THE CONSTRUCTOR AND THE REPLICATOR: PREREQUISITES FOR THE CONSTRUCTION OF A SYNTHETIC CELL

What are the theoretical tools most useful for understanding biological systems?

IHES, 13 november 2007

Genetics of Bacterial Genomes
http://www.pasteur.fr/recherche/unites/REG/

ACKNOWLEDGEMENTS

The University of Hong Kong
Dpt of Mathematics and HKU-Pasteur Research Centre

- Stanislas Noria (collective name, working seminar in « Conceptual Biology »)

Génétique des Génomes Bactériens (in silico)

- Gang Fang
- Etienne Larsabal
- Géraldine Pascal
- Eduardo Rocha

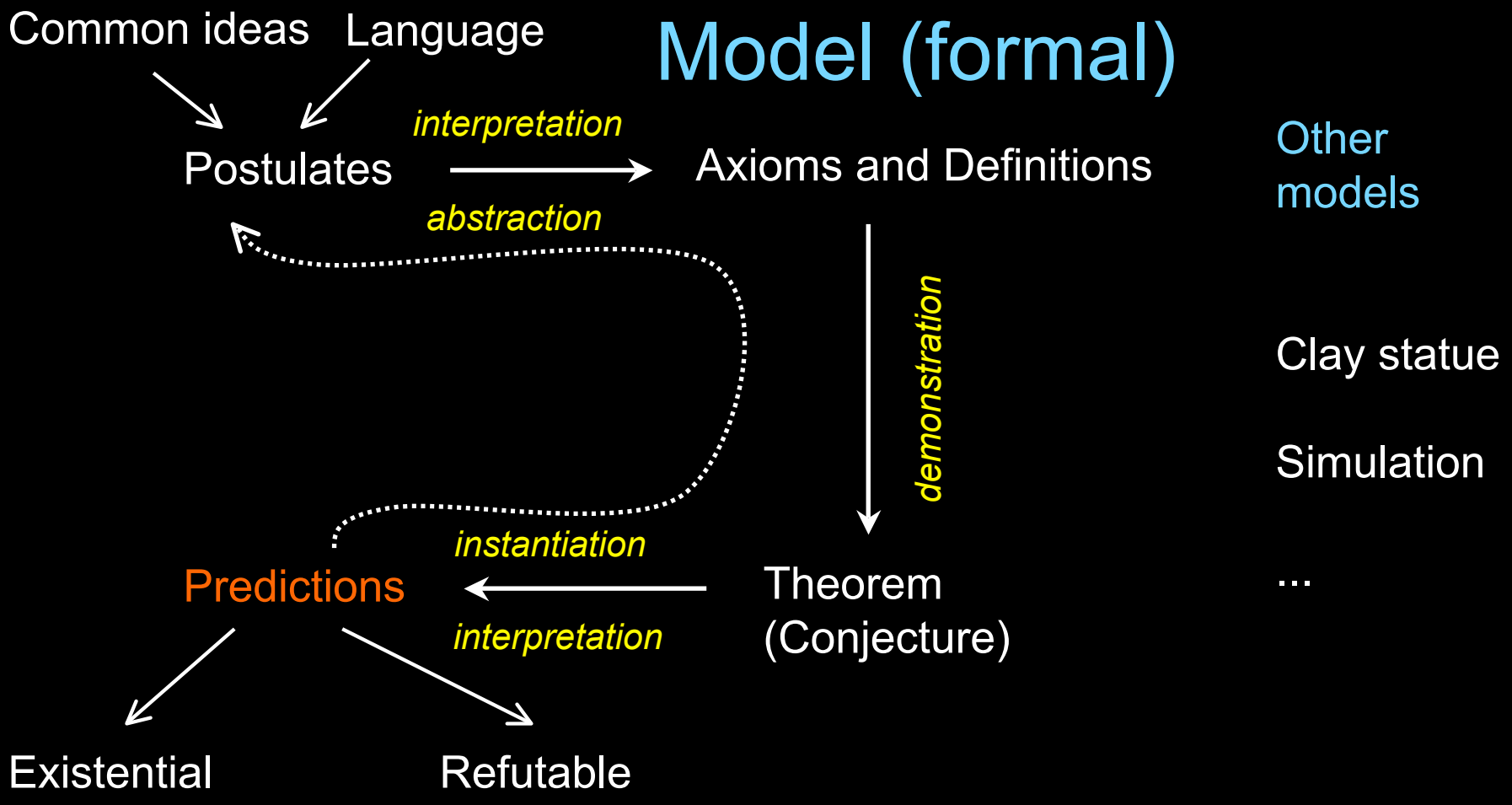
Génétique in silico

- Marc Bailly-Béchet
- Massimo Vergassola

Génoscope AGC

- Claudine Médigue

THE EXTINCTION OF POPPERISM: THE CRITICAL GENERATIVE METHOD



INSTITUT PASTEUR

MODELS AND REALITY

- Against Popper: coupling a model with reality is not straightforward (abstraction and instantiation)
- Several models can account for the same reality (e.g. quantic and classical views in Nuclear Magnetic Resonance)
- « Falsification » is not a direct outcome of a contradiction with an experimental prediction
- Biology is, in depth, particularly abstract; however properties of life are implemented as specific components which have idiosynchratic properties; this results in the multiplication of « anecdotes »

INSTITUT PASTEUR

THREE REVOLUTIONS

- 1944 - 1985 MOLECULAR BIOLOGY
- 1985 - 2005 GENOMICS
- 2005 - … SYNTHETIC (SYMPLECTIC) BIOLOGY (highly multidisciplinary !)

« Symplectic » is in Greek (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

- LIFE AND COMPUTATION
- SOME SIMPLE PHYSICAL CONSTRAINTS
- TRANSLATION ORGANIZES THE BACTERIAL GENOME
- THE PALEOME: CONSTRUCTOR AND REPLICATOR
- THE CENOME: THE “PURPOSE” OF THE MACHINE
- REPRODUCTION vs REPLICATION: THE ESSENTIALITY OF METABOLISM

INSTITUT PASTEUR

WHAT LIFE IS

Three co-existing processes constitute life:

- **Metabolism** | a
- **Compartmentalization** | machine

- **Information transfer** | a "program" (declarative, not prescriptive)

The cell is the atom of life

INSTITUT PASTEUR

THE "GENETIC PROGRAM"

- Physics: *matter, energy, time*
- Statistical physics: Physics + *« information »*
- Biology: Physics + *information, coding, control...*
- Arithmetics: *sequences of integers, recursivity, coding…*
- Computation: Arithmetics + *programs* + *machine...*

The « genetic program » metaphor has practical consequences: we know how to manipulate genes and gene products, do we have the conceptual tools to push the metaphor to its ultimate consequences?

INSTITUT PASTEUR

WHAT COMPUTING IS

Two entities permit computing:

- A machine able to read and write
- A program on a physical support, split (in practice, but not conceptually) into two entities:
  - Program (providing the “goal”)
  - Data (providing the context)

The machine is distinct from the program

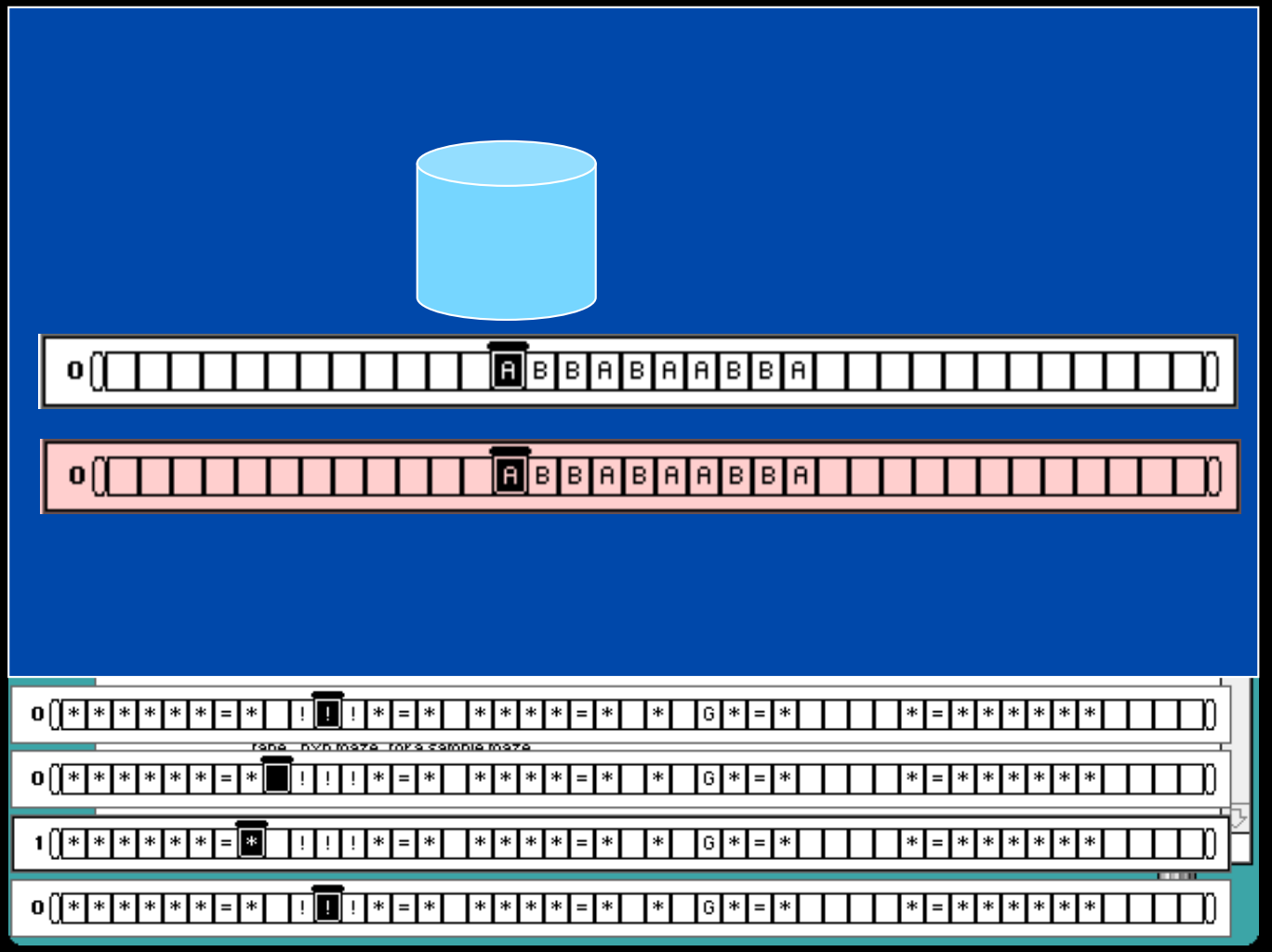
INSTITUT PASTEUR

THE TURING MACHINE

the machine
(read/write)

is physically distinct from

the program
(data)
as a linear
sequence
of symbols



INSTITUT PASTEUR

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»

Need: conceptual analysis of biological information (algorithmic complexity, logical depth...)

AN ALGORITHMIC VIEW OF BIOLOGICAL ACTIONS

Replication, transcription, translation: high parallelism

"Begin, Check Control Points, Repeat, End"

The action is always oriented, with a beginning and an end

The processes of time control (check points) are rarely taken into account (except for the replication/division processes), but their role is essential to allow coordination of multiple actions in parallel

Need: conceptual analysis of check points; experimental identification

INSTITUT PASTEUR

IS THERE A MAP OF THE CELL IN THE CHROMOSOME?

John von Neumann, trying to understand the brain, suggested that were a computer both to behave as a computer and to construct the machine itself, it should harbour an image of the machine somewhere

That special computer had to be split into a replicator and a constructor, which expresses the program for construction of both the replicator and the constructor

The metaphor does not appear to apply to the brain, does it apply to the cell?

INSTITUT PASTEUR

A GENETIC COMPUTER?

- In a computer the machine is distinct from data and program
- In the cell, data and program play the same role (they are ‘ declarations ’ not prescriptions); **they can be modified by the machine (Pol IV, Pol V...)**
- General reflection (Number Theory) considers the actions of the machine, but not the way it is constructed

INSTITUT PASTEUR

PARADOX

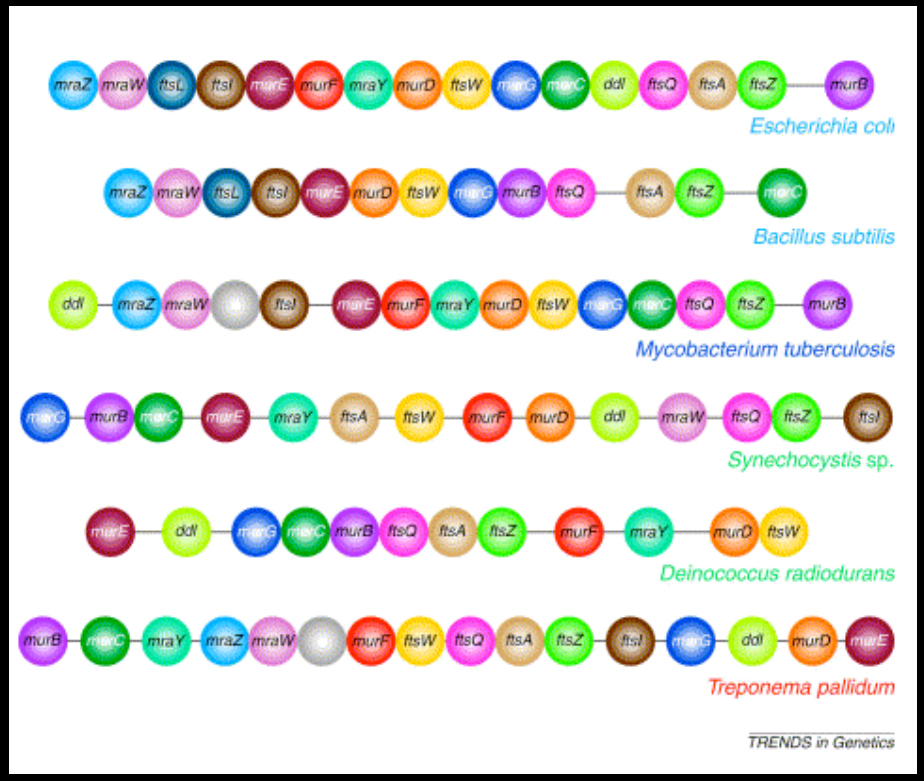
If the machine has not only to behave as a Turing machine but also to make the machine, one must find a geometrical program somewhere in the machine (J. von Neumann)

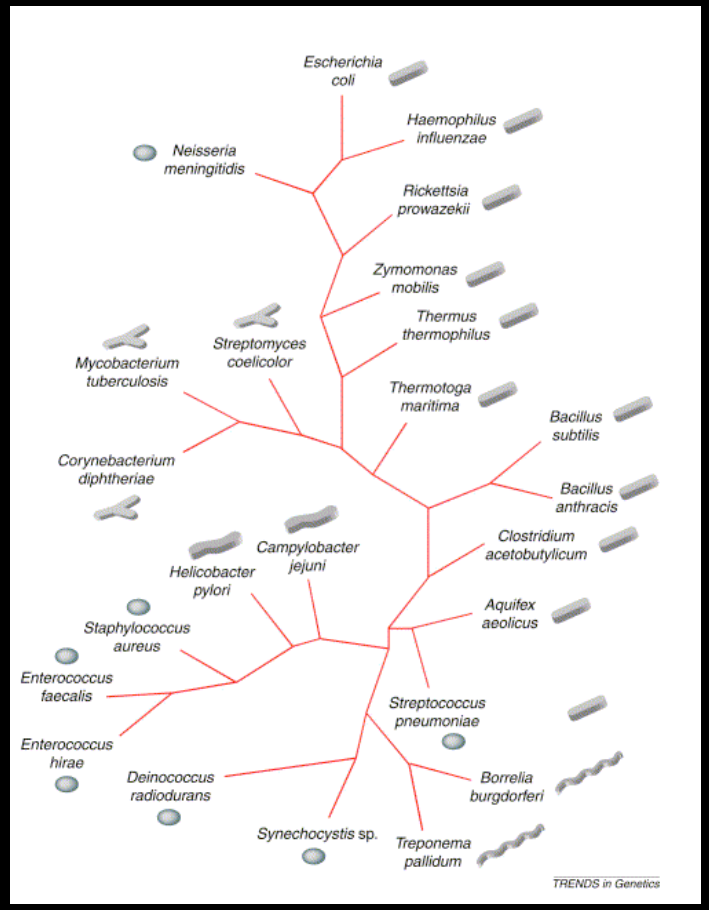
Is there an image of the organism in the genome?

