metabolic and spatial frustration as constraints for synthetic biology

antoine danchin 唐善・安東

amabiotics sas

tarpol summer course
basel, september 27, 2010

a future for microbial research

- reconstructing and understanding: forgetting the "black box" sb reconstructs life to explore whether we understand what life is and learn missing entities from our failures

- abstracting: sb keeps the laws defining life, and applies them using objects of a different physico-chemical nature (orthogonality) => managing information

- engineering: sb designs and standardises « biobricks » to construct a « cell factory » with man's interests' drive

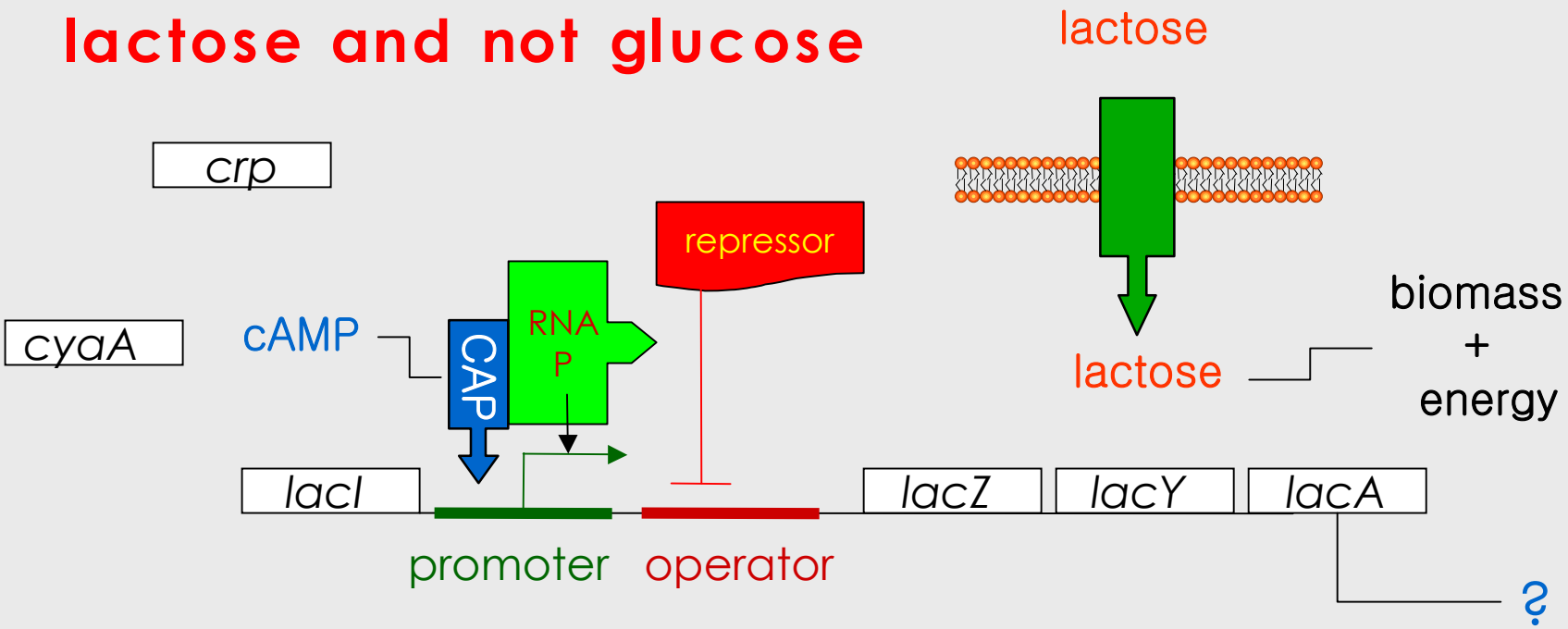
- evolving: sb combines design and evolution to use (poorly understood) principles that drive adaptation

reasoning as engineers do

. . . . . .

"the missing staircase"

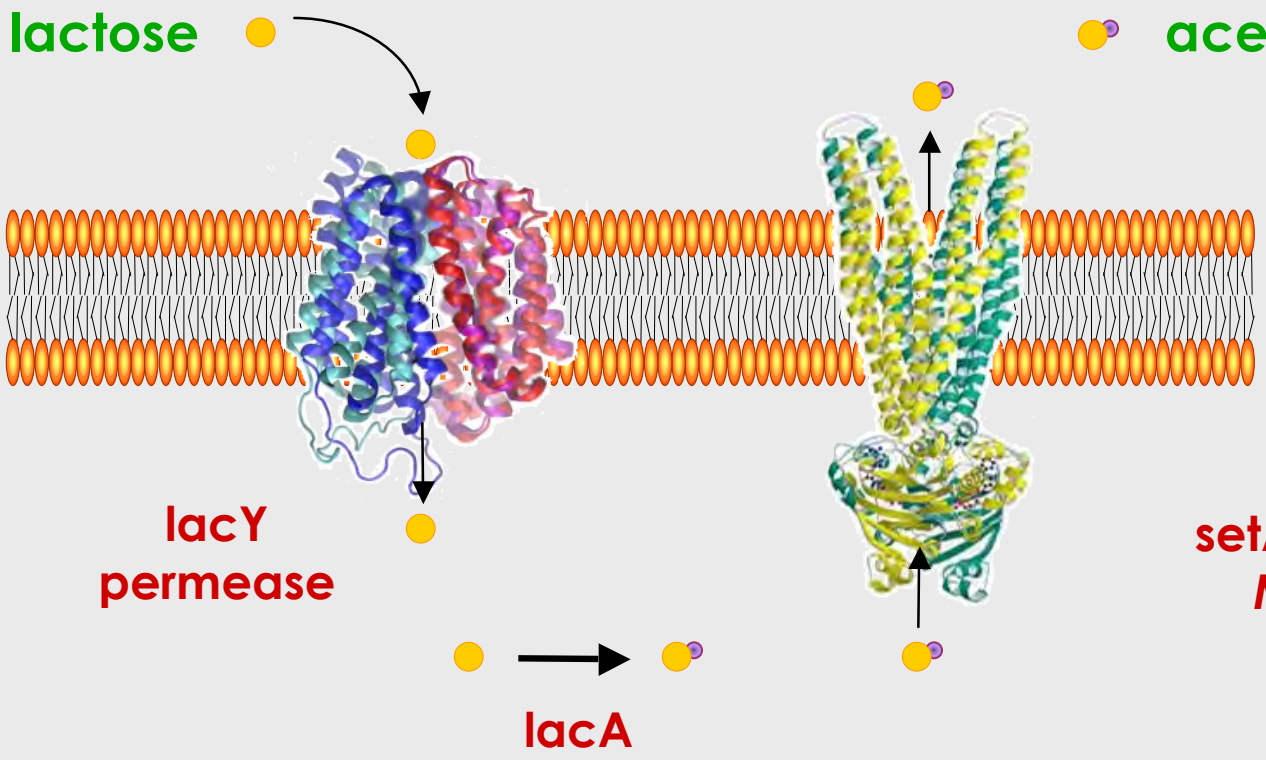
lactose and not glucose



what is the function of lacA?

why did we need 60 years to ask the question?

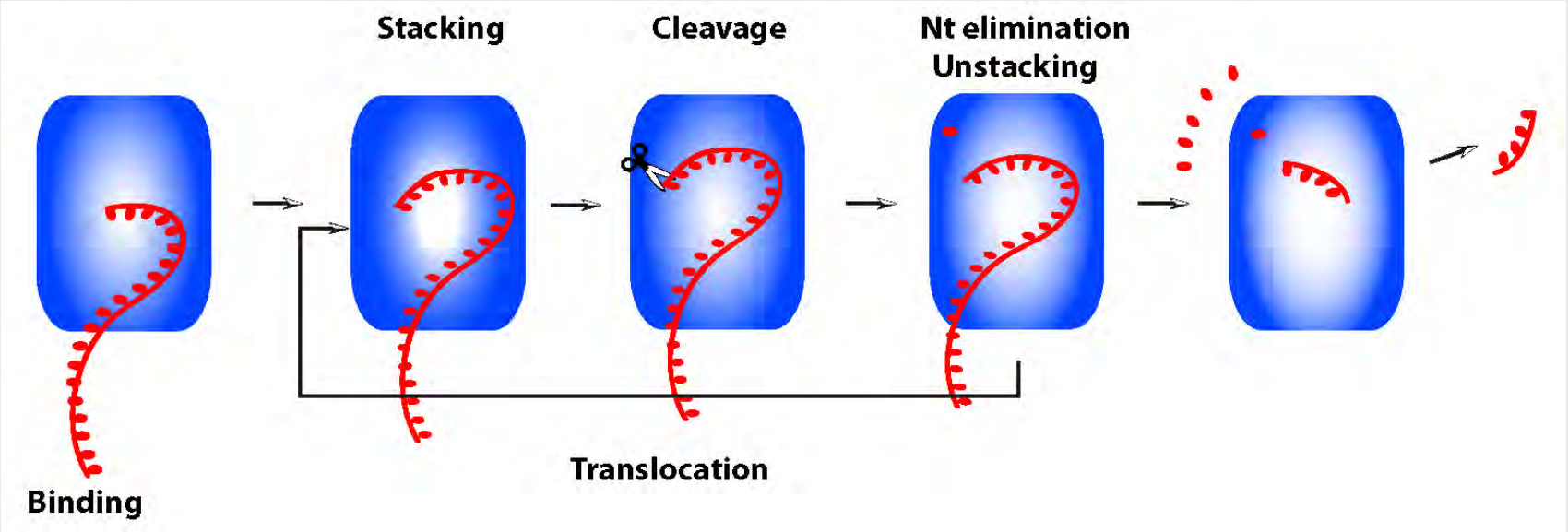
cells need safety valves, not a leaks



lactose acetyl-transferase (very inefficient K_M: 1 M)

