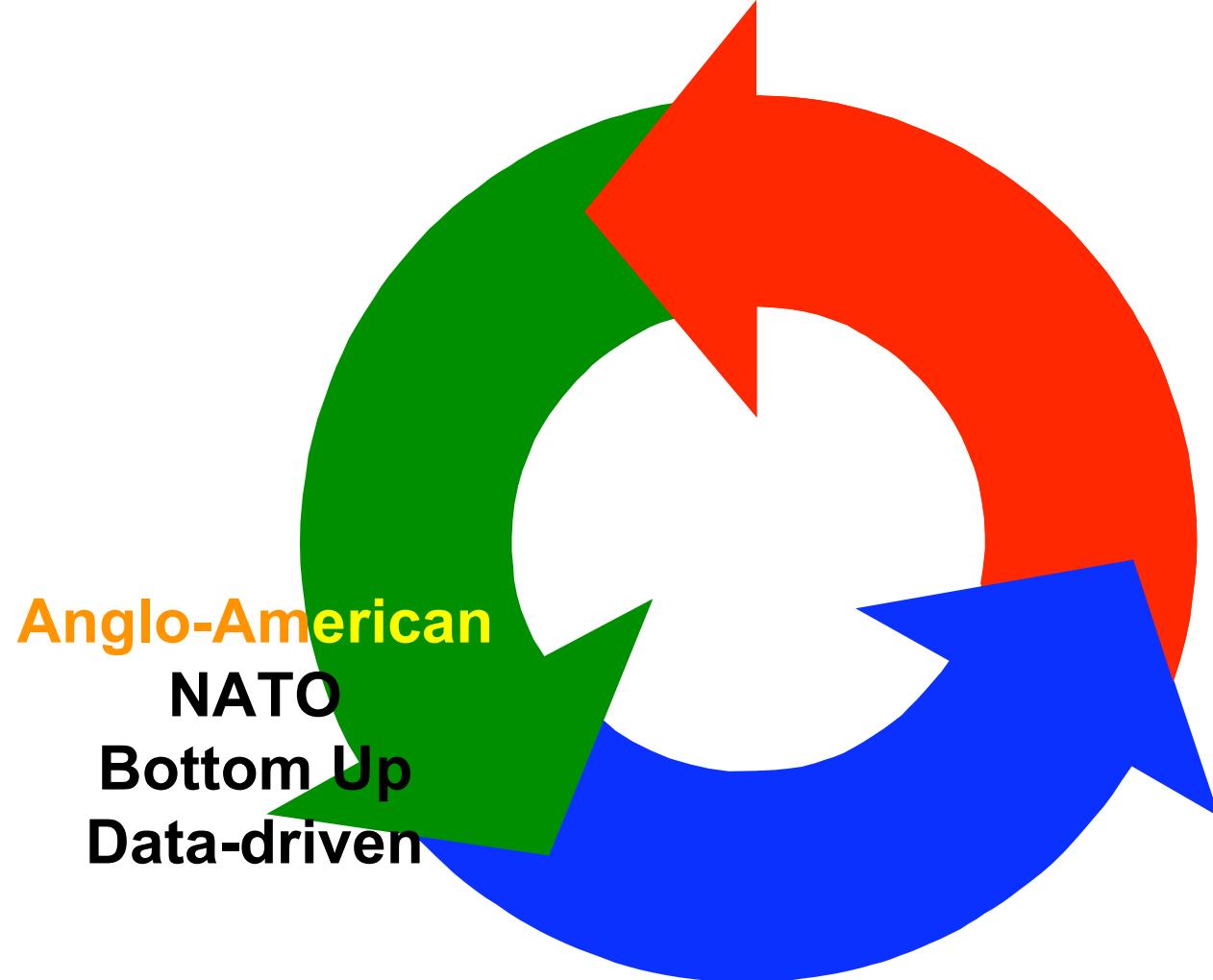
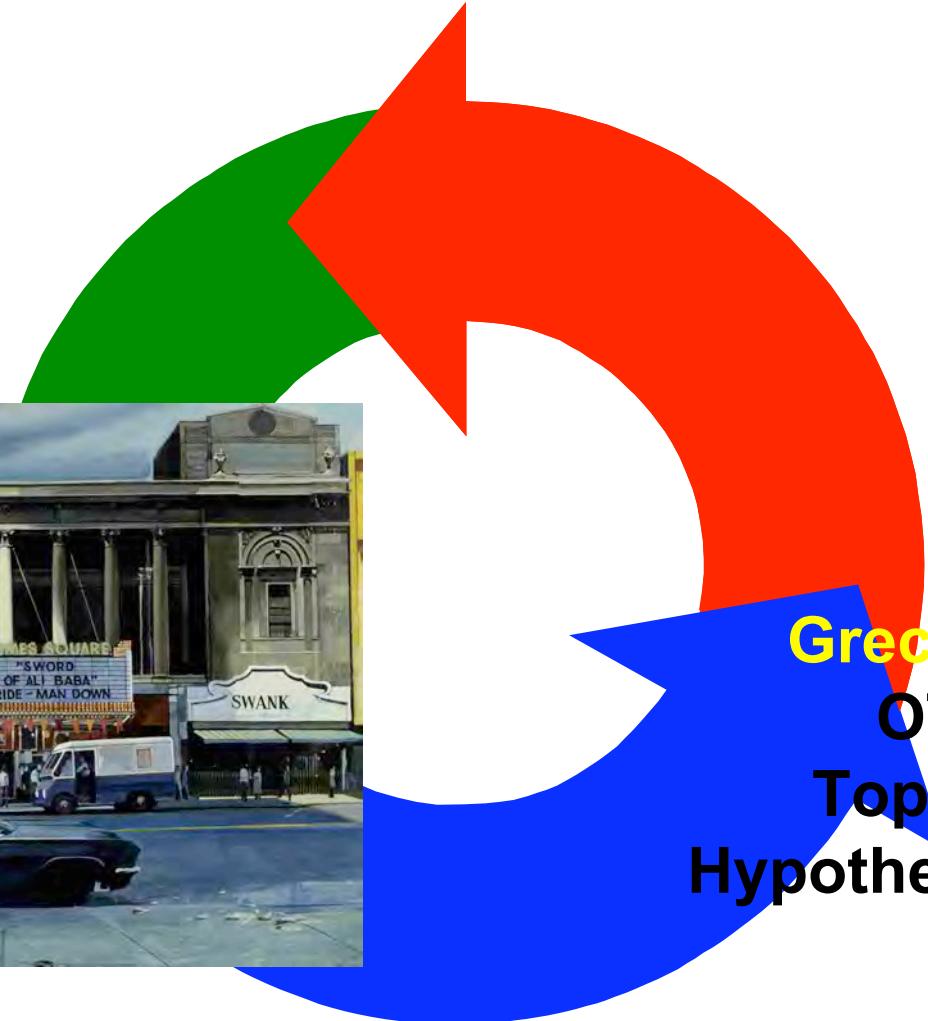
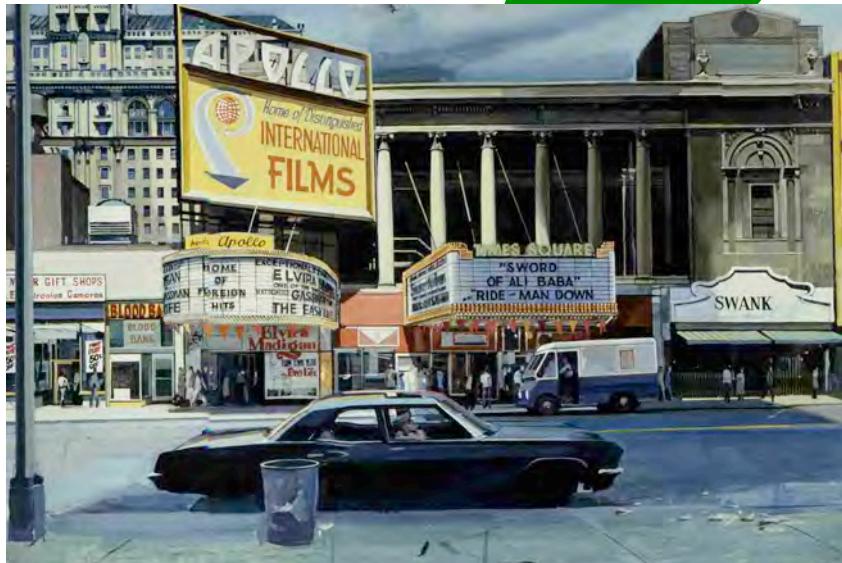