INSTITUT PASTEUR

GENE ORDER AND CELL SHAPE

The mur-fts cluster





Tamames J, Gonzalez-Moreno M, Mingorance J, Valencia A, Vicente M
Bringing gene order into bacterial shape
Trends in Genetics (2001) **17**: 124-126

Genetics of Bacterial Genomes
http://www.pasteur.fr/recherche/unites/REG/

- LIFE AND COMPUTATION
- SOME SIMPLE PHYSICAL CONSTRAINTS
- TRANSLATION ORGANIZES THE BACTERIAL GENOME
- THE PALEOME: CONSTRUCTOR AND REPLICATOR
- THE CENOME: THE "PURPOSE" OF THE MACHINE
- REPRODUCTION vs REPLICATION: THE ESSENTIALITY OF METABOLISM

INSTITUT PASTEUR

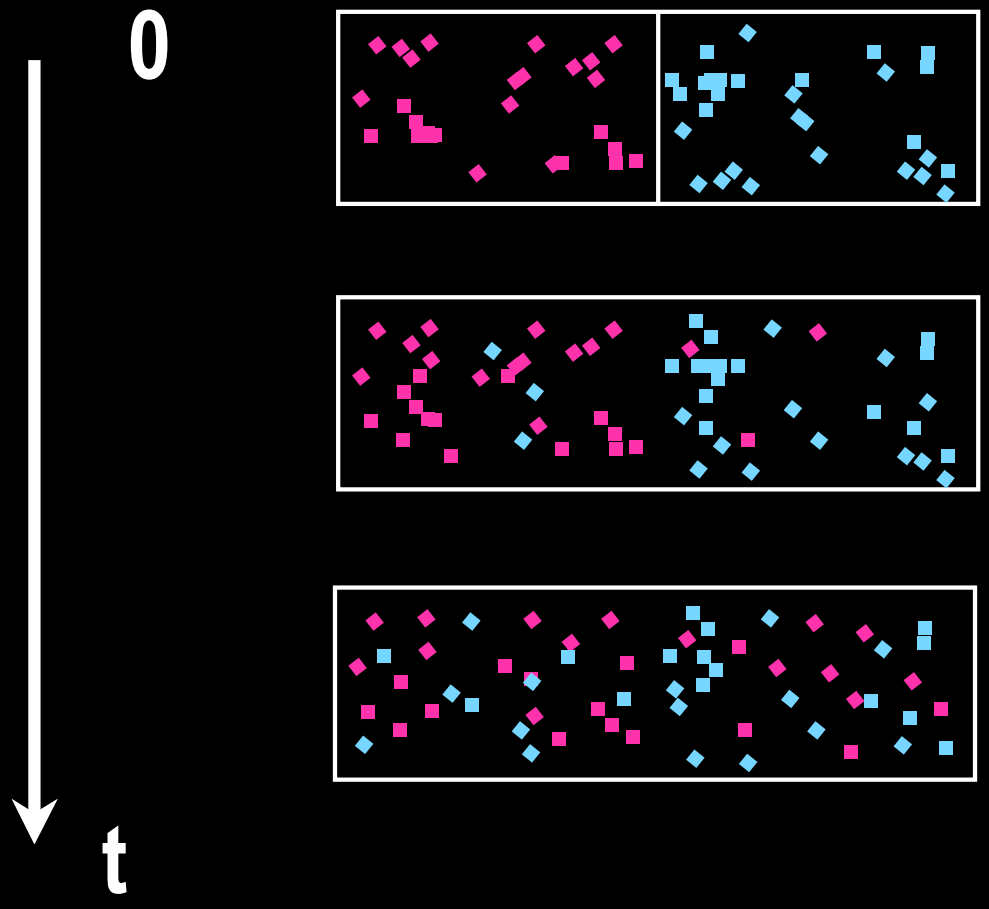
PHYSICS OF REPLICATION

- DNA forms a long folded thread: how do the daughter molecules separate?
- Are physical constraints reflected in the sequence?
- [Replication is oriented: the physics of a strand cannot be that of its complement]

A correct use of physics helps!

INSTITUT PASTEUR

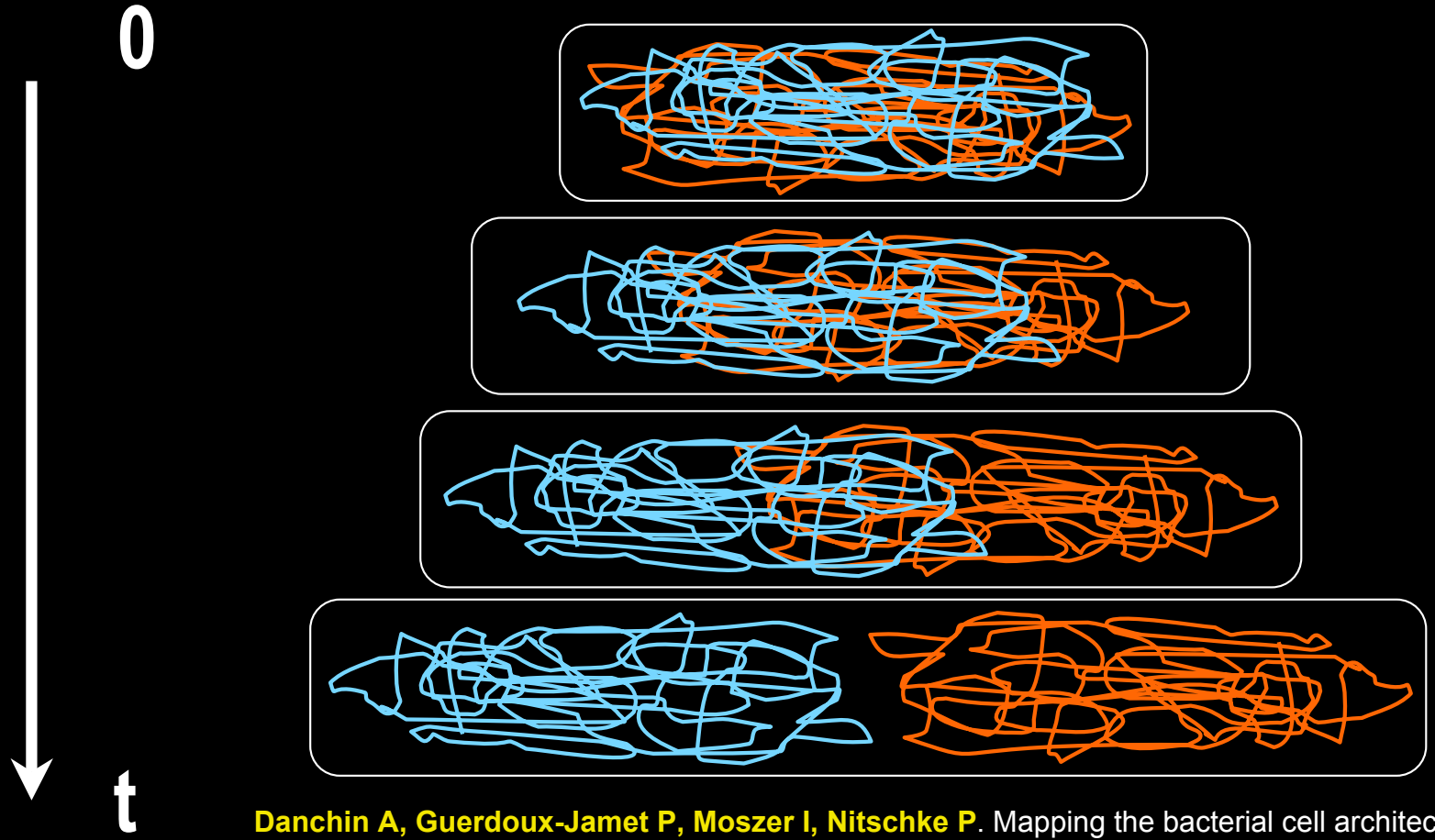
A TEXTBOOK VIEW OF ENTROPY



$$S = k \log \Omega$$

INSTITUT PASTEUR

AN INCREASE IN ENTROPY IS ENOUGH TO SEPARATE CHROMOSOMES



Danchin A, Guerdoux-Jamet P, Moszer I, Nitschke P. Mapping the bacterial cell architecture into the chromosome. *Philos Trans R Soc Lond B Biol Sci* 2000, **355**:179-190

Jun S, Mulder B. Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome .Proc Natl Acad Sci U S A. 2006 103:12388-12393

INSTITUT PASTEUR

CONCEPTS AND PATCHES

The processes constituting life can be analyzed conceptually. They need however to be implemented with concrete objects, having idiosynchratic properties. The DNA sequence cannot be a smooth linear double helix, simply because of the chemical nature of its nucleotides; it winds, turns and bends. However it needs to be recognized by control or structural elements. How can these divergent constraints be reconciled?

INSTITUT PASTEUR

RECURSIVE MODELLING

- Realistic Model 1 <=> Real sequence
  Prediction 1
- Realistic Model 2 <=> Real sequence
  Prediction 2
- Realistic Model 3 <=> Real sequence
  Prediction 3

.....

INSTITUT PASTEUR

CONSTRAINTS IN THE DNA SEQUENCE

Evolution optimises replication, while DNA needs also to support gene sequences

This is witnessed by:

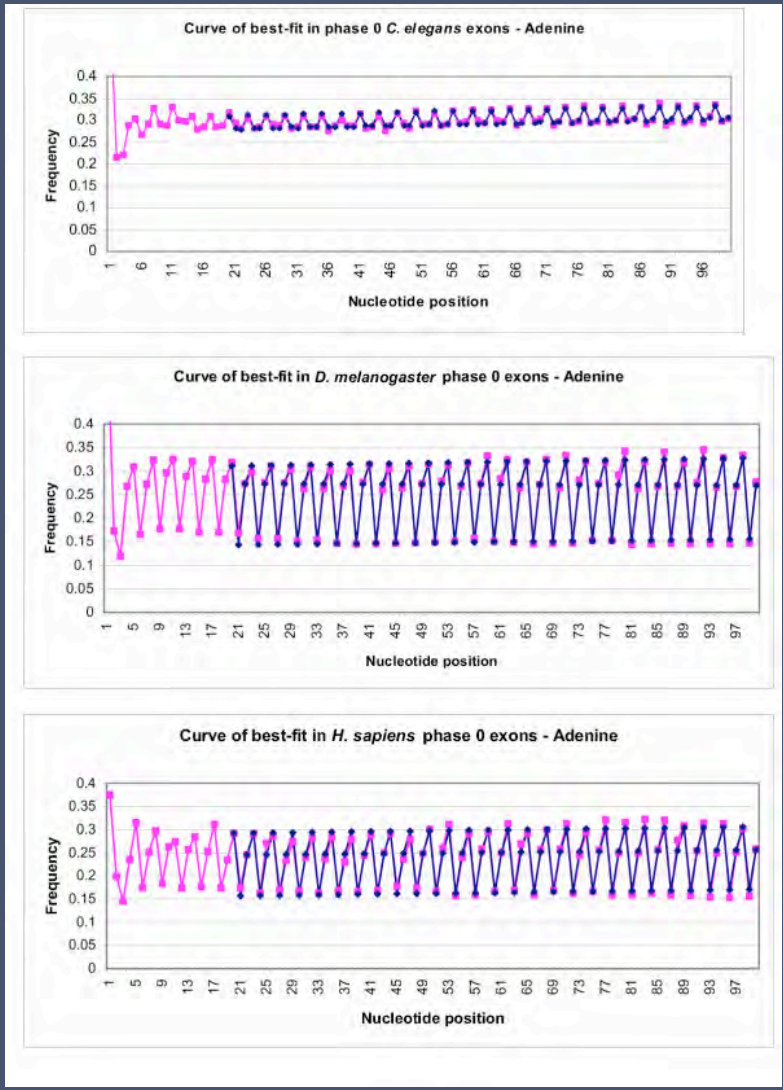
- A period 3, signature of the codon succession in genes (constrained by the genetic code rule)

- A period 10-11.5 of yet unknown function...