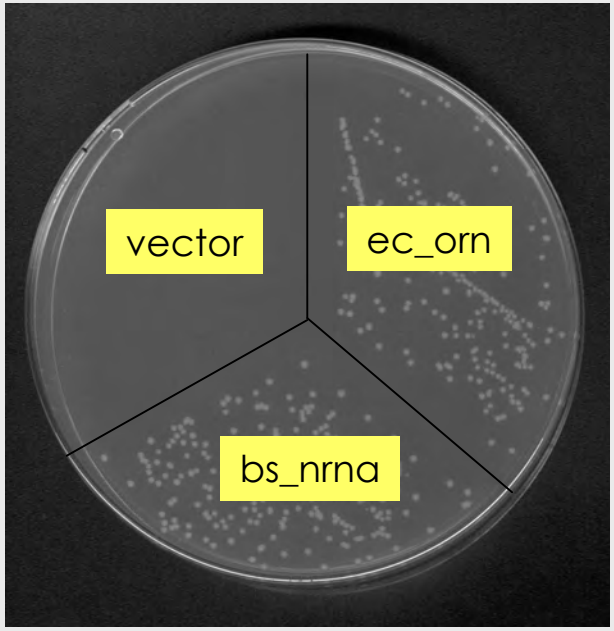
coping with leftovers

nanornase is an essential function



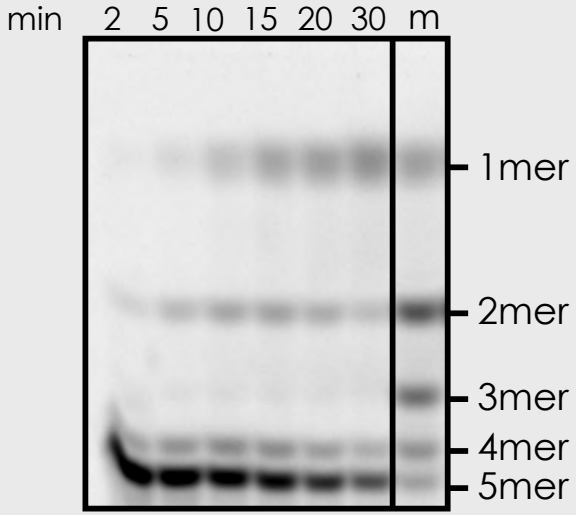
functional ubiquity is not structural ubiquity

in vivo



b. subtilis nrna complements
e. coli orn^{-}, no common origin

in vitro



degradation of nanorna 5mers
(cy5-ccccc-3')

an unlimited list of functions

- storage (location, address)
- coping with errors
  - metabolic interference (alpha-dicarbonyl)
  - misfolding
  - modifications: programmed or accidental
- robustness and promiscuity (functional leaks…)
- aggregation (what about crystals?)
- lubrication

e n g i n e e r i n g l i f e

computers making computers

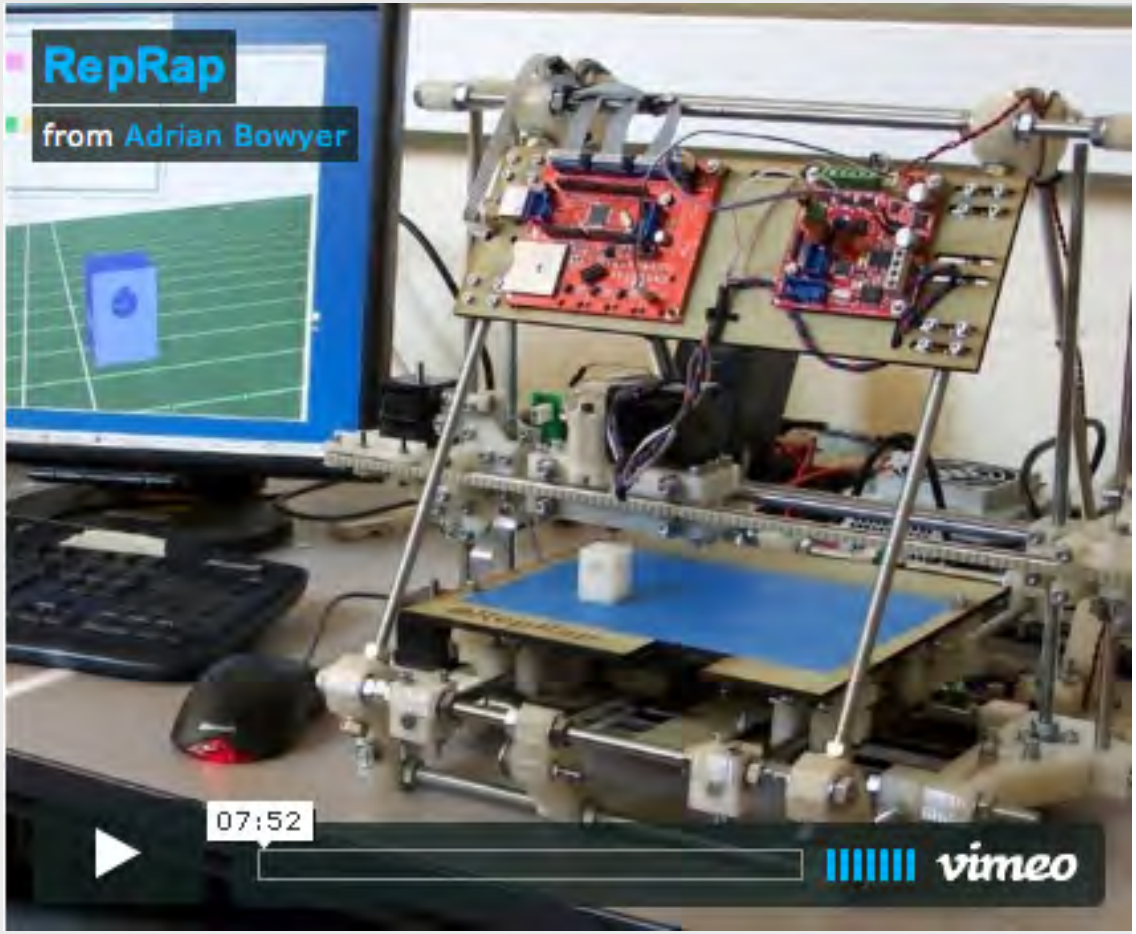
reprap (replicating rapid prototyper, 2004) aims at creating an auto - reproducing laser 3d printer:

the machine produces most of its components (= "biobricks")

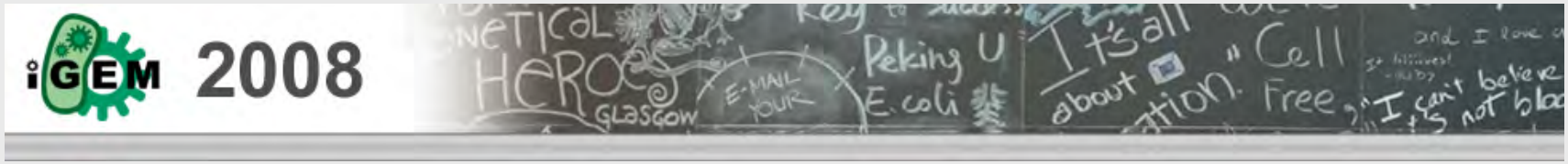
missing:

- the program
- the assembly line (management of time and space, and specific functions such as lubrication)

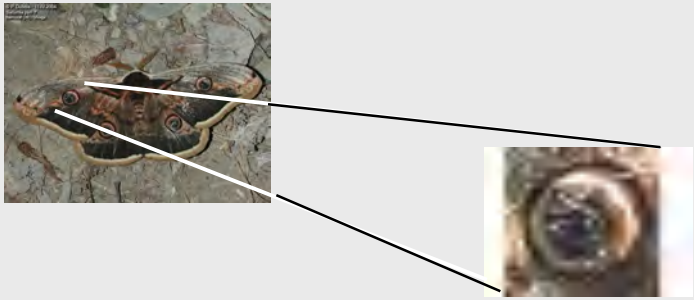
http://reprap.org/



mit : igem and its « biobricks »



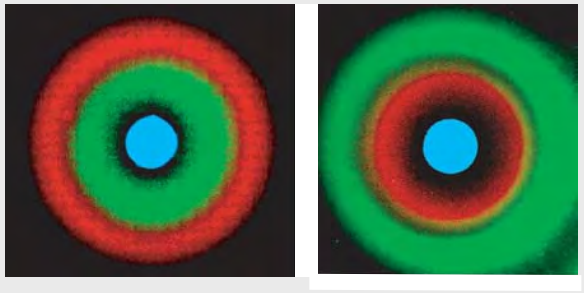
igem (international genetically engineered machines) **asks the question:**



can we construct construct biological systems from standardised elements, placed in living cells?

or

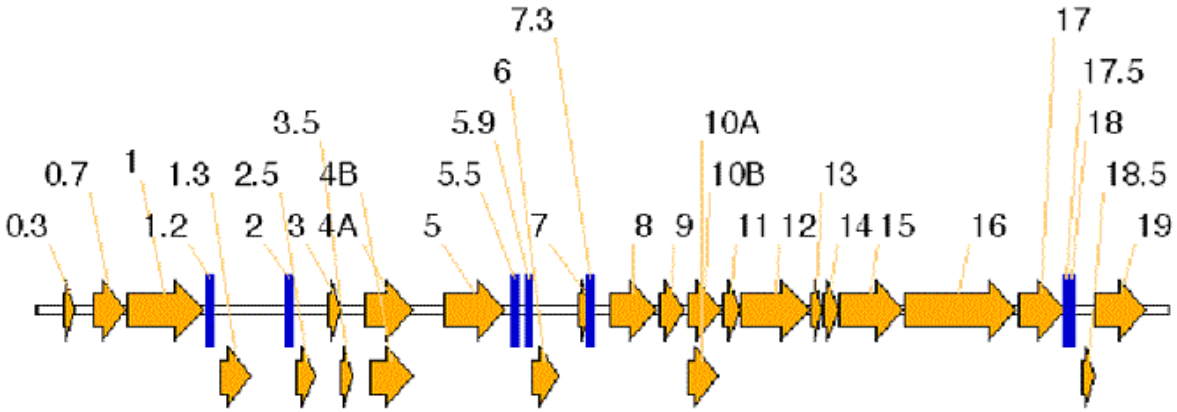
is biology too complicated to be reconstructed that way?



basu s, gerchman y, collins ch, arnold fh, weiss r.
a synthetic multicellular system for programmed pattern formation.
nature. 2005 434:1130-1134

reprogramming bacteriophage T7

known genes of bacteriophage T7



•taking control•

• destruction •

• replication •

• synthesis of the capsid •

•encapsidation•

• getting out of the cell lysis •

drew endy: http://online.kitp.ucsb.edu/online/infobio01/endy/

reprogramming bacteriophage t7

- control regions overlap
- they are not standardised
- they may be redesigned according to engineering rules, and tested using mathematical models
- models predict the synthetic phage behaviour and compare it with that of the natural phage
- the synthetic phage forms smaller lysis plaques than its natural counterpart
- the evolution the synthetic phage to more virulent forms erases the human construct…

w h a t l i f e i s

life requires:

a machine ("chassis") allowing the program to be expressed (reproduces)

1. metabolism (a dynamic process)
2. compartmentalisation (casings, defining inside and outside)

a program (a "book of recipes": replicated)

3. recursive information transfer and trapping => coding from one level to a second level introduces an essential asymmetry (**fundamentally different from feedback)**

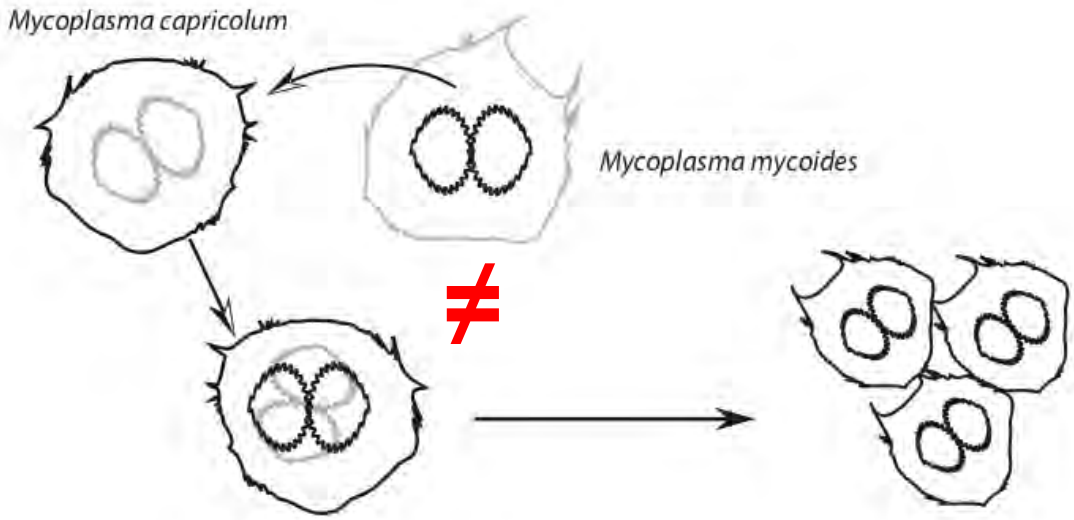
the cell is the atom of life

replication and reproduction are not the same

the program replicates (makes an identical copy)

the cell reproduces (makes a similar copy)

this split is the basis of evolution




spatial constraints

cells and computers

genetics rests on the description of genomes as texts written with an alphabet: but do cells behave as computers?

