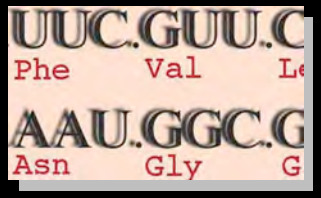


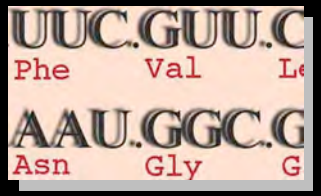
INSTITUT PASTEUR

Symplectic biology:
Universals in microbial genomes

24 april 2006

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

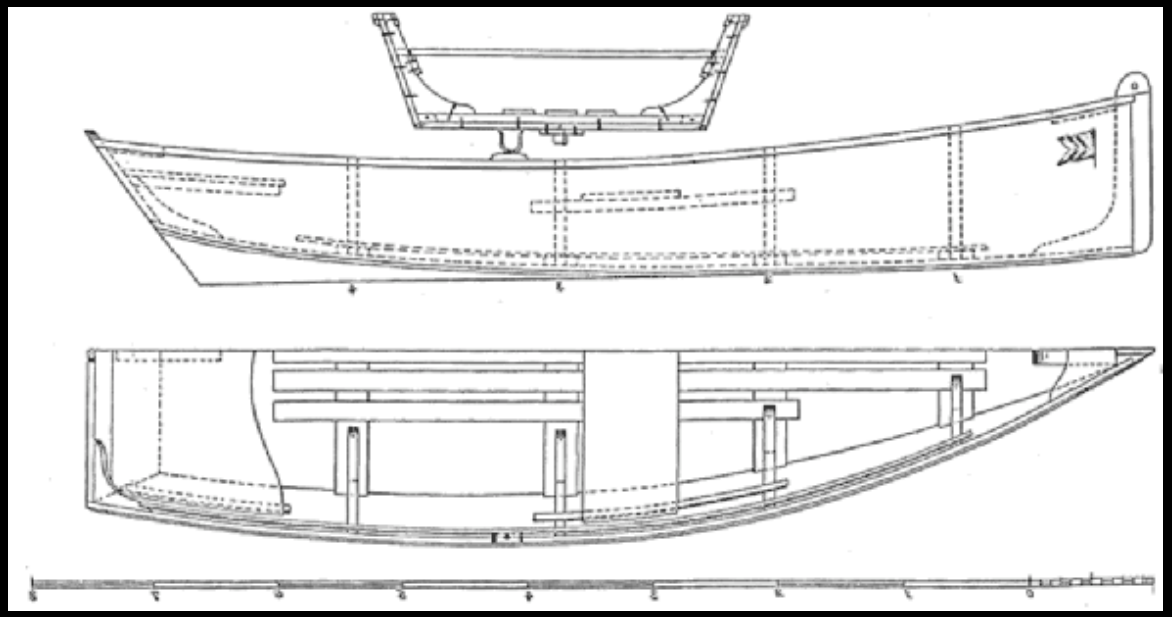
adanchin@pasteur.fr



Symplectic biology: The Delphic Boat

INSTITUT PASTEUR

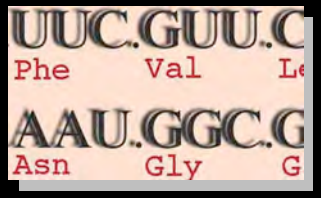
- Biology is a science of relationships between objects rather than from objects: from συν together, πλεκτειν, to weave
- Proteins are part of complexes, as are parts in an engine
- As for constructing a boat, failing to understand their relationships will result in ultimate failure of synthetic objects



The Delphic Boat: Harvard University Press, february 2003

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

adanchin@pasteur.fr



What is Life?

INSTITUT PASTEUR

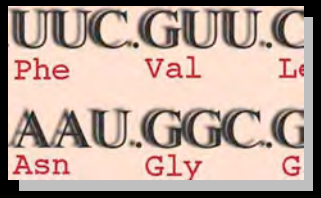
Three processes are needed for Life:

- **Information transfer** (Living Computers?) => **the goal of genomics is to decipher the blueprint of the "read-only" memory of the machine**

Driving force for a coupling between the genome structure and the structure of the cell:

- **Metabolism**
- **Compartmentalisation**

adanchin@pasteur.fr



What is computing?

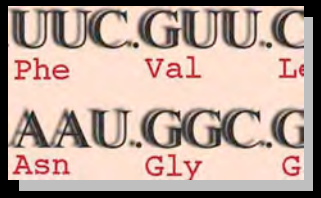
INSTITUT PASTEUR

Two processes are needed for computing:

- A read/write machine
- A program on a physical support (typically, a tape illustrates the sequential string of symbols that makes up the programme), split (in practice) into two entities:
  - Programme (providing the goal)
  - Data (providing the context)

The machine is distinct from the programme

adanchin@pasteur.fr



Cells as computers

INSTITUT PASTEUR

Genomics rests on an alphabetic metaphor, that of a text written with a four-letter alphabet, acting as a programme

Conjecture: do cells behave as computers?

Genetic engineering

Viruses

Horizontal gene transfer

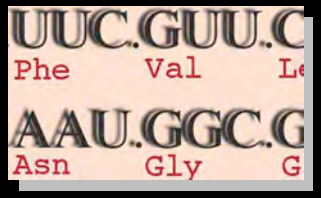
Cloning animal cells

all point to separation between

Machine

Data + Programme

adanchin@pasteur.fr

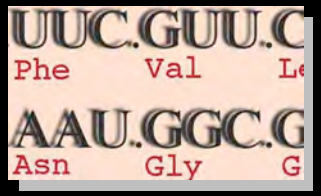


Is there a map of the cell in the chromosome?

INSTITUT PASTEUR

If the machine has not only to behave as a computer but has also to construct the machine itself, one must find an image of the machine somewhere in the machine (John von Neumann)

adanchin@pasteur.fr



Genome organisation

INSTITUT PASTEUR

Is the gene order random in the chromosomes?

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

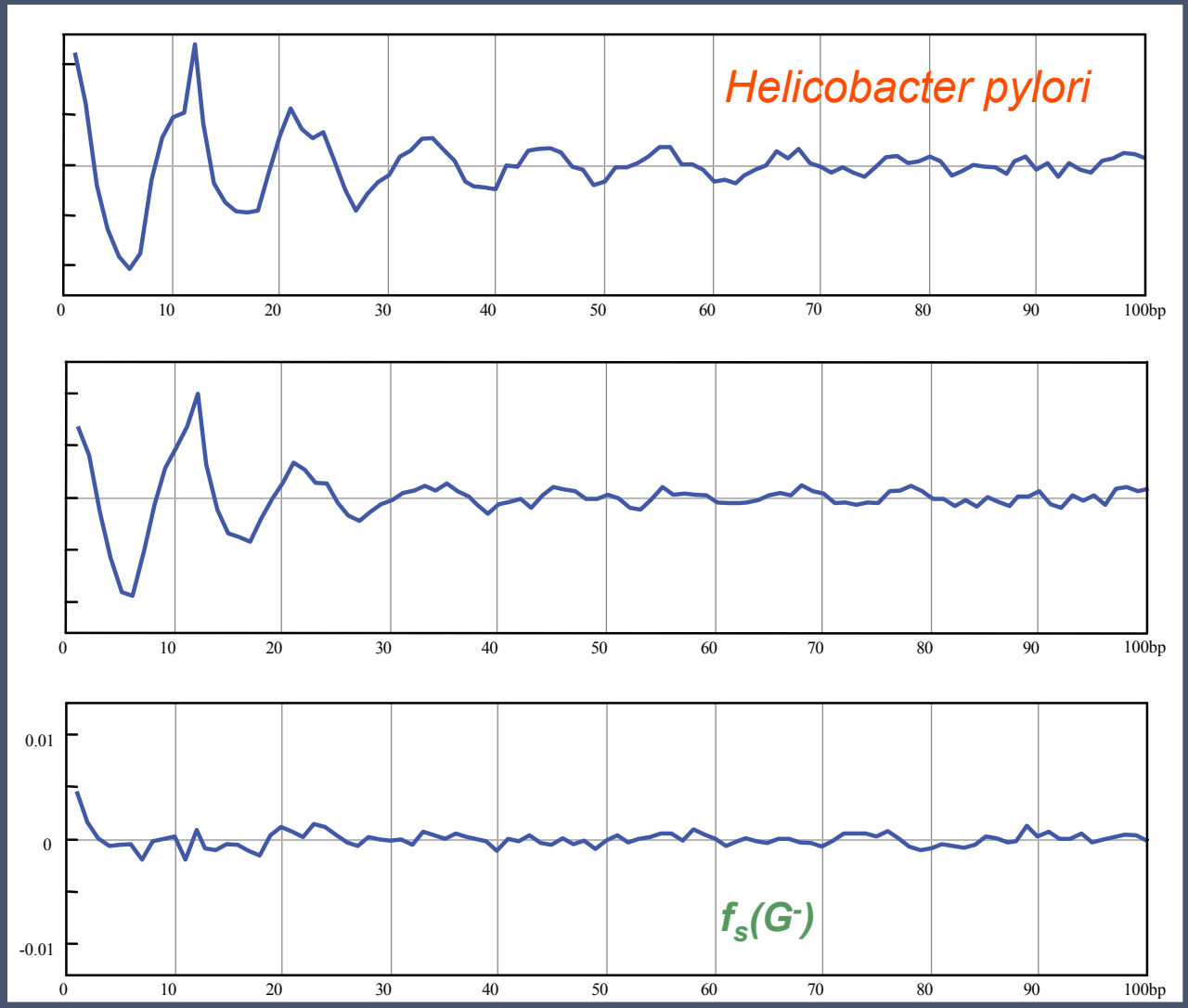
However, groups of genes such as operons or pathogenicity islands tend to cluster in specific places, and they code for proteins with common functions. « Persistent » genes are clustered together