PERIODS IN GENOMES

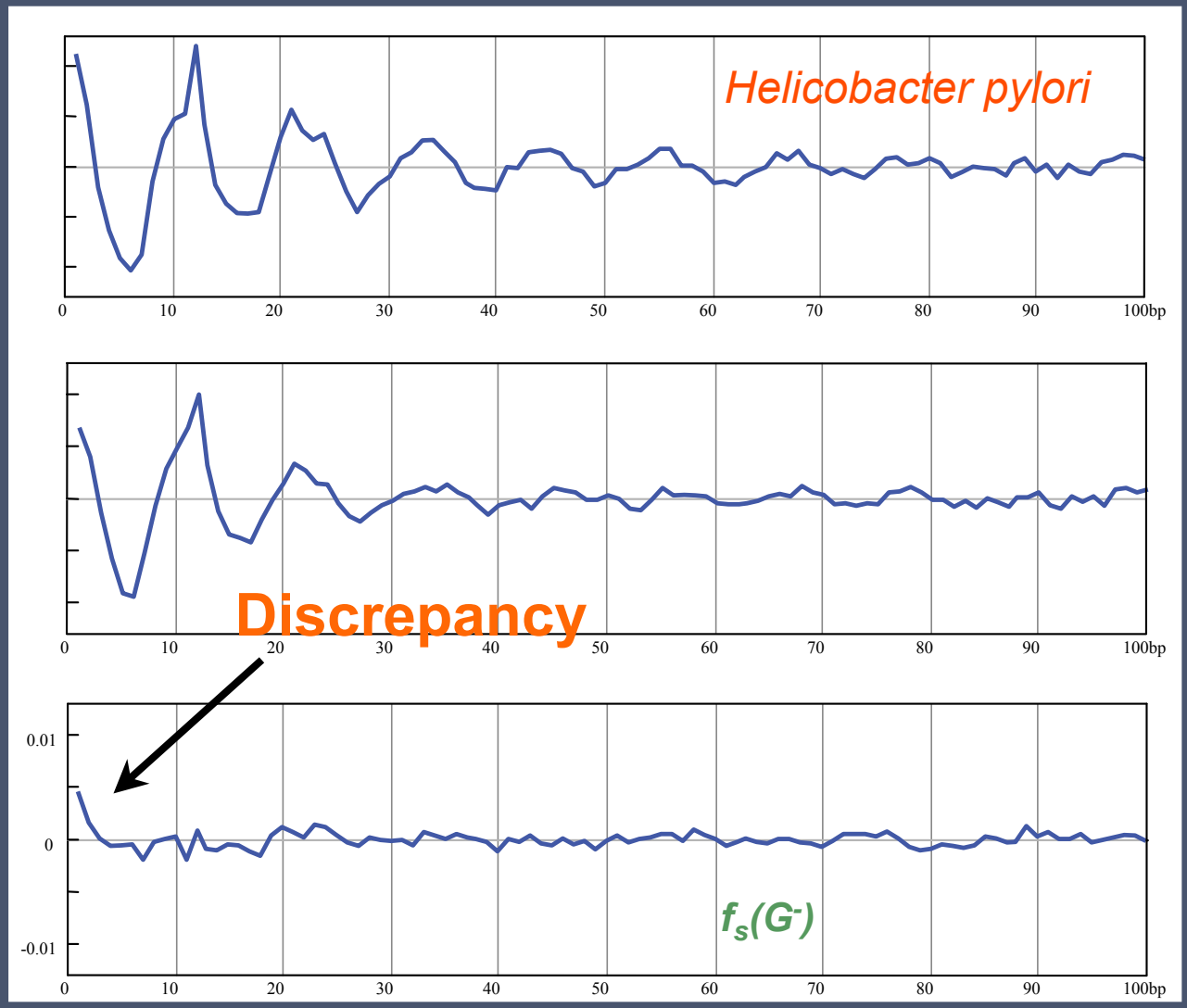
One observes a correlation between base pairs with period three

After deconvolution of this period there remains a somewhat fuzzy period of 10 to 11.5 base pairs



Eskesen et coll. BMC Molecular Biology Volume 5, 12, 2004

A UNIVERSAL FEATURE OF THE GENOME TEXT: 10-11.5



Genetics of Bacterial Genomes
http://www.pasteur.fr/recherche/unites/REG/

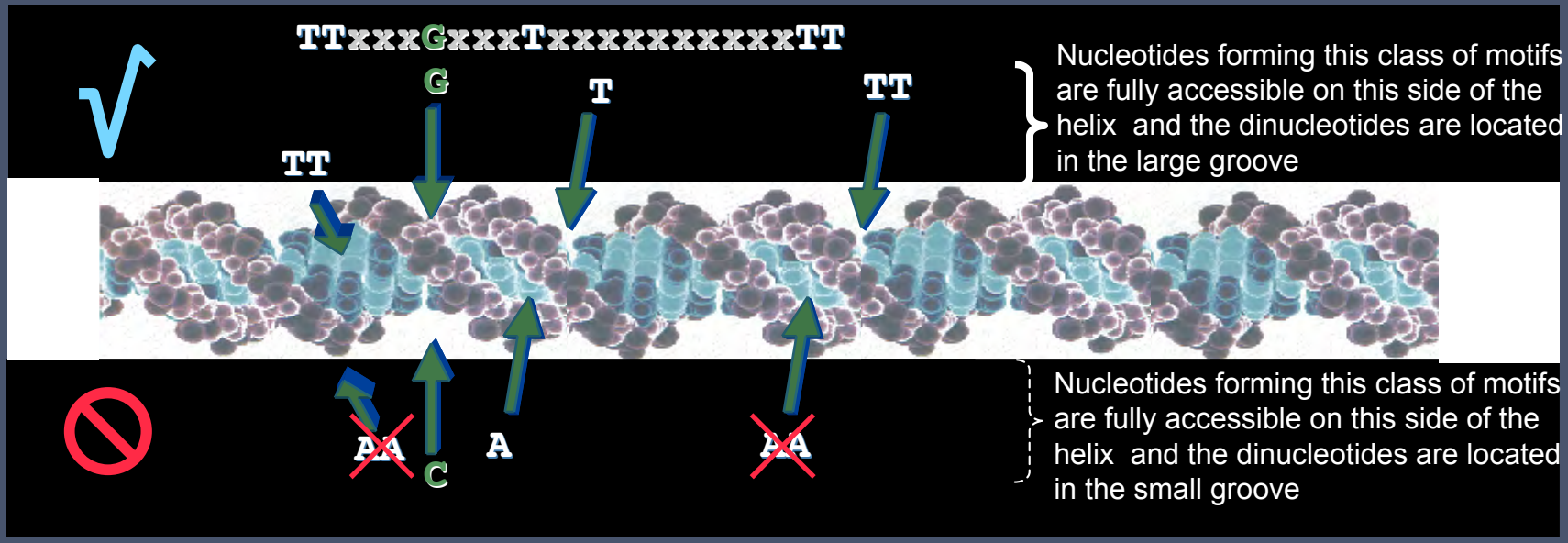
TYPE A FLEXIBLE MOTIFS

1-xAxxxxTxxxxAxxxxTTxxxxxAxxxxTxxxxAxxx: All domains

2-xxxxxxxxxxxGxxxxTTxxxCxxxxxTxxxxxxxxx: Proteobacteria

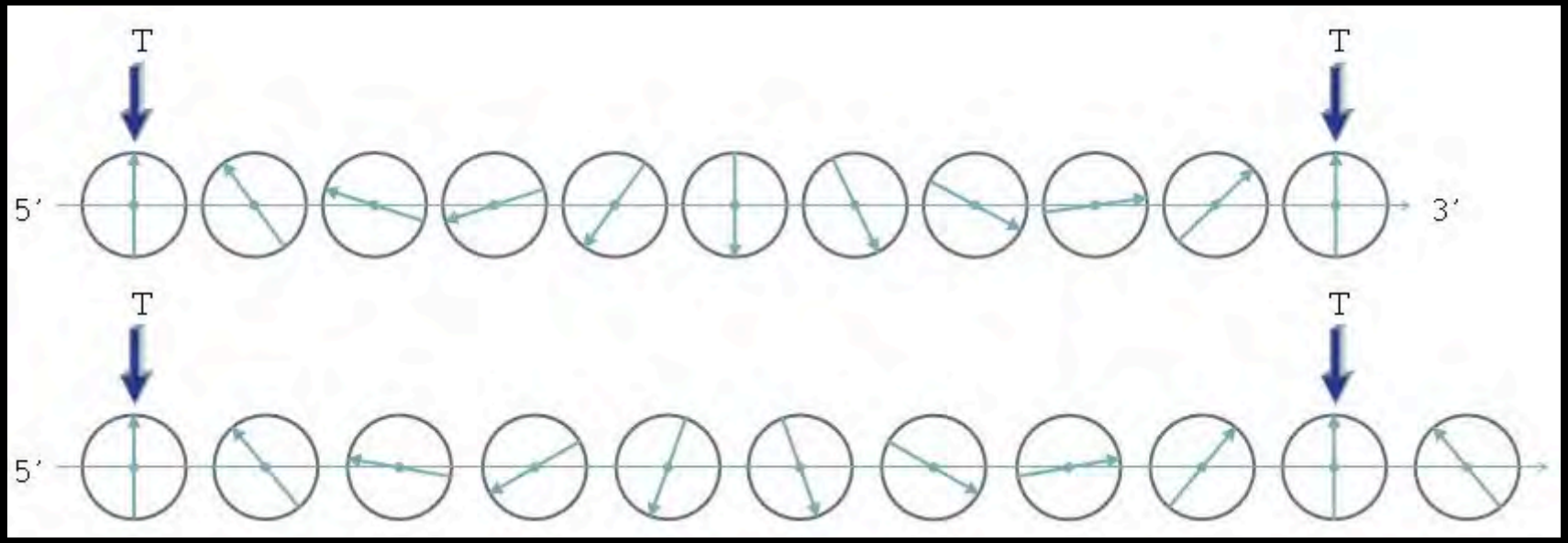
4-xxxxxxTxxxxAGxxxTTxxxxxxxxTxxxxxxxxxx: Archaea

5'-xxx-10xxxxxxxxx0xxxxxxxx10xxxxxxbp-3'



INSTITUT PASTEUR

FLEXIBLE MOTIFS ACCOMMODATE LOCAL VARIATIONS OF THE DNA STRUCTURE



The flexiblility of these motifs allow DNA to take into account superturns and bends

Larsabal, E, Danchin, A
Genomes are covered with ubiquitous 11 bp periodic patterns, the "class A flexible patterns »
BMC Bioinformatics. 2005 6:206

INSTITUT PASTEUR

OPEN QUESTIONS

- The constraints resulting from the presence of flexible motifs is so large that it should be visible in gene products
- It may result in non random distribution of genes if some functions are associated to regularities in proteins (alpha helices, beta sheets, beta turns etc)

- LIFE AND COMPUTATION
- SOME SIMPLE PHYSICAL CONSTRAINTS
- TRANSLATION ORGANIZES THE BACTERIAL GENOME
- THE PALEOME: CONSTRUCTOR AND REPLICATOR
- THE CENOME: THE "PURPOSE" OF THE MACHINE
- REPRODUCTION vs REPLICATION: THE ESSENTIALITY OF METABOLISM

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)

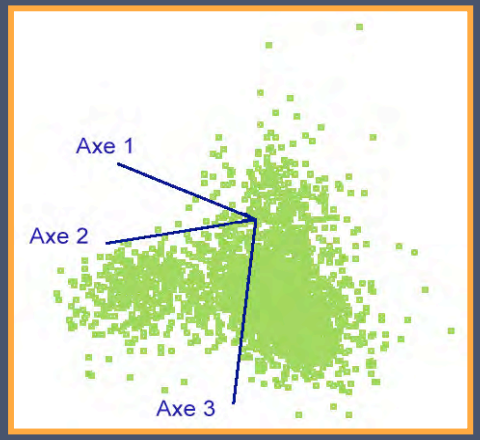
Independent Component Analysis uses the non gaussian character of the values associated to descriptors; it characterizes objects belonging to common independent clusters (the « cocktail party » theorem), (Hérault, 1984)

Further methods need to be developed

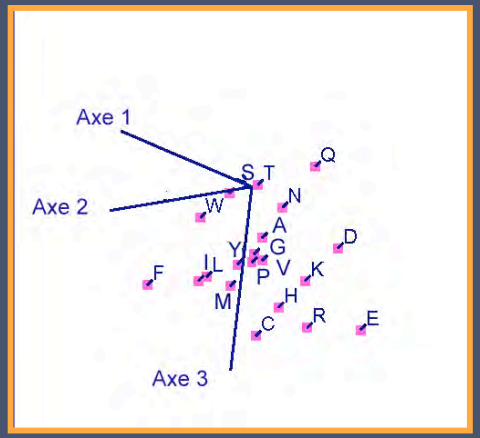
INSTITUT PASTEUR

Correspondence Analysis (CA)

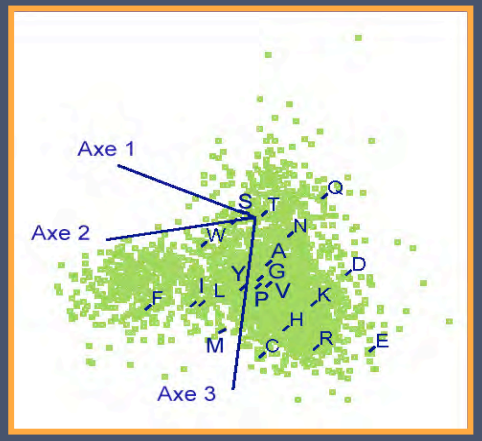
Factorial space of the proteins



Factorial space of the amino acids



Superimposition of both spaces (clouds)



INSTITUT PASTEUR

UNIVERSAL BIASES IN PROTEIN AMINO ACID COMPOSITION

- **First axis:** separates Integral Inner Membrane Proteins (IIMP) from the rest; driven by opposition between charged and large hydrophobic residues

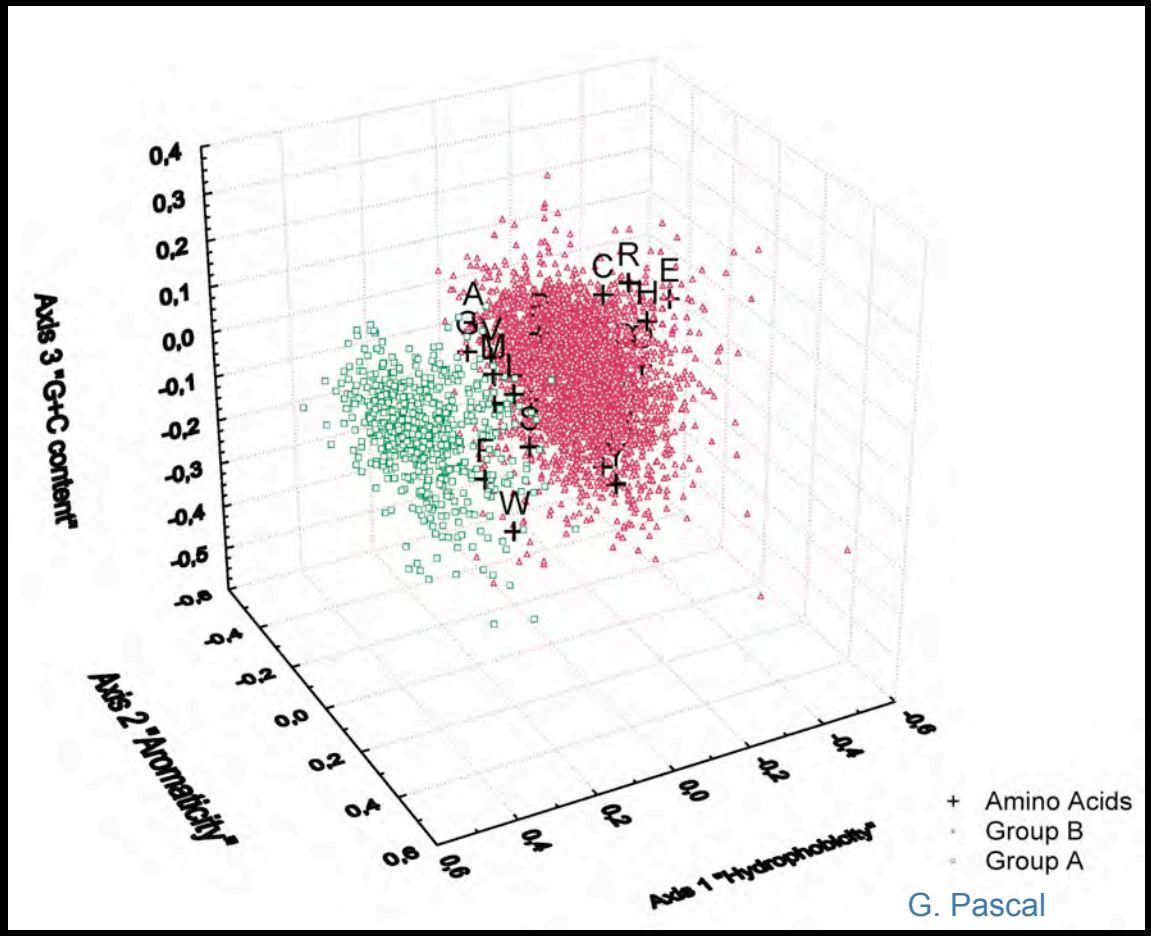
- **Second axis:** separates proteins by their content in aromatic amino acids; enriched in orphan proteins

- **Third axis:** separates proteins according to an opposition driven by the G+C content of the *first* codon base