horizontal gene transfer

viruses

genetic engineering

transplantation of a naked genome in a recipient cell changing the host recipient into a new one (2007)

everything separates

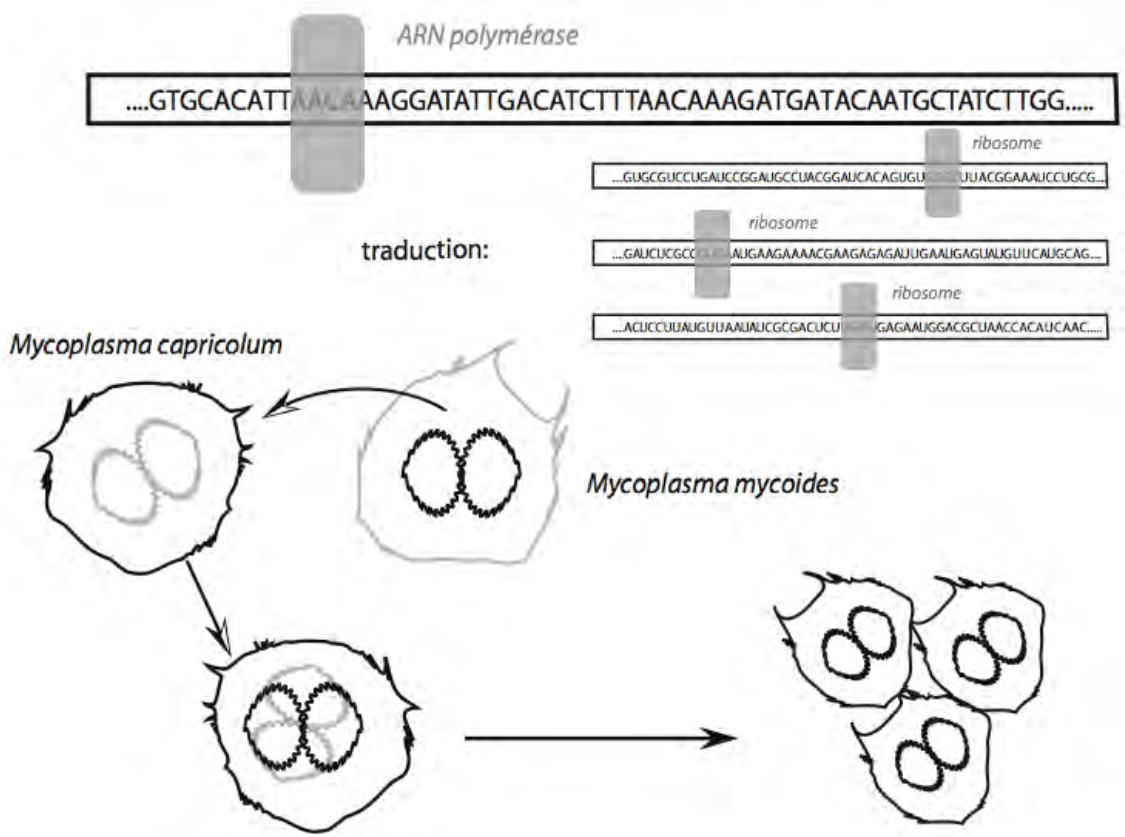
"machine" (chassis) and "data/program" (genome)

need for an operating system, and for constraints in the chassis

cells as computers

computers implement the turing machine concept, with a finite program

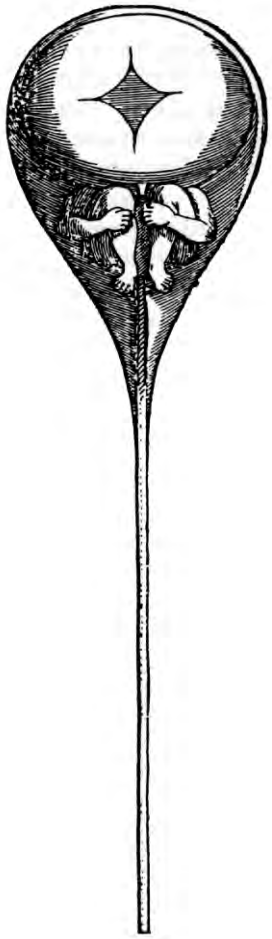
genome transplantation



genome transplantation in bacteria: changing one species to another
lartigue c, glass ji, alperovich n, pieper r, parmar pp, hutchison ca 3rd, smith ho, venter jc
science (2007) 317: 632-638

program and chassis

- combining bonnet's preformationism and harvey's epigenesis: a construction algorithm, not an organism, is replicated through generations
- the machine (chassis), which reproduces over generations, runs the algorithm; it is physically separated from the algorithm's support, as in computers
- if computers were to make computers, how would the aging problem be tackled?
- comparative analysis of genetic programs uncovers genes (maxwell's demon's genes) necessary to make young organisms from old ones; it presents living organisms as information traps



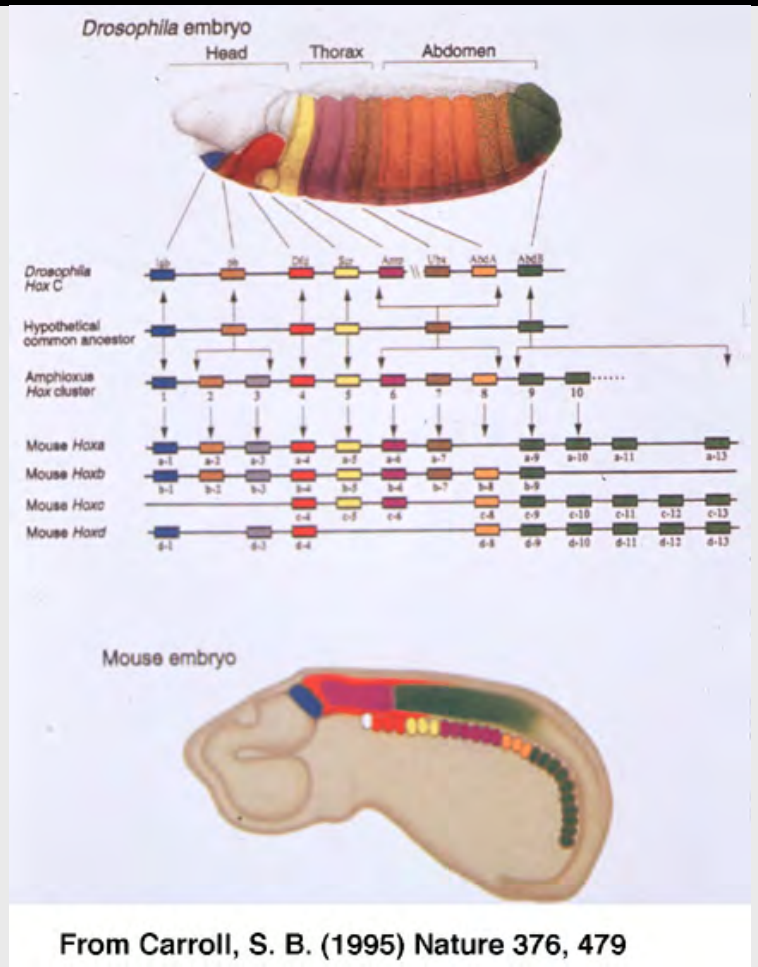
a standing enigma: homeogenes

drosophiloculus ?

homunculus ?

celluloculus ?

conservation of a functional hierarchy between mammalian and insect hox/hom genes.bachiller d, macias a, duboule d, morata g.embo j. 1994 13:1930-1941



From Carroll, S. B. (1995) Nature 376, 479

a research program

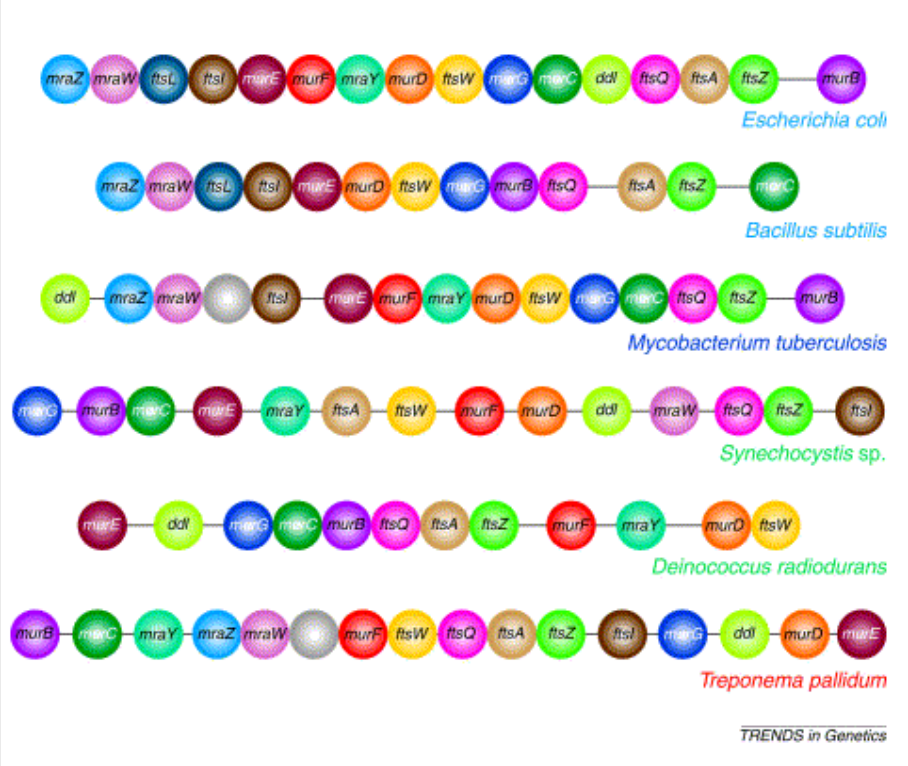
is there a map of the cell in the chromosome?