Also, some motifs are ubiquitously present, suggesting general rules constraining genome organisation

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

adanchin@pasteur.fr

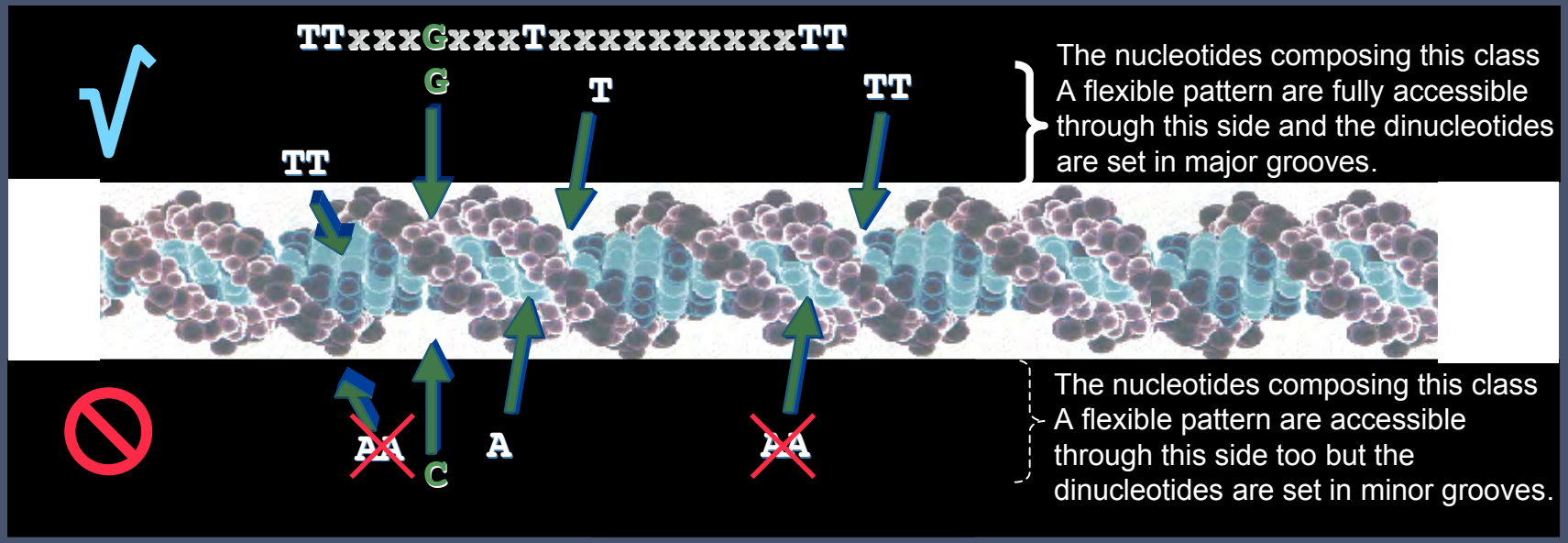
A universal feature of the program: the period of 10-11.5



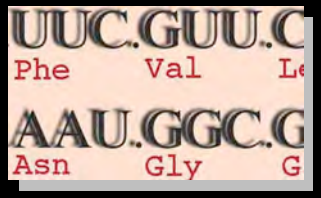
Genetics of Bacterial Genomes
http://www.pasteur.fr/recherche/unites/REG/
adanchin@pasteur.fr

Flexible motifs of type A

```
1-xAxxxxTxxxxAxxxxTTxxxxxAxxxxTxxxxAxxx: All kindoms

2-xxxxxxxxxxxGxxxxTTxxxCxxxxxTxxxxxxxxx: Proteobacteria

4-xxxxxxTxxxxAGxxxTTxxxxxxxxTxxxxxxxxxx: Archaea

5'-xxx-10xxxxxxxxx0xxxxxxxx10xxxxxxbp-3'
```



adanchin@pasteur.fr

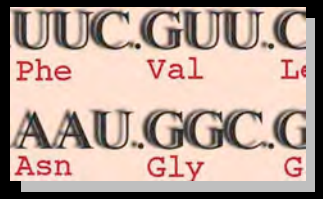


To lead or to lag...

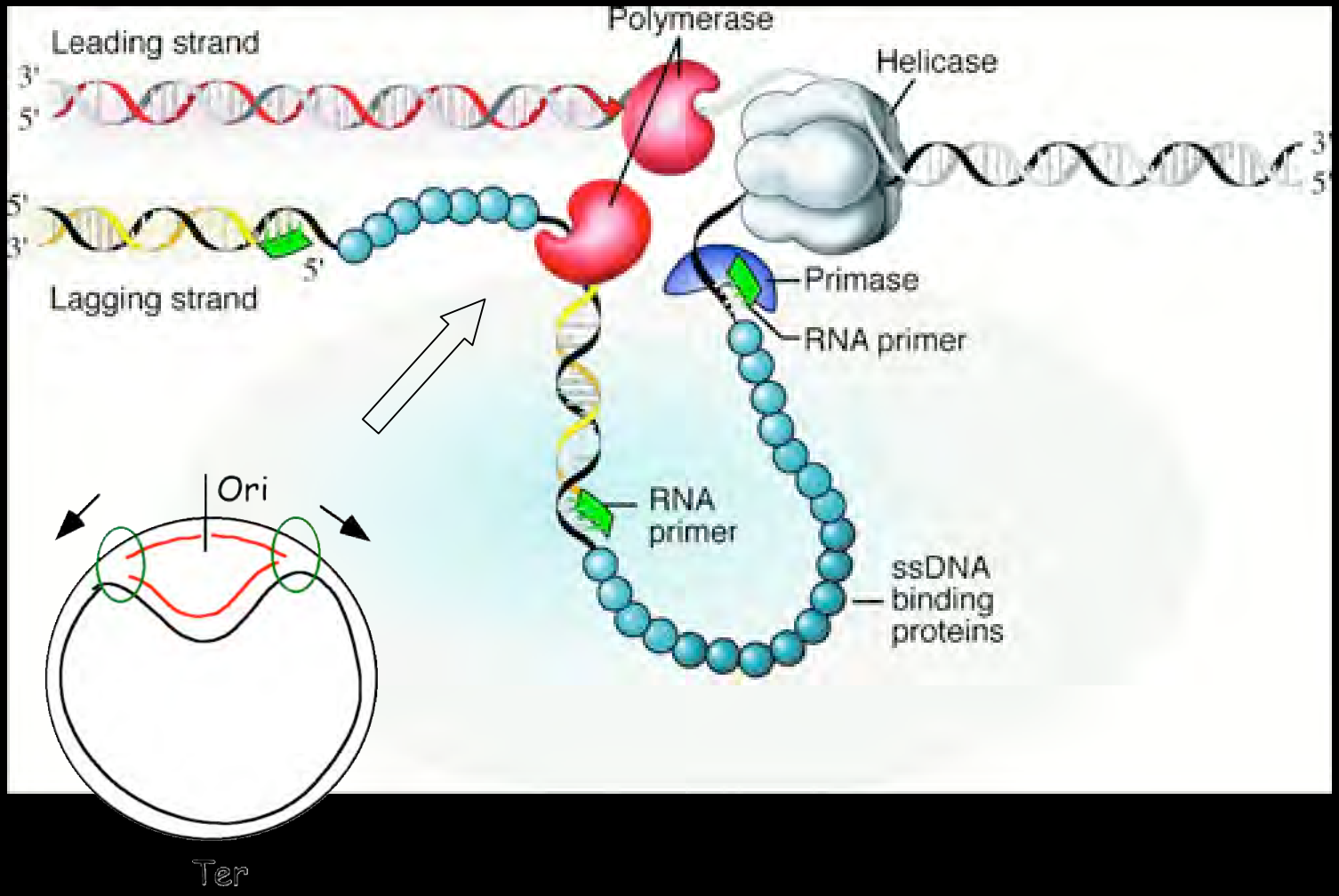
INSTITUT PASTEUR

Is it possible to see whether the position of genes in the chromosome is randomly distributed on the leading and lagging strand?

