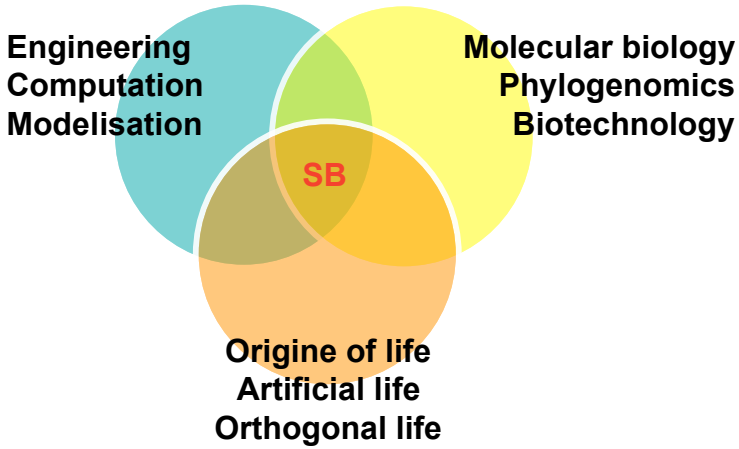
In vivo, in vitro, in silico: The future of biology, from genomes to synthetic cells



Antoine Danchin 唐善 安東
Director, Department Genomes and Genetics, Institut Pasteur
Hong Kong University, october 14th, 2008

INSTITUT PASTEUR

In vivo, in vitro, in silico

- Biology
  - Biochemistry
  - Genetics

  → Molecular Biology

- Number Theory and Computer Sciences
  - Turing Machines
  - Algorithmics

  → Information Theory

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

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

Goals of synthetic biology

- A first aim of SB is to reconstruct life, in an endeavour to explore whether we understand what life is and learn missing entities from our failures
- A second aim is to keep the laws defining life, and to apply them using objects of a different physico-chemical nature
- A third aim is to see life from an engineering standpoint, trying to class and normalise « biobricks » to construct a « cell factory »

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

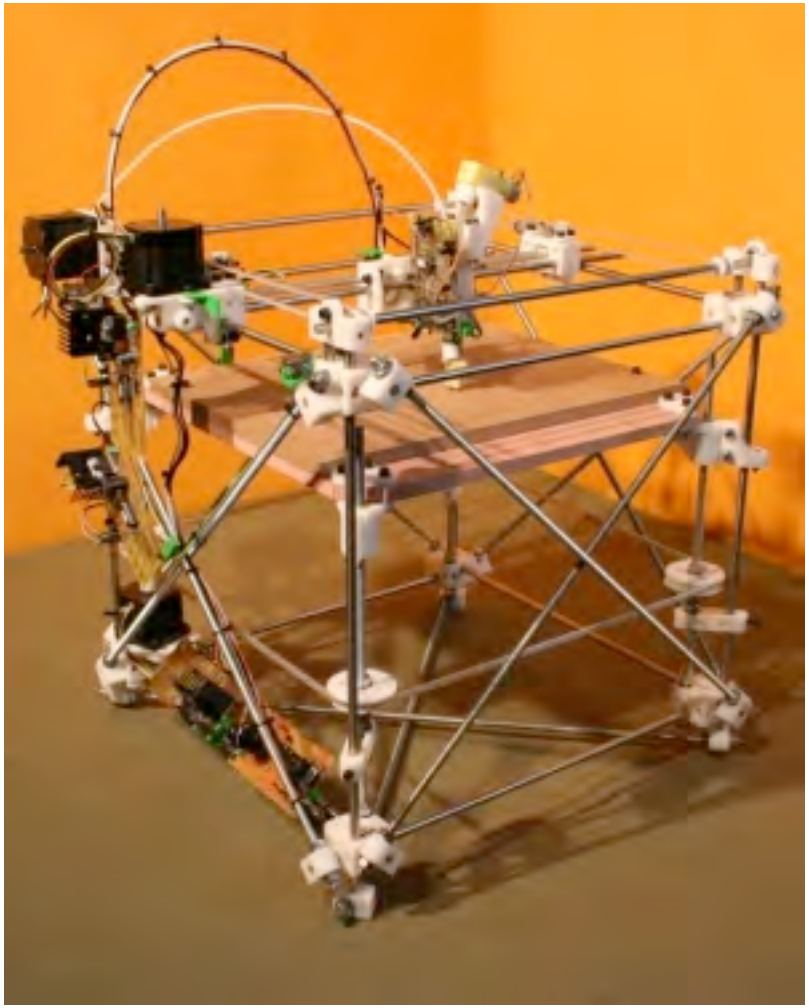
However, here is the situation …

A 3D self-reproducing printer

Project RepRap ((Replicating Rapid-prototyper, 2004) **aims at creating a laser 3D self-reproducing printer :**

- **The machine produces most of its components (= “biobricks”)**
- **What is missing :**
  - **The program**
  - **The assembly (managing space and time - sequence of events, and specific functions such as lubrication)**

http://reprap.org/



“Living organisms as information traps”

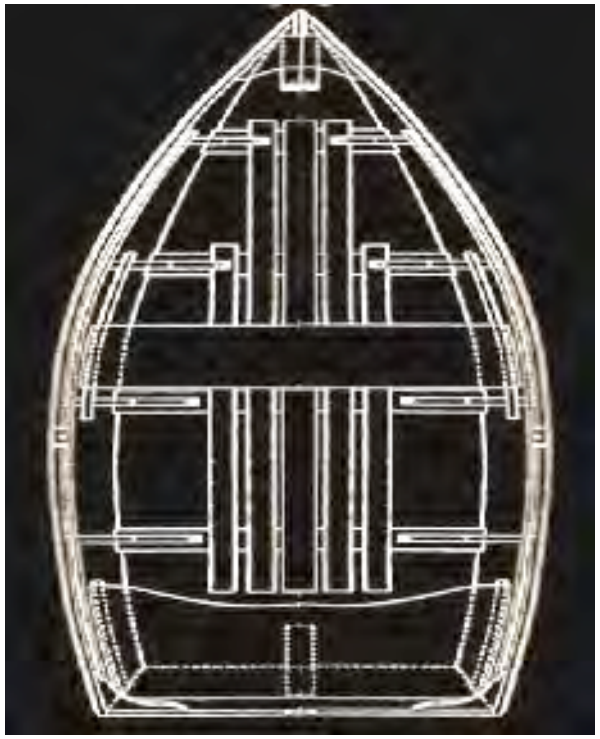
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Biology is « symplectic »