# **Maxwell's demon's genes: information of the program and information of the machine**

**Antoine Danchin** 唐善 • 安東

**When evolution evolves  
AgroParisTech, Paris, november 12<sup>th</sup>, 2009**



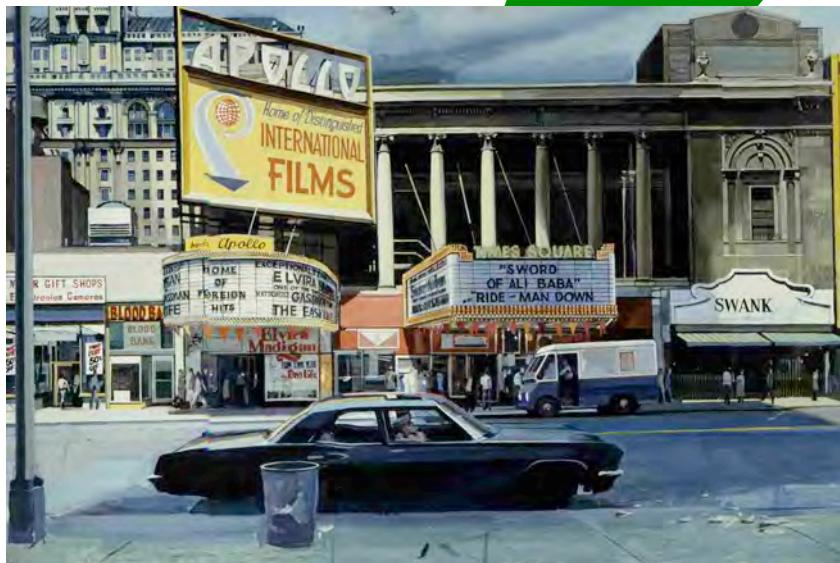


**Greco-Latin  
OTAN  
Top Down  
Hypothesis-driven**

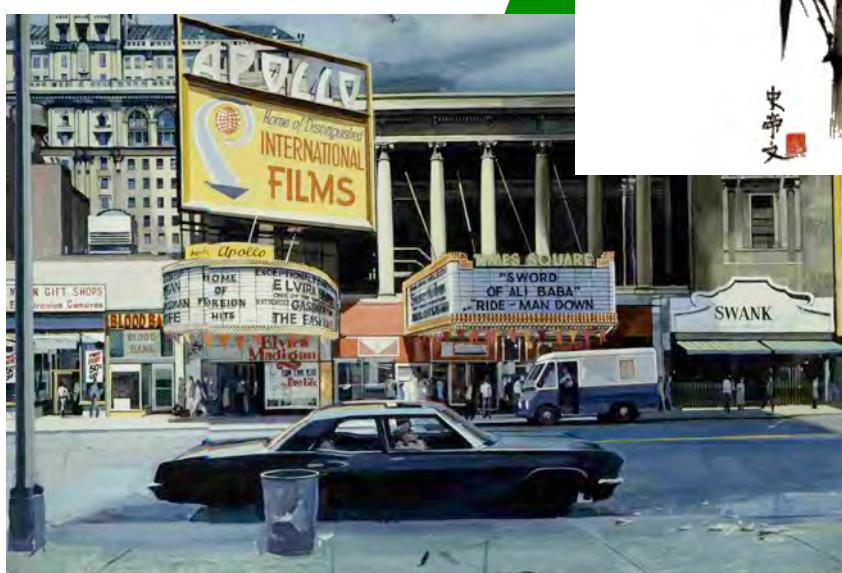
Chinese

# « Bombardment of the Chinese Embassy in Belgrade »

## Sideways Context-driven



## « Bombardmer

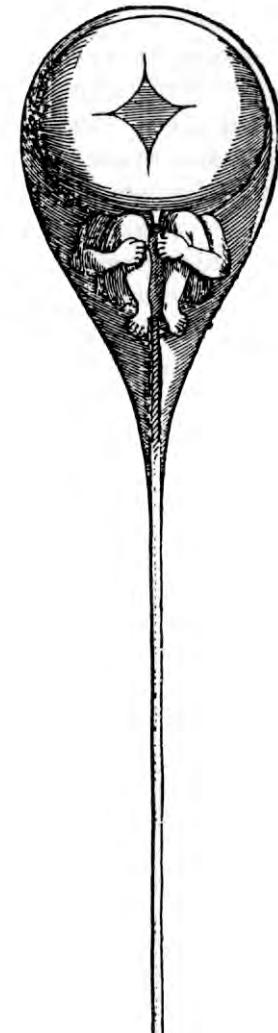


## Dassy in Belgrade »



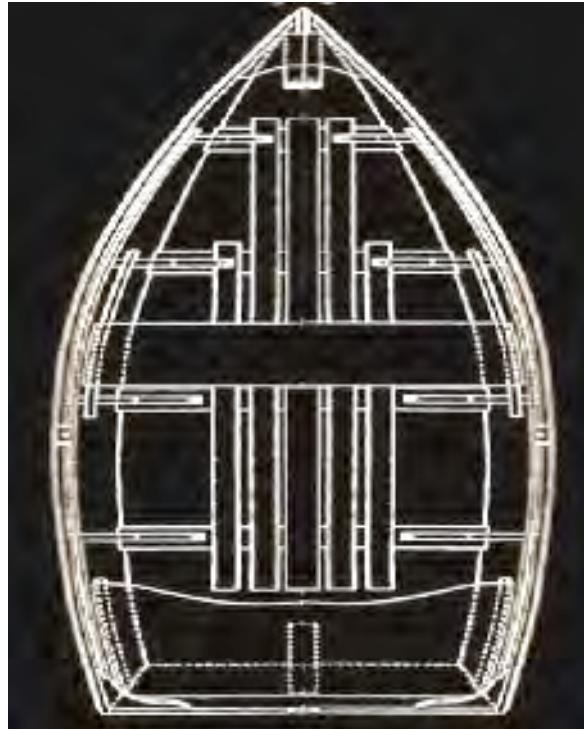
# Synopsis

- A solution for Bonnet's preformationism and Harvey's epigenesis: a **construction algorithm**, not an organism, is **replicated** through generations, together with a machine in its particular states
- The **machine (chassis)**, which **reproduces** over generations, runs the algorithm and 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**



