from eec genes to maxwell's demon's genes
antoine danchin 唐善・安東

AMAbiotics SAS

piotr slonimski, from mitochondria to genomes
gif sur yvette, july 8, 2010
in 1991, europe was pioneering microbial genomics with two programmes supported by the european commission: sequencing of the genomes of \textit{saccharomyces cerevisiae} and \textit{bacillus subtilis}.

at the elounda meeting, the first discovery of genomics was presented: the sequence of yeast chromosome III and 100 kb of the genome of \textit{b. subtilis} revealed that > 50 \% of the cds did not look like anything known yet; this was a complete surprise as the adversaries of the genomic programmes argued that nothing unknown would be discovered.

piotr slonimski proposed to name these genes "elusive, esoteric, conspicuous" genes, \textit{eec genes}, to celebrate the support of the eec.
a minimal set of functions
2003: 60 conserved proteins

the number of conserved genes tends to zero!
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 and adaptation to transient phenomena ; a fraction manages the evolution of the organism

a common structural feature: persistent genes are located in the leading DNA strand
is « ubiquitous » synonymous with « essential »?

genes essentiels and ....

too many genes!

energy-dependent degradation metabolic patches

is « ubiquitous » synonymous with « essential »?

persistent genes

genes essentiels and ....

too many genes!

energy-dependent degradation metabolic patches

g fang, ep rocha, a danchin
how essential are non-essential genes?
mol biol evol (2005) 22: 2147-2156
clustering frequency in genomes

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

paleome (from παλαιος, ancient): the cell factory

Pseudomonas putida

Kuiper's Test Statistic

known
unknown
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
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" (cell factory) and "data/program" (genome)
Genome transplantation in bacteria: changing one species to another


caveat: reproduction vs replication

the program replicates
(makes an identical copy)

the cell reproduces
(makes a similar copy)

this split is the basis of evolution
the machine reproduces
  reproduction can improve over time: it is always an old organism that gives birth to a young one (this implies creation and accumulation of information)

the program replicates
  replication progressively accumulates errors

which genes permit accumulation of information?
intuition tells us that creation of information asks for energy

**wrong**: creation of information is reversible (Landauer, 1961; Bennett, 1982, 1988; Zurek, 1989); yet, to accumulate information requires an energy-dependent process to "make room", without erasing valuable information

**open question:**
- can we identify in genomes the genes coding for the functions that permit this process?
- can we find a ubiquitous and stable energy source?
classical models of information do not take "meaning" into account, nor the value of an 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 come from?

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

The demon accumulates information or reverses time if it can measure the speed and the position of the atoms of gas, collecting an information to calculate when it must close the trap.
ATP \rightarrow \text{poly}(P)_n \rightarrow ADP + Pi \rightarrow \text{poly}(P)_{n-1} + Pi \leq \text{in the paleome}

the degradation machinery uses energy to reject unaltered a functional entity

non functional entities are recognised and degraded
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
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
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 at 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
predictions

- bacterial persistence in a host depends on persistent non-essential genes

- initiation of cancer comes from cells (stem cells) that discovered adaptive mutations that permit them to generate an immortal progeny

- accumulation of information in the brain (memory and learning) implies the existence of processes to make room while preserving functional connections, in a way which must be energy-dependent
• the engineering view of SB precludes that artificial cells be innovative
• we can exclude the genes permitting 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

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