- Biology is a science of relationships between objects:
- It is **symplectic** (from συν together, πλεκτειν, to weave) the same word as « Complex » in Latin; used here to avoid the unwanted fuzzy connotations associated to « Complexity »; a connotation in Geometry will not interfere…
- As in the construction of a boat, lack of understanding of interactions between the planks will lead the boat to sink



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

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Information is a fifth category of nature

Classical Physics

$E = mc^2$

energy matter space/time

Quantum Physics

$\Delta x \Delta p \geq h/4\pi$

uncertainty = lack of information

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Revisiting the hierarchy of the sciences

Matter / Energy / Space / Time

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

Information

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

What life is

Life requires:

- **A machine allowing the program to be enacted (reproduces)**
  - **1. Metabolism (a dynamic process)**
  - **2. Compartmentalisation (defining an inside and an outside)**
- **A program (a "book of recipes", which replicates)**
  - **3. Recursive information transfer => coding from one level to a second one as an essential element**

The cell is the atom of life

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

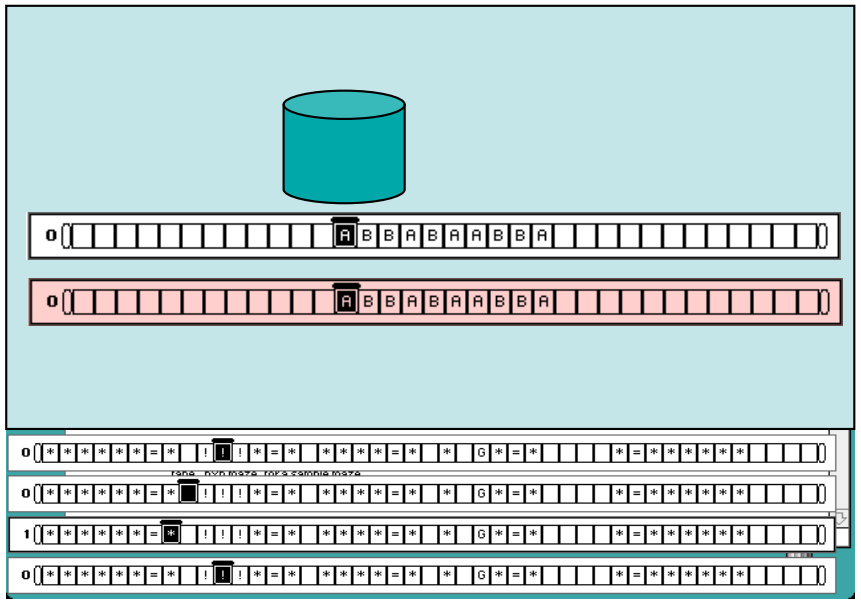
INSTITUT PASTEUR

CNRS

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 the context)



The machine is distinct from the data/program

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Cells and computers

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

Horizontal Gene Transfer

Viruses

Genetic engineering

Direct transplantation of a naked genome into a recipient cell with subsequent change of the recipient machine into a new one (2007)

all points to separation between

«Machine» (the cell factory) and «Data/Program» (the genome)

CONJECTURE: living organisms are information traps, and we can identify the concrete processes permitting accumulation of information

