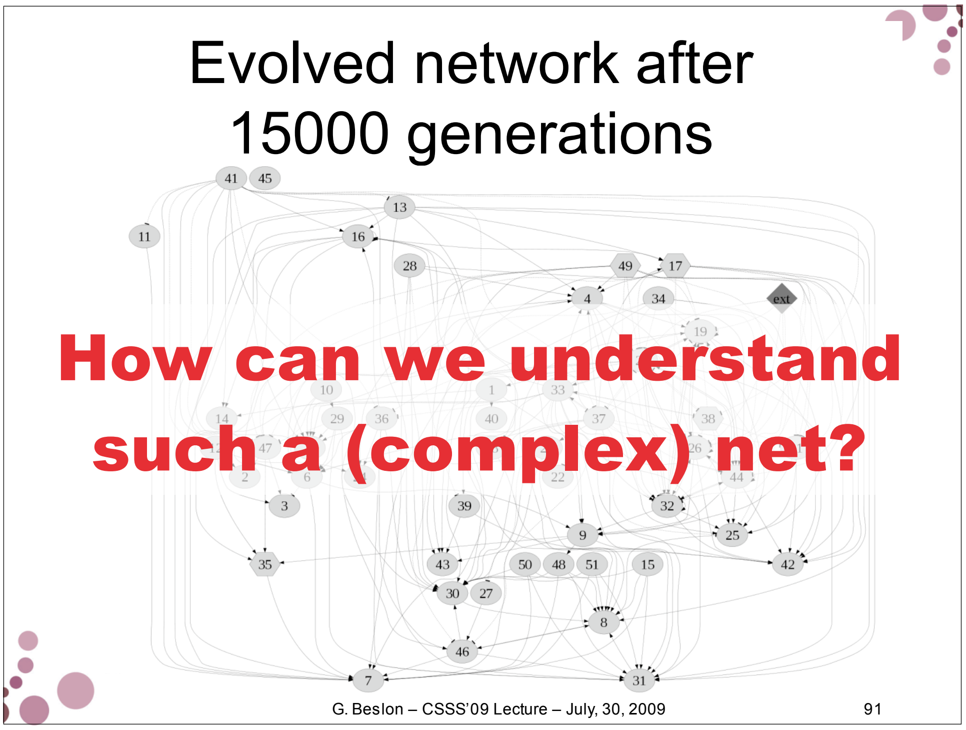
- The organism's phenotype becomes a function of t
  - Organisms have a "life"