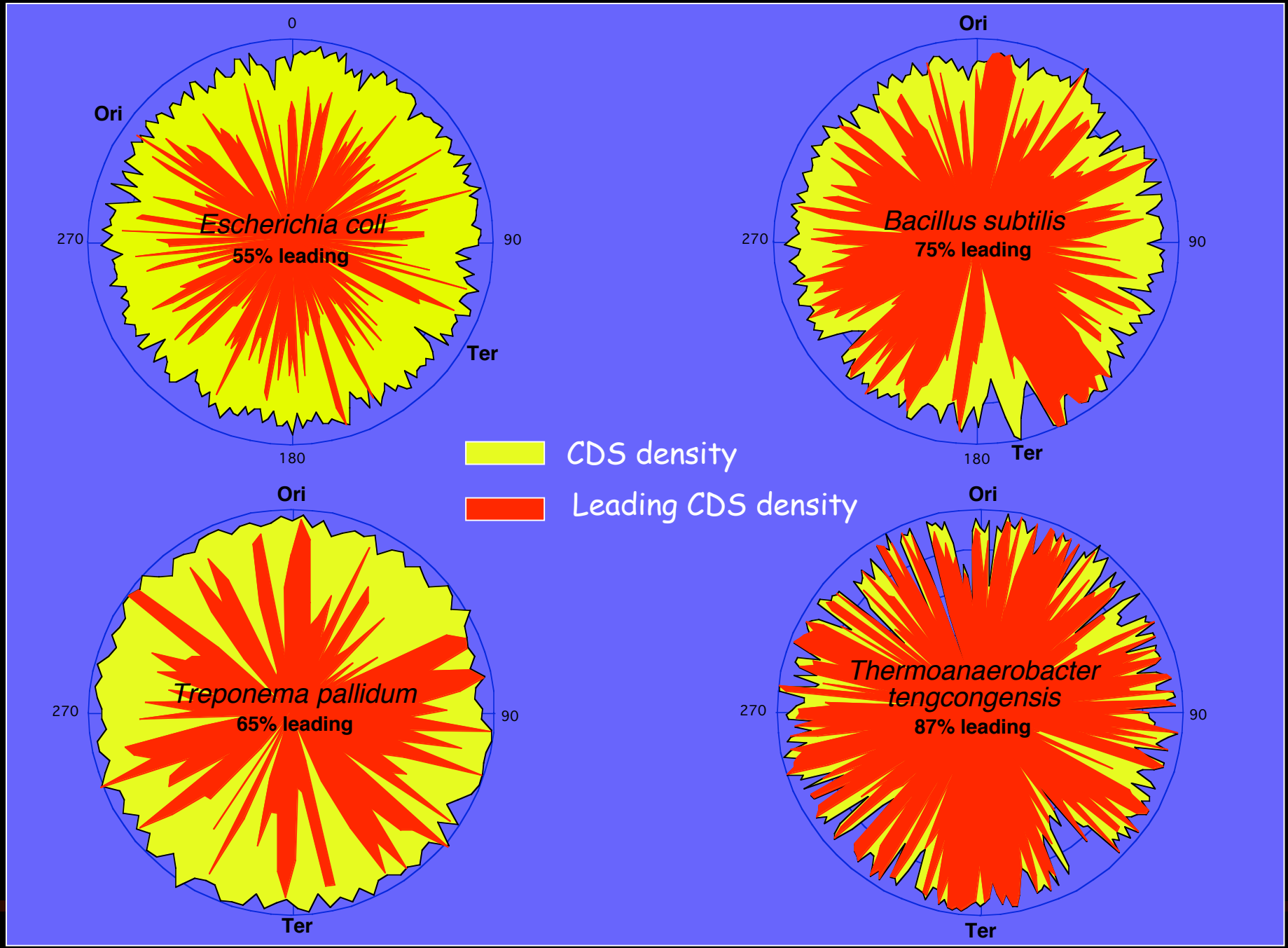
adanchin@pasteur.fr



INSTITUT PASTEUR

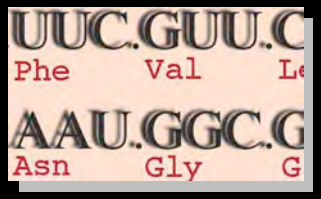


adanchin@pasteur.fr



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

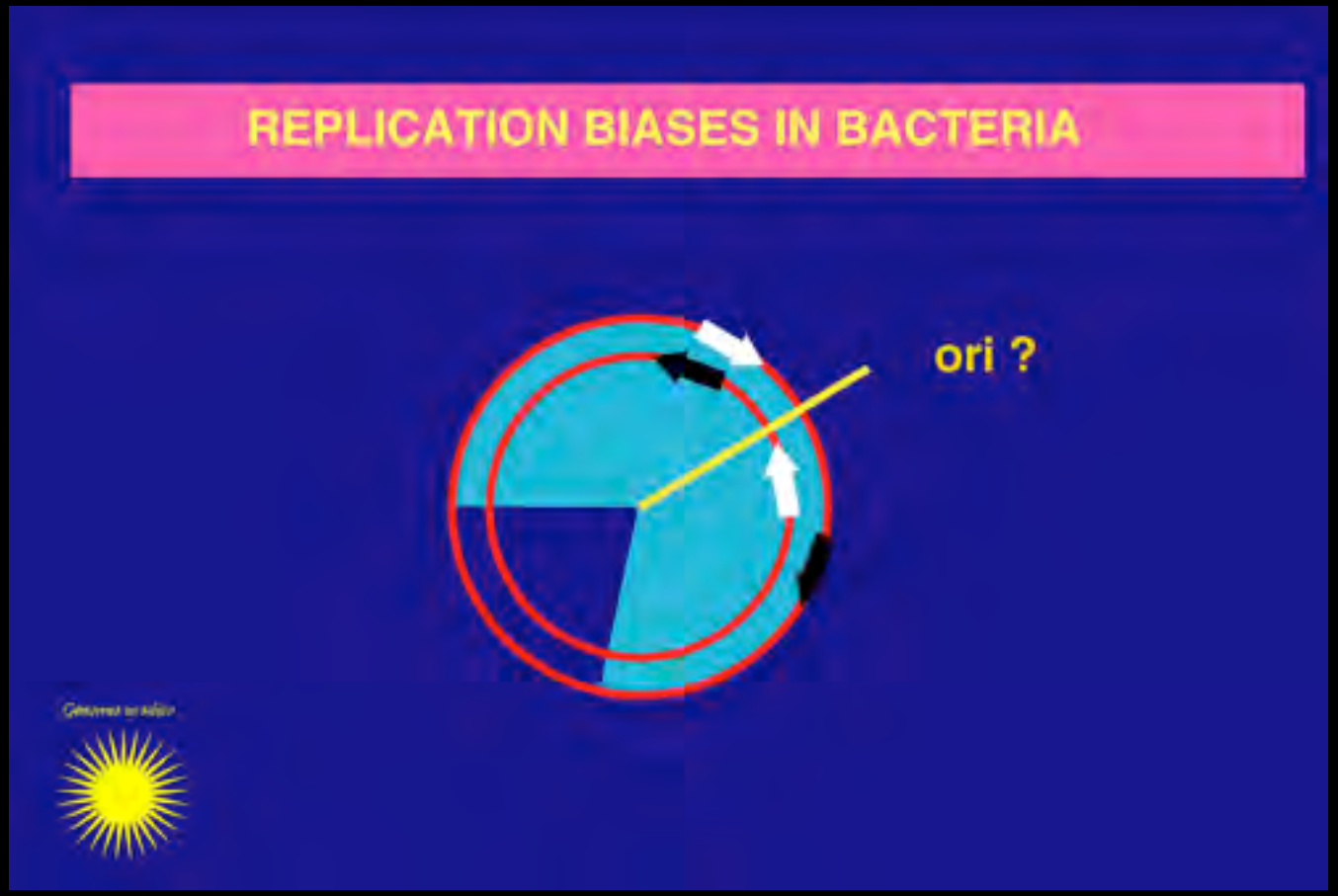
adanchin@pasteur.fr



To lag or to lead...

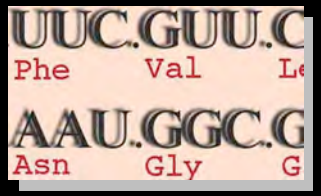
INSTITUT PASTEUR

Chosing arbitrarily an origin of replication and a property of the strand (base composition, codon composition, codon usage, amino acid composition of the coded protein…) one can use discriminant analysis to see whether the hypothesis holds.