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

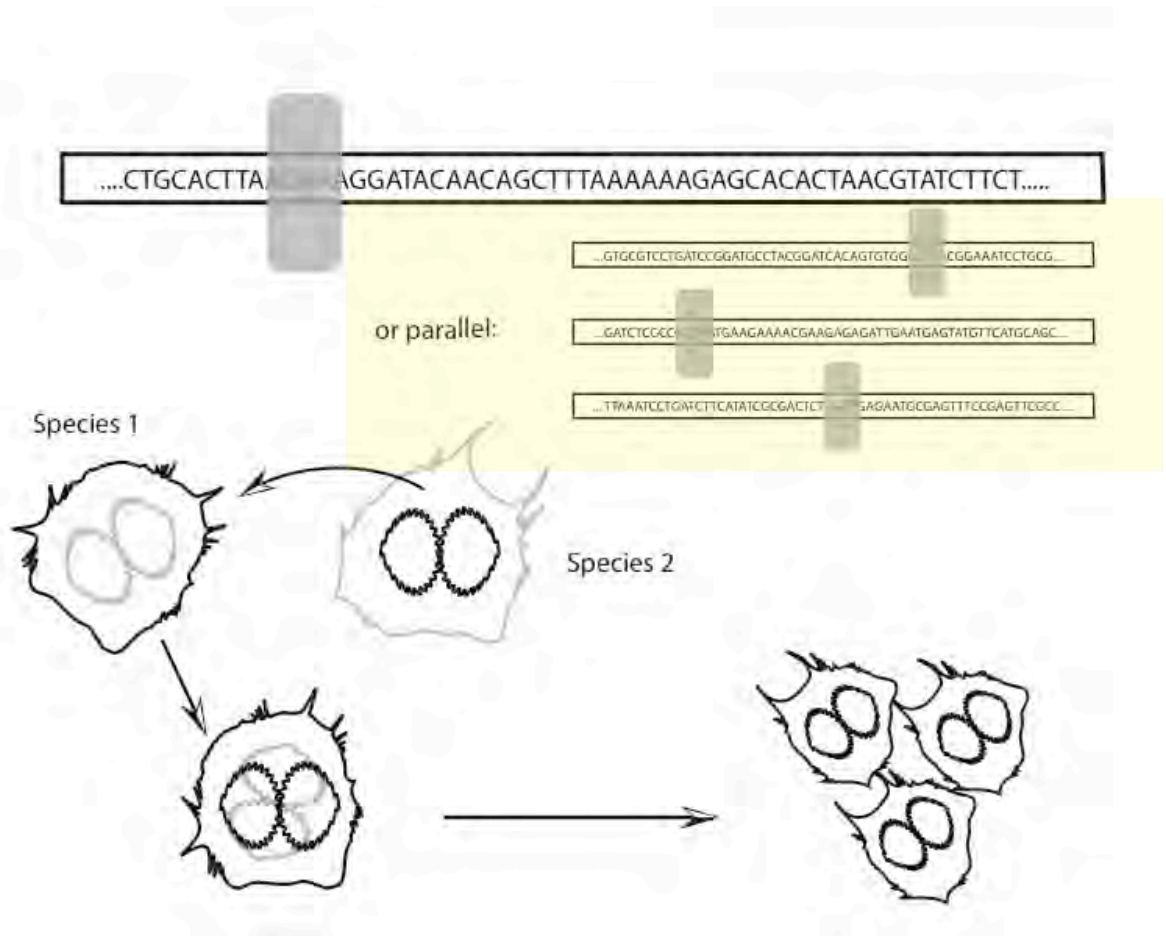
INSTITUT PASTEUR

CNRS

Venter's demonstration

The Turing machine

May exist in a parallel set up



Genome transplantation

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Objection to the computer model of the cell

“Beside the genetic program, the cell carries a considerable amount of information…”

“Living organisms as information traps”

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Even in authentic computers, mind the physical support!

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

Reproduction vs replication

- The program *replicates*
- The machine *reproduces*

- Replication accumulates errors
- Reproduction can improve over time

This process implies creation of information

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Ageing is a ubiquitous constraint in the genome, in the proteome, in metabolism…

Yet « babies are born very young » !

This implies that creation of information is a ubiquitous process of life

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

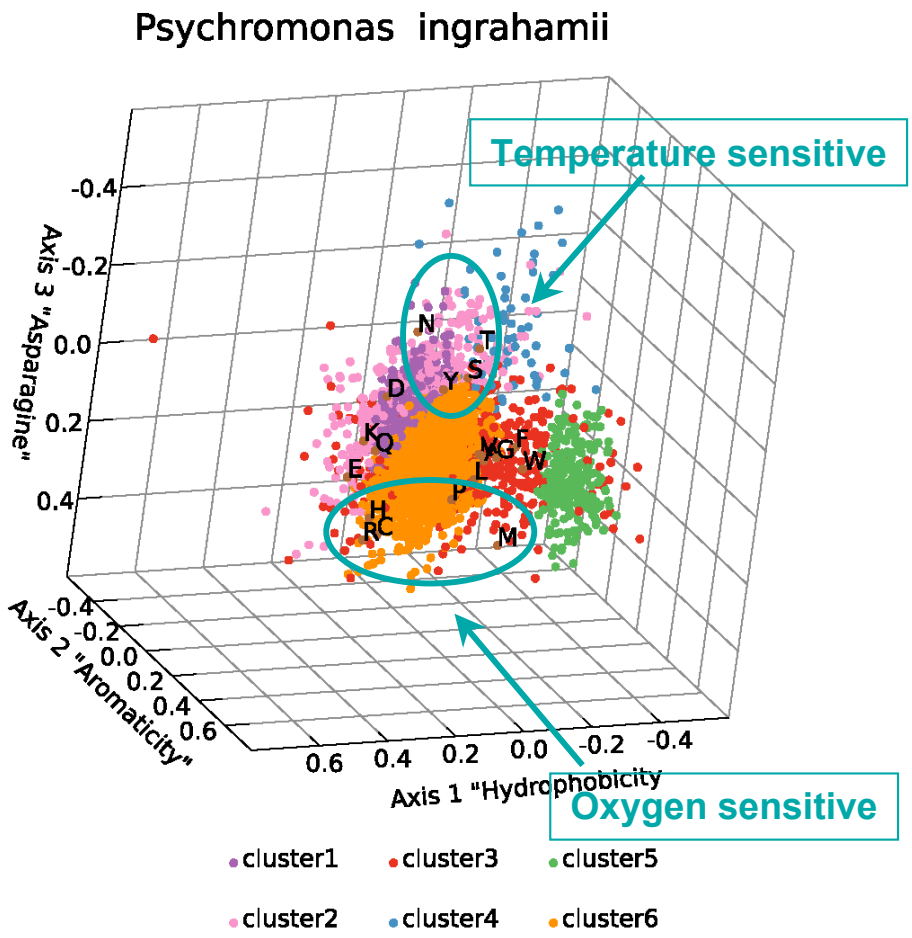
INSTITUT PASTEUR

CNRS

Looking for universal biases

Distribution of amino acids in the proteome of *Psychromonas ingrahamii*

(-12°C, 100 h gt)



Riley M, Staley JT, Danchin A, Wang T, Brettin TS, Hauser LJ, Land ML, Thompson LS. Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. BMC Genomics. 2008 May 6;9(1):210