INSTITUT PASTEUR

BIAS IN AMINO ACID DISTRIBUTION

Neighborhood: distribution of aminoacids in the proteome



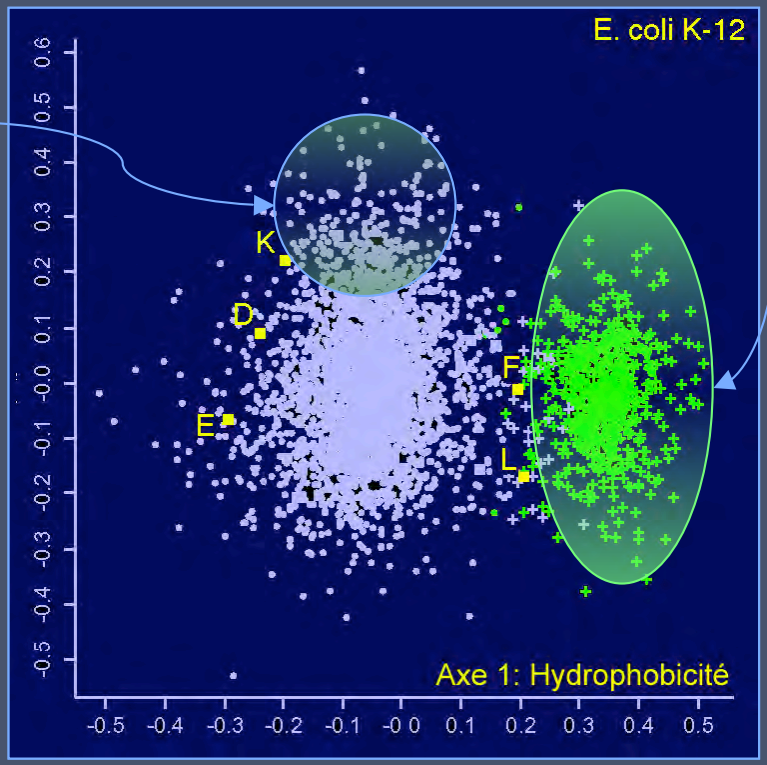
G. Pascal

INSTITUT PASTEUR

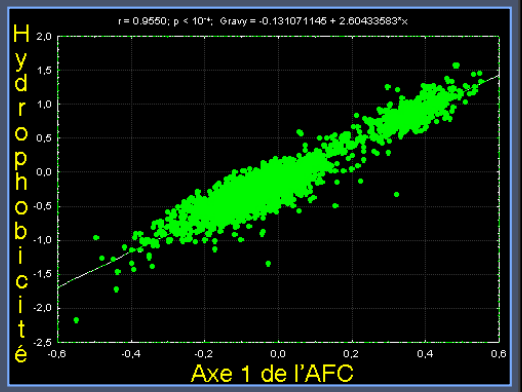
BIASES IN HYDROPHOBICITY AND AROMATICITY OF PROTEINS

A strong bias opposing charged residues to hydrophobic residues

Proteins of the outer membrane: OmpT. OmpL…



Proteins of the inner membrane: LacY. SecE…



INSTITUT PASTEUR

TEMPERATURE-DEPENDENT BIASES IN PROTEIN AMINO ACID COMPOSITION

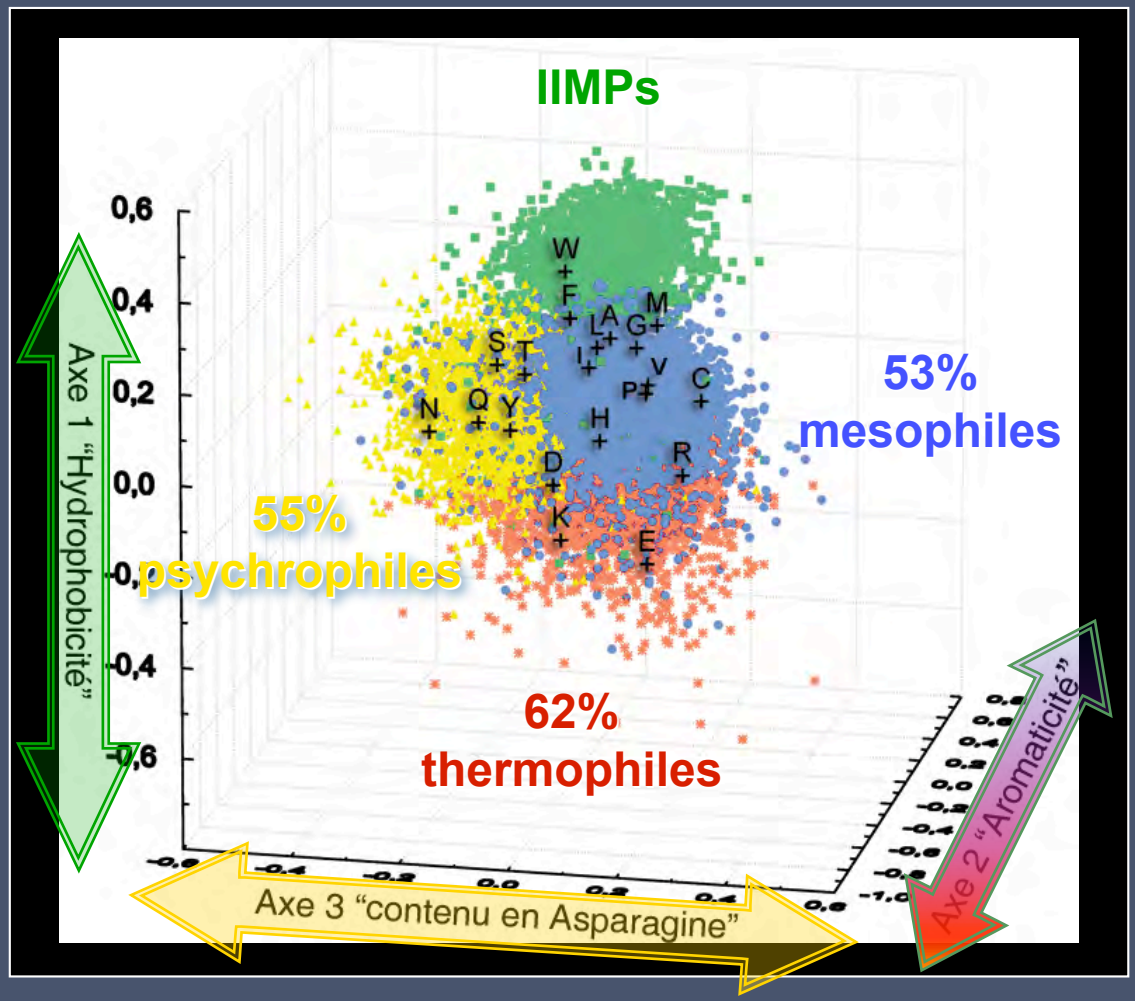
- The general trend of amino acid composition bias is to avoid some aminoacids at higher temperatures (associated to aging processes)
- Mesophilic bacteria belong to at least two different classes (in a 5-clusters analysis)
- Biases are always dominated by the IIMP clustering

INSTITUT PASTEUR

COMPARATIVE PROTEOMICS

A specific asparagine bias in psychrophiles

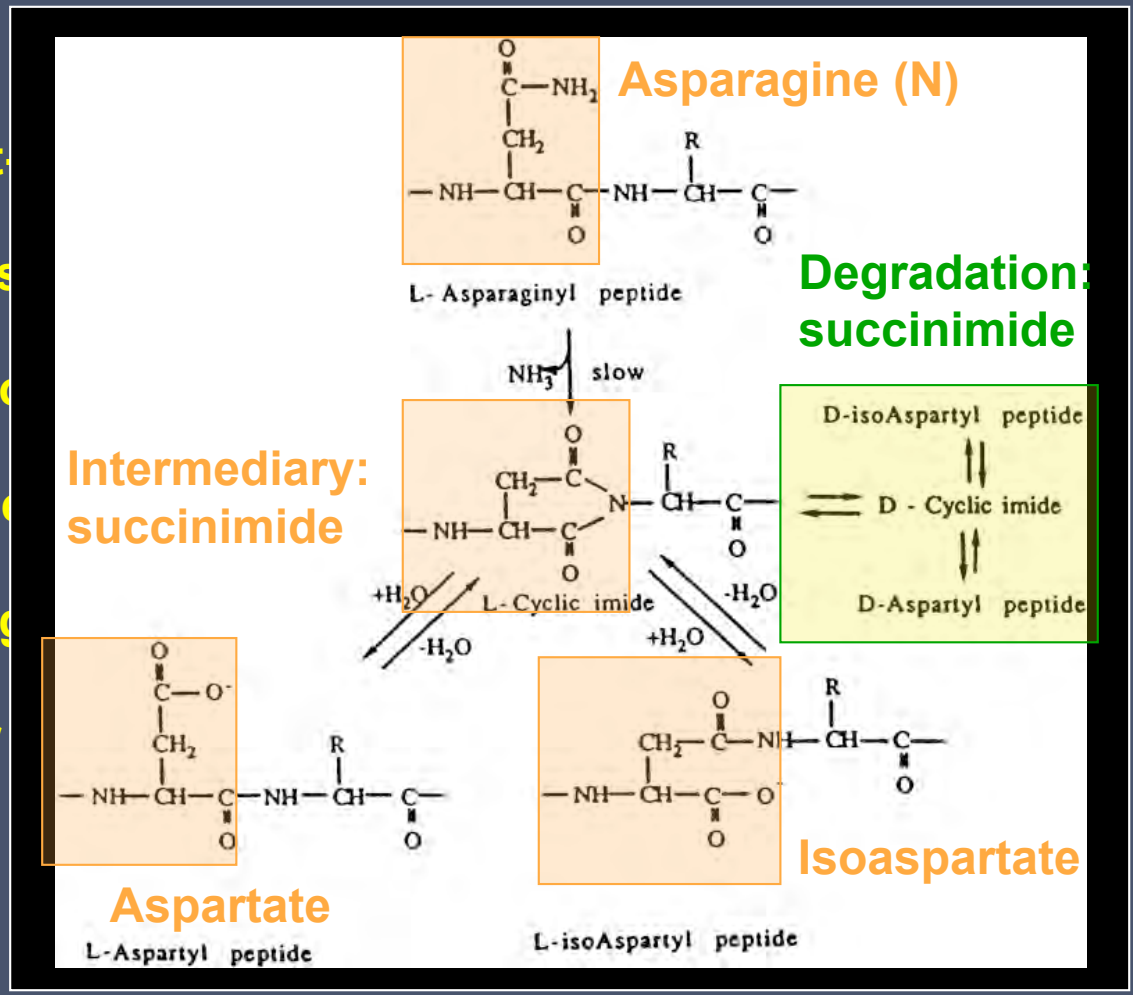
- Motility
- Cell wall, outer membrane
- Transport (TonB), secretion
- Adaptation to stress
- Metabolism of DNA and RNA



A CHEMICAL ANECDOTE

Asparagine deamidates: a major contribution to protein aging

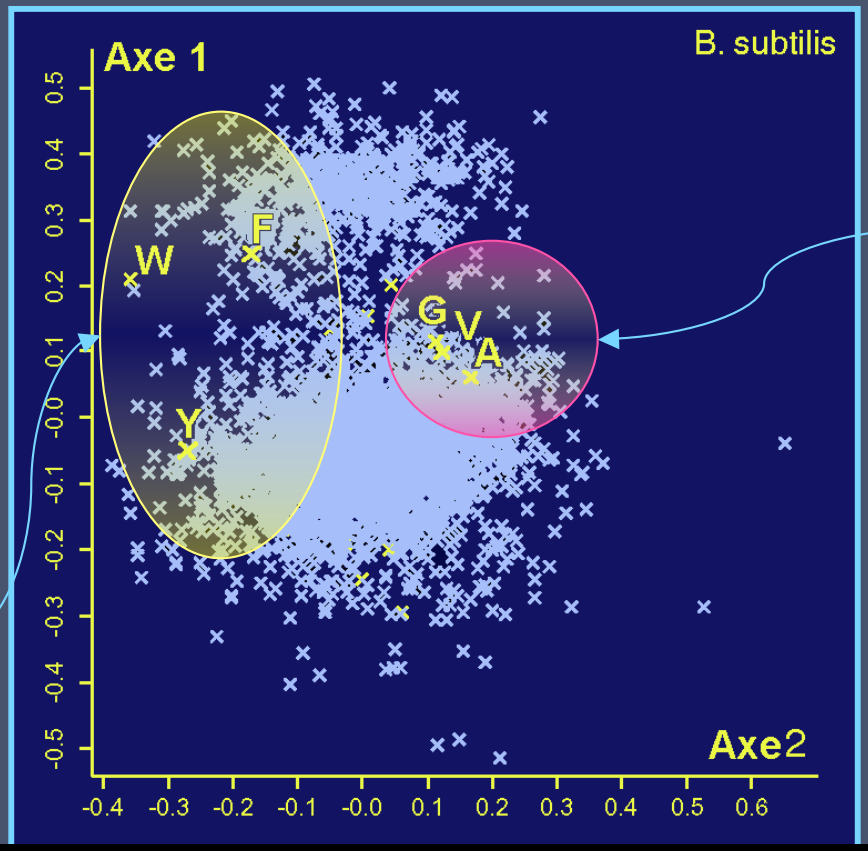
- Main post
- Reaction s
- Spontaneo
- Affects th
- Role in reg
- Signal for



INSTITUT PASTEUR

SECOND UNIVERSAL BIAS

Bias driven by protein aromaticity

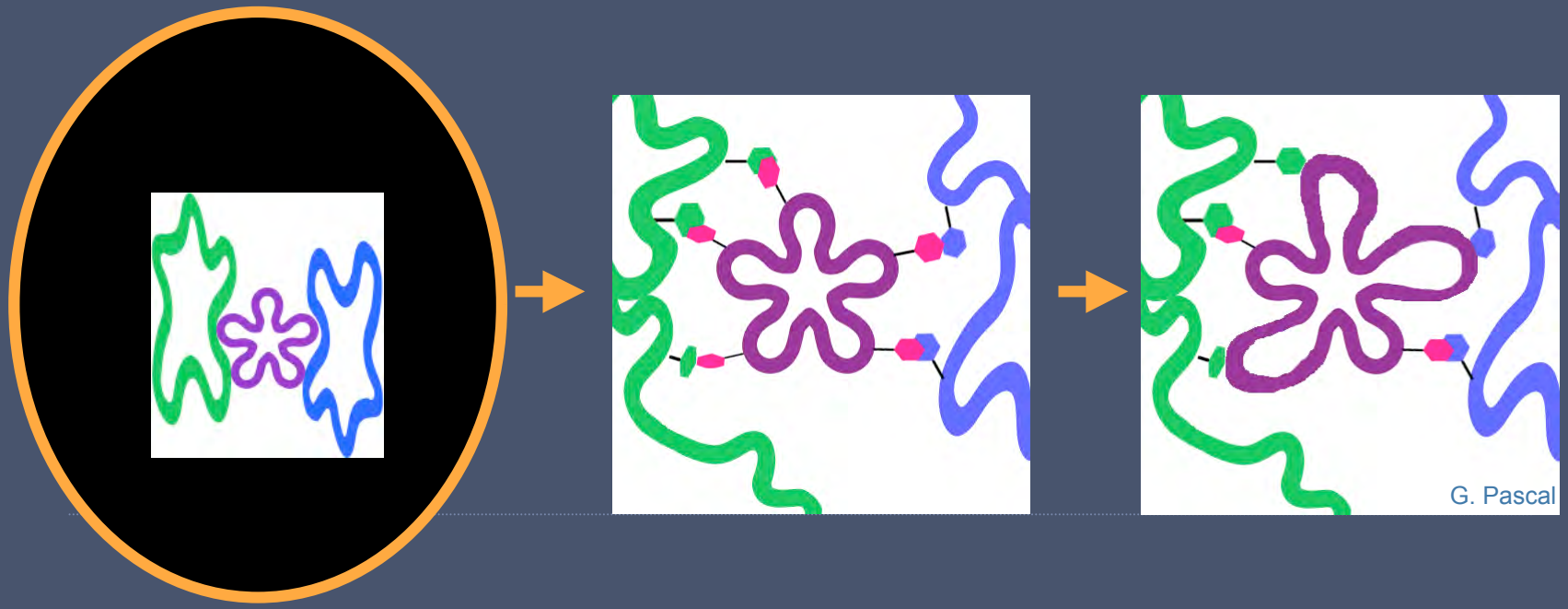


Small amino acids (low metabolic cost)

Aromatic amino acids

WHY AROMATIC RESIDUES IN ORPHAN PROTEINS?