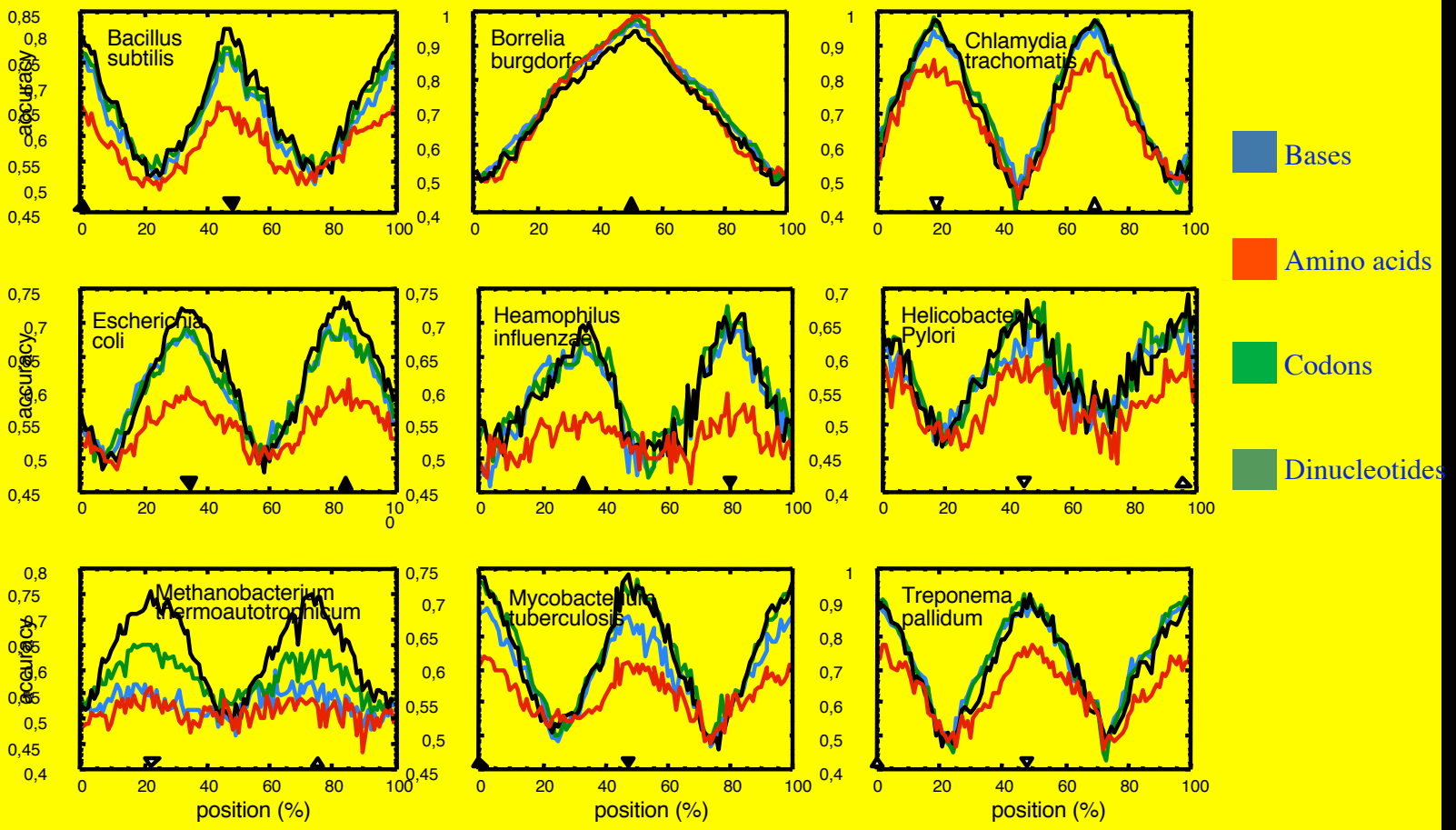
E. Rocha, A. Danchin & A. Viari Universal replication biases in bacteria. Mol. Microbiol. (1999) 32: 11-16

adanchin@pasteur.fr

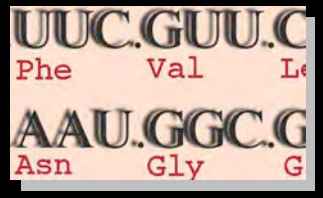


To lag or to lead, that is the question

INSTITUT PASTEUR

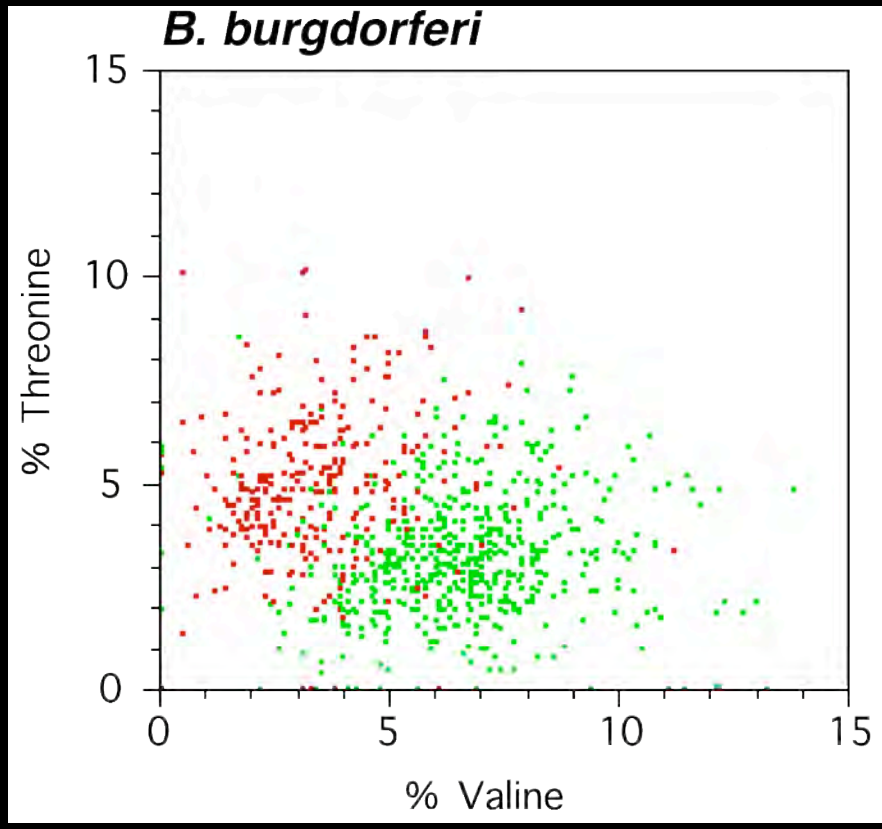


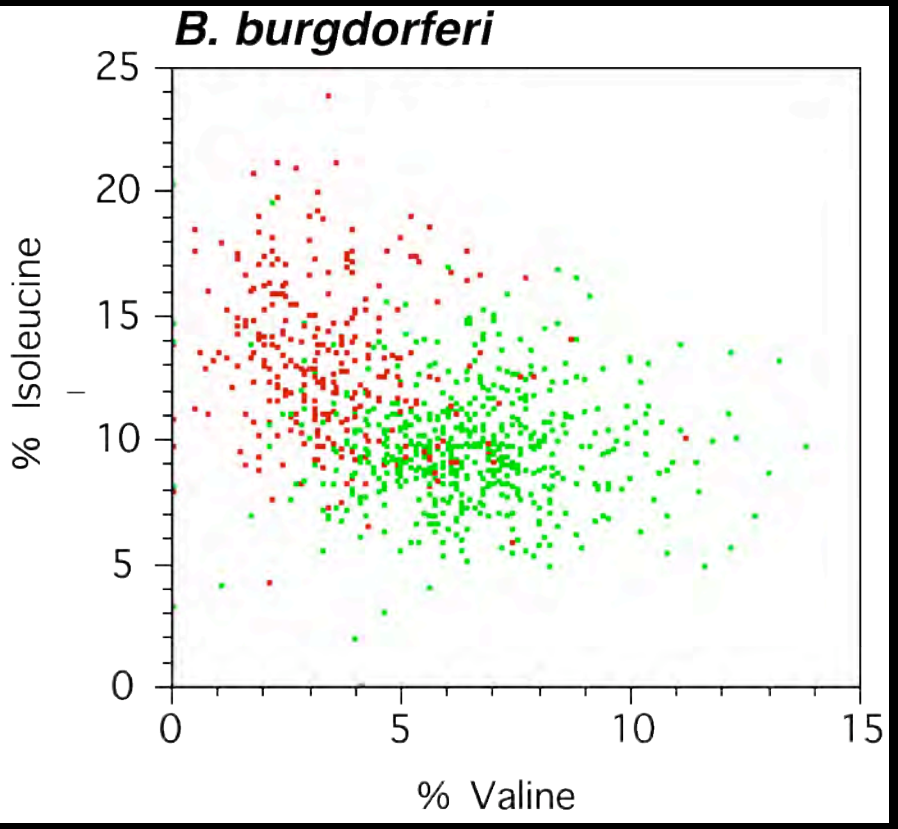
Genetics of Bacterial Genomes
http://www.pasteur.fr/recherche/unites/REG/
adanchin@pasteur.fr