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

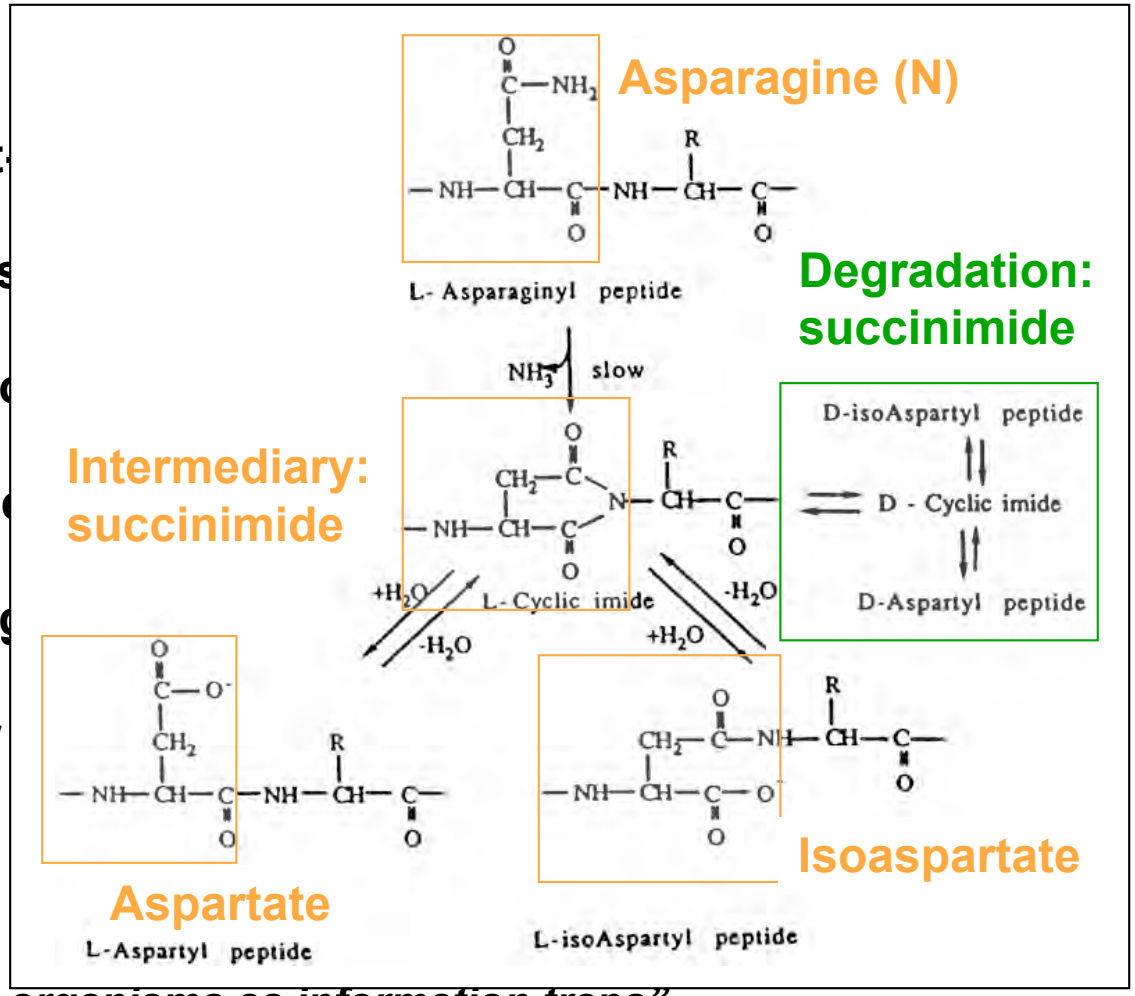
INSTITUT PASTEUR

CNRS

A chemical anecdote?

Asparagine deamidates: a major contribution to protein ageing

- Main post
- Reaction s
- Spontaneo
- Affects the
- Role in reg
- Signal for



"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Looking for ubiquitous functions

Variation / Selection / Amplification

↳ Stabilisation ↱

Evolution

↓ *creates*

Function

↓ *captures (recruits)*

Structure

↕ *codes*

Sequence

Functional ubiquity does not imply structural ubiquity

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

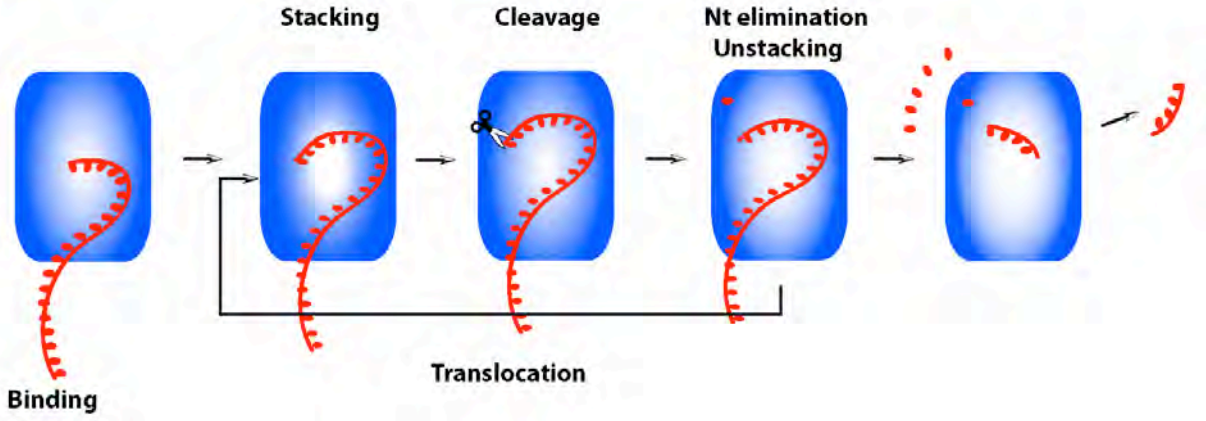
INSTITUT PASTEUR

CNRS

NanoRNase is an essential function

Exoribonucleases are processive

This allows them to stick to their substrate and chew it up until they reach an end point where they can no longer proceed, yielding a leftover 2-5 nt long (usually 3nt)



These « nanoRNAs » are toxic, as they can enter transcription and replication « bubbles »; they must be degraded

***Orn* in *E. coli*, SFN (coded by REXO2) in *H. sapiens* are indeed essential genes**

"Living organisms as information traps"

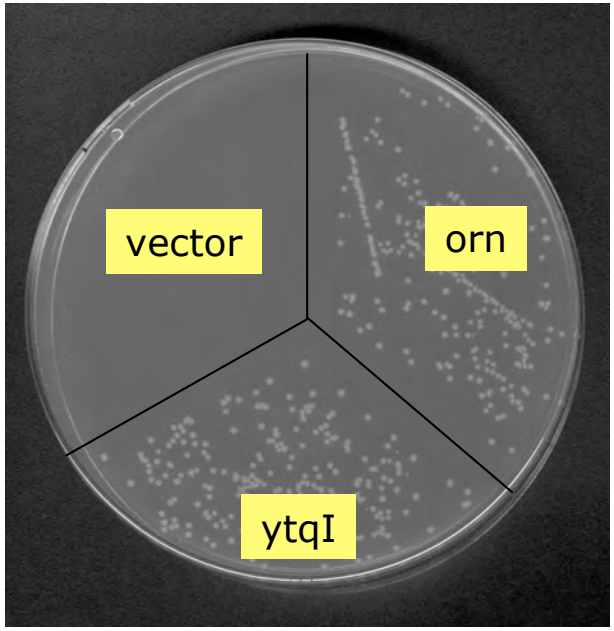
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