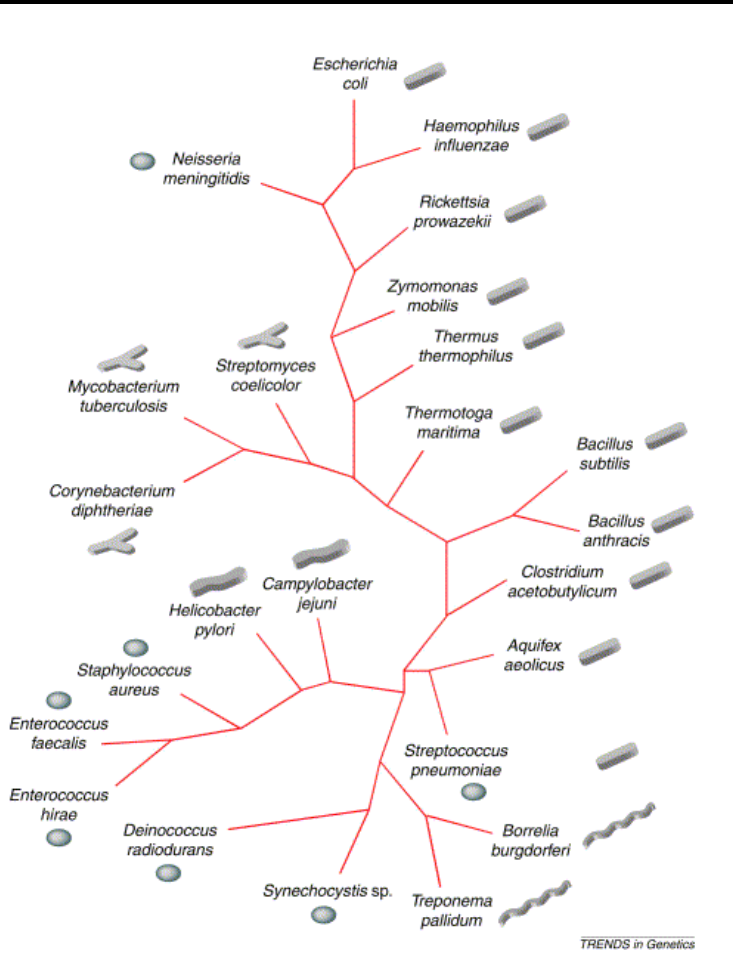
if the machine must both behave as a computer and construct the machine itself, one must find an image of the machine somewhere in the machine (john von neumann)

this idea is the root of the concept of « operating system »

gene order and cell shape

mur-fts islands



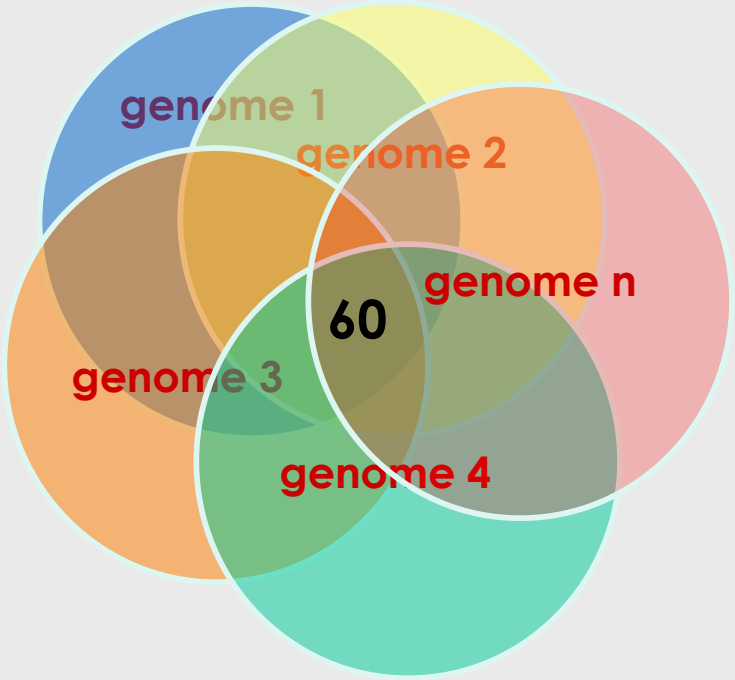


tamames j, gonzalez-moreno m, mingorance j, valencia a, vicente m
bringing gene order into bacterial shape
trends in genetics (2001) **17**: 124-126

a minimal set of functions

. . . . . .

2003: 60 conserved proteins



the number of conserved genes tends to zero!

looking for ubiquitous functions

variation / selection / amplification

↳ stabilisation ↱

evolution

↓ *creates (here comes information)*

function

↓ *traps ("recruits")*

structure

↕ *encodes*

sequence

functional ubiquity does not imply structural ubiquity
(think about transparency; lens, jellyfish, fish…)

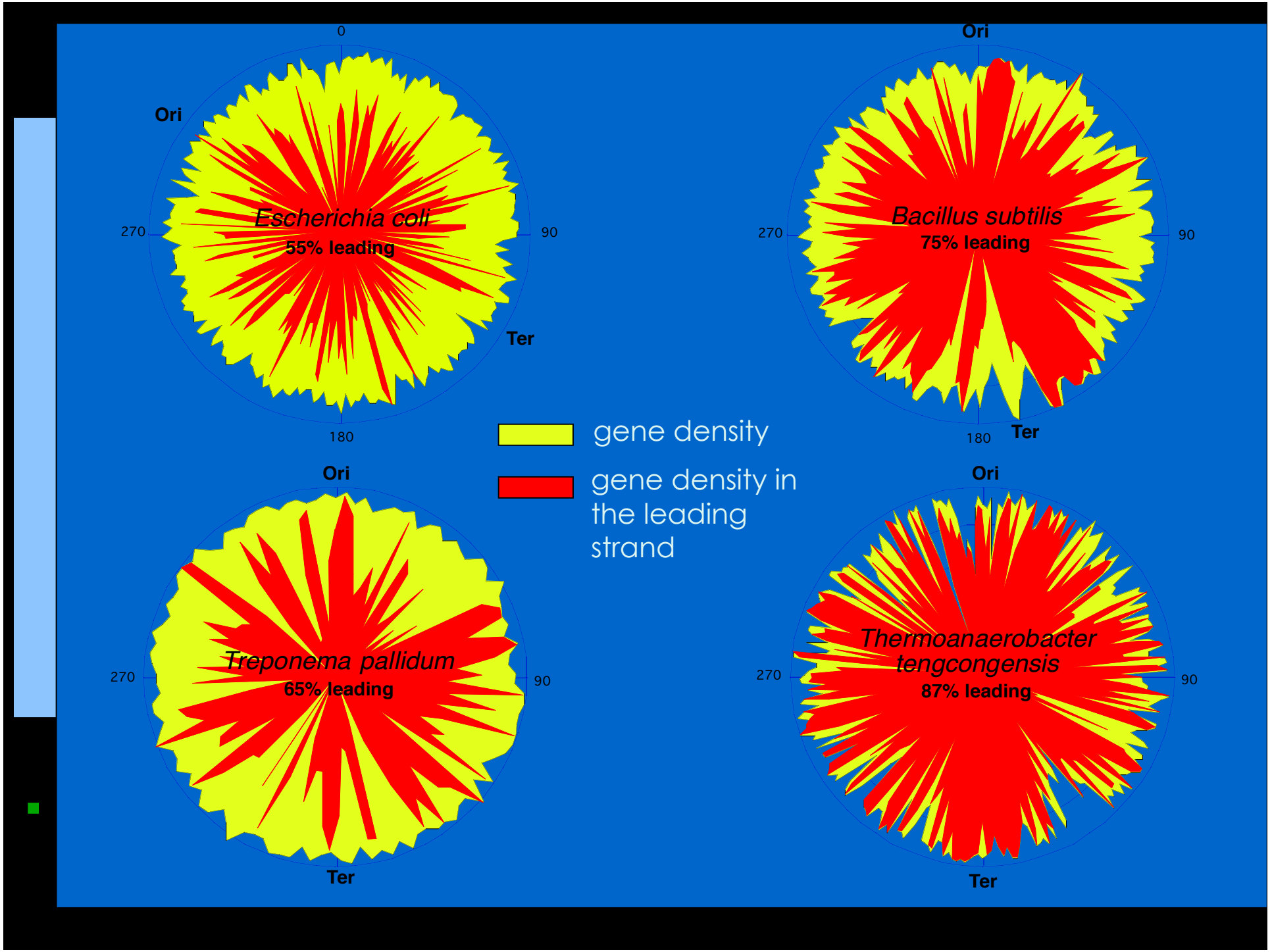
from functional ubiquity to gene persistence

functional gene ubiquity does not imply gene ubiquity

yet, efficient entities tend to persist through generations

→ looking for « persistence » identifies most ubiquitous functions

~ 500 genes persist in bacterial genomes; they are involved not only in the three processes required for life but also in maintenance, adaptation to transient phenomena and evolution of the organism, via energy-dependent degradative processes


0
Ori
270
90
180
Ter
Escherichia coli
55% leading
Ori
270
90
180
Ter
Bacillus subtilis
75% leading
gene density
gene density in the leading strand
Ori
270
90
Ter
Treponema pallidum
65% leading
Ori
270
90
Ter
Thermoanaerobacter tengcongensis
87% leading

multivariate analyses

multivariate analyses try to extract information by reducing as much as possible the number of descriptors of the objects of interest

laplace-gauss statistics

principal component analysis uses the centered average and a simple distance (identity); it is the reference method

correspondence analysis belongs to the same family, but it uses the χ^2 measure as a distance (benzécri, 1965)

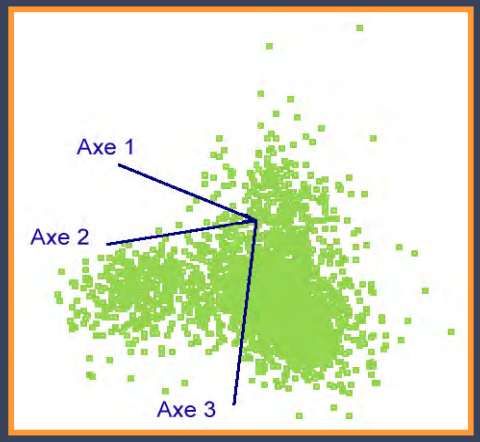
absence of normality (or log-normality)

independent component analysis uses the non gaussian character of the values associated to descriptors; it characterizes objects belonging to common independent

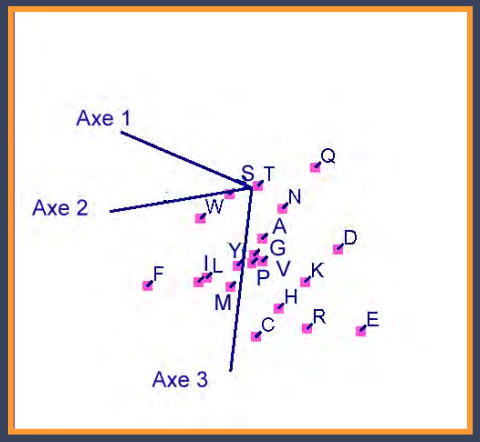
further methods need to be developed....

correspondence analysis (ca)

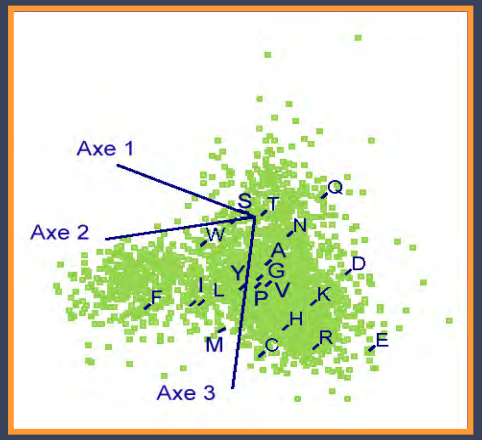
factorial space of the proteins



factorial space of the amino acids



superimposition of both spaces (clouds)