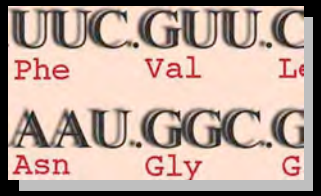
Visible even in proteins…

INSTITUT PASTEUR



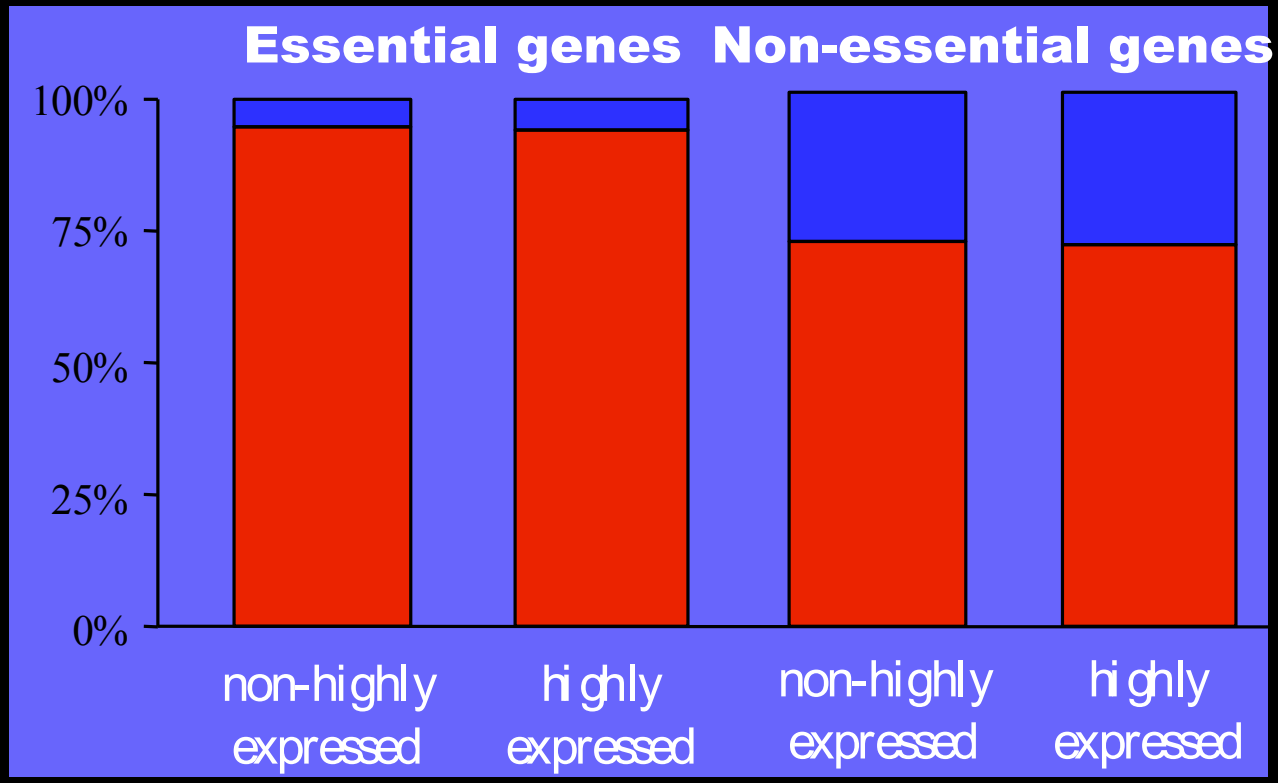


adanchin@pasteur.fr



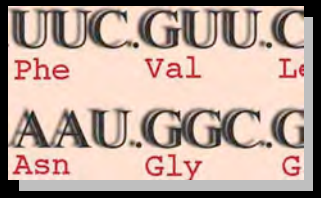
Essentiality in *B. subtilis*

INSTITUT PASTEUR



EPC Rocha, A Danchin
Essentiality, not expressiveness, drives gene-strand bias in bacteria
Nature Genetics (2003) **34**: 377-378

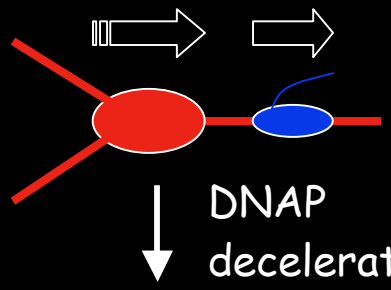
adanchin@pasteur.fr

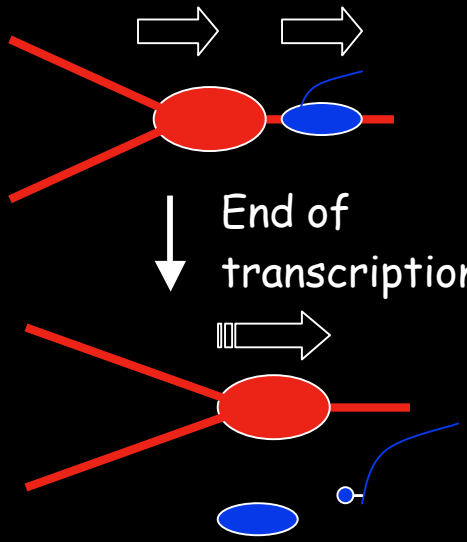


When polymerases collide

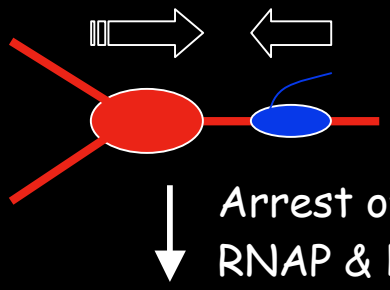
INSTITUT PASTEUR

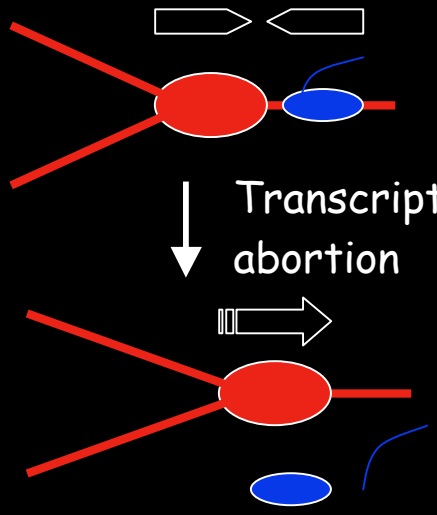
Co-oriented





Head-on





~~Consequences:~~

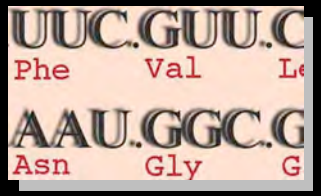
~~1. Replication slow-down~~

~~2. Loss of transcripts~~

Consequences:

1. Aborted transcripts
2. Truncated essential proteins

adanchin@pasteur.fr



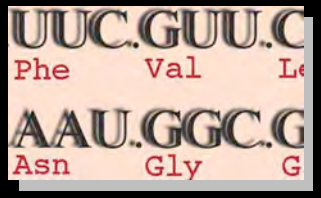
Three examples of the role of the context

INSTITUT PASTEUR

- Microbial genes are of infinite diversity but there exists universals; only about 10% of their genes are of persistent and recognized function; we do not have yet a fair idea of the number of microbial species; the number of genes in a given species is highly variable (horizontal gene transfer)

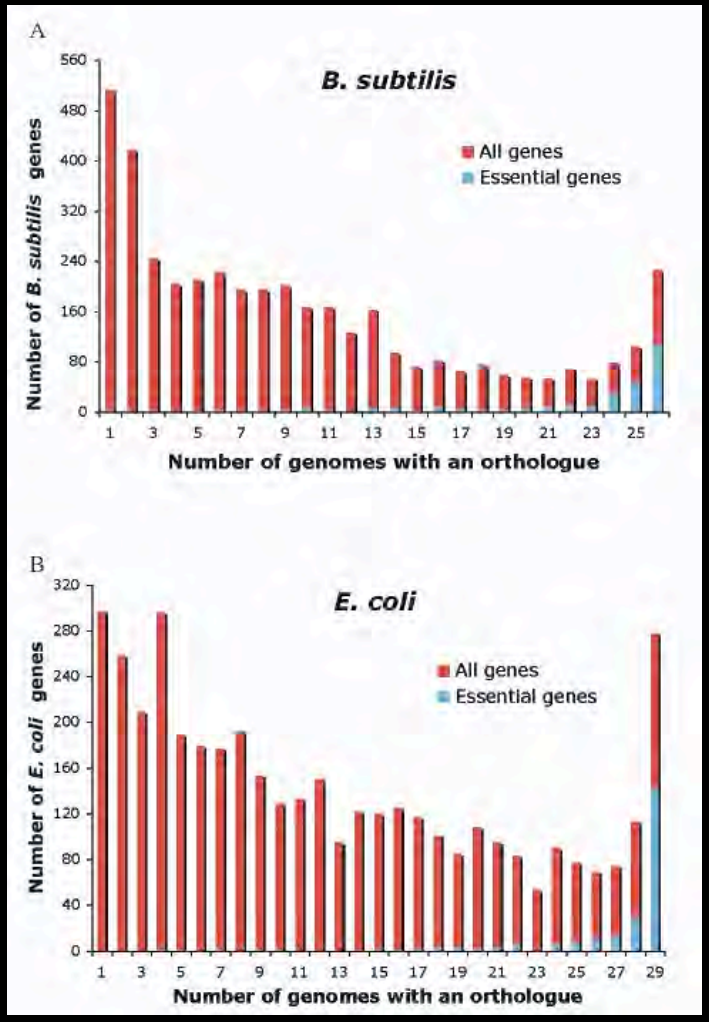
- Example 1: persistent genes
- Example 2: orphan genes and universal amino acids
- [Example 3: a new metabolic pathway]
- ....