# The Delphic boat

- Biology is a science of relationships between objects
- It is an information that expresses what is conserved in the boat, not the matter of its planks !



A. Danchin **The Delphic Boat, What genomes tell us.** Harvard University Press, 2003  
V. de Lorenzo, A. Danchin **Synthetic Biology: discovering new worlds and new words 9:** 822-827. EMBO Reports, 2008  
A. Danchin **Bacteria as computers making computers.** *FEMS Microbiology Reviews* 2009, 33: 3-26

# Information

*Historically, much of fundamental physics has been concerned with discovering the fundamental particles of nature and the equations which describe their motions and interactions. It now appears that a different programme may be equally important: to discover the ways that nature allows, and prevents, information to be expressed and manipulated, rather than particles to move*

**Andrew Steane (1998) Oxford University**

*Engineered biological systems have been used to manipulate information, construct materials, process chemicals, produce energy, provide food, and help maintain or enhance human health and our environment*

**Drew Endy (2005) MIT**

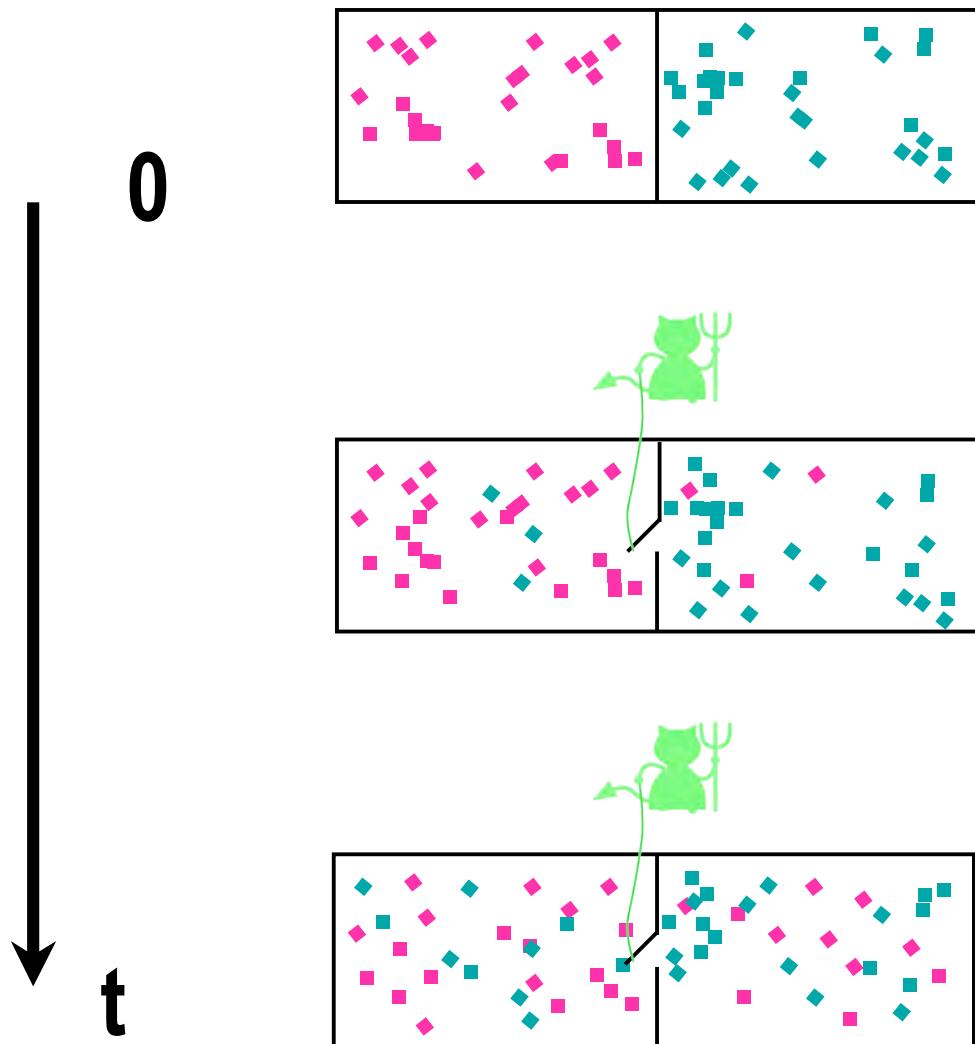
**“Les organismes vivants comme pièges à information”**

<http://www.normalesup.org/~adanchin>

# Information theories in physics

- 1929 **Leo Szilard**, (wrong) intuition of the relationship between energy and information: creation of 1 bit requires  $1/2 kT$ , while analysing Maxwell's demon
- 1949 **Claude Shannon**, theory of communication
- **1961 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 and reappraises Landauer's theorem
- 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 of the energy cost of memory erasure

# Maxwells's demon



The demon is able to reverse time, if it can **measure** speed and position; collecting **information** while computing to close the trap when needed



# Making information concrete: infotaxis



*Saturnia pyri*

<http://pdubois.free.fr/>

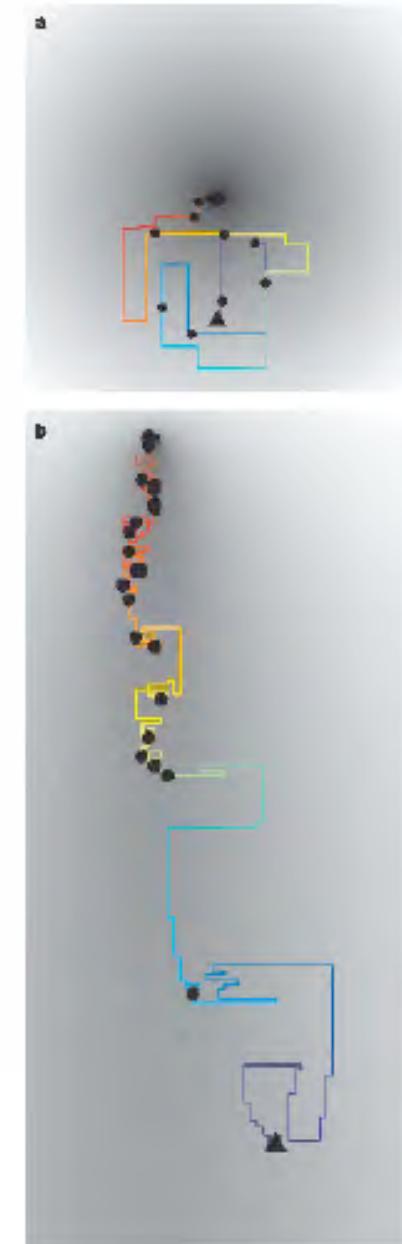
How does a moth find a partner one kilometer away?