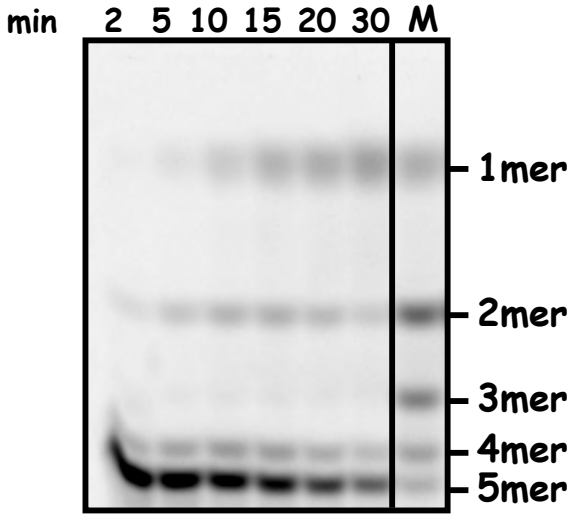
B. subtilis ytqI is a functional analogue of *E. coli orn*

in vivo



***ytqI* complements *E. coli orn*$^-$**

in vitro



Degradation of nanoRNA 5mers (Cy5-CCCCC-3')

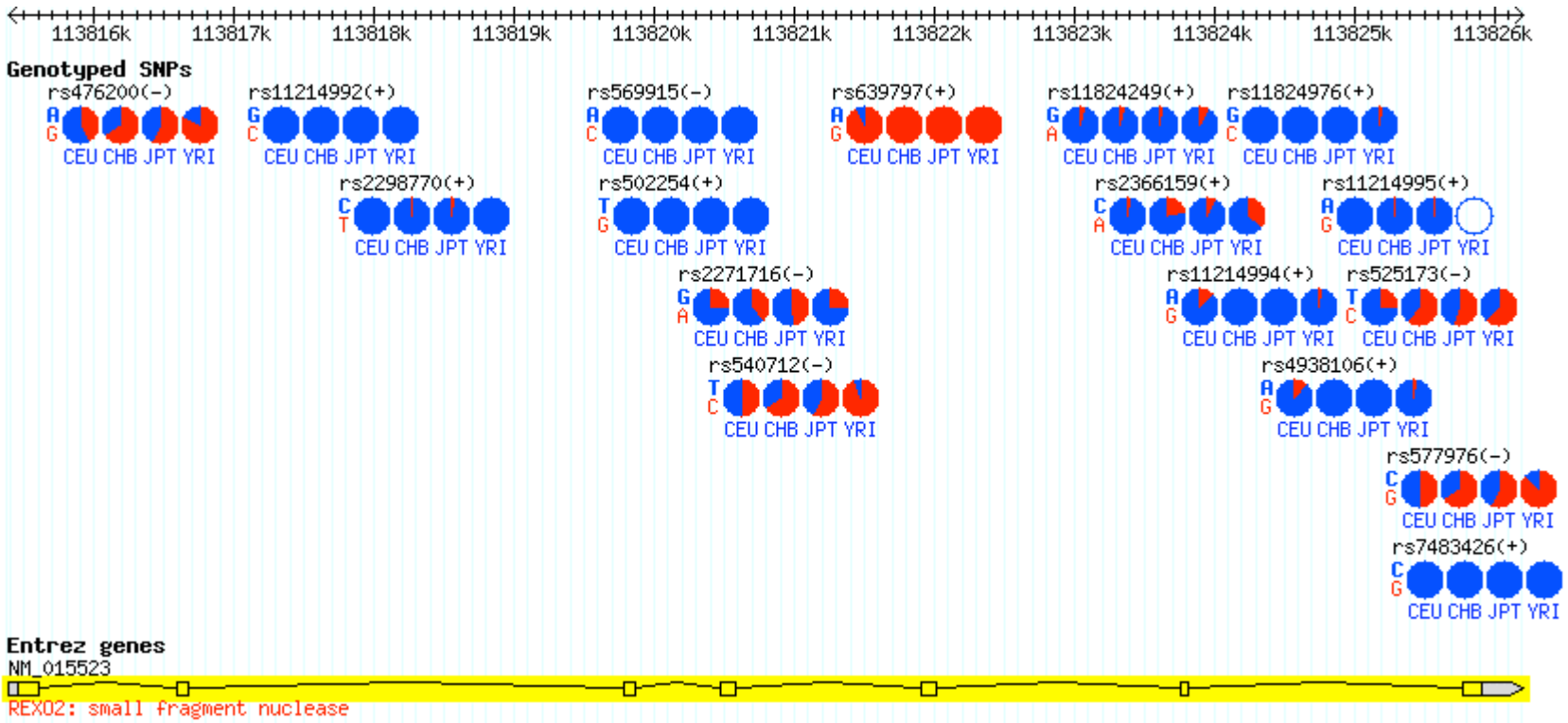
"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

REX02 evolution in human populations



Genotyped SNPs and frequency distribution in HapMap b21a

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

From functional ubiquity to gene persistence

Functional ubiquity does not imply structural ubiquity
Efficient objects tends to persist through generations:
➡ Looking for « persistence » permits identification of (most) ubiquitous functions
➡ Is « ubiquitous » a synonym of « essential »?
« Laboratory essential » genes are located in the DNA leading strand

~ 500 genes persist in bacterial genomes; they are not only involved in the three processes needed for life, but in maintenance and in adaptation to transient phenomena; a fraction manages the evolution of the organism