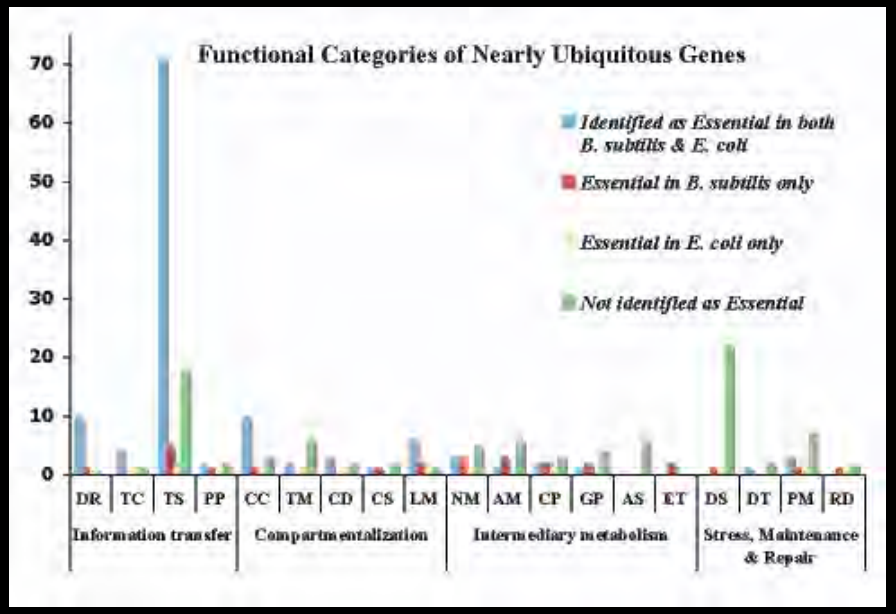
adanchin@pasteur.fr



An extension of essentiality: Gene persistence

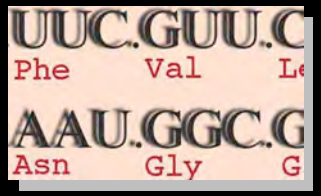
INSTITUT PASTEUR





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

adanchin@pasteur.fr

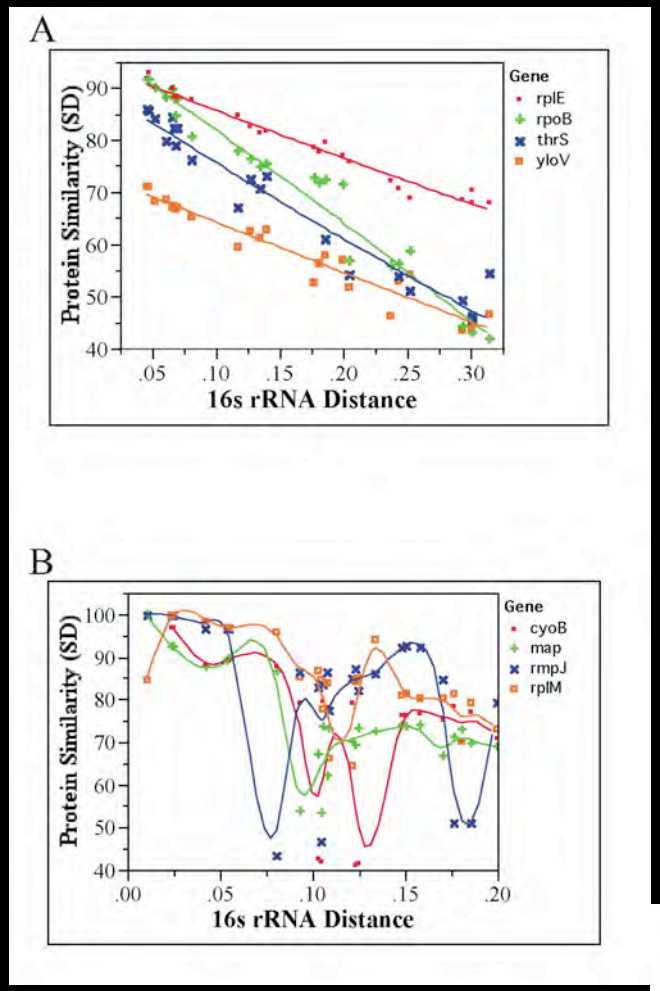


Gene persistence

INSTITUT PASTEUR

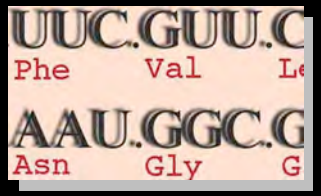
Some of the genes missing from the list of persistent genes have diverged considerably. To assess the contribution of this effect we measured for each pair of genomes the correlation between the similarity of orthologous pairs and that of the 16S rRNA. The correlations were high. For example (A), 38% (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.

In contrast, some genes (B) evolve in an erratic way. This may be due to horizontal gene transfer, local adaptations leading to faster or slower evolutionary pace, or simply wrong assignments of orthology. The latter can be a significant problem, especially in large protein families. The genes presenting such an erratic pattern are rare in the persistent set.



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

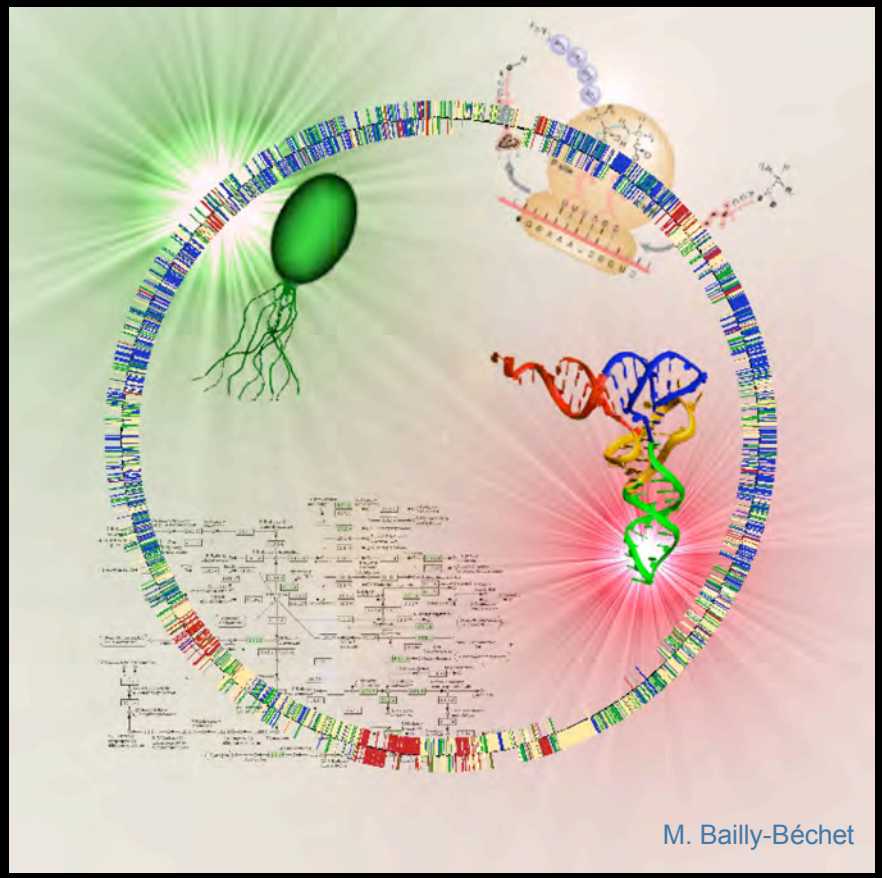
adanchin@pasteur.fr



Genomic islands

INSTITUT PASTEUR

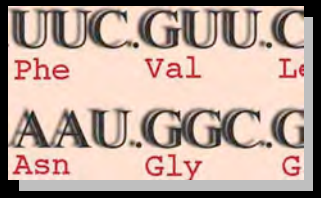
A clustering method based on the analysis of codon usage biases, using an information theory leads to group the genes into homogeneous clusters, which are not distributed randomly in the chromosome. One cluster corresponds to highly expressed genes. Other clusters are linked to specific functions or processes: horizontally transferred genes, motility or intermediary metabolism.



M. Bailly-Béchet

M Bailly-Bechet, A Danchin, M Iqbal, M Marsili, M Vergassola
Codon usage domains over bacterial chromosomes
PLoS Computational Biology (2006) **2**: april 20th

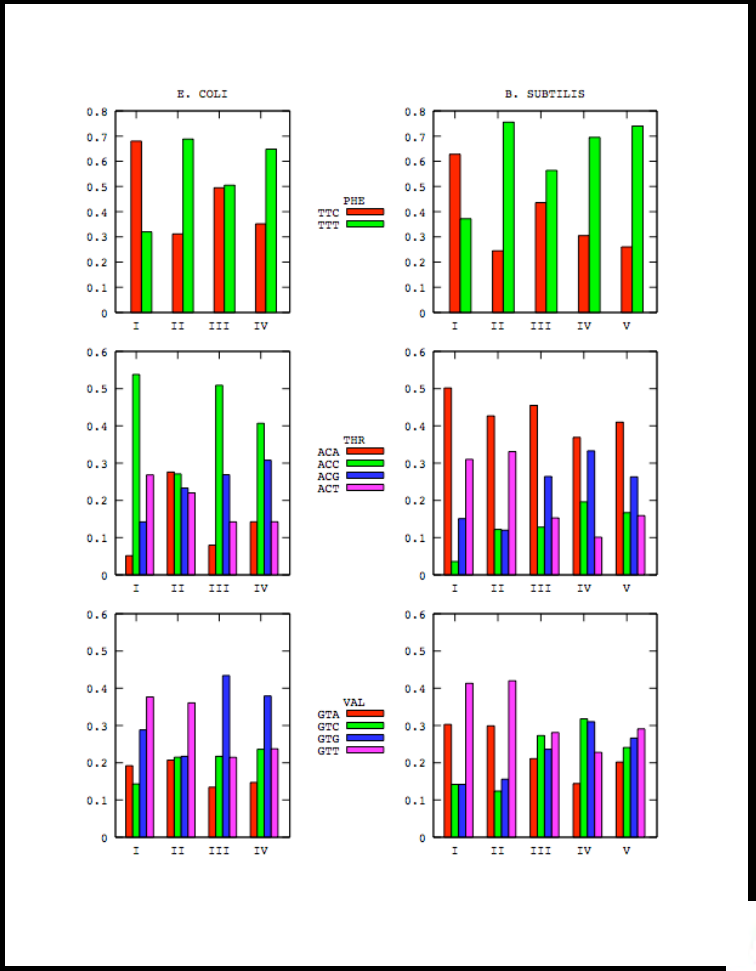
adanchin@pasteur.fr



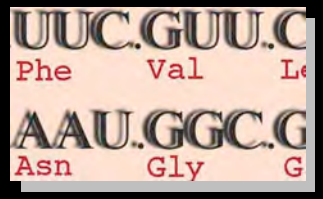
Genome islands

INSTITUT PASTEUR

One cluster is related to expression levels. Other groups feature an over-representation of genes belonging to different functional groups: horizontally transferred genes, motility and intermediary metabolism. Genes with a similar bias are close on the chromosome and organized in coherent domains, more extended than operons, demonstrating a role of translation in structuring bacterial chromosomes. A sizeable contribution to this effect comes from the dynamic compartimentalization induced by the recycling of tRNAs, leading to gene expression rates dependent on their genomic and expression context