Climbing up a chemical gradient is impossible at such a distance (air turbulence, obstacles...)

Vergassola and co-workers have shown that maximising **information** collection (via a simple computation) permits the insect to reach that goal...

'Infotaxis' as a strategy for searching without gradients

Vergassola M, Villermaux E, Shraiman BI  
*Nature* (2007) **445**: 406-409

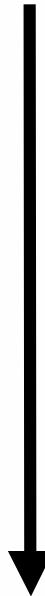


# Heuristic stance: a fifth category of reality

## Matter / Energy / Space / Time

- Classical physics
- Quantum physics
- Chemistry
- Biology
  - Development
  - Neurobiology
  - Linguistics
- Mathematics

Information



We choose to believe that at least some of the functions of life are computable

# Reproduction vs Replication

Dyson has shown that we need to separate reproduction — making a similar copy — from replication — making an exact copy

Reproduction may improve over time. Replication is generally doomed to accumulate errors

Hence, reproduction must have predated replication

Reproduction supposes networks of interactions, and it is most easily accounted for as a type of compartmentalised metabolism

**Which functions (coded by genes) permit accumulation of information?**

# What life is

Life requires:

- ⇒ A machine ("chassis") allowing the program to be enacted (reproduces)
  - ⇒ 1. Metabolism (a dynamic process)
  - ⇒ 2. Compartmentalisation (casings, defining inside and outside)
- ⇒ A program (a “book of recipes”, which is replicated)
  - ⇒ 3. Recursive information transfer and trapping => coding from one level to a second one introduces an essential asymmetry (fundamentally different from feedback)

The cell is the atom of life

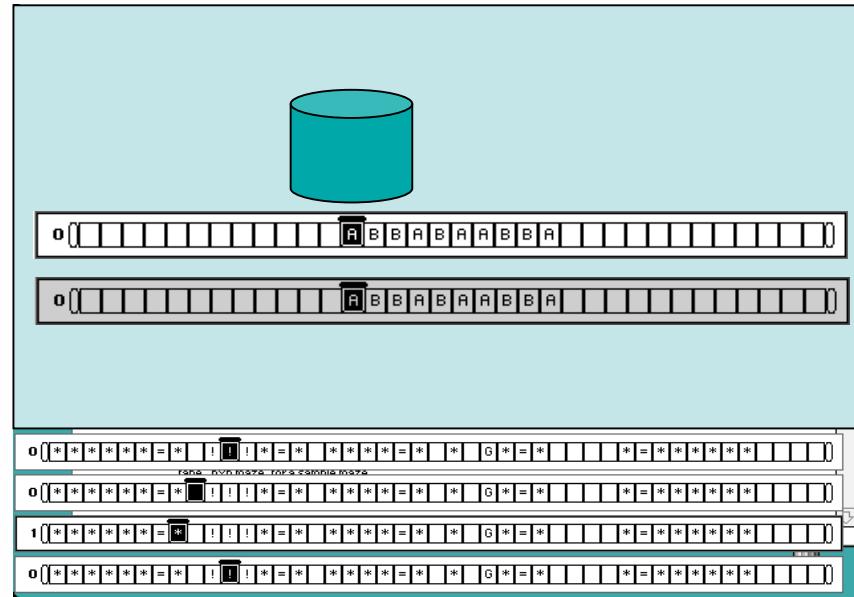
“Living organisms as information traps”

<http://www.normalesup.org/~adanchin>

# What computing is

Two entities permit computing:

- A machine able to read and **write**
- A program on a physical support, split by the human mind (not conceptually!) into two entities:
  - **Program** (providing the “goal”)
  - **Data** (providing some context)
- For the machine, Data and Program are indistinguishable



The machine is distinct from the data/program

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Cells and computers

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

- Horizontal Gene Transfer
- Viruses
- Genetic engineering
- Transplantation of a naked genome into a host cell followed by the replacement of the host by a novel host (2007)

all this separates

«Machine» (the "chassis") and «Data/Program» (the genome)

***“Les organismes vivants comme pièges à information”***

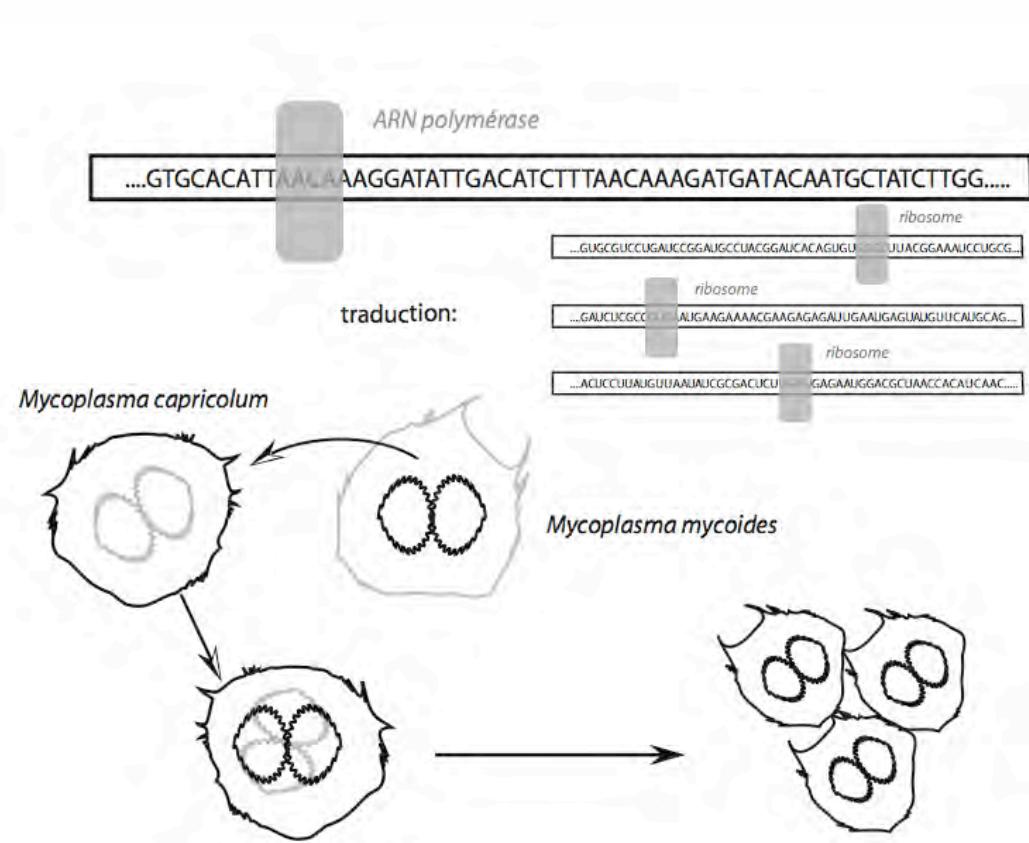
<http://www.normalesup.org/~adanchin>

# Lartigue-Venter's demonstration

## The Turing Machine

May exist in a parallel set up

## Genome transplantation



Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC  
Genome transplantation in bacteria: changing one species to another  
*Science* (2007) **317**: 632-638

**“Living organisms as information traps”**  
<http://www.normalesup.org/~adanchin>

# 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”**

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Information of the physical support!