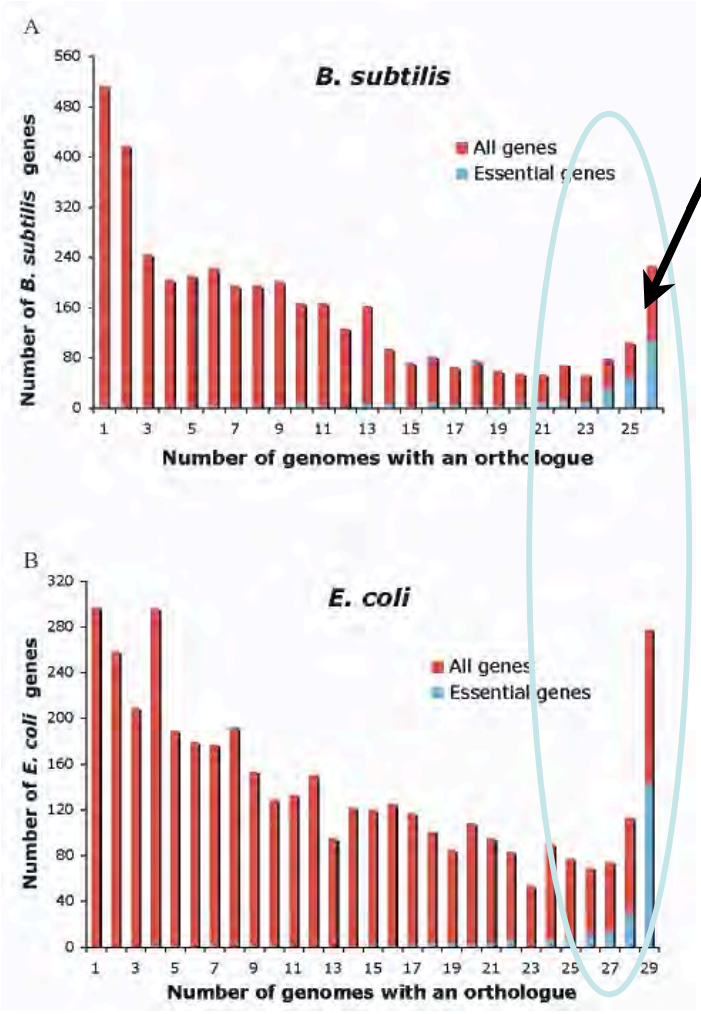
"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Gene persistence: too many genes



Persistent genes

Essential genes and ….

Stress, maintenance and repair

Energy-dependent degradation

Metabolic patches (serine effect)

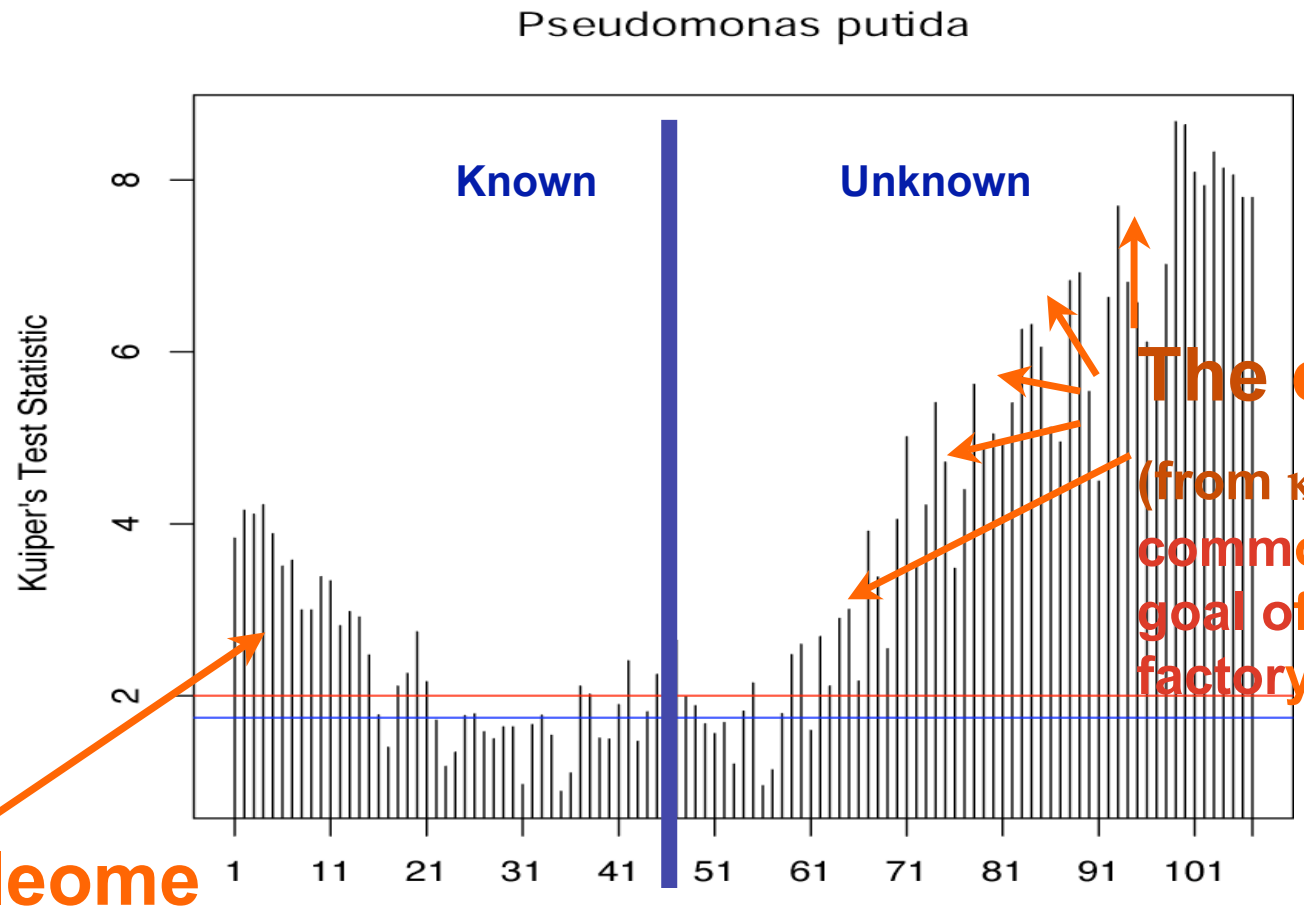
"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Organisation of bacterial genomes



Clustering frequency

Frequency in genomes

The cenome

(from κοινος, common): the goal of the cell factory

The paleome

(from παλαιος, ancient): the cell factory

Genome core
<2,000 genes

Variable genes
already > 50,000 genes

Persistent genes are clustered together

Persistent genes are functionally defined. They are located in the DNA replication leading strand

Depending of their tendency to remain clustered in genomes (in > 250 bacteria with genome length > 1,500) they form three families that reflect a scenario of the origin of life