From orphans to « **Gluons** »; how are genes created?



- Orphan loose their status in the course of evolution: *Rocha. 2002. Pedulla. 2003*

INSTITUT PASTEUR

LOCAL BIASES OF CODON USAGE

Correspondence Analysis shows that genes with similar biases are functionnally related. How is this reflected in the chromosome?

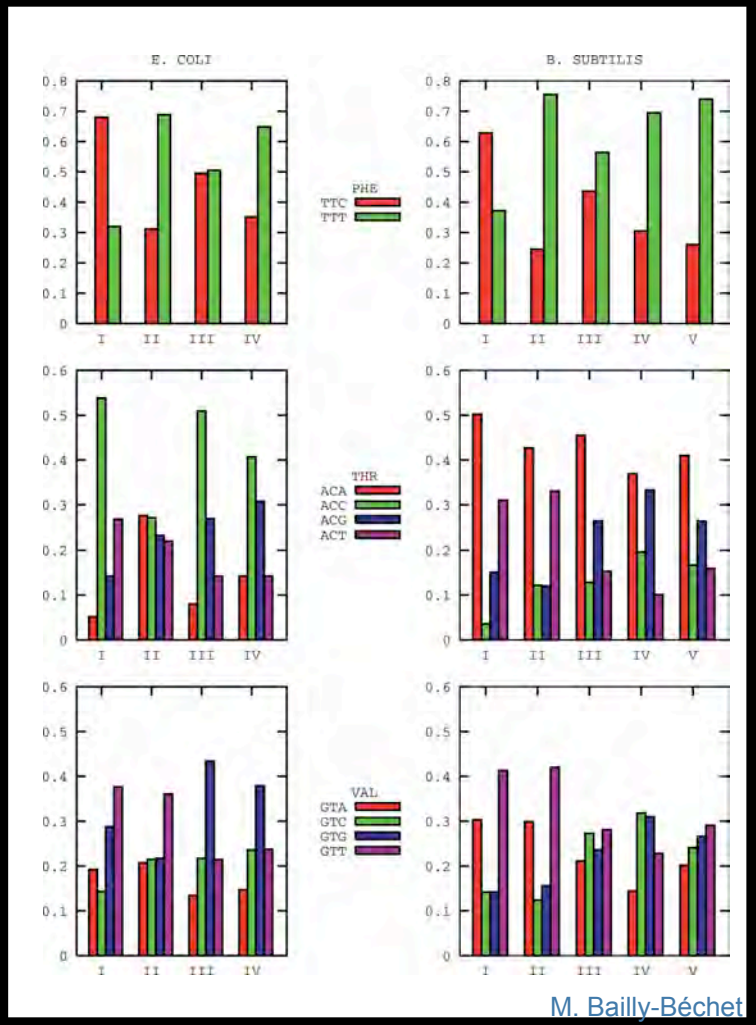
A clustering method (Vergassola et al.) based on information theory groups the genes into homogeneous families, which appear not to be randomly spread in the chromosome. The method identifies 4 classes in *E. coli* and 5 in *B. subtilis*. Genes sharing similar codon bias tend to be close to each other on the chromosome, in coherent patches extended on average ten times the extent of transcriptional units

INSTITUT PASTEUR

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 concetration!) are important in the cell



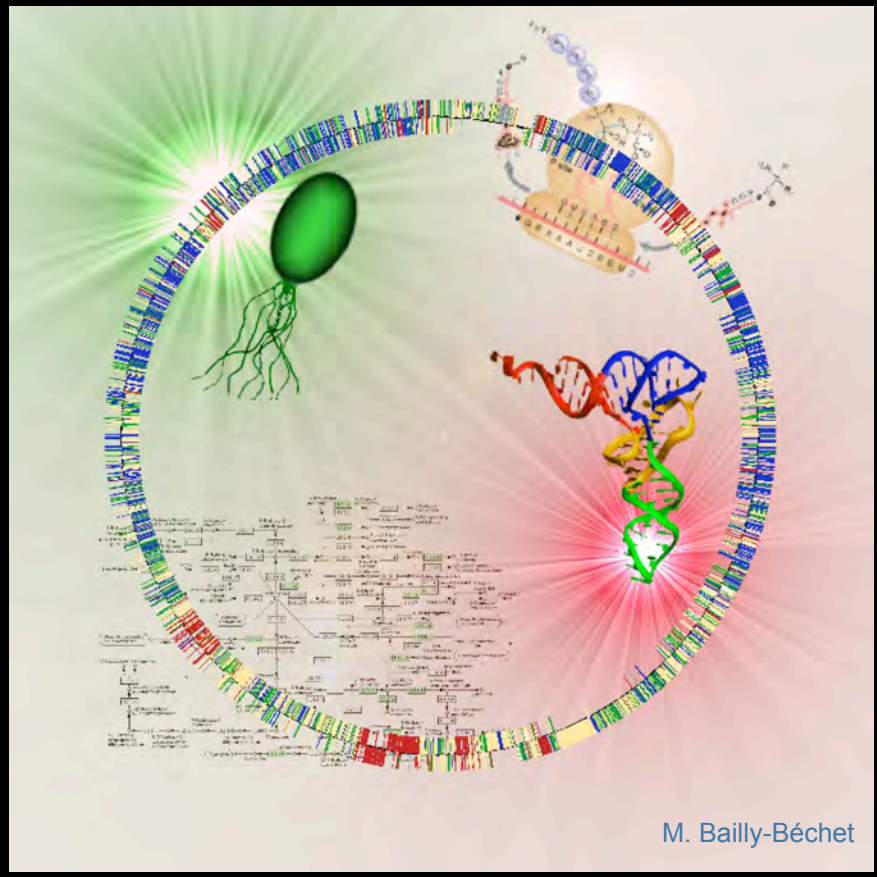
M. Bailly-Béchet

INSTITUT PASTEUR

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**: april 20th

SEQUENCES AND ARCHITECTURES

The non-random distribution of genes in the genome suggests strong constraints of the 3D distribution of molecules in the cell. *Escherichia coli* has to accomodate in less than one cubic micrometer 20,000 ribosomes, 150,000 tRNAs, 1,000 mRNAs (each 3 times longer than the cell), and a DNA molecule 1,000 longer than the length of the cell, together with a huge number of proteins. Occupation of space is therefore a major question combining constraints related to the physics of diffusion and the physics of polymers. Furthremore, the « concentration » of many small molecules is meaningless (1 μM = 600 molecules in *E. coli*)…

- LIFE AND COMPUTATION
- SOME SIMPLE PHYSICAL CONSTRAINTS
- TRANSLATION ORGANIZES THE BACTERIAL GENOME
- THE PALEOME: CONSTRUCTOR AND REPLICATOR
- THE CENOME: THE “PURPOSE” OF THE MACHINE
- REPRODUCTION vs REPLICATION: THE ESSENTIALITY OF METABOLISM

INSTITUT PASTEUR

LOOKING FOR THE REPLICATOR AND THE CONSTRUCTOR

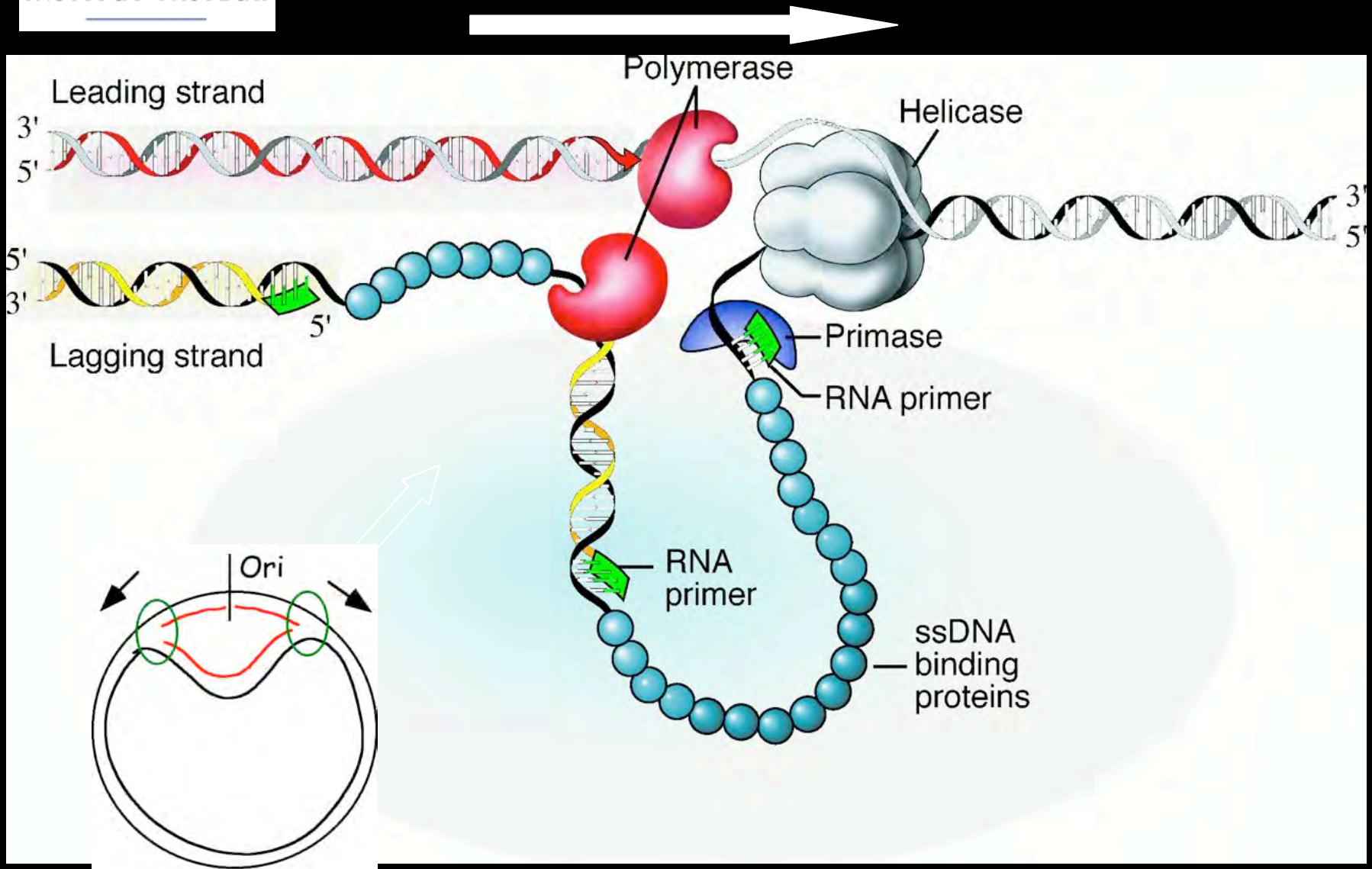
Are genes grouped randomly in the chromosomes?

Do we find different gene categories, in terms of the way they are organized?

At first sight, consistent with different DNA management processes in different organisms not much is conserved, while genes transferred from other organisms are distributed throughout genomes

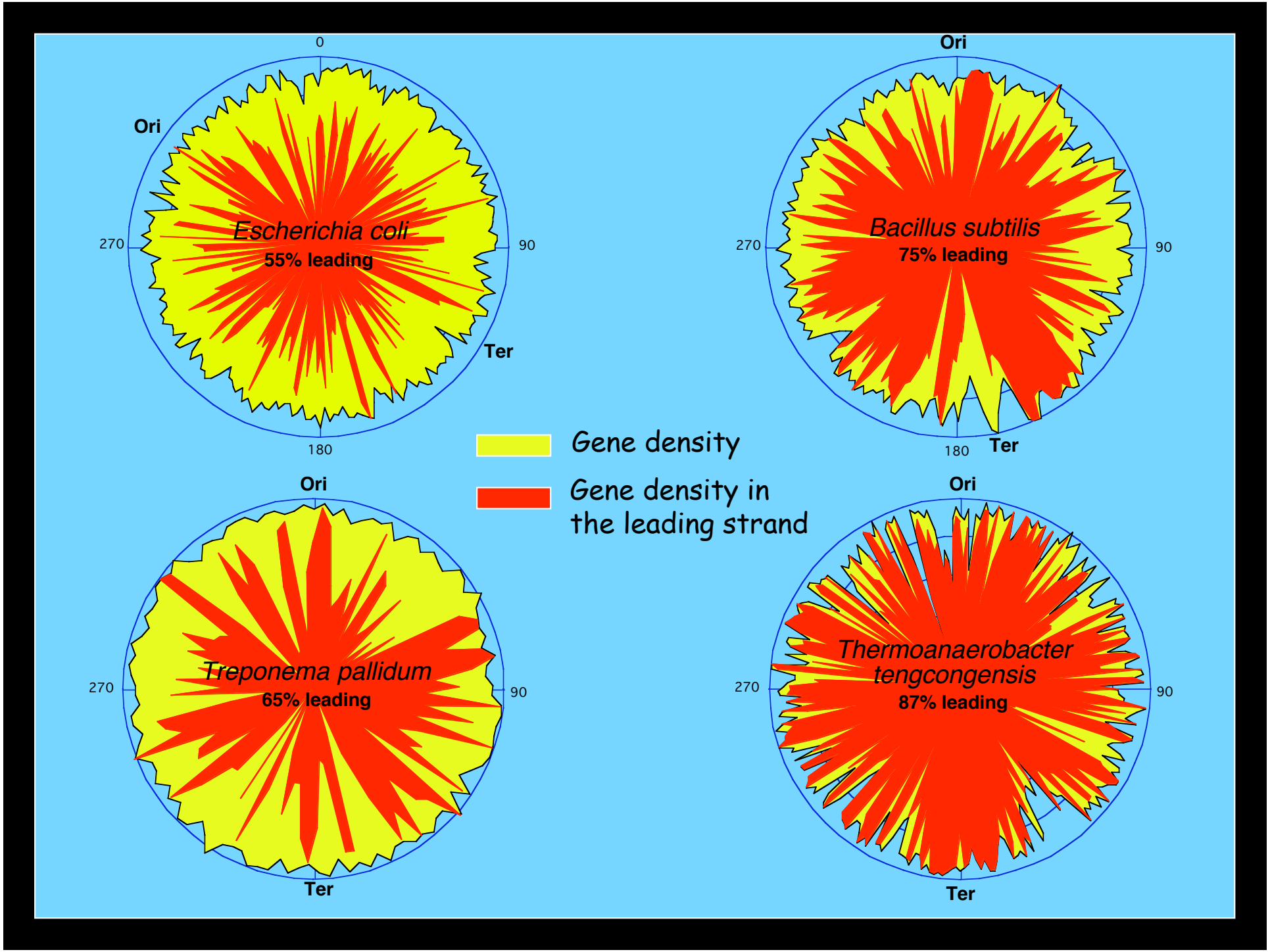
However, groups of genes such as operons or pathogenicity islands tend to cluster in specific places, and they code for proteins with common functions. « Persistent » genes are clustered together