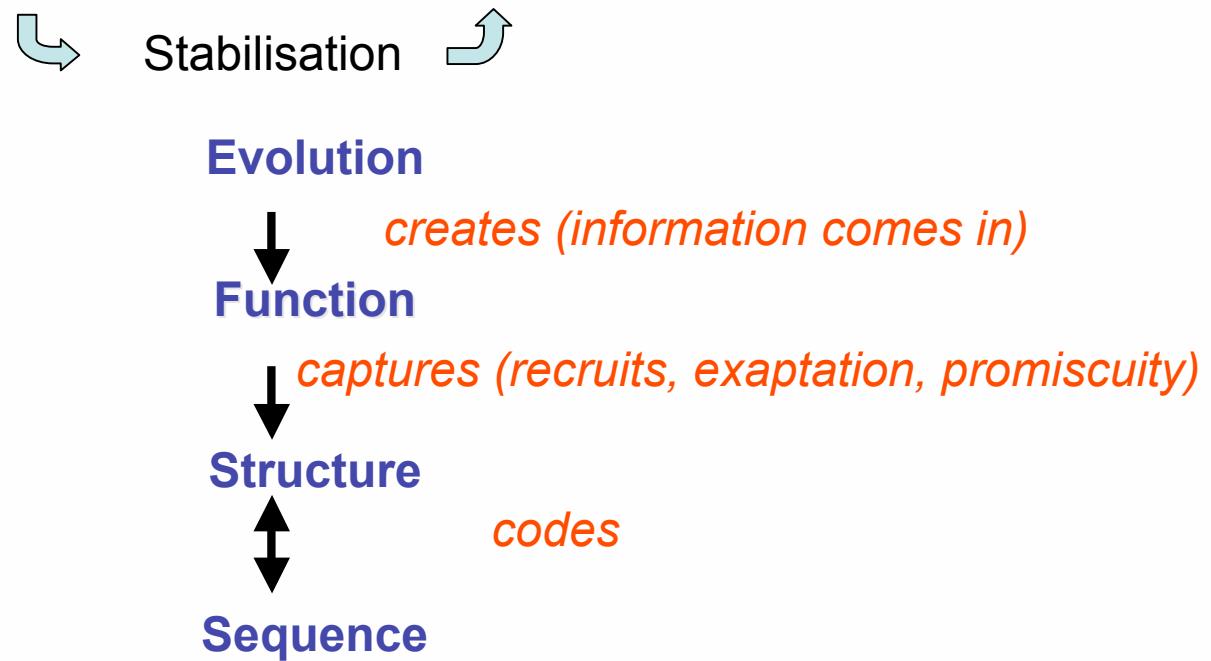
**It is not enough to have a DNA molecule with the right sequence, it needs to be correctly folded!**

**Also, this implies that animal cloning is a further support of the Turing Machine hypothesis**

Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A.* (2008) 105:20404-20409

# Looking for ubiquitous functions

Variation / Selection / Amplification



Functional ubiquity does not imply structural ubiquity

# From functional ubiquity to gene persistence

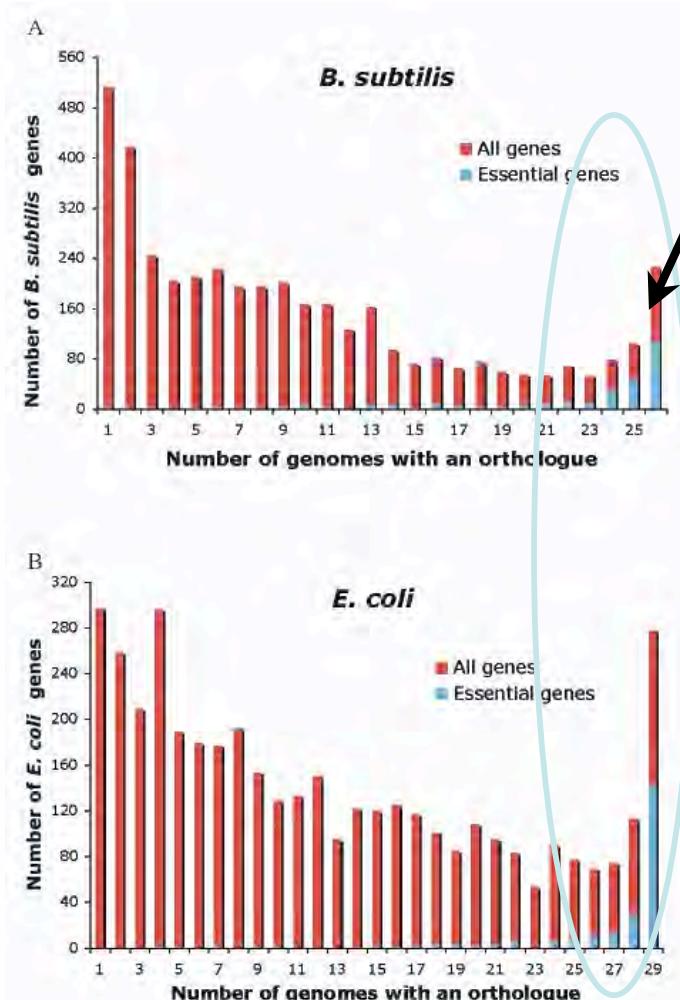
Functional ubiquity does not imply structural ubiquity

Fortunately, efficient objects tend to persist over generations:

- ➔ Looking for « persistence » permits identification of (most) ubiquitous functions
- ➔ Is « ubiquitous » a synonym of « essential »? **NO**

~ 500 genes persist in bacterial genomes; about 250 cannot be inactivated without loss of capacity to live;  
**what about the other half?**

# Gene persistence: too many genes



Persistent genes

Essential genes and ....