Protein concentrations over time



Phenotype over time



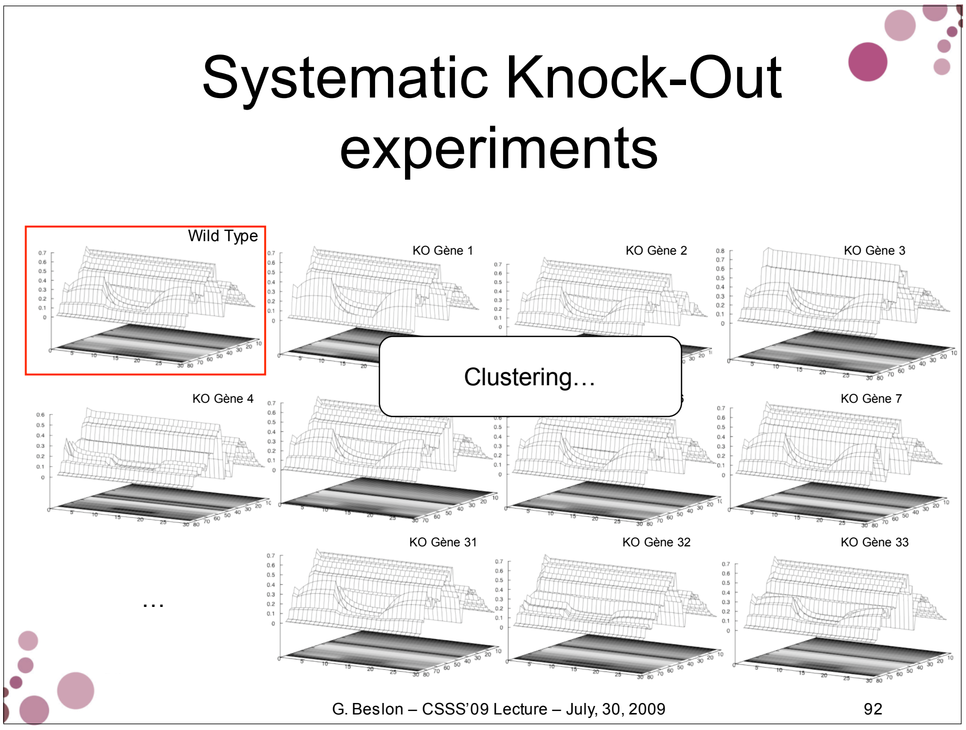


Evolved network after 15000 generations

How can we understand such a (complex) net?