This group of core genes form the paleome (from παλαιος, ancient)

A mineral scenario for the origin of life

- The surface of charged solids (e.g. pyrite (Fe-S)) selects and compartimentalises charged molecules; this first step forms some aminoacids, the main coenzymes, fatty acids and ribonucleotides; polymerisation with elimination of water molecules increases entropy

- Compartmentalised metabolism creates surface substitutes via polymerisation of ribonucleotides in the presence of peptides (the RNA world, with tRNA ancestors)

- RNAs discover the complementarity law, and the genetic code is invented (from substrates to templates). Nucleic acids are stabilised by the invention of deoxyribonucleotides, a the time when the rules controlling information transfer are discovered, first within the RNA world where vesicles carrying the ancestors of genes split and fuse randomly, before formation of the first genomes

"Living organisms as information traps"

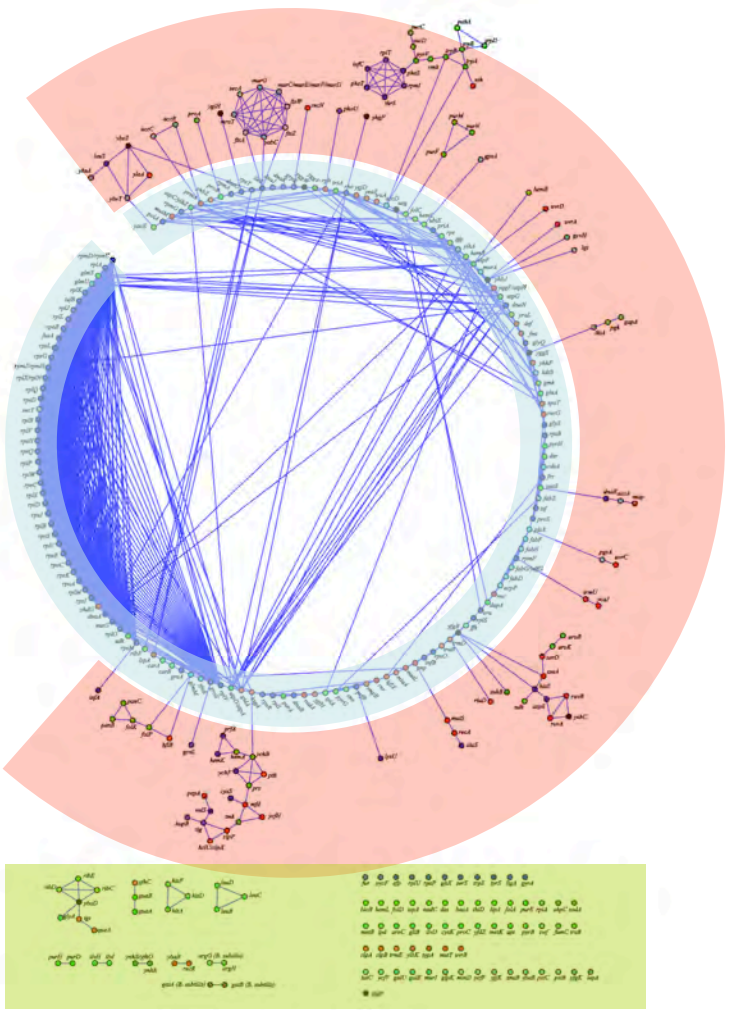
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Persistent genes recapitulate the origin of life

The external network, made of genes of intermediary metabolism (nucleotides and coenzymes, lipids), is highly fragmented; the middle network is built around class I tRNA synthetases, and the inner network, almost continuous, organized around the ribosome, transcription and replication manages information transfers



A Danchin, G Fang, S Noria
The extant core bacterial proteome is an archive of the origin of life
Proteomics. (2007) **7**:875-889

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

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
  - *Staphyloccoccus epidermidis* (Fudan University, Shanghai)
  - *Photorhabdus luminescens* and *Escherichia* sp. (ColiScope)
  - *Bartonella birtlesii* (Beijing Genome Institute, F. Biville)

"Living organisms as information traps"

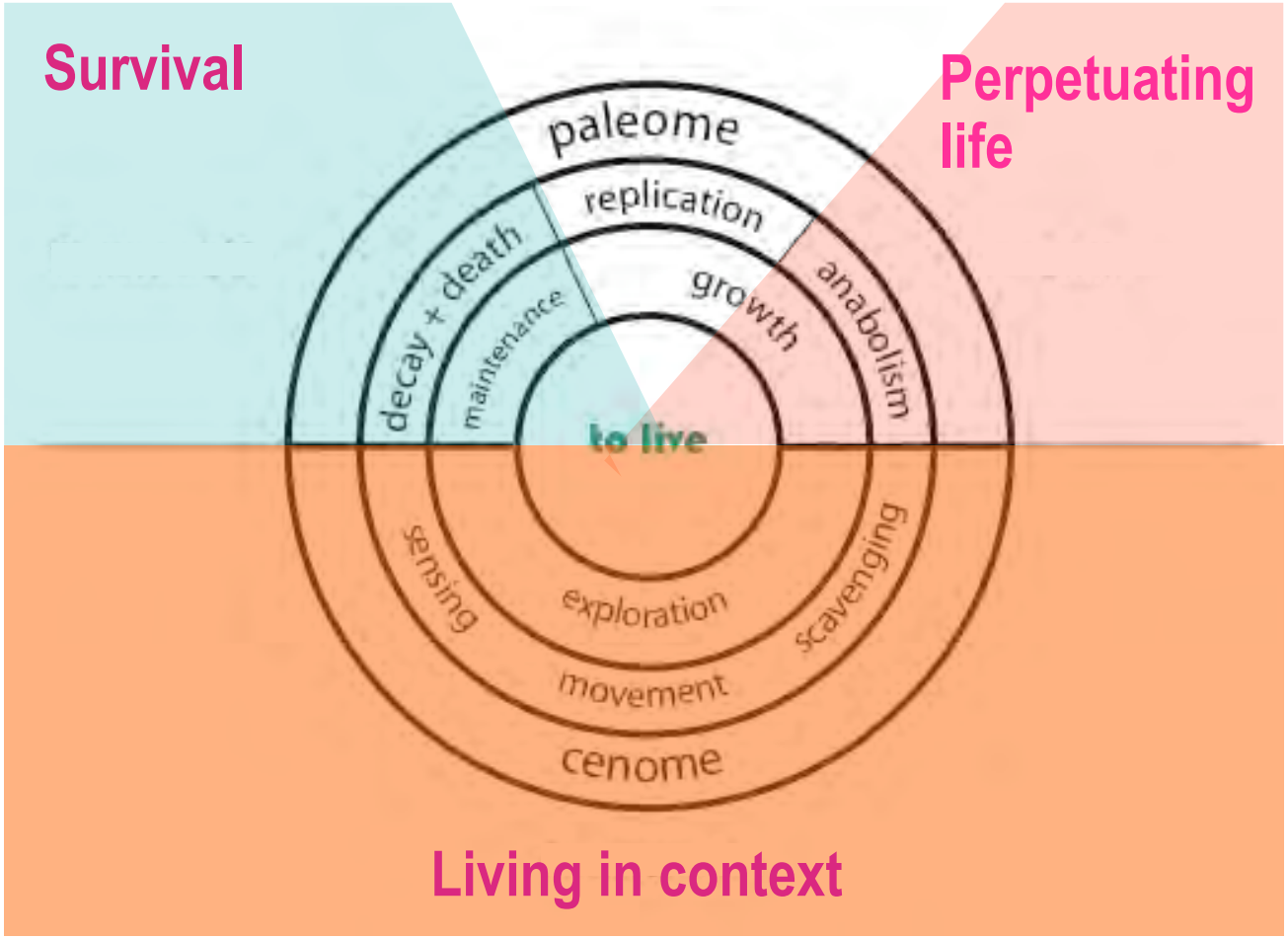
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