Stress, maintenance and repair

Energy-dependent degradation

Metabolic « frustration »  
(chemical constraints)

Gang Fang & Eduardo Rocha

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

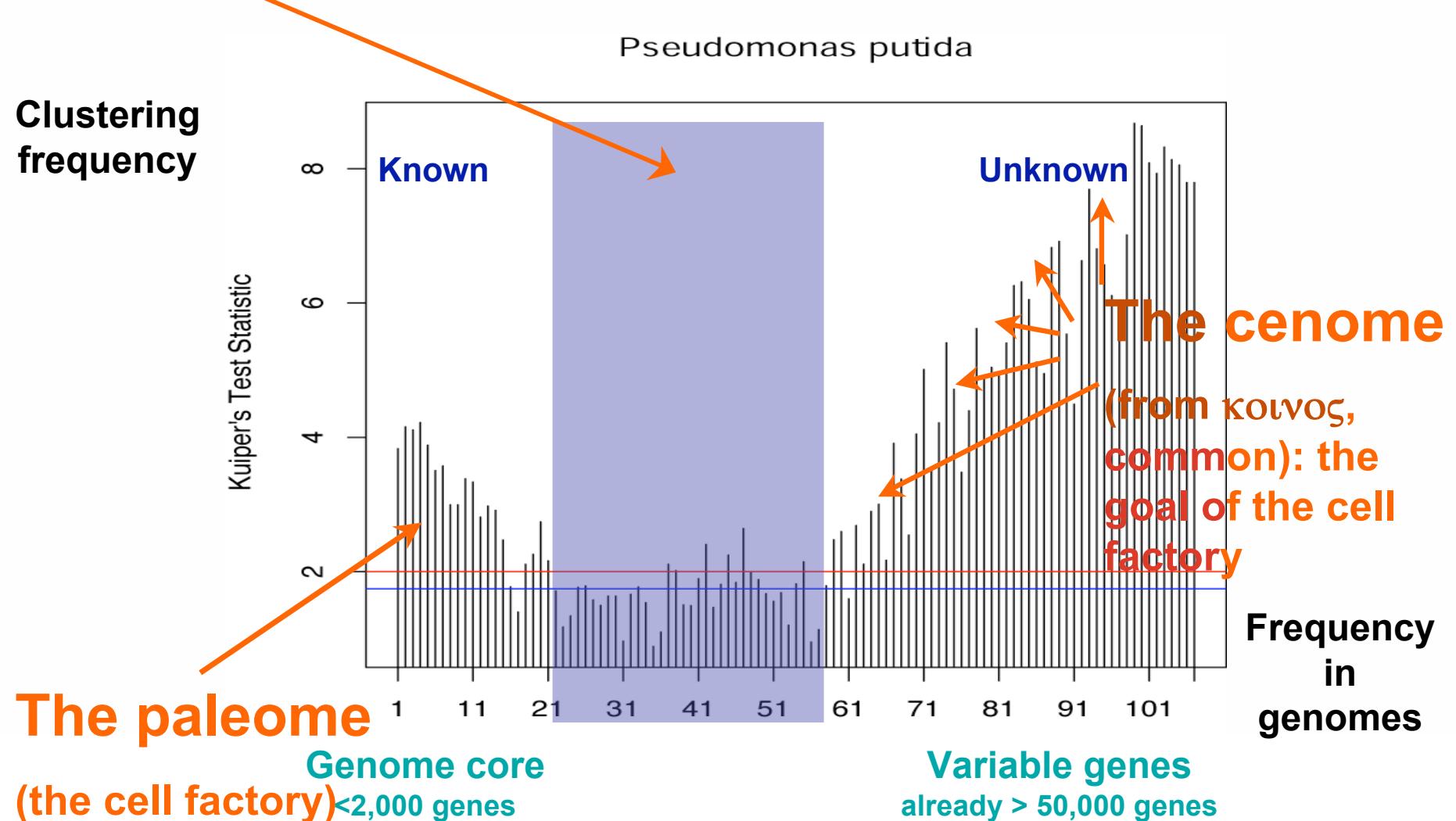
# Persistent genes connectivity

With 228 genomes (made of >1500 genes) and accurate annotations, we identified genes that stay close together

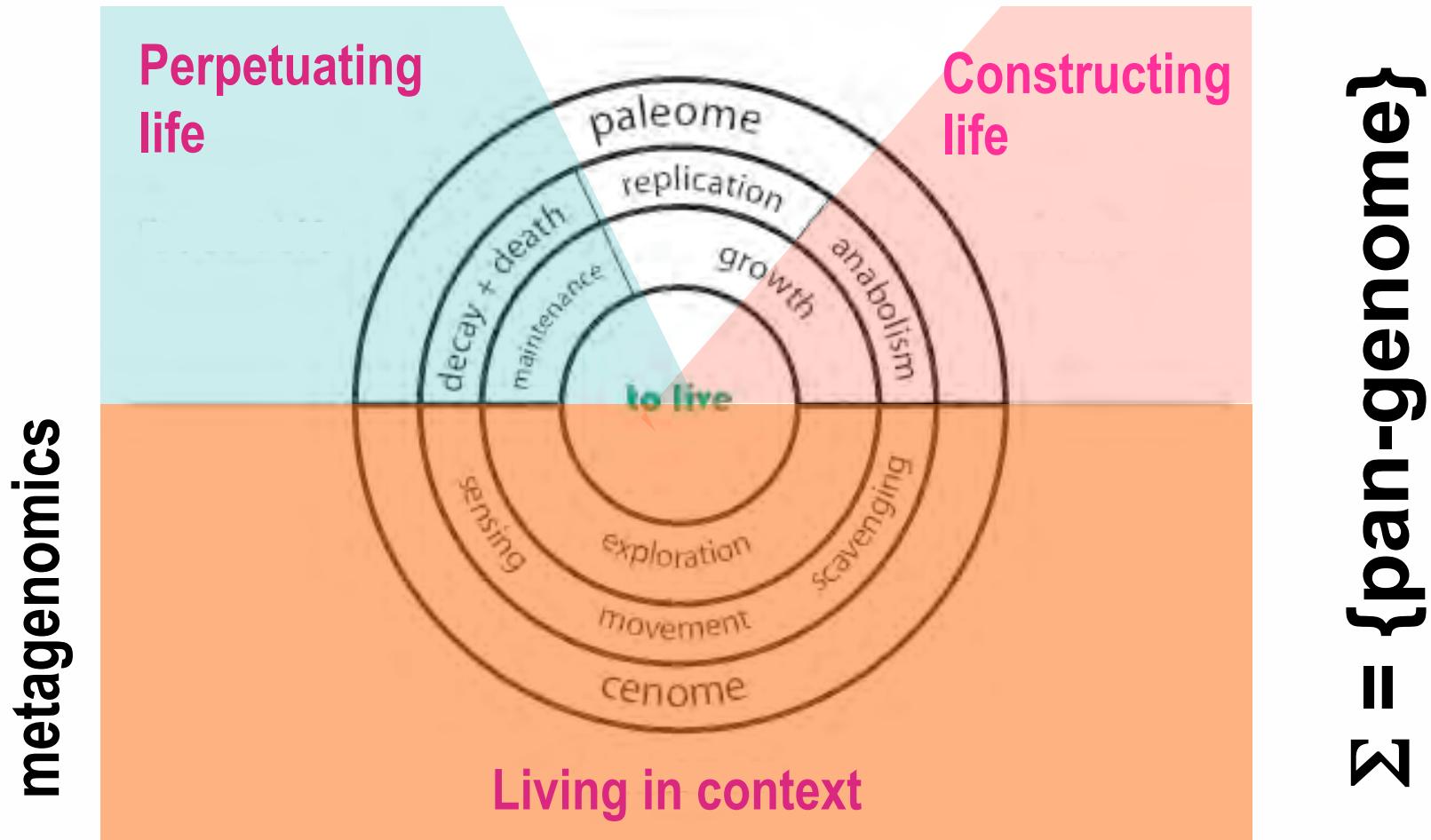
Two main domains were found: persistent genes and rare genes