INSTITUT PASTEUR



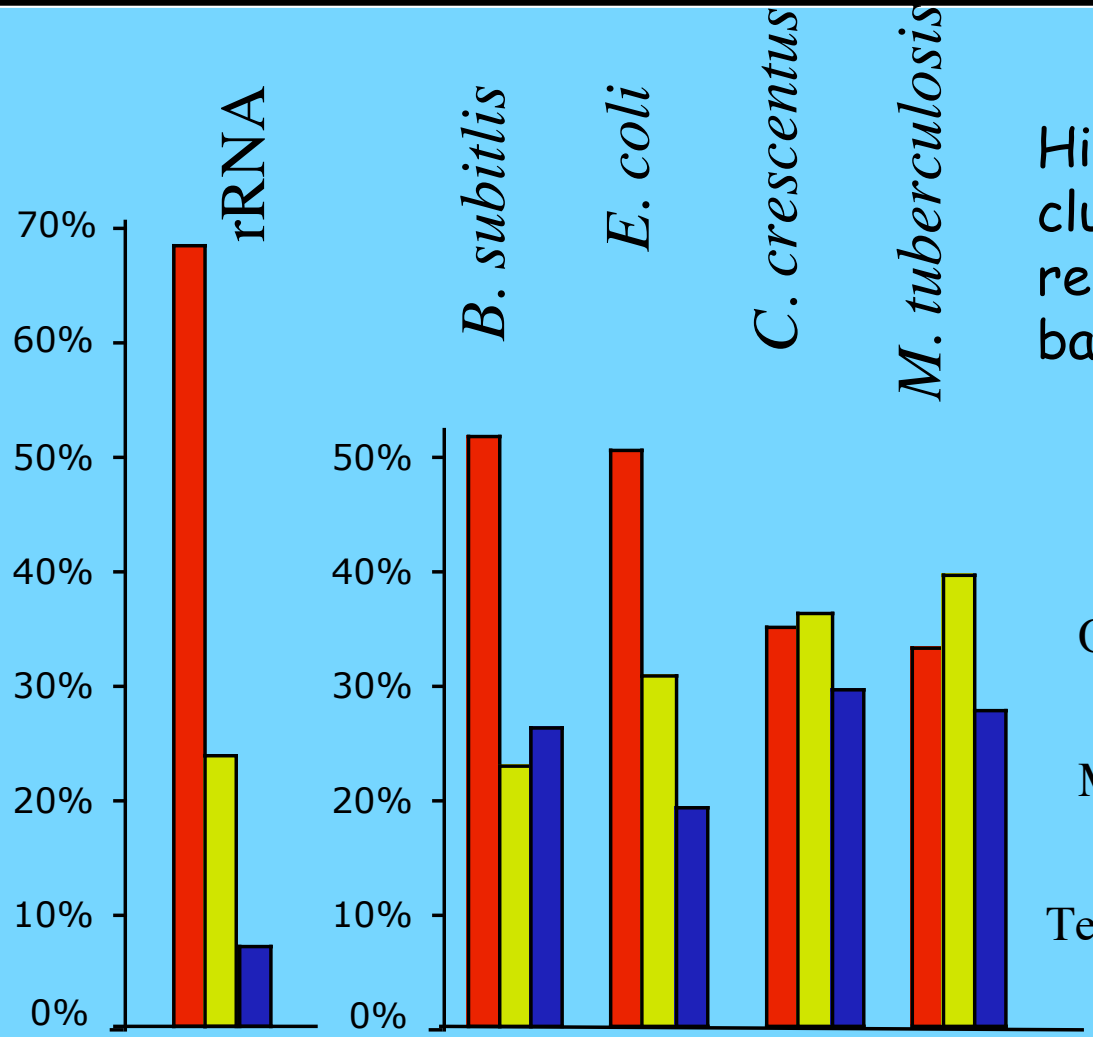
INSTITUT PASTEUR

Genes are preferentially located in the leading replication strand in Bacteria. There is however much variation, depending on the organism, with a considerable bias in A+T-rich Gram-positive organisms

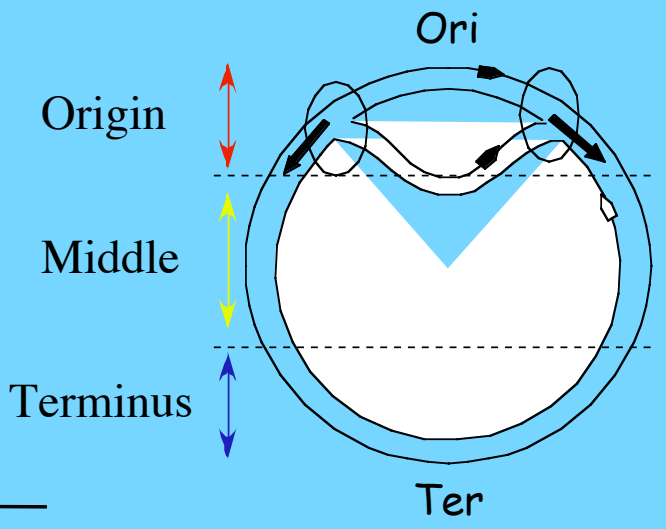

0
Ori
270
90
180
Ter
Escherichia coli
55% leading
Ori
270
90
180
Ter
Bacillus subtilis
75% leading
Gene density
Gene density in
the leading strand
Ori
270
90
Ter
Treponema pallidum
65% leading
Ori
270
90
Ter
Thermoanaerobacter
tengcongensis
87% leading

INSTITUT PASTEUR

DISTRIBUTION OF HIGHLY-EXPRESSED GENES



Highly-expressed genes are clustered near the origin of replication in fast growing bacteria



INSTITUT PASTEUR

CONCLUSION

The genome organization is so rigid that the overall result of selection pressure on DNA is visible in the genome text, which differentiates the leading strand from the lagging strand

INSTITUT PASTEUR

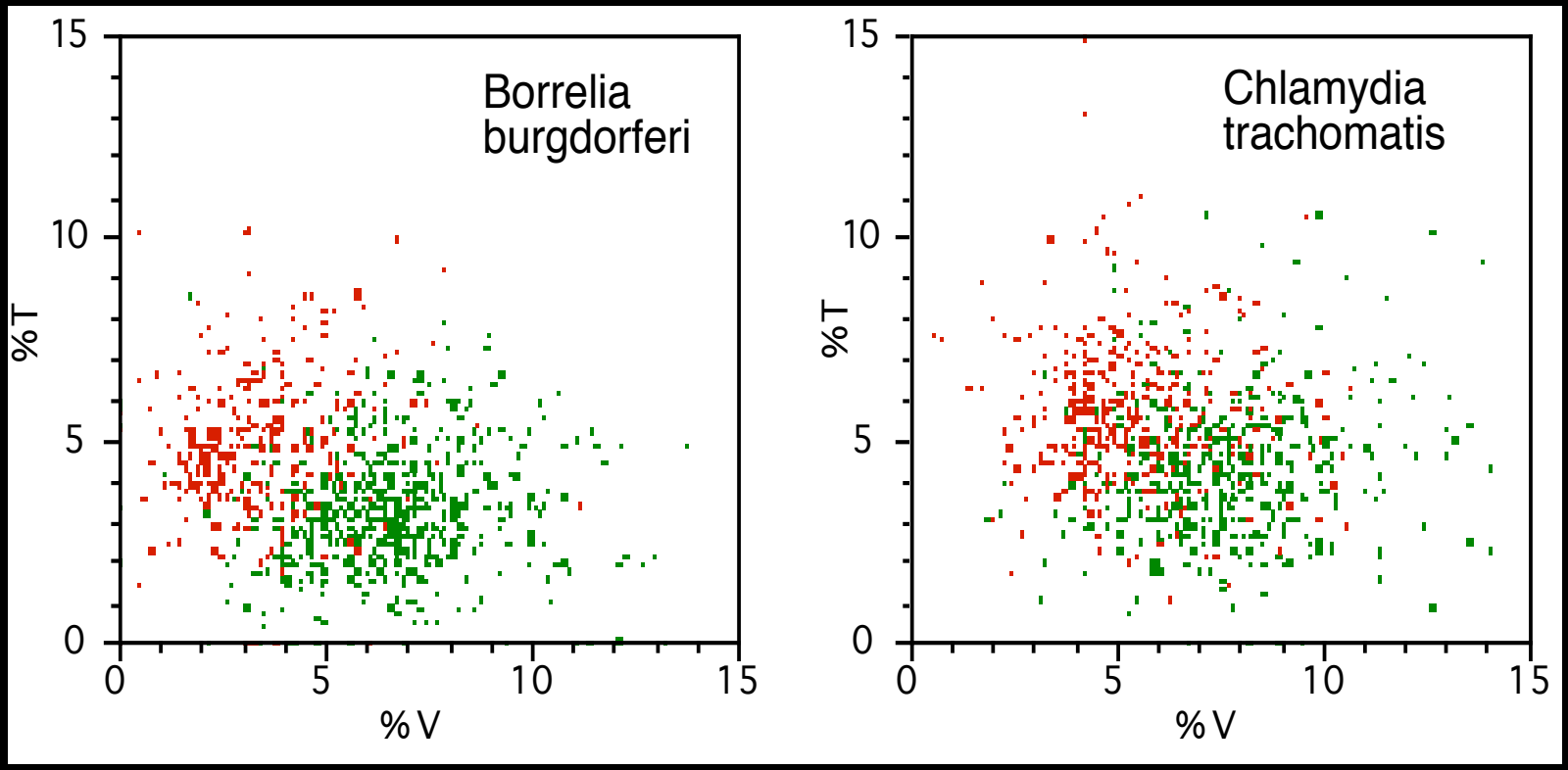
TO LEAD OR TO LAG...

Is it possible to see whether there is a difference in the nucleotide composition, between the leading and the lagging strand? Does that have a consequence on the codon biases? Does that have a consequence for the protein amino acid sequence?

INSTITUT PASTEUR

TO LEAD HAS A COST: BIAS VISIBLE IN PROTEINS…

GT in the leading strand, CA in the lagging strand...



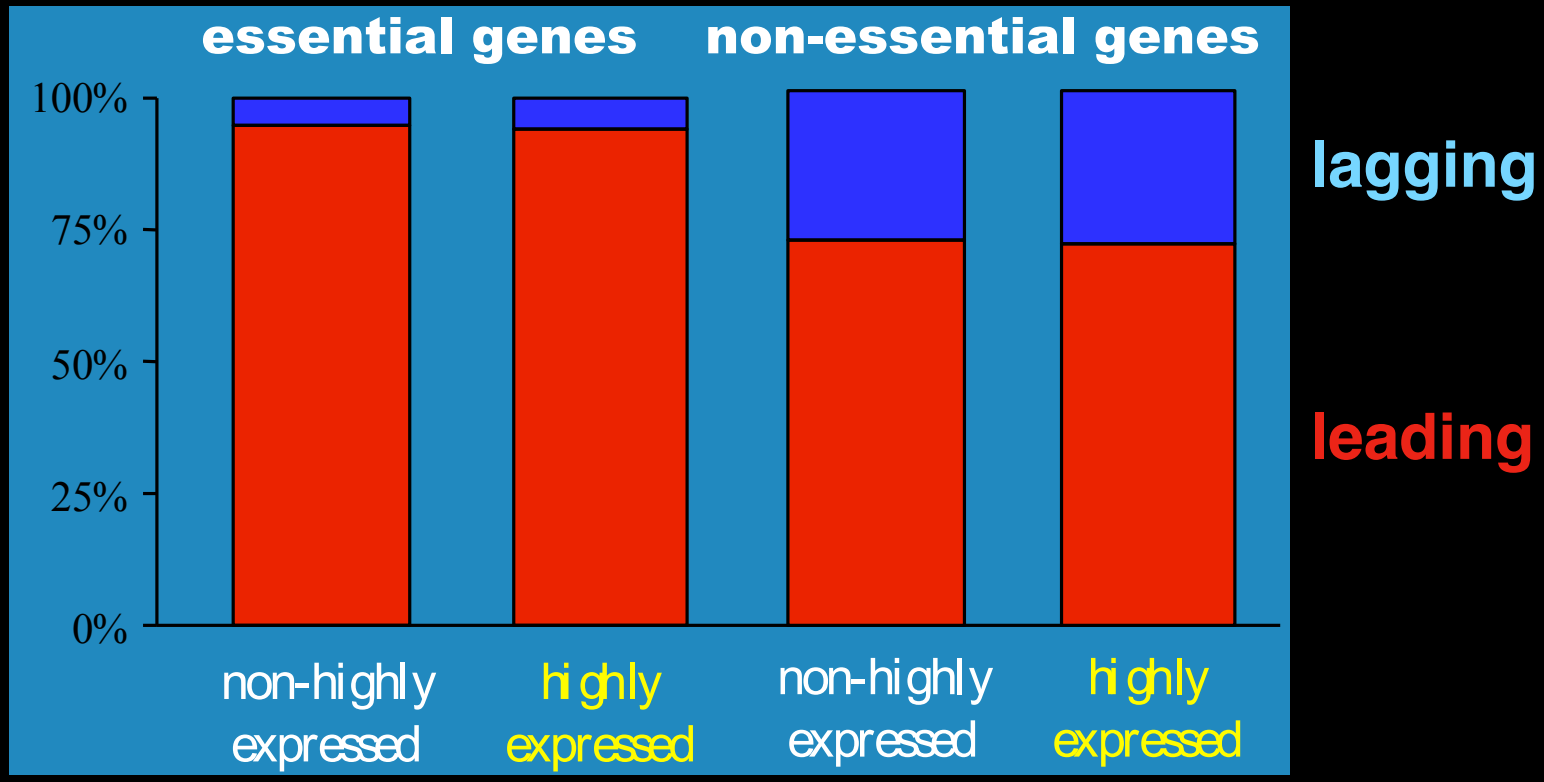
INSTITUT PASTEUR

Proteins are made of 20 amino acid types, among which Valine and Threonine, and one observes that Valine-rich protein are on the leading strand while Threonine-rich proteins are on the lagging strand! Isologous proteins replace preferentially one residue for the other when their gene change strand