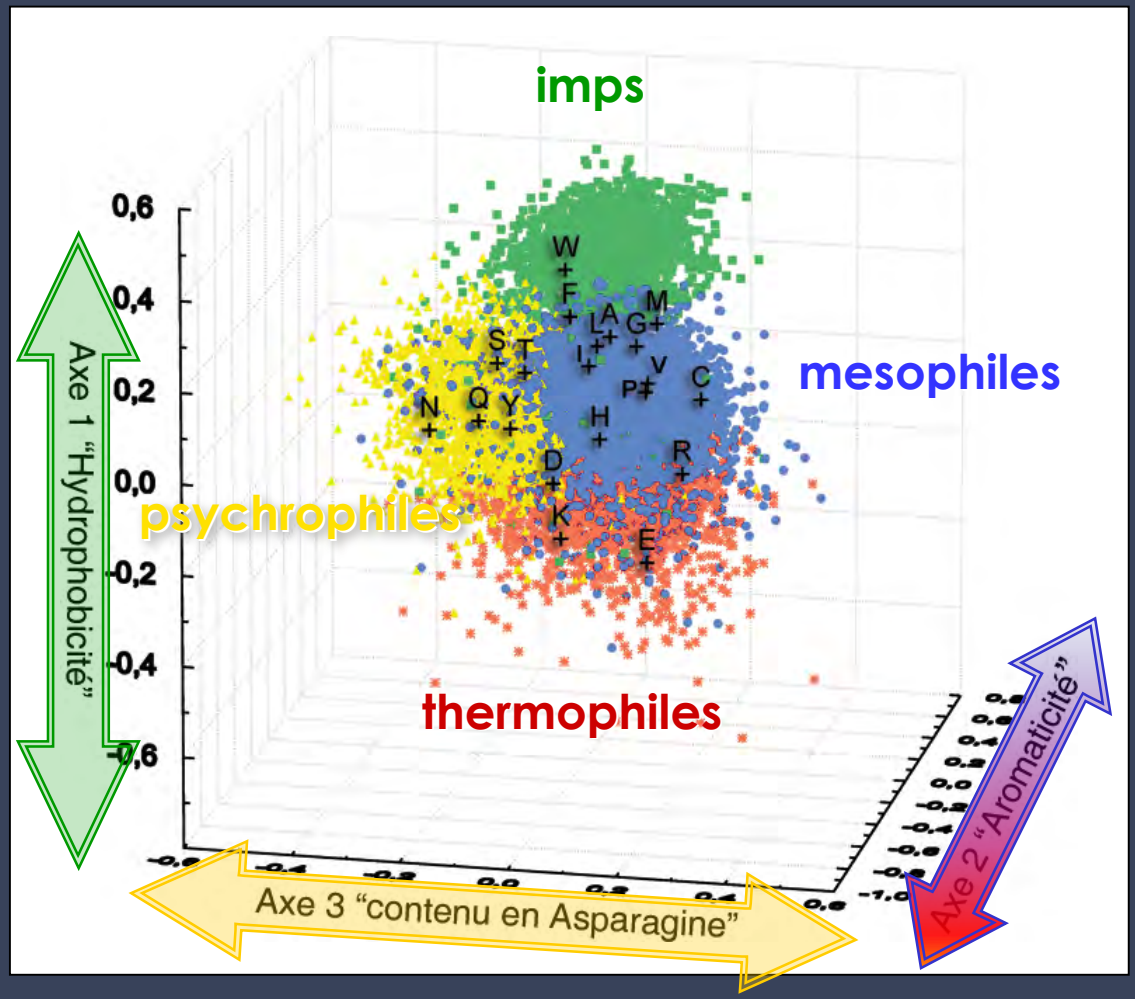
universal biases in protein amino acid composition

- **first axis:** separates integral membrane proteins (imp) from the rest; driven by opposition between charged and large hydrophobic residues
- **second axis:** separates proteins by their content in aromatic amino acids; enriched in orphan proteins
- **third axis:** separates proteins according to an opposition driven by the g+c content of the *first* codon base

comparative proteomics

a specific asparagine bias in psychrophiles

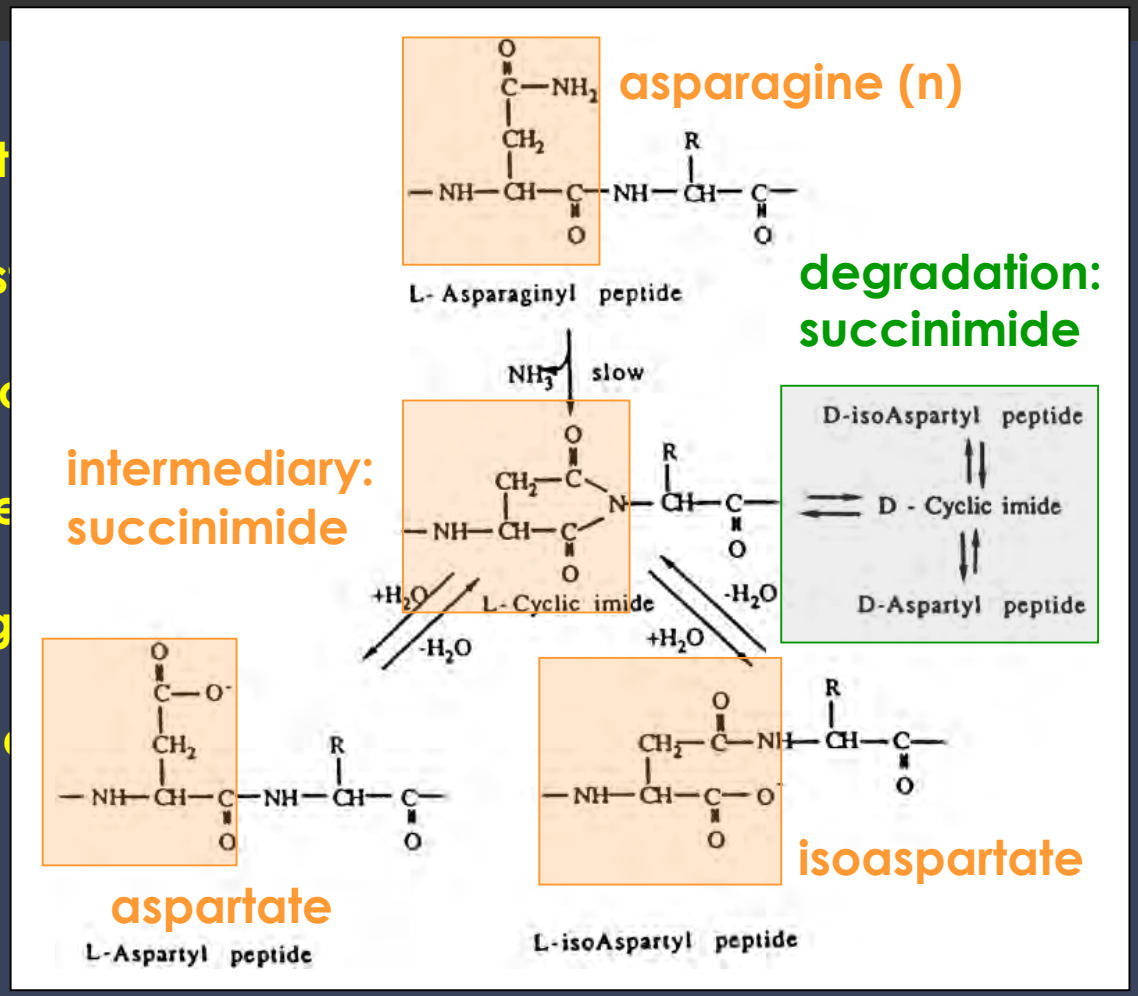
- motility
- cell wall, outer membrane
- transport (tonb), secretion
- adaptation to stress
- metabolism of dna and rna



a chemical anecdote

asparagine deamidates: a major contribution to protein aging

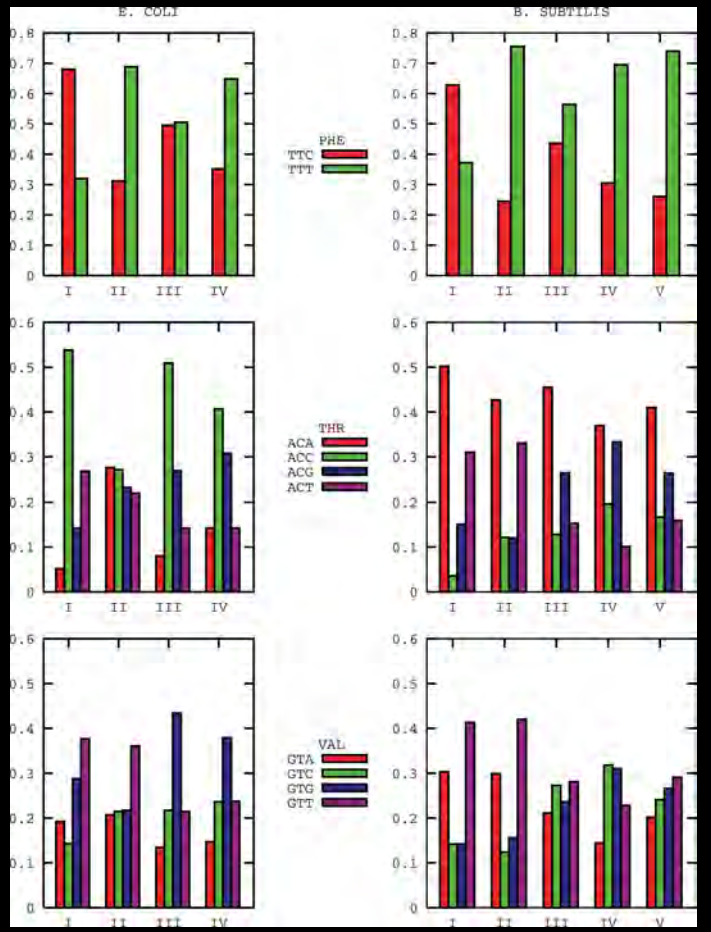
- main post
- reaction s
- spontaneo
- affects the
- role in reg
- signal for



genomic translation islands

genes with similar bias are organized into groups longer than operons, showing some translation-driven organization of the chromosome