# Organisation of bacterial genomes

**mixome: functional assimilation**



# 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

# The paleome and the cenome

## → The structure of the paleome

- Essential functions; evolution of the gene expression machinery
- Energy-dependent degradation
- Sulfur metabolism (anabolism, salvage, catabolism)
- Chemical frustration (metabolic « patches »)

## → The cenome: from commensalism to virulence

- The methionine salvage pathway
- *Staphylococcus epidermidis* (Fudan University, Shanghai)
- ColiScope: *E. coli*, *E. fergusonii* (E. Denamur et al.)

“Les organismes vivants comme pièges à information”

<http://www.normalesup.org/~adanchin>

# Necessity of reproduction: Everything ages

**Aging is a ubiquitous constraint in the genome, in the proteome, in metabolism...**

**Yet « babies are born very young » !**

**This implies that creation (or recovery) of information is a ubiquitous process of life; resting on ubiquitous functions**

*“Les organismes vivants comme pièges à information”*

<http://www.normalesup.org/~adanchin>

# The physical status of natural selection

Intuition tells us that creation of information requires energy

**WRONG!**

In an endeavour to calculate the limits of practical computation, Landauer demonstrated that **creation of information is reversible** (*i.e.* does **not** require energy: Landauer, 1961; Bennett, 1982, 1988); however, accumulating information requires an **energy-dependent process to make room for this accumulation**

**Open question:** if « making room » is needed to accumulate information, how is it obtained? Can we identify in genomes the genes coding for the functions required to put this process in action? Can we find a ubiquitous and stable energy source?

# Too many persistent genes....

- **Paleome 1 (genes that are essential for a synthetic bacterium)**
  - **Constructor:** DNA specifies proteins which form the machine that constructs the cell (reproduction)
  - **Replicator:** DNA specifies proteins that replicate DNA (replication)
- **Paleome 2 (persistent non-essential genes)**
  - Perennisation of life, **energy-dependent processes, preventing degradation of functional entities**
  - Metabolic patches (chemical frustration)

*“Les organismes vivants comme pièges à information”*

<http://www.normalesup.org/~adanchin>

# Maxwell's demon's genes

**Conjecture** (may work at different scales and levels):

- Energy-dependent degradative processes make room for newly synthesised entities; energy is used to **prevent** degradation of functional entities
- This process accumulates information, **whatever its origin**, in a ratchet-like manner
- As this process is ubiquitous, we expect that the corresponding functions are encoded in the paleome, including management of the major energy sources postulated here

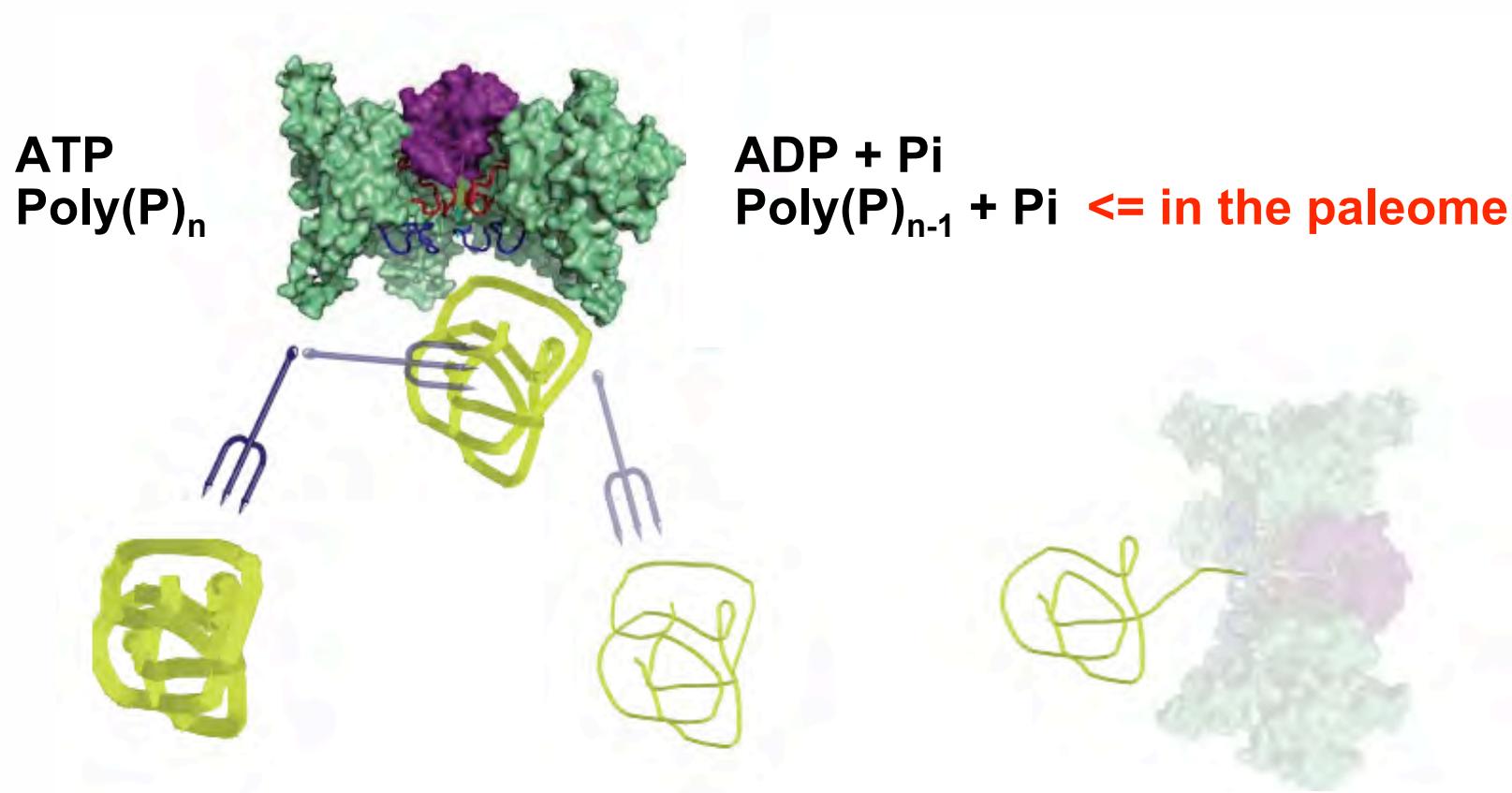
# A ubiquitous source of energy: Polyphosphates

- Poly-P synthesis and turnover belong to the non-essential paleome genes; overall the process is poorly known and associated to degradation of RNA
- Poly-P is a mineral, hence extremely stable; it is ubiquitously present in cells
- NTPs can be regenerated from NMP and poly-P; Protease Lon can use poly-P instead of ATP; NADP (anabolism) can be generated from NAD and poly-P...