G. Beslon – CSSS'09 Lecture – July, 30, 2009

91



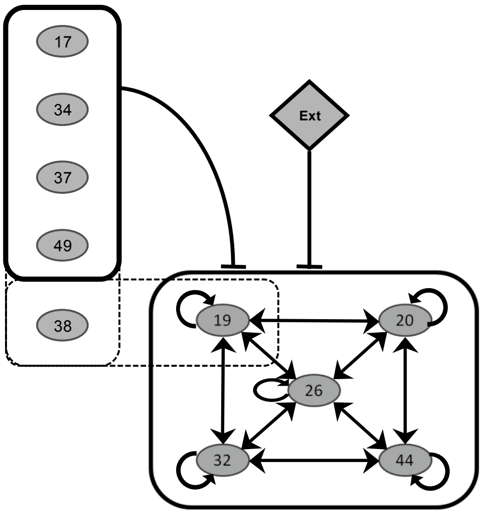
Systematic Knock-Out experiments

Clustering…

G. Beslon – CSSS'09 Lecture – July, 30, 2009

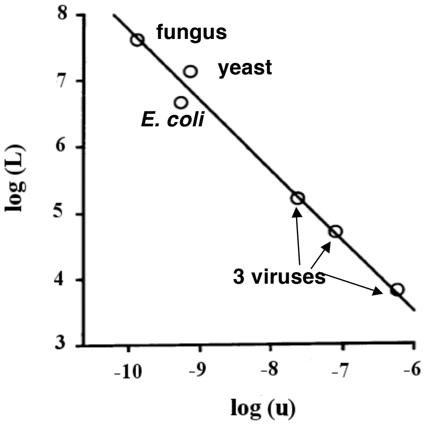
92

Reduced "schematic" network with two modules

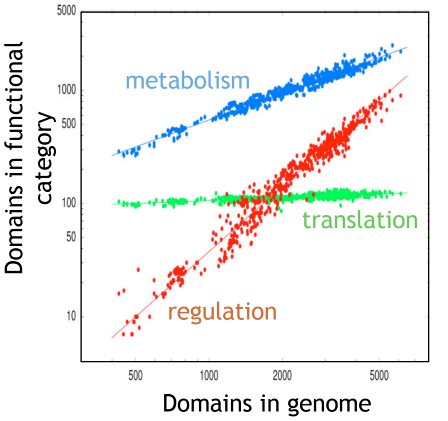


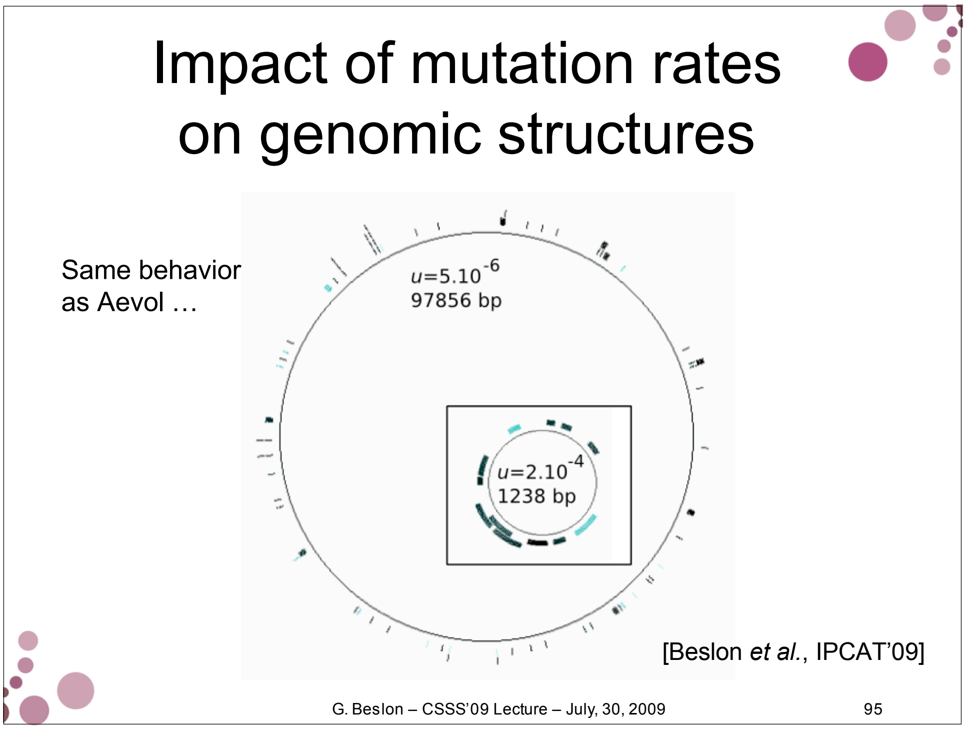
Scaling laws in real micro-organisms

Genome size
(Drake, 1991)



Domain content of genomes
(Molina & Van Nimwegen, 2008)