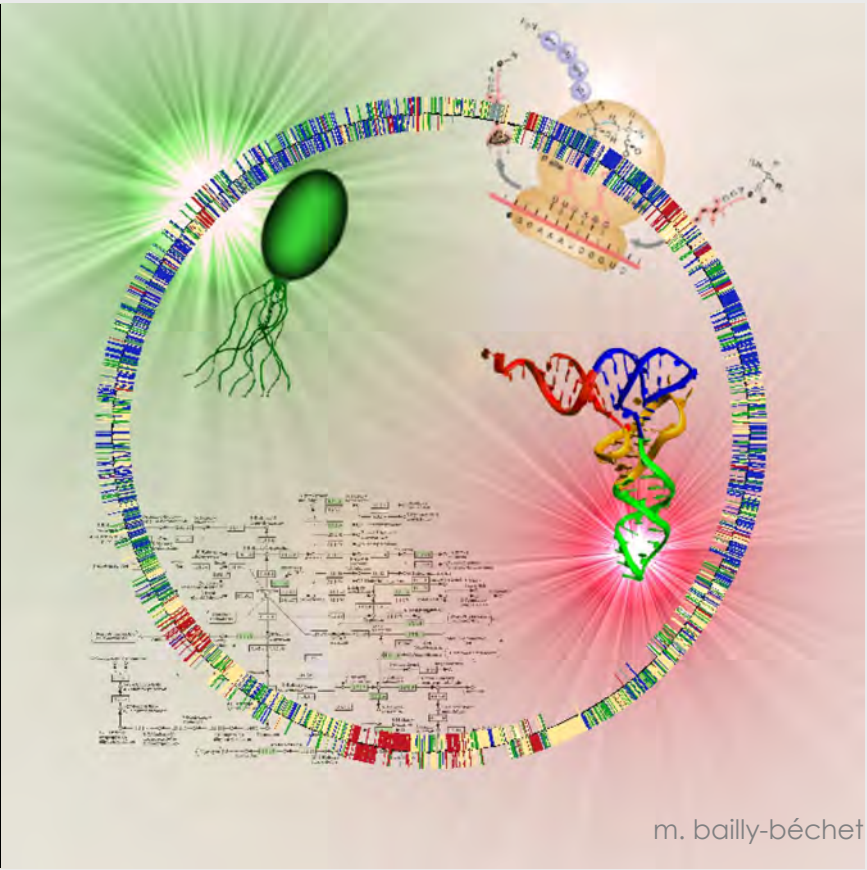
a major part of this effect comes from the recycling or rare transfer rna molecules. it is essential to understand that individual molecules (not concentration!) are important in the cell



translation islands

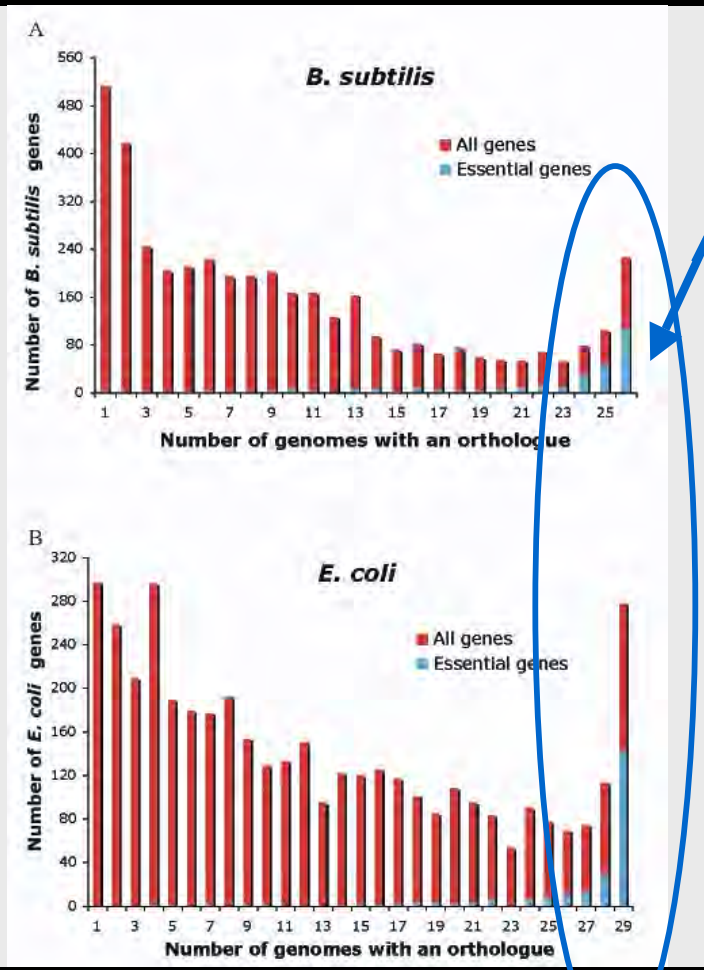
one groups is associated to high expression (blue)

the other groups are also functionnally consistent: horizontally transferred genes (red) motility (yellow) and intermediary metabolism (green).



m bailly-bechet, a danchin, m iqbal, m marsili, m vergassola
codon usage domains over bacterial chromosomes
plos computational biology (2006) **2**: e37

persistence: too many genes!



persistent genes

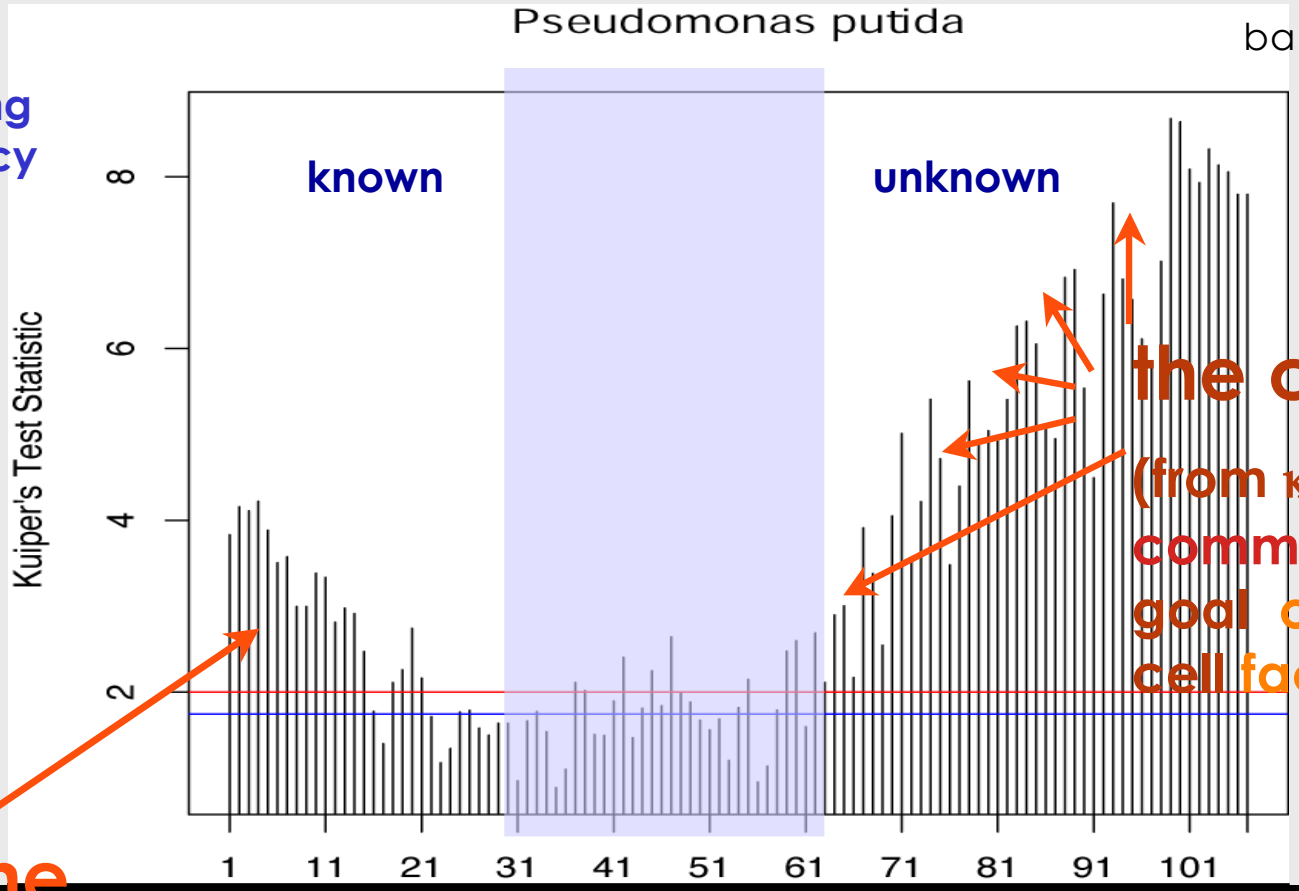
essential genes and ....

energy-dependent degradation

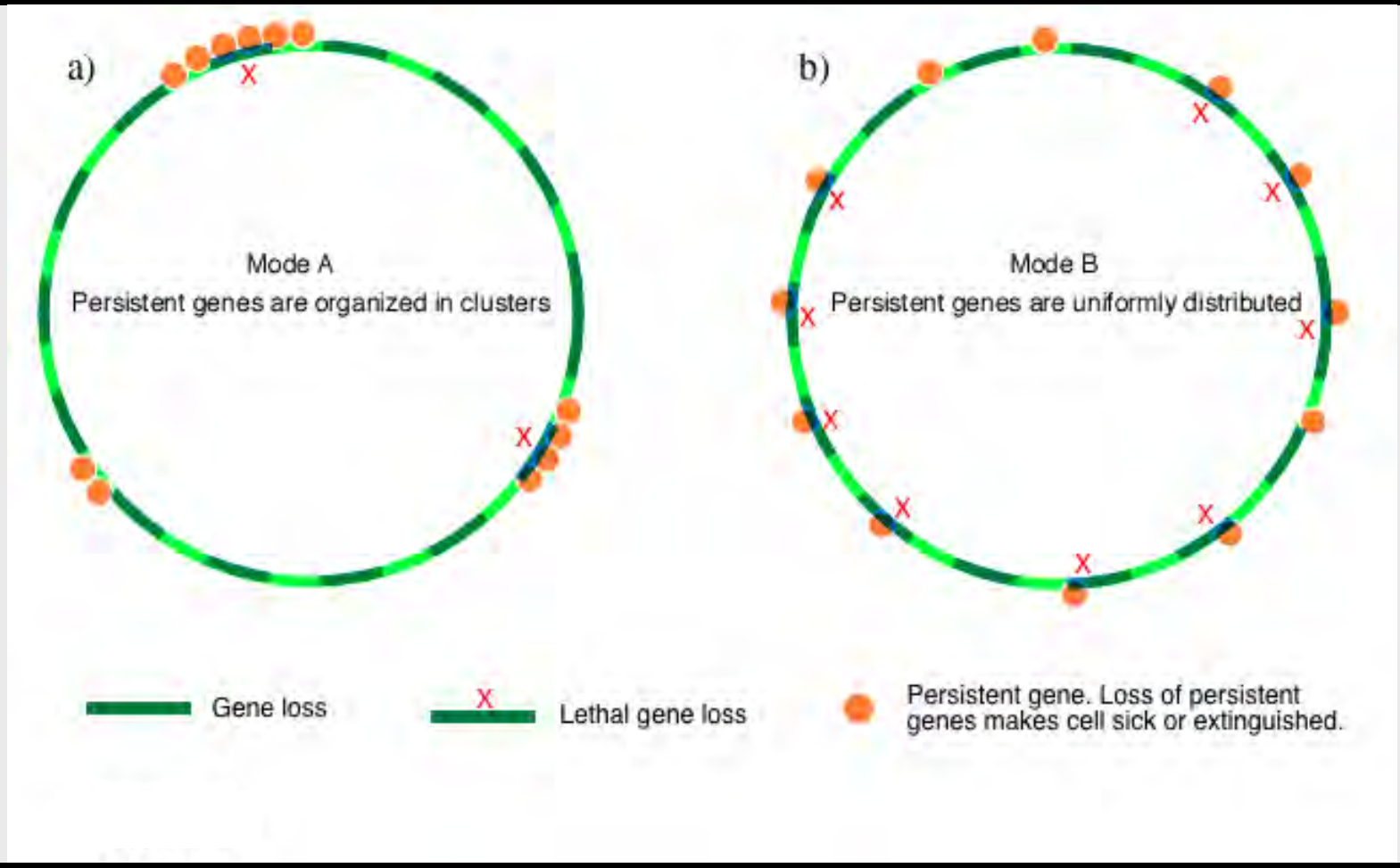
metabolic patches

G Fang, EP Rocha, A Danchin
How essential are non-essential genes?
mol Biol Evol (2005) 22: 2147-2156