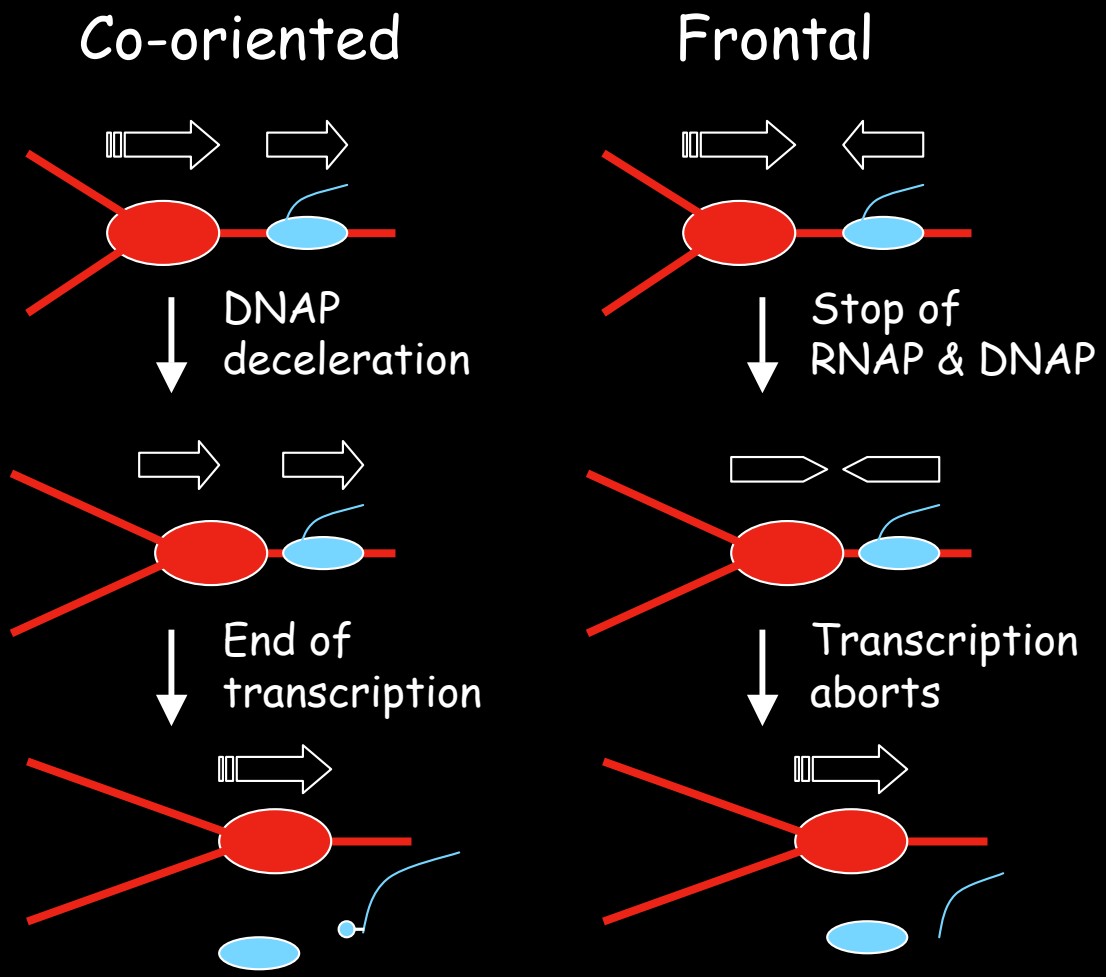
This should be taken into account in models of evolution

INSTITUT PASTEUR

ESSENTIAL GENES LOCATE IN THE LEADING STRAND



PHYSICAL CAUSALITY: AVOIDING COLLISIONS BETWEEN RNA AND DNA POLYMERASES



~~Consequences:~~

~~1. Slowing down of replication~~

~~2. Loss of transcripts~~

Consequences:

1. Truncated transcripts
2. Truncated essential proteins

INSTITUT PASTEUR

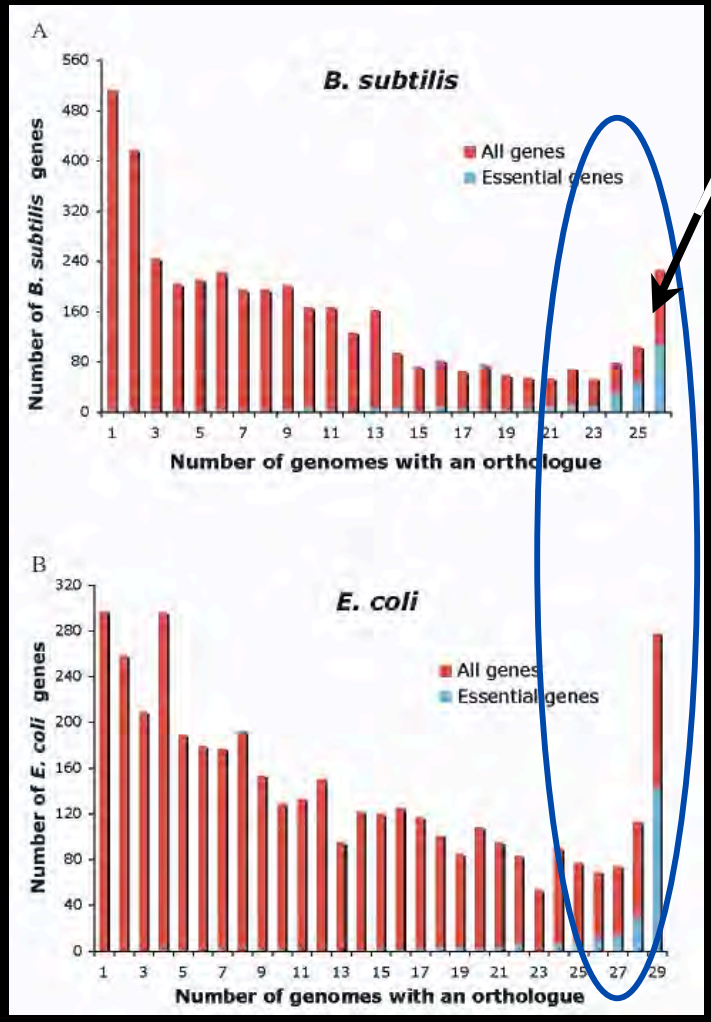
PERSISTENT GENES

Laboratory essential genes are located in the DNA leading strand. They are conserved in a majority of genomes. By contrast the genes that are conserved and located in the leading strand make a particular category, which doubles the number of « essential » genes.

These genes make a universal category; 400-500 genes persist in a majority of 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.

INSTITUT PASTEUR

GENE PERSISTENCE



Persistent genes

Which functional category?

- Information transfer
- Compartmentalization
- Intermediary metabolism
- Stress, maintenance and repair

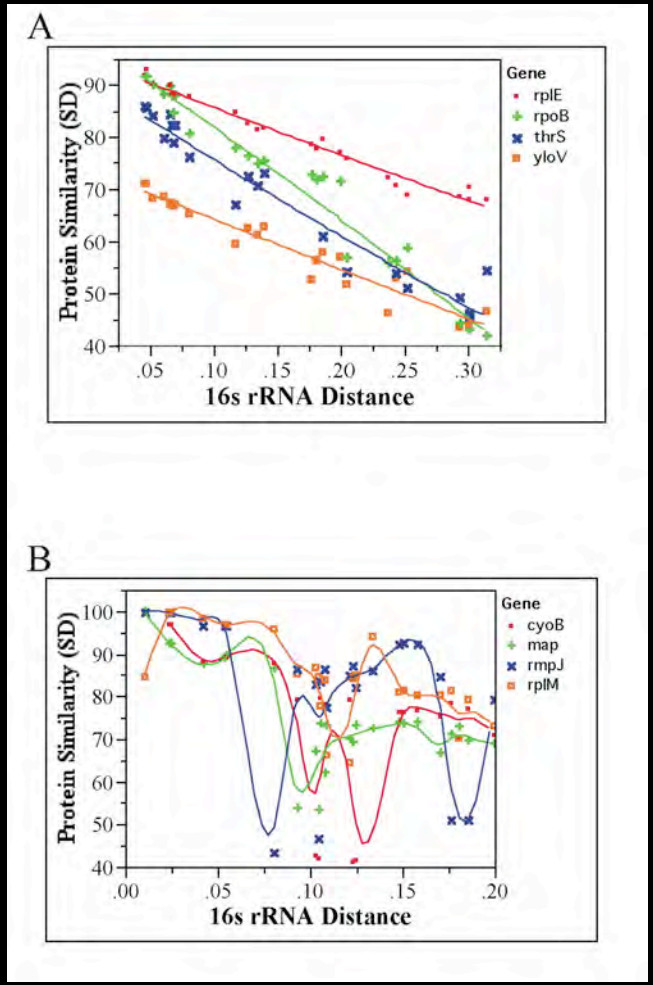
Highly non random!

GENE PERSISTENCE

The contribution of gene divergence was measured for each pair of genomes the correlation between the similarity of orthologous pairs and that of the 16S rRNA

(A), 38% (resp. 48%) of *B. subtilis* (resp. *E. coli*) persistent genes showed a correlation coefficient >0.9 between the sequence similarity of the pair of orthologs and the 16S RNA

Some genes (B) evolve in an erratic way. This may be due to horizontal gene transfer, local adaptations leading to change in evolutionary pace, or simply wrong assignments of orthology. The latter is a significant problem, especially in large protein families



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

PERSISTENT GENES ARE CLUSTERED TOGETHER

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

The way they group along chromosomes in more than 250 bacteria (genome length > 1,500) displays three clusters that reflect a scenario of the origin of life. This is why it is proposed to name paleome (from παλαιος, ancient) this group of core genes

INSTITUT PASTEUR

EXISTENCE IMPLIES CLUSTERED PERSISTENCE

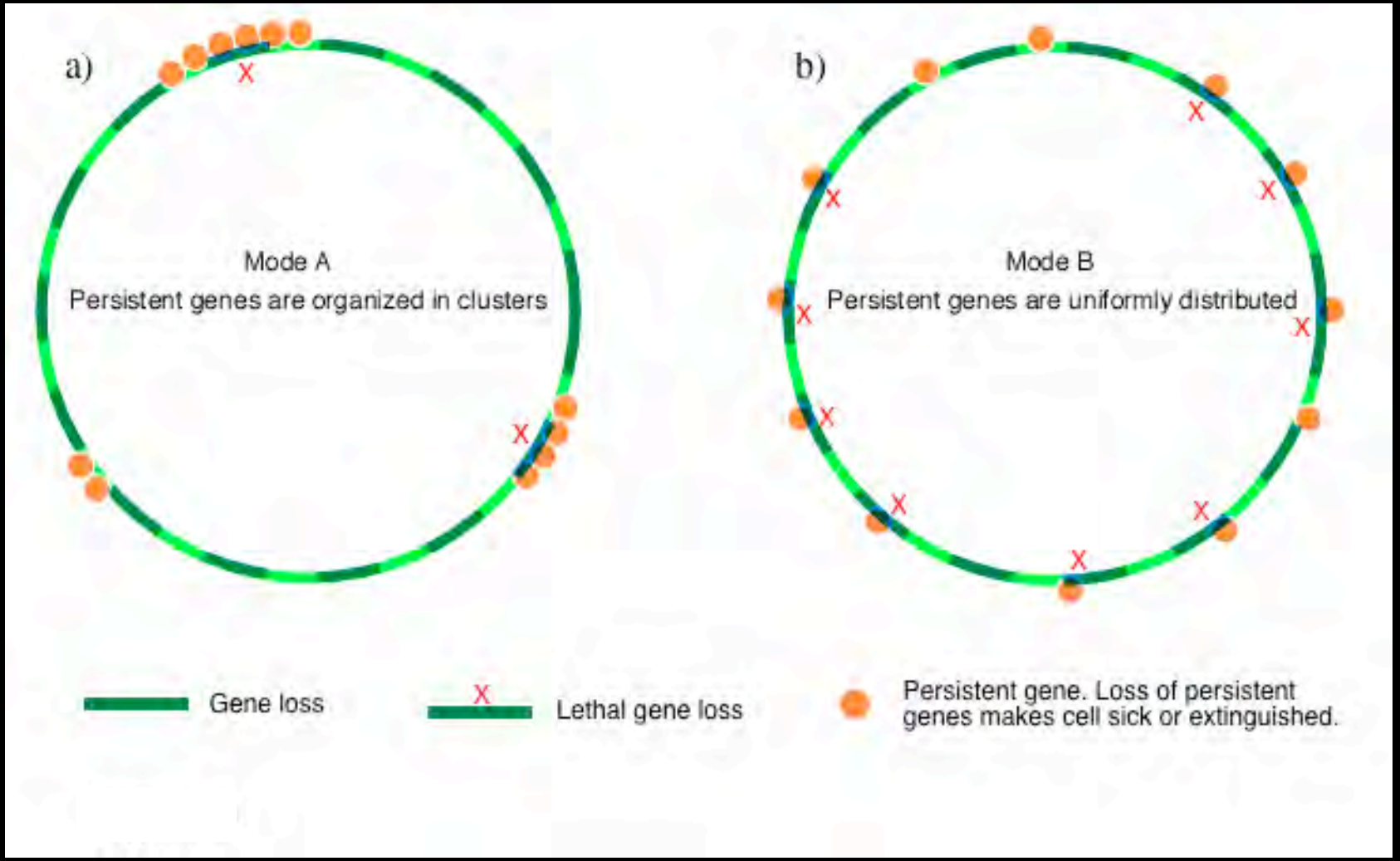
Why are persistent genes clustered? A simple model shows that if, in addition to horizontal gene transfer, there is a process deleting genes in groups in genomes, then any gene contributing to fitness frequently enough over generations will tend to cluster with other genes with similar properties. This accounts for clustering of essential genes, but most probably also for clustering of antibiotic resistance genes in bacteria found in hospitals....

As a consequence gene clustering will precede, not derive from co-transcription or protein-protein interaction (no intelligent design!)