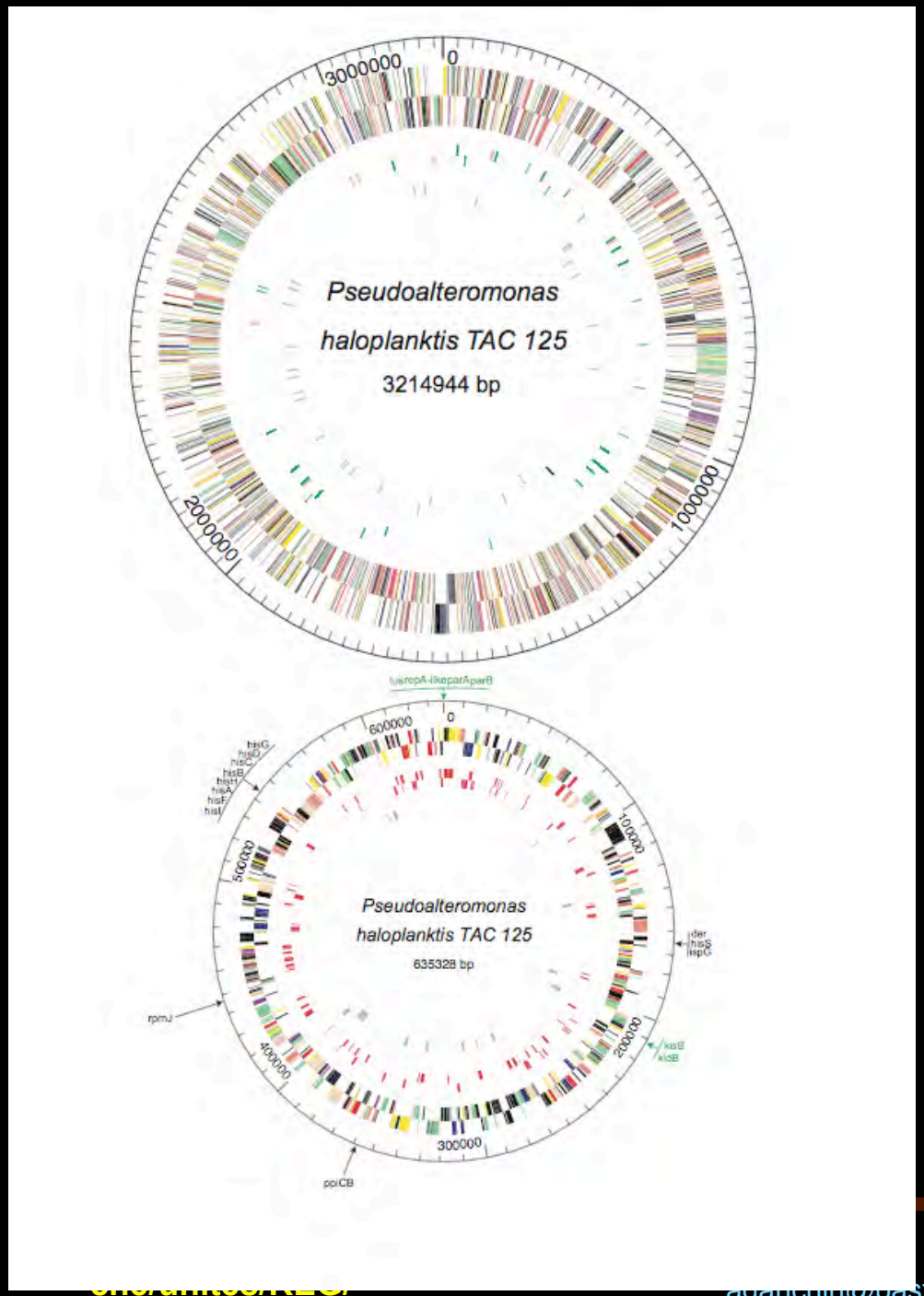


Genetics of Bacterial Genomes
http://www.pasteur.fr/recherche/unites/REG/
adanchin@pasteur.fr