A tale of two genomes

metagenomics



$\Sigma = \{pan\text{-}genome\}$

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Reproduction and replication

- **The machine reproduces**
  - Reproduction can improve over time: it is always an aged organism that gives birth to a young one (this implies creation of information)
- **The program replicates**
  - Replication keeps accumulating errors

Which genes permit accumulation of information?

Revisiting information

Improvement of metabolism can be conceptually tolerated as creation of information is reversible (Landauer, 1961; Bennett, 1982, 1988); accumulating information requires an energy-dependent process for making room

Open question: « 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?

A split paleome

- **Paleome 1 (essential genes)**
  - **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, (ATP)-dependent degradation
  - Metabolic patches (chemical frustration)

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

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.

Coevolution of ribonucleases

	RNases and putative RNases	Co-evolving accessory proteins
Co-evolution with 16S rRNA (*B. subtilis*)	**RnhB**, **Rnc**, **PnpA**, **Rnr**	**YpfD**, SmpB, **Eno**, TpiA, Pgk, GapA, GlmM, Zwf
Co-evolution with RNase J (*B. subtilis*)	RnmV, MrnC(YazC), YacP, YmdA, YhaM, YkzG, YlbM, YlmH, YloA, RnjB(YmfA), RnhC(YsgB), NrnA(YtqI), YusF, YybT	Cca, CshB(YqfR), YjbKLMN, YsnB GlcU, GlcT, GlcK, Pgi, GntK CggR, FruR, UgtP,
Co-evolution with 16S rRNA (*E. coli*)	**Rnr**, **Rnc**, **Pnp**, **RnhA**	SpoT, **RpsA**, Hfq GlpK, Zwf, TpiA, Pgk, **Eno**
Co-evolution with RNase E (*E. coli*)	Orn, Rng, Rnd	PcnB, Cca, RelA, Ppx, YhcM PykA, Glk

Polyphosphates

- Poly-P synthesis and degradation appear in non-essential persistent 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…

VBNCs and adaptive mutations

- The status of « viable but not culturable cells » is predicted by the model (lack of regerative potential)
- 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

Miscellaneous predictions

- Bacterial peristence in hosts is dependent on non-esssential persistent genes
- Cancer stem cells are (stem) cells that discovered adaptive mutations allowing them to create a immortal process of progeny
- Accumulation of information in the brain (learning and memory) implies processes making room while preserving functional connections in an energy-dependent manner

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

A synthetic cell?

- The engineering view of SB precludes innovation in synthetic cells
- It is possible to exclude genes permitting accumulation of information
- The consequence is that, as factories, cell factories will age and have to be systematically reconstructed
- This as the considerable societal advantage that the associated risks are minimised

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS

Acknowledgements

The paleome

TingZhang Wang proteome analysis and ageing
Gang Fang, Eduardo Rocha gene persistence
Isabelle Martin-Verstraete et al sulfur anabolism and regulation
Anne-Marie Gilles nucleotide anabolism
Undine Mechold, *Sandra Cescau* et al RNA metabolism

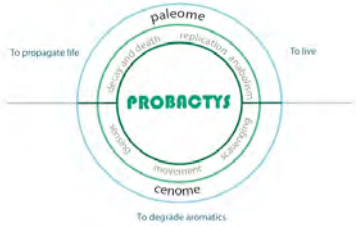
The paleome and the cenome

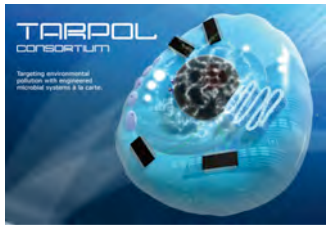
Agnieszka Sekowska, *Conghui You*, *Jean-Baptiste Masson,* Andrew Martens sulfur salvage and metabolic patches and physics of the individual cell in its collective behaviour

The cenome
Evelyne Krin, Evelyne Turlin, *Sabina Chalabaev,* from commensalism to pathogenicity
Sandra Cescau et al intracellular colonisation of red blood cells

Collaborations
Francis Biville, Massimo Vergassola et al
GenoScope
Beijing Genome Institute, Fudan University, The University of Hong Kong
Hong Kong University of Science and Technology

Funding





FONDATION
SCIENTIFIQUE
FOURMENTIN-GUILBERT

ANR *Bartonella birtlesii* (F Biville)
ANR PNANO (M Vergassola)

"Living organisms as information traps"

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/

INSTITUT PASTEUR

CNRS