Note: the model needs to be refined. It may yield interesting chaotic behaviours

INSTITUT PASTEUR

EXISTENCE IMPLIES CLUSTERING



INSTITUT PASTEUR

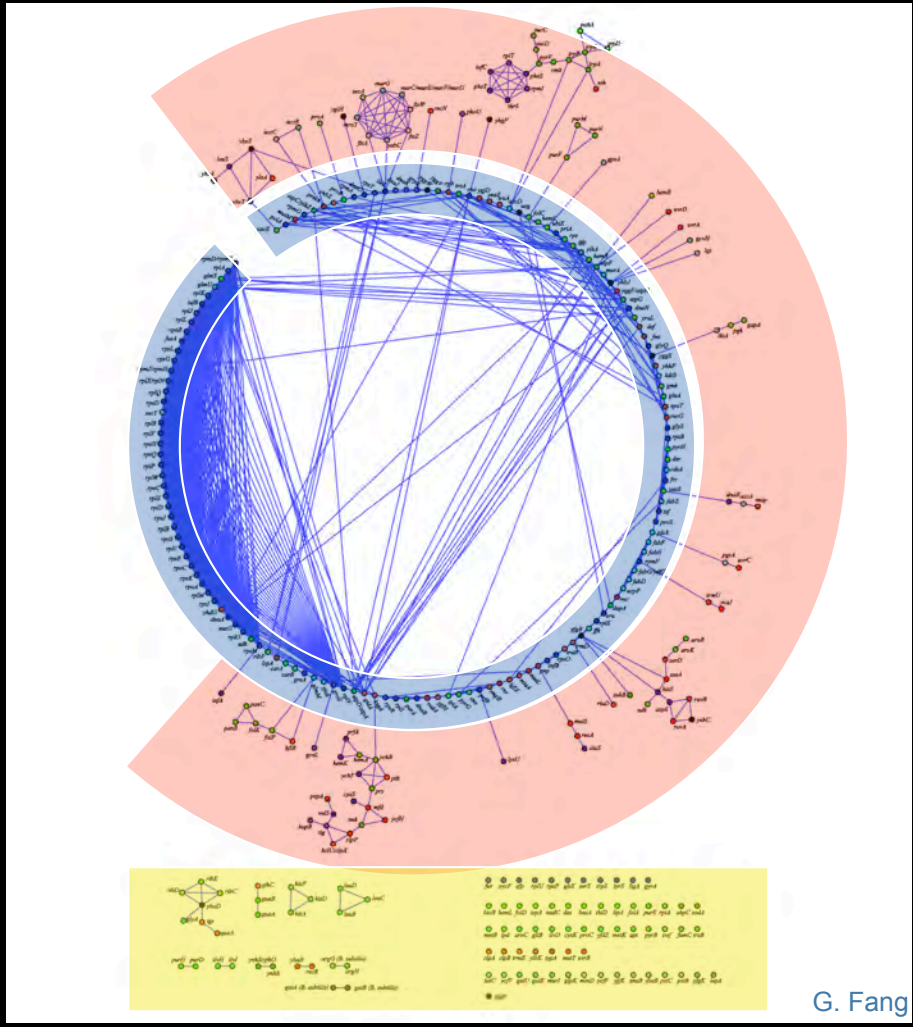
PERSISTENT GENES CONNECTIVITY

Using 228 genomes with more than 1500 genes and « correct » annotations, we have identified genes that tend to remain close to one another; this « mutual attraction » constructs a remarkable network made of three layers

INSTITUT PASTEUR

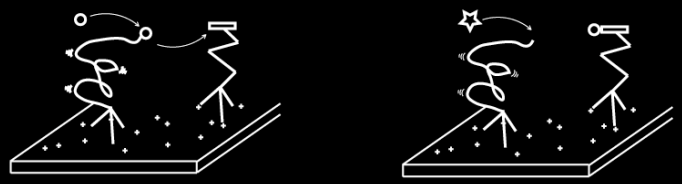
PERSISTENT GENES RECAPITULATE THE ORIGIN OF LIFE

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

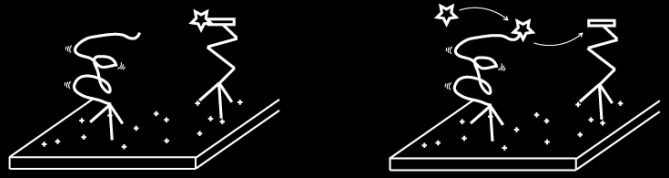


G. Fang

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



HOMEOTOPIC TRANSFORMATION



Origin

Origine

tRNA OUT OF TRANSLATION

FORMATION OF A PEPTIDIC LINK

- Gly~ARNt^{Gly} murein
- Phe~ARNt^{Phe} N-terminal
- Leu~ARNt^{Leu} N-terminal
- Arg~ARNt^{Arg} Ubiquitine

HOMEOTOPY

- Met~ARNt^{FMet} —> FMet~ARNt^{FMet}
- Glu~ARNt^{Gln} —> GlnARNt^{Gln}
- Ser~ARNt^{SeCys} —> SeCys~ARNt^{SeCys}

DIVERSE

- Lys~ARNt^{Lys} Lipids
- Glu~ARNt^{Glu} Aminolevulinate

OTHER MODIFICATIONS

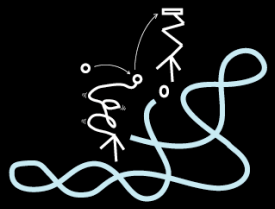
- Base moodifications

PATHWAYS USING PHOSPHATE RESIDUES WITH NO "REASON"

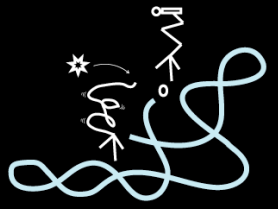
- SERINE (SERINE~PHOSPHATE)
- PYRIDOXAL PHOSPHATE
- ISOPRENYL ~ PYROPHOSPHATE
- 4-PHOSPHOPANTETHEINE
- THIAMINE ~ PYROPHOSPHATE

CONTAINING NUCLEOTIDES :

- CYTIDYLATE (LIPIDES)
- ADENYLATE, GUANYLATE, URIDYLATE (SUGARS)
- COENZYMES (NAD, CoA…)



TRANSFER RNA SUBSTITUTES TO SURFACES



Origine

Origin

INSTITUT PASTEUR

METABOLISM AND REPLICATION

This scenario emphasizes the separation between metabolism and replication, the latter being a secondary invention of prebiotic systems:

Building blocks => nucleotides => tRNA => ribosome => DNA

- LIFE AND COMPUTATION
- SOME SIMPLE PHYSICAL CONSTRAINTS
- TRANSLATION ORGANIZES THE BACTERIAL GENOME
- THE PALEOME: CONSTRUCTOR AND REPLICATOR
- THE CENOME: THE "PURPOSE" OF THE MACHINE
- REPRODUCTION vs REPLICATION: THE ESSENTIALITY OF METABOLISM

INSTITUT PASTEUR

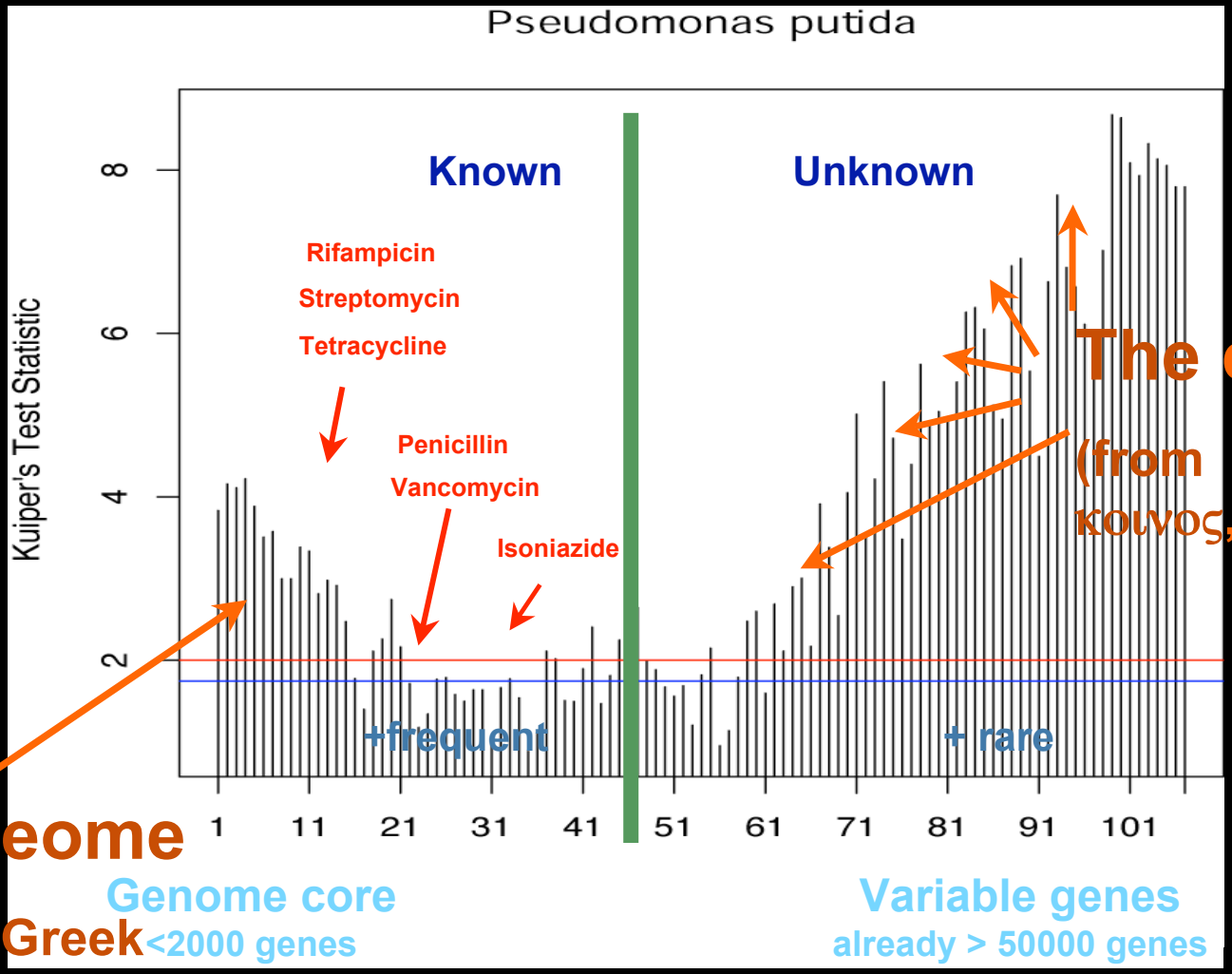
THE COMPOSITE GENOME

- Expecting two genome components, coding for the machine and for the "purpose" of the machine, we need to separate between the replicator/constructor and secondary functions.

- Extant genomes should comprize ubiquitous functions (not genes!) which would correspond to the former (here named the paleome) and functions specific to the environment of the organism (named the cenome — as in "biocenose" — to express the fact that these genes correspond to a specific niche)

INSTITUT PASTEUR

CONSERVATION OF GENE CLUSTERING



The paleome

(from the Greek παλαιος, ancient)

The cenome

(from the Greek κουνος, common)

Frequency in genomes

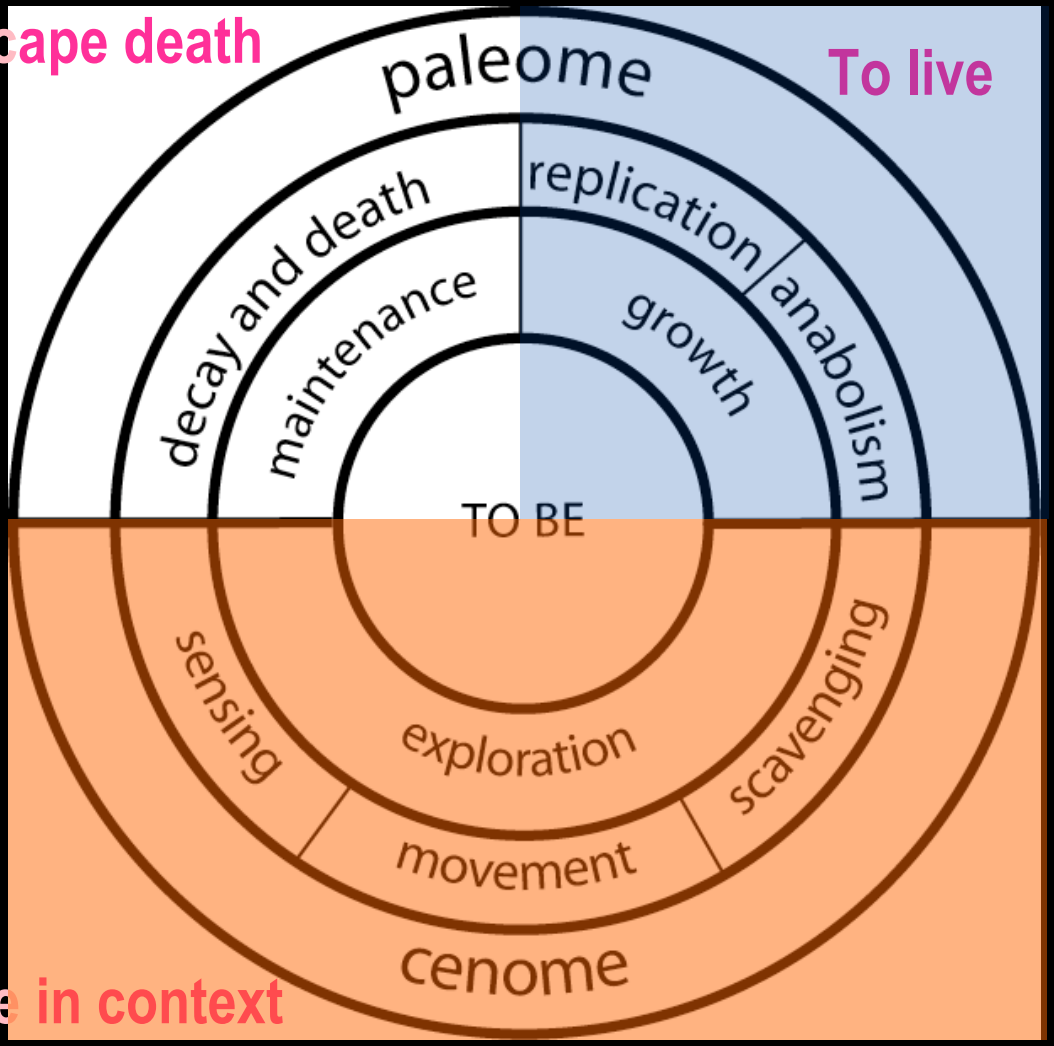
INSTITUT PASTEUR

A RECURSIVE MACHINE

- **Replicator**: DNA specifies proteins that replicate DNA
- **Constructor**: DNA specifies proteins which form the machine that constructs the cell
- However, DNA can only accumulate errors… the machine needs to cope with errors

SYNTHESIS: A TALE OF TWO GENOMES

Life manifests first by growth and repair of weathering: the corresponding genome exists since the origin, it is the paleome. Exploration of the environment is an inevitable consequence of existence, it results from continuous creation and exchange of the genes which form the cenome



A. Danchin. Archives or Palimpsests? Bacterial Genomes Unveil a Scenario for the Origin of Life Biological Theory (MIT Press) (2007) 2: 52-61.

INSTITUT PASTEUR

THREE PARTS IN THE GENOME'S ORGANIZATION

- Anabolism and Replication
- Maintenance and Repair: coping with errors
- Life in context (the cenome)

- LIFE AND COMPUTATION
- SOME SIMPLE PHYSICAL CONSTRAINTS
- TRANSLATION ORGANIZES THE BACTERIAL GENOME
- THE PALEOME: CONSTRUCTOR AND REPLICATOR
- THE CENOME: THE “PURPOSE” OF THE MACHINE
- REPRODUCTION vs REPLICATION: THE ESSENTIALITY OF METABOLISM

INSTITUT PASTEUR

METABOLISM AND REPLICATION

Freeman Dyson's « origins of life » revisited

- Replication accumulates errors (Muller's ratchet and Orgel's error catastrophe)
- Reproduction: can metabolism reproduce in an error-prone context, and improve on unperfect components?

INSTITUT PASTEUR

INFORMATION AGAIN

Metabolism improvement can be conceptually tolerated as creation of information is reversible (Landauer, 1961; Bennett, 1988)

Open question: « room » is needed to accomodate innovation; how is it obtained? experiments are needed to identify the corresponding processes

INSTITUT PASTEUR

GENOME EXPLORATION

- Is the structure of the paleome homogeneous?
- How do we see the dialogue between the young and the old (creation of information, in practice)?

INSTITUT PASTEUR

ESSENTIAL METABOLISM

- Recovery from an aged state requires:
  - Stepwise improvement of the « quality » of biological objects
  - Energy-dependent selection of what needs to be destroyed (ratchet mechanism)
- Persistent genes of unknown function evolve following a tree that differs from that of the anabolic pathways…

INSTITUT PASTEUR

CAVEAT: PATCHES FOR ANECDOTES

- Each component of the cell has idiosynchratic properties, some are incompatible with other components
- The paleome codes for anabolic and maintenance features, except for a few purely catabolic steps
- One example: serine catabolism (accounts for serine toxicity)

This results in the « anecdotal » appearance of biological systems