SIB 10th Anniversary Conference

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

Antoine Danchin
Directeur du Département Génomes et Génétique, Institut Pasteur
Bern 24 September 2008
IN VIVO, IN VITRO, IN SILICO

• Biology
  • Biochemistry
  • Genetics

• Number Theory and Computer Sciences
  • Turing Machines
  • Algorithmics

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
## SPEED AND DATABASES

<table>
<thead>
<tr>
<th>Year</th>
<th>Speed or Storage</th>
<th>1986 &gt; 1 GigaFlops</th>
<th>1986 &gt; 9 million bps at EMBL/GenBank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>1 TeraFlops</td>
<td></td>
<td>1997 &gt; 1 billion bps at DDBJ/EMBL/GenBank</td>
</tr>
<tr>
<td>2008</td>
<td>1 PetaFlops</td>
<td></td>
<td>2008 &gt; 200 billion bps at the INSDC</td>
</tr>
<tr>
<td>2019</td>
<td>1 ExaFlops?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"Pioneers at the Heart of Science" 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
S I B

databases

literature

INFORMATION

knowledge

http://www.pasteur.fr/recherche/unites/REG/literature/databases
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

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
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  La barque de Delphes, Odile Jacob, 1998
The Delphic Boat, Harvard University Press, 2003

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

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
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 ».

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
However, here is the situation ...
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/
INFORMATION IS A FIFTH CATEGORY OF NATURE

Classical Physics

\[ E = mc^2 \]

energy  matter  space/time

Quantum Physics

\[ \Delta x \Delta p \geq \frac{\hbar}{4\pi} \]

uncertainty = lack of information

“Pioneers at the Heart of Science” 1998 – 2008

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

Matter / Energy / Space / Time

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

“This Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
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
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

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
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
VENTER’S DEMONSTRATION

The Turing machine

May exist in a parallel set up

Genome transplantation

“Pioneers at the Heart of Science” 1998 – 2008

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
“Beside the genetic program, the cell carries a considerable amount of information…”
It is not enough to have a DNA molecule with the right sequence, it needs to be correctly folded!

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
REPRODUCTION vs REPLICATION

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

- Replication accumulates errors
- Reproduction can improve over time

> This process implies creation of information
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
BIAS IN AMINO ACID DISTRIBUTION

Distribution of amino acids in the proteome of *Psychromonas ingrahamii* (-12°C, 100 h gt)


“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
Main post-translational modification

Reaction still poorly understood

Spontaneous reaction (untargeted?)

Affects the protein structure (and function?)

Role in regulating protein folding

Signal for degradation of intracellular proteins

Asparagine deamidates: a major contribution to protein ageing

Asparagine (N)

Intermediary: succinimide

Degradation: succinimide

Aspartate

Isoaspartate

“Pioneers at the Heart of Science” 1998 – 2008

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

Variation / Selection / Amplification

Stabilisation

Evolution

creates

Function

captures (recruits)

Structure

codes

Sequence

“Pioneers at the Heart of Science” 1998 – 2008

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
Functional ubiquity does not imply structural ubiquity. Efficient objects tend 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.

400-500 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.
CONSERVATION OF GENE CLUSTERING

The paleome (from the Greek παλαιος, ancient)

The genome (from the Greek κοινος, common)

Genome core <2000 genes

Variable genes already > 50000 genes

Clustering frequency

Known

Unknown

Rifampicin
Streptomycin
Tetracycline
Penicillin
Vancomycin
Isoniazide

Frequency in genomes

Pseudomonas putida
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 παλαίος, ancient).
A MINERAL SCENARIO FOR THE ORIGIN OF LIFE

• The surface of charged solids (e.g. pyrite (Fe-S)) selects and compartmentalises 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
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

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
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)
A TALE OF TWO GENOMES

Survival

Perpetuating life

Σ = \{pan-genome\}

Living in context

“Pioneers at the Heart of Science” 1998 – 2008

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

“Pioneers at the Heart of Science” 1998 – 2008
Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
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)

---

“Pioneers at the Heart of Science” 1998 – 2008

Genetics of Bacterial Genomes, Institut Pasteur / URA2171, CNRS
http://www.pasteur.fr/recherche/unites/REG/
• 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.
• 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…
The status of « viable but not culturable cells » is predicted by the model (lack of regenerative 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
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
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 metabolism  
Undine Mechold, Sandra Cescau et al RNA metabolism

The paleome and the cenome

Agnieszka Sekowska, Conghui You, Andrew Martens sulfur salvage and metabolic patches

The cenome

Evelyne Krin, Evelyne Turlin, Sabina Chalabaev, Jean-Baptiste Masson

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  
Probactys, Tarpol, Microme