syntenies of orthologs

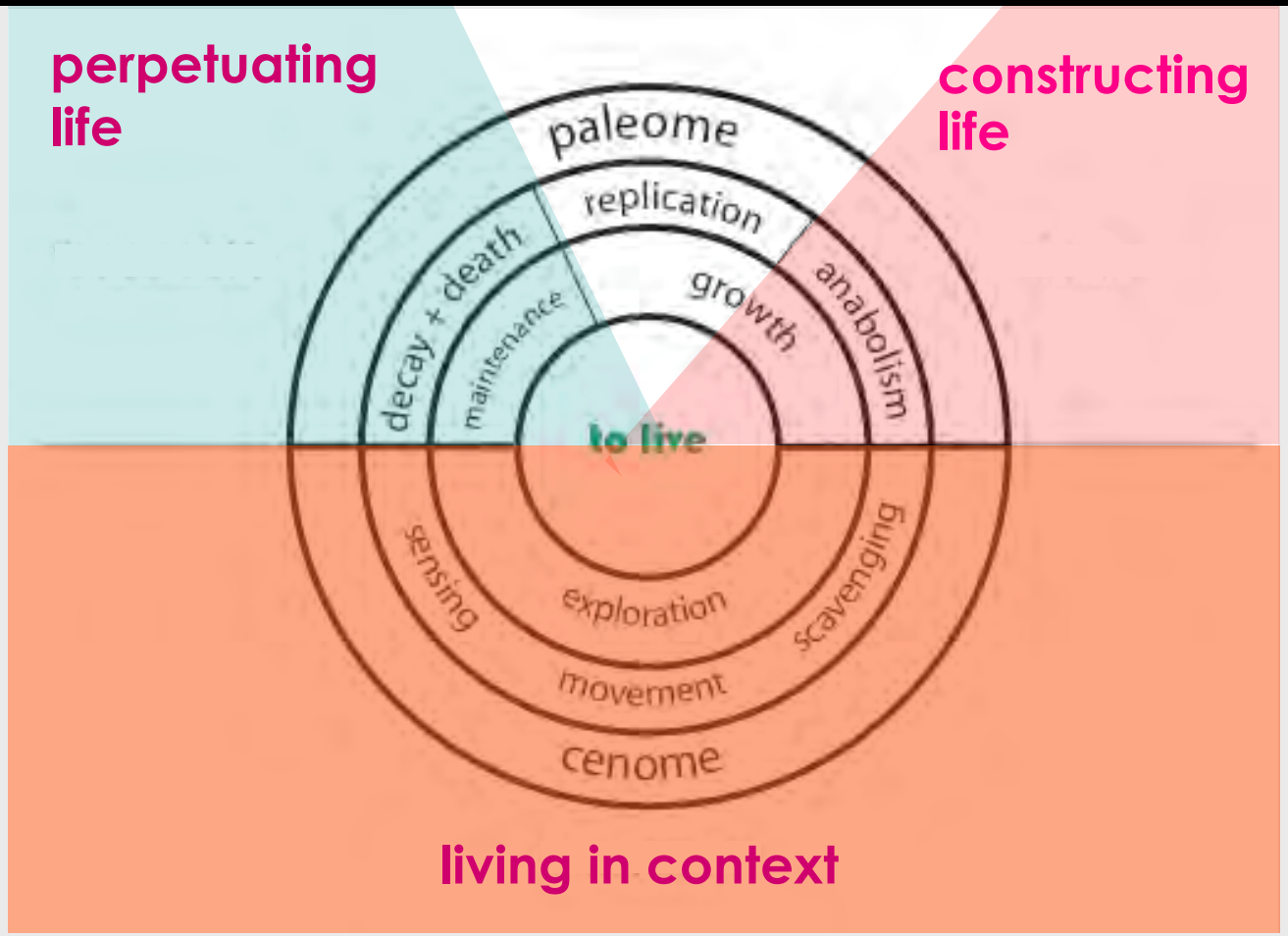


existence implies clustering



fang g, rocha ep, danchin a.
persistence drives gene clustering in bacterial genomes.
bmc genomics. (2008) **9**:4.

a tale of two genomes



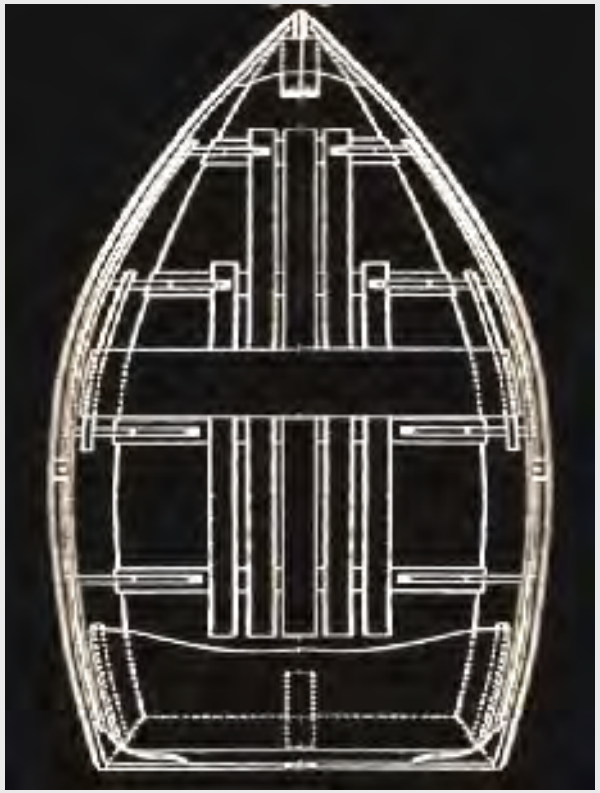
organised genome dynamics in the *escherichia coli* species results in highly diverse adaptive paths
touchon m, hoede c, tenaillon o, barbe v, ..., medigue c, rocha ep, denamur e.
plos genet. 2009 jan;5:e1000344

maxwell's demon's genes

.....

the delphic boat

- biology is a science of relationships between objects
- it is symplectic (συν together, πλεκτειν, to weave), same word as « complex » in latin;
- an information expresses what is conserved in the boat, not the matter of its planks !



A. Danchin The Delphic Boat, Harvard University Press, 2003
la barque de Delphes, Odile Jacob, 1998

V. de Lorenzo, A. Danchin synthetic Biology: discovering new worlds and new words 9: 822-827. EMBO reports, 2008

exploring information: infotaxis

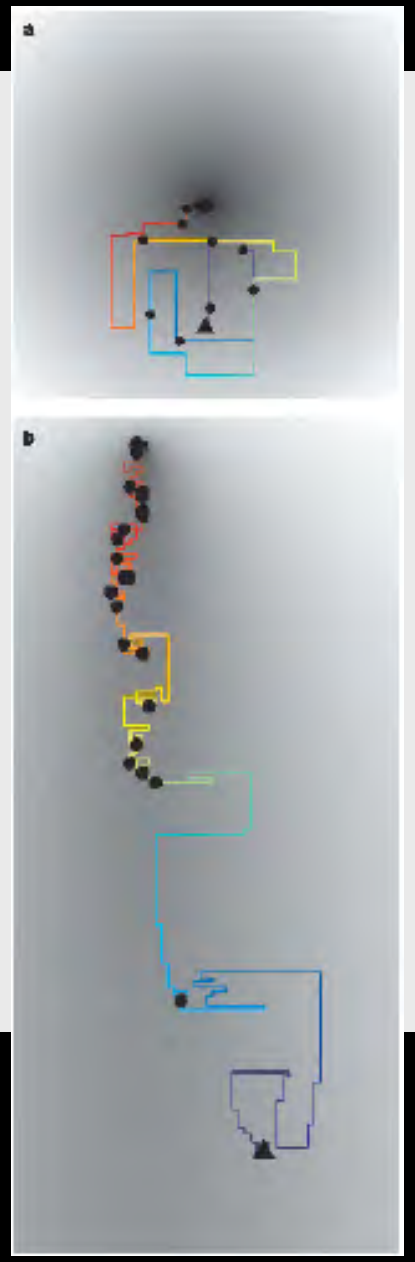
large peacock *saturnia pyri*
http://pdubois.free.fr/

how can a moth find a partner 1,000 meters away?

to climb a chemical gradient is impossible (air turbulence, obstacles…)

vergassola and co-workers have shown that maximising the collection of information allows the animal to achieve this goal…

'Infotaxis' as a strategy for searching without gradients
Vergassola M, Villermaux E, Shraiman BI
Nature (2007) 445: 406-409



a new heuristics

matter / energy / space / time

- classic physics
- quantum physics
- chemistry
- biology
  - development
  - neurobiology
  - linguistics
- mathematics (informatics)

information as a new currency of reality

- 1929 leo szilard (wrong) intuition of the link between energy and information: creation of 1 bit requires kTlog2, while analysing maxwell's demon
- 1949 claude shannon theory of communication
- **1**961 rolf landauer proof of computation reversibility (no energy is required for creation of information); energy is required for erasing memory
- ~1974 andrey kolmogorov, gregory chaitin, ray solomonoff define algorithmic complexity
- 1988 charles bennett defines logical depth (links time and algorithmic complexity) to define value of information
- 1989 wojciech zurek links algorithmic complexity and energy, reflecting on maxwell's demon
- 2007 scott muller defines information as any attribute that helps determine the state of a system, via asymmetry
- 2009 takahiro sagawa and masahito ueda reassess landauer's theorem

a bridge with matter and energy

- information does not derive from matter
  - orthogonal synthetic biology
- information does not derive from energy
  - creation of information is reversible
  - accumulation of information requires energy
- information bridges digital and analogic engineering
- information bridges the cell and its genetic program

two types of information

standard information: carries its own forces along with it (e.g. information in DNA replication)

contextual information: the presence of a cell is necessary to build up a cell

it is in this second case that we can best see information as a category distinct from mass or energy; it is similar to something like « situation »; the theory does not exist yet…

information of the machine

"beside the genetic program, the cell carries a considerable amount of information…"

true: but in a computer as well

this requires construction of an entirely novel theory of "machine-information"

revisiting information

living organisms are information gathering and using systems (igus) that aim at maintaining their activity by building a *record* of the relevant measurements they have performed (zurek, 1989) => genetic and epigenetic heredity

to say that a system occupies a certain state implies that one has the information necessary to generate a complete description of that state: the information gathering process is reversible (i.e. does not use energy) provided it is allowed to save a copy of the input (landauer, 1961)

landauer's theorem: to erase a bit of information in an environment at temperature T requires dissipation of energy $\geq k_B T \ln 2$; this demands that information be granted a physical status as a negative contribution to free energy

$$\mathcal{F} = F - K_B T \ln 2 \, I = E - k_B T \ln 2 \, (H+I)$$

value of information

the information of the program is transmitted "as is" during replication, with no value associated to particular sequences: where does the information of the machine (and of the environment) come from?

to accumulate information requires an energy-dependent processs to "make room", without erasing valuable information

can we imagine the genes of a maxwell's demon which would select among what is functional or young (locally) and what does not work?