Genome organization

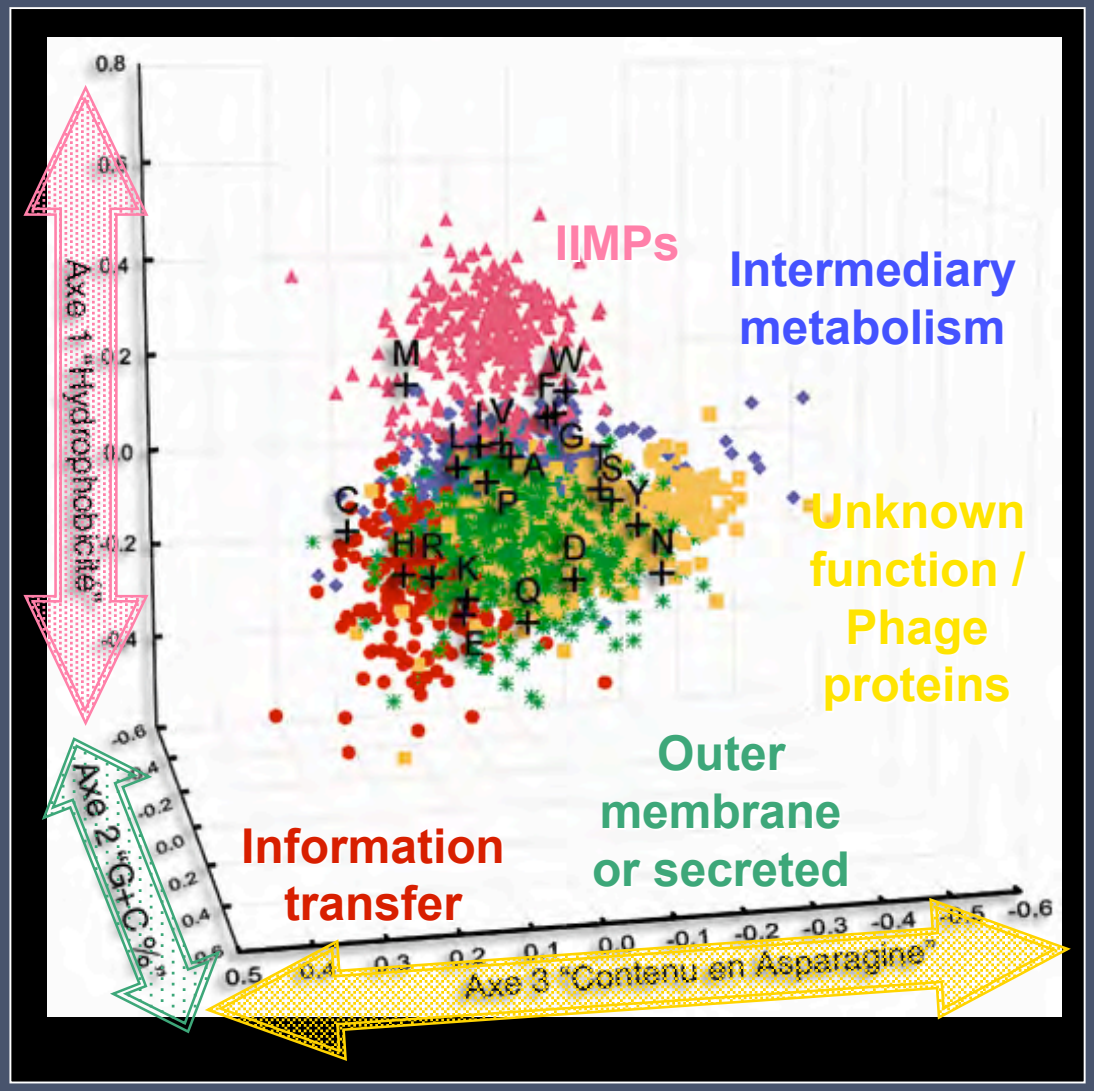
P. haloplanktis



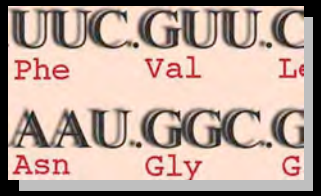
adanchin@pasteur.fr

Pseudoalteromonas haloplanktis

P. haloplanktis : Correspondence Analysis



adanchin@pasteur.fr



Universal biases in amino acid composition

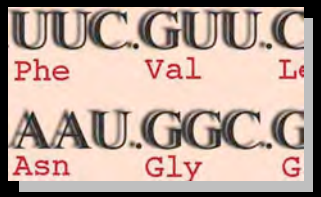
INSTITUT PASTEUR

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

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

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

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



Temperature-dependent biases in protein amino acid composition

INSTITUT PASTEUR

- 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

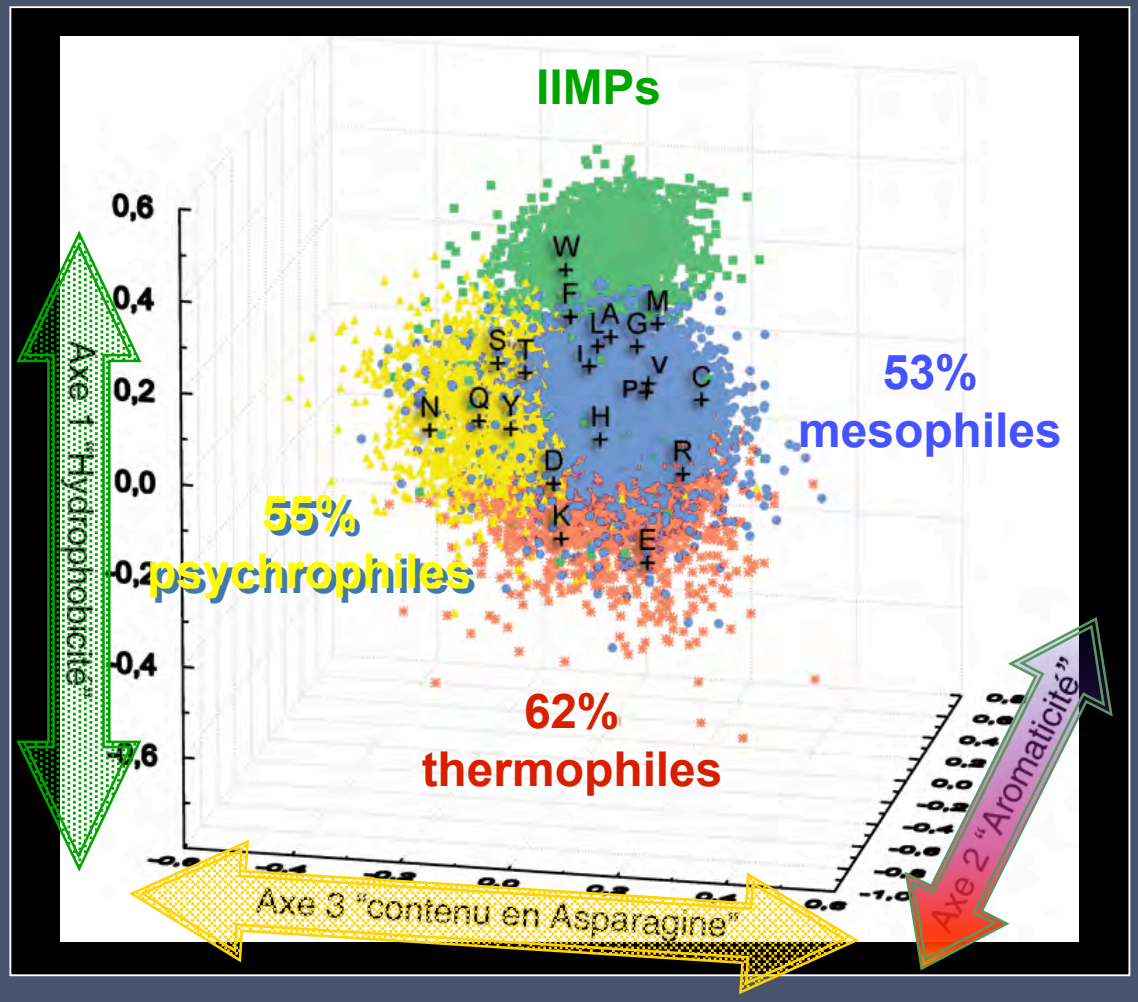
C Médigue, E Krin, G Pascal, V Barbe, A Bernsel, PN Bertin, F Cheung, S Cruveiller, S D'Amico, A Duilio, G Fang, G Feller, C Ho, S Mangenot, G Marino, J Nilsson, E Parrilli, EPC Rocha, Z Rouy, A Sekowska, ML Tutino, D Vallenet, G von Heijne, A Danchin
Coping with cold: the genome of the versatile marine Antarctica bacterium Pseudoalteromonas haloplanktis TAC125
Genome Research (2005) **15**: 1325-1335

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

Comparative proteomics

A specific asparagine bias in psychrophiles

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

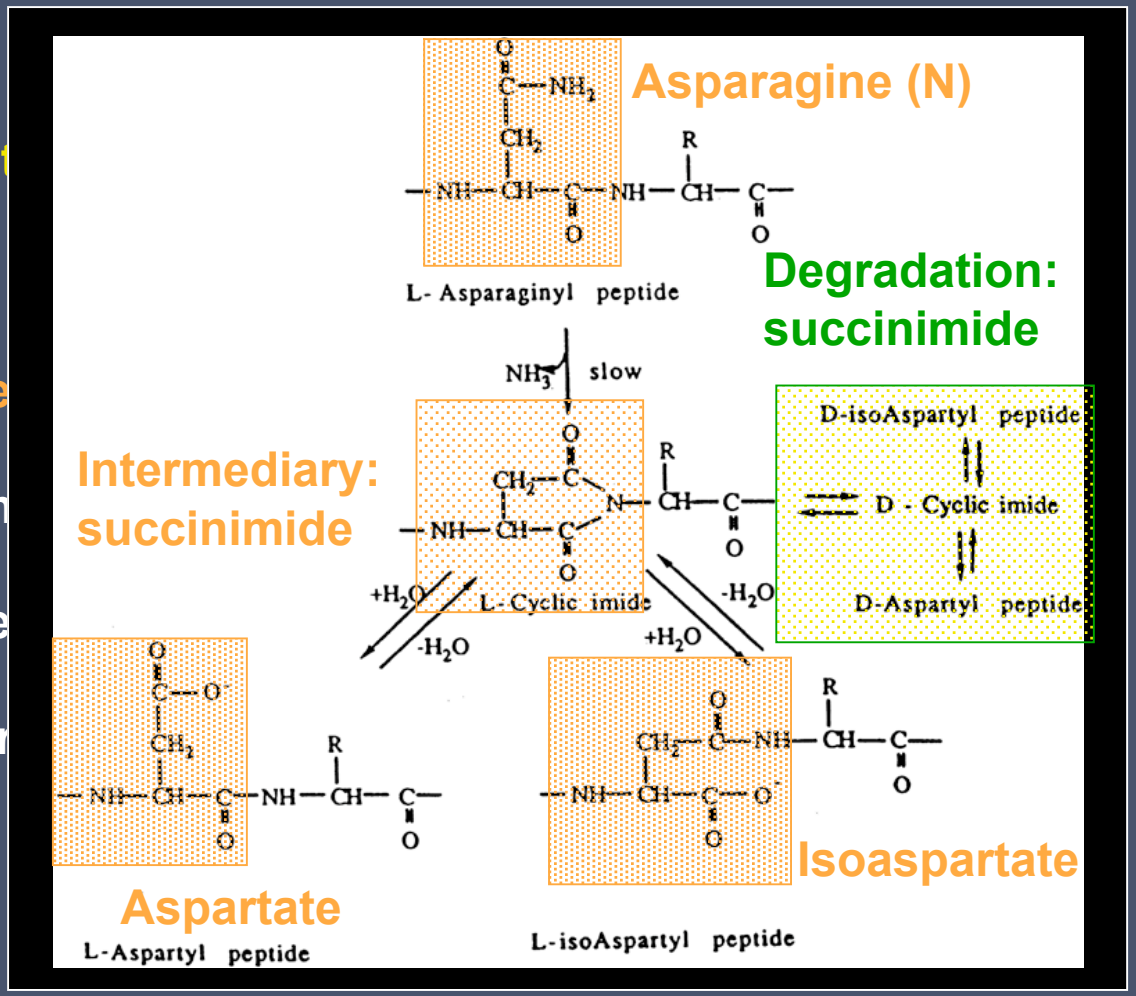


adanchin@pasteur.fr

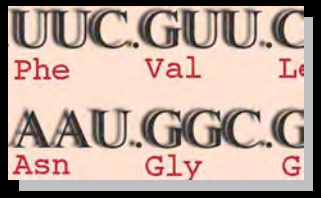
Chemistry

Asparagine deamidates: a major contribution to protein aging

- Main post
- Reaction
- Spontane
- Affects th
- Role in re
- Signal for



adanchin@pasteur.fr

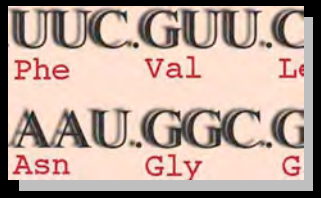


The first discovery of genomics

INSTITUT PASTEUR

In 1991, at the EU meeting on genome programs in Elounda, Greece, the presentation of the yeast chromosome III and the first 100 kb of the *Bacillus subtilis* genome revealed that, contrary to expectation (the only cases where this had been observed were phages, for obvious reasons), at least half of the genes uncovered were totally unknown, whether in structure or in function

adanchin@pasteur.fr



Orphans: the gluons

INSTITUT PASTEUR