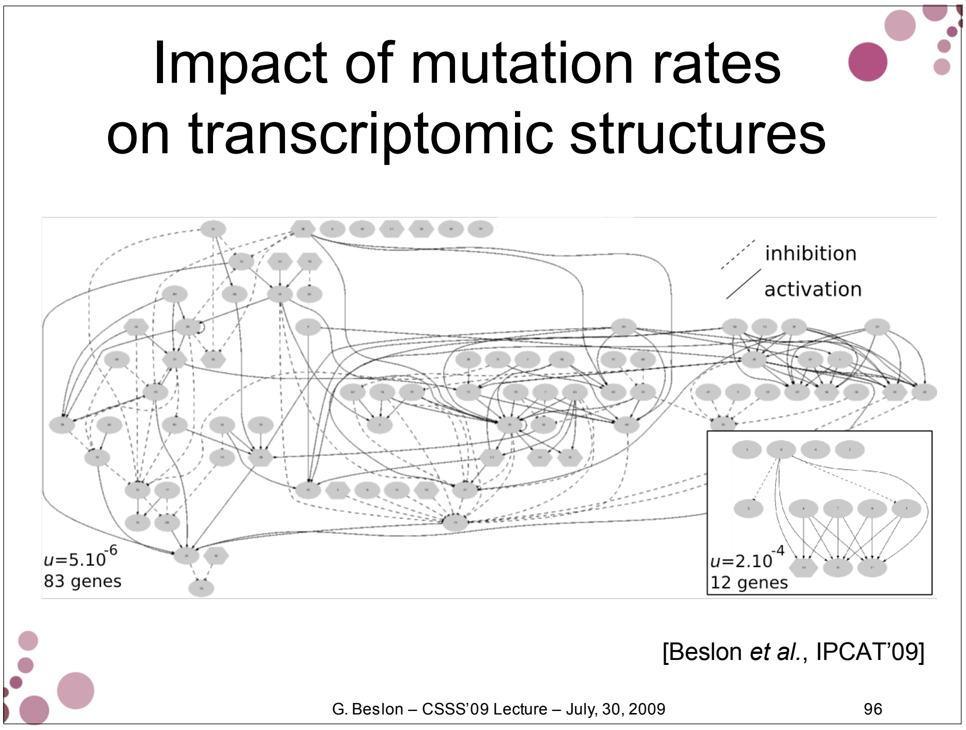




Impact of mutation rates on genomic structures

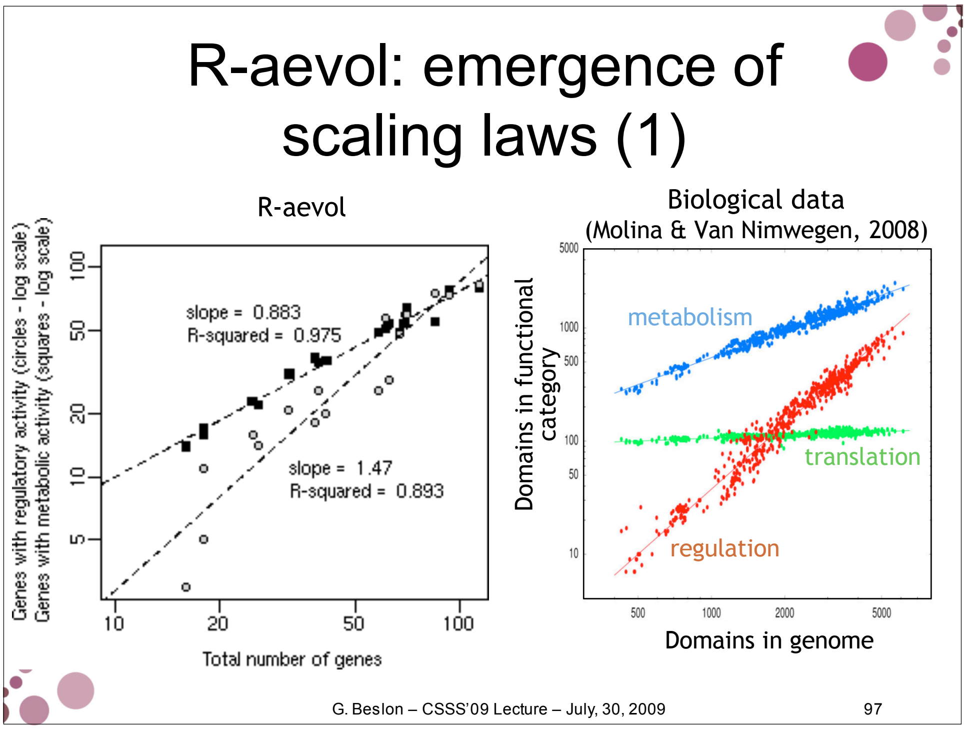
Same behavior as Aevol …

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



Impact of mutation rates on transcriptomic structures

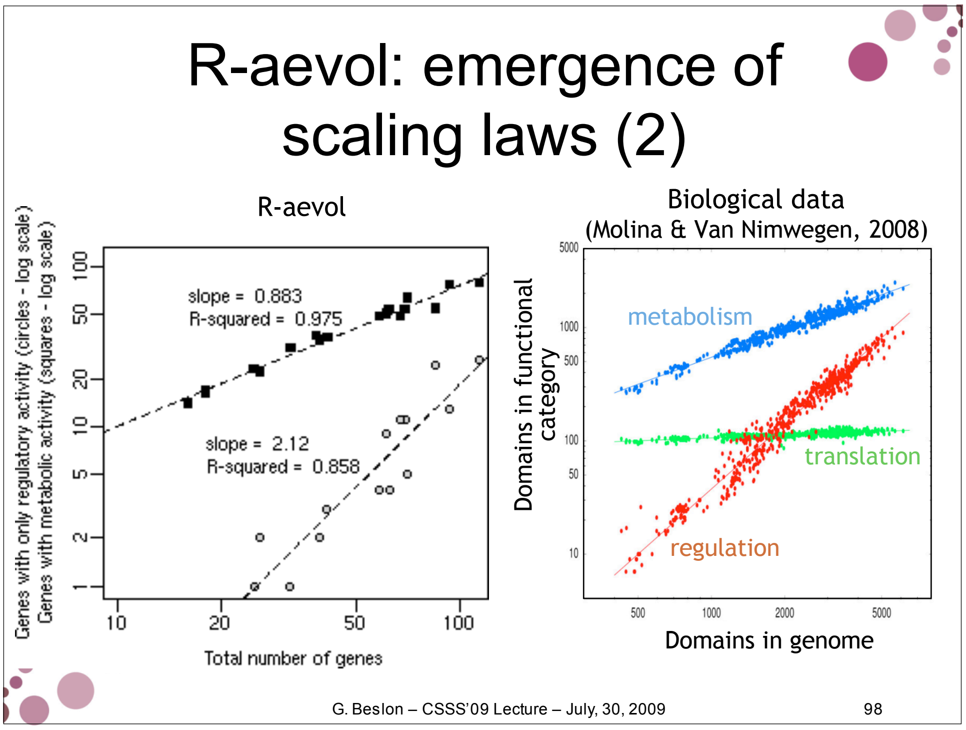
[Beslon *et al.*, IPCAT'09]



R-aevol: emergence of scaling laws (1)

R-aevol

Biological data
(Molina & Van Nimwegen, 2008)

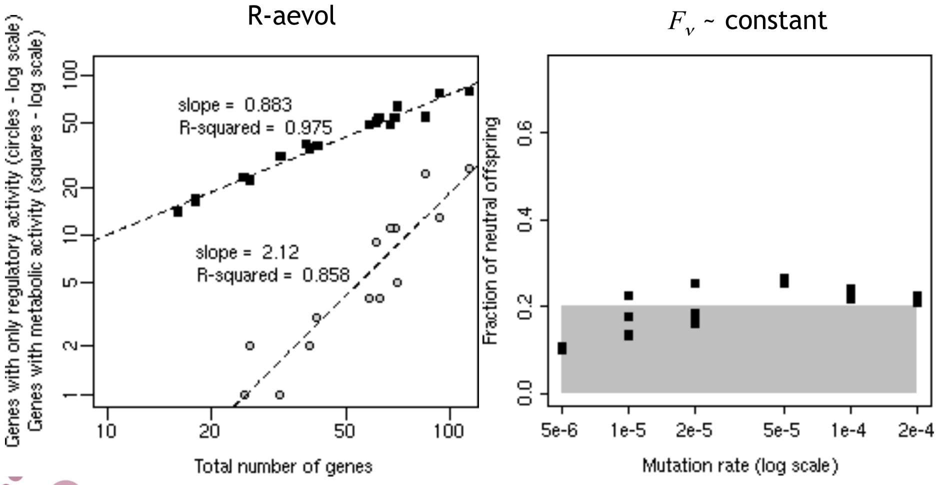


R-aevol: emergence of scaling laws (2)

R-aevol

Biological data
(Molina & Van Nimwegen, 2008)

R-aevol: emergence of scaling laws (3)



R-aevol

F_v ~ constant

Conclusion

- Aevol: *In silico* evolution of virtual « creatures »
  - with a variable-length genome, a variable gene number, a set of proteins and a phenotype
  - with or without a regulatory network
  - small mutations and large rearrangements
- Reproduces many real data
  - scaling laws, relationship between genome size and network size… other experiments are on the way…
- Enables to « open the box » to understand how these laws emerged
  - no direct cost of complexity
  - evolution of genome organization is driven by pressures for robustness and evolvability

References

G. Beslon, D. P. Parsons, Y. Sanchez-Dehesa, J.-M. Peña, C. Knibbe. Scaling Laws in Bacterial Genomes: A Side-Effect of Selection of Mutational Robustness? Submitted to: *BioSystems*

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

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 Genetics* 7(9):729-735.

thanks …