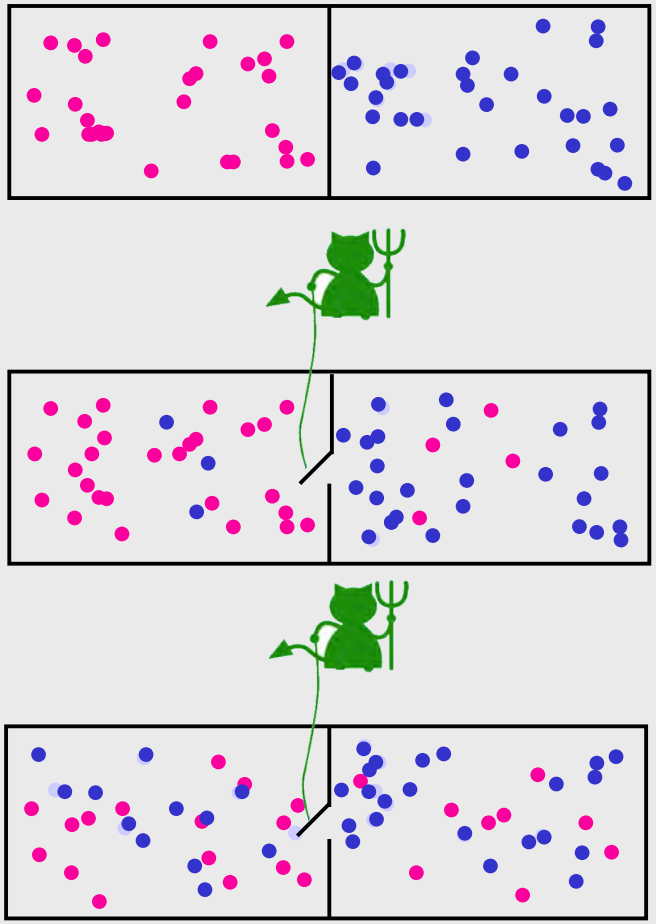
maxwell's demon

hot

cold

0

t



0

t

the demon reverses time while measuring the speed of the atoms of gas, recording an information to calculate when it must close the trap, thus permitting temperature-dependent generation of energy

maxwell's demon's genes

ATP
poly(P)$_n$

ADP + Pi
poly(P)$_{n-1}$ + Pi **<= in the paleome**

the degradation machinery uses energy to reject unaltered a functional entity

non functional entities are recognised and degraded

innovation: adaptive mutations

- energy-dependent accumulation of information is blind; it ignores the source of information
- information can come from a memory, that of the pre-existing genome; it can also be created de novo

- adaptive mutations are de novo creations of information; therefore they dependent on genes involved in accumulation of information

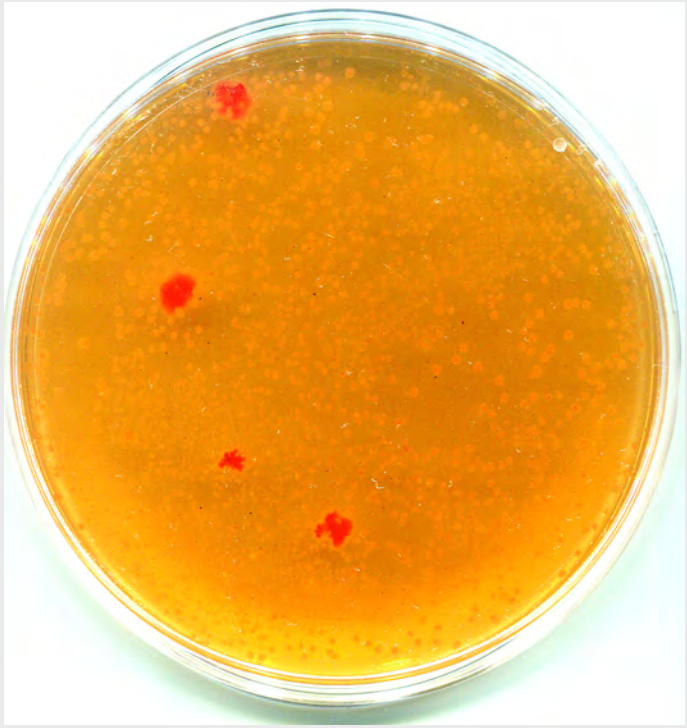
adaptive mutations

construction of "intelligent" bacteria

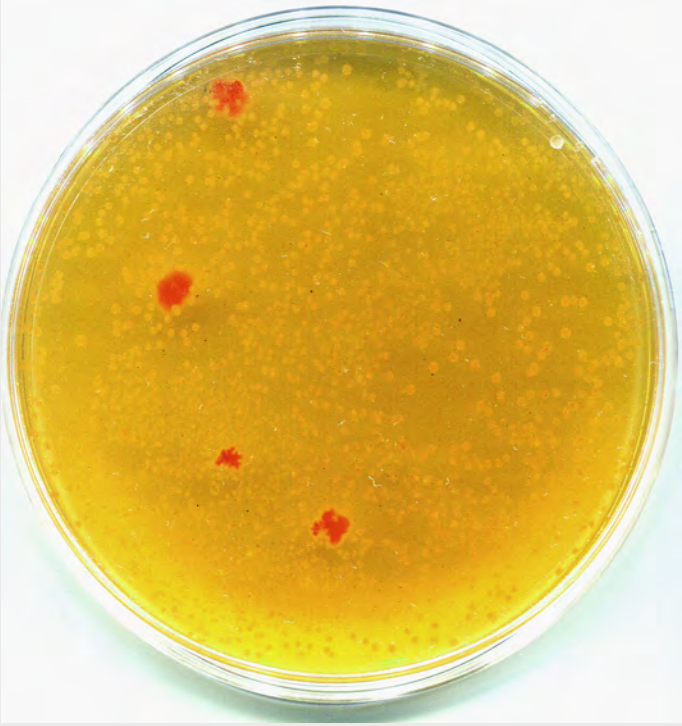
placed to grow on a medium with limited nutrient supply; form colonies of approximately 10^7 bacteria; the medium also contains nutrients that they cannot use

after a few days/weeks time, papillae appear that begin to grow and invade the plate, using supplied "unusable" nutrients; they derive from adaptive mutations

they did not pre-exist, and this supposes creation and recording of information



adaptive mutations



sequencing seven genomes + 30 pcrs

the total number of mutations is higher in older colonies

mutations are spread throughout the chromosome, and concentrated in one gene => pcr of many colonies

in this particular gene one finds different mutations in different papillae, 2 mutations in 30% of the cases

in some cases one of the two mutations is silent

on a particular carbon source, there is a least one other gene involved

natural selection is a principle of physics

- **natural selection**: making room using energy to avoid erasing context-dependent functional information
- energy-dependent degradative processes make room for newly synthesised entities; energy is consumed to **prevent** degradation of functional entities
- this process accumulates information, whatever its origin, in a ratchet-like process
- this process is **myopic**: it cannot have a design, hence the "tinkering" feature of life and its evolution

a synthetic cell?

- the engineering view of sb precludes that artificial cells be innovative
- we can exclude the genes permetting accumulation of information
- the consequence is that the cell factory will age and will need to be systematically rebuilt
- this has a in-built societal benefit, as risks are minimised
- but this poses problems when applications require that industrial processes are scaled-up: this may not be possible, unless we can harness the function of the maxwell's demon's genes to the human goals

contributions

in silico

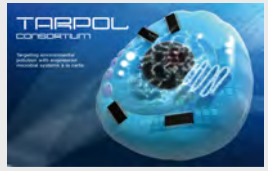
gang fang, eduardo rocha

in vivo

undine mechold, agnieszka sekowska

collaborations

genoscope, beijing genome institute, fudan university, the university of hong kong, hong kong university of science and technology



a novel publication concept

symplectic biology

rapid research notes in systems and synthetic biology

Home | About Us | Copyright | 1st Max Planck Society - Chinese Academy of Science Exploratory Round Table Conference (ERTC) on Synth

Log in

News

International Symposium on Synthetic Biology, october 18-19, 2010, Orchard Hotel, Singapore
Jul 10, 2010

Synthetic Biology Copenhagen 2010 International Workshop, August 25-27, 2010, Nano-Science Center, University of Copenhagen
Jul 03, 2010

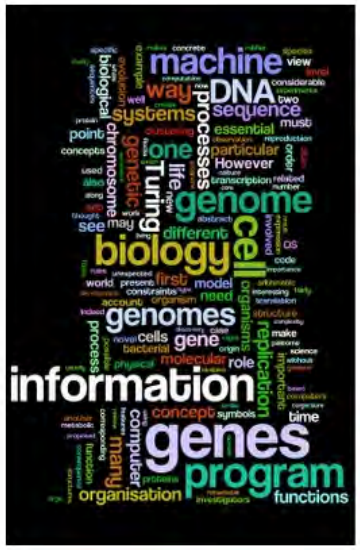
EC-US task Force on Biotechnology Research, Workshop on Standards in Synthetic Biology, 4-6 june 2010 - Spain
May 10, 2010

Workshop on Synthetic Biology, Denmark, August 25-27, 2010
May 10, 2010

More news...

Welcome to Symplectic Biology

A peer-reviewed journal for innovative publications in Systems and Synthetic Biology.



Featured Articles

The "missing" experiment showing separation of the program from the chassis has been performed by Craig Venter and his co-workers

Challenges

Challenge n°1: Is the synteny around the pyrH gene in bacterial genomes significant?

Aims and Scope

Symplectic Biology is a peer reviewed journal fostering the integration of Synthetic Biology with the more traditional Systems Biology. It aims at rapid publication of novel experiments and concepts. As its name indicates (« symplectic » is the greek equivalent of the latin « complex » without its fuzzy connotations) it endeavours to integrate physics, chemistry, information sciences and other mathematics-based disciplines into a rapidly accessible network of experiments and models permitting to combine in a challenging way the various aspects of what is traditionally named the complexity of living phenomena. Its primary aim is to promote construction of synthetic life via the quantitative characterization and

Search

Search Site | Search

Advanced Search...

Board of Editors

Editors in chief

Antoine Danchin
Victor de Lorenzo

Editors

Sigal Ben-Yehuda
Philippe M. Binder
Jesús Blázquez
Nediljko Budisa
Jason W. Chin
Gang Fang
Jian-Dong Huang
Jose Ignacio Jimenez
Frédérique Lisacek
Philippe Marlière
Kenta Nakai
Christos Ouzounis
Sven Panke
Manuel Porcar
Alfonso Valencia
Anil Wipat
J. Tze-Fei Wong
Huanming Yang

http://knol.google.com/k/symplectic-biology#