*“Les organismes vivants comme pièges à information”*

<http://www.normalesup.org/~adanchin>

# Maxwell's demon's genes



The degradation machinery uses energy to reject a functional entity

Non functional entities are recognised and degraded

# A test for identifying Maxwell's demon's genes

- Energy-dependent accumulation of information is blind; it cannot know where information will come from
  - Information can just come from memory; it can also be created de novo
- 
- **Prediction:** adaptive mutations are a de novo creation of information; they should depend on the genes involved in accumulation of information

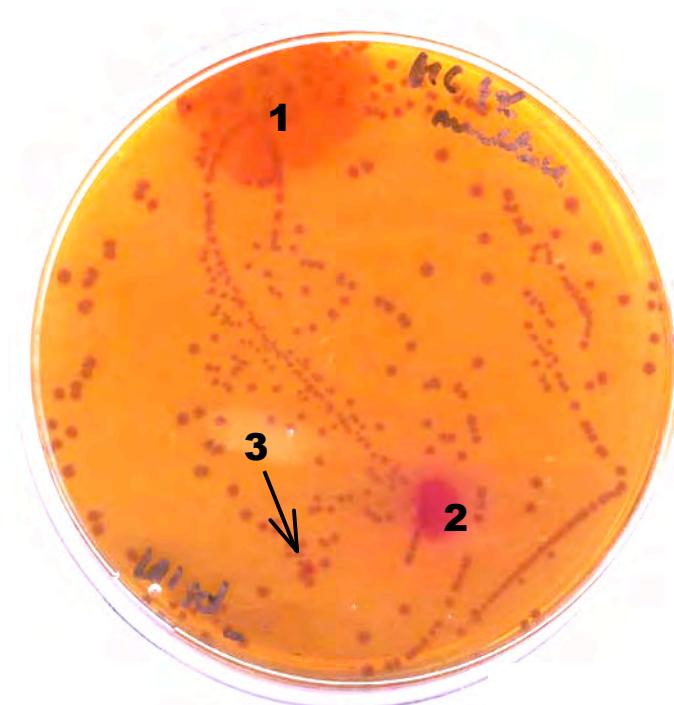
# Putting the conjecture to test: Adaptive mutations

## 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 weeks time, papillae appears that begin to grow and invade the medium, using supplied "unusable" nutrients. They derive from **adaptive mutations**

They did not pre-exist, and this supposes **creation of information**



Agnieszka Sekowska  
Evelyne Turlin  
Andrew Martens

**"Les organismes vivants comme pièges à information"**

<http://www.normalesup.org/~adanchin>

# Mutations appear during the process of ageing

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 at least one other gene involved



# Natural selection traps information

- Energy-dependent degradative processes make room for newly synthesised entities; energy is used to **prevent** degradation of functional entities
- This process accumulates information, whatever its origin, in a ratchet-like manner
- Because the process is ubiquitous, the corresponding functions are expected to be coded in the paleome, including the possible energy source
- This process is **myopic**: it cannot have any grand design, hence the “tinkering” feature of the evolution of life

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Contextual information and reproduction

Exploring reproduction led Doron Lancet and his colleagues at the Weizmann Institute to explore the emergence of a primitive metabolism via fusion / scission of active metabolic groups that assembled and reproduced while progressively enriching some ensembles, via mutual catalytic effects. This scenario is similar to the scenario proposed by Freeman Dyson for metabolism reproduction

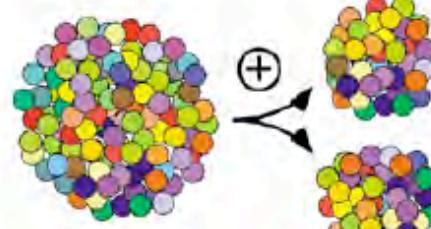
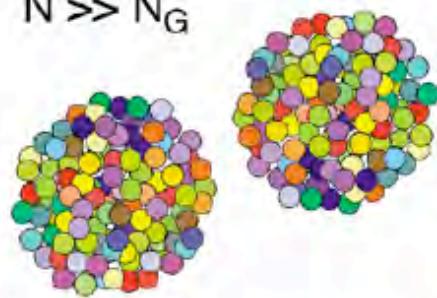
Composing life. Segré D, Lancet D. *EMBO Reports* (2000) 1: 217-222

In Rome, Pier Luigi Luisi and his colleagues build up synthetic minimal cells, using “liposomes” able de synthétiser les lipids that make their surface

A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells.Kuruma Y, Stano P, Ueda T, Luisi PL. *Biochim Biophys Acta* (2008) Nov 5.

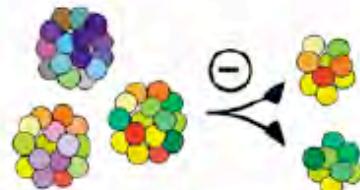
# Contextual information and reproduction

$$N \gg N_G$$



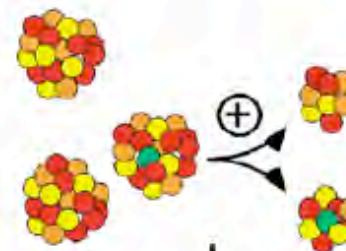
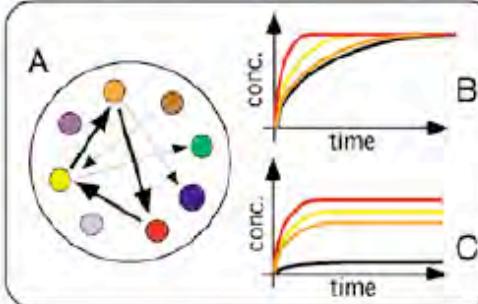
No selection possible

$$N \ll N_G$$



High diversity of assemblies

Selection by mutual catalysis



Natural selection and Evolution

## 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 flame elicits a response, such as the flight of a moth, in the part of an external, autonomous agent which provides all necessary forces and energies

It is in this second case that we can best see information as a category distinct from mass or energy. **The theory does not exist yet...**

# ACKNOWLEDGEMENTS

## In silico

Gang Fang, Eduardo Rocha, TingZhang Wang

## In vivo

Agnieszka Sekowska, Evelyne Turlin, Andrew Martens

## Collaborations

### GenoScope

Beijing Genome Institute, Fudan University, The University of Hong Kong  
Hong Kong University of Science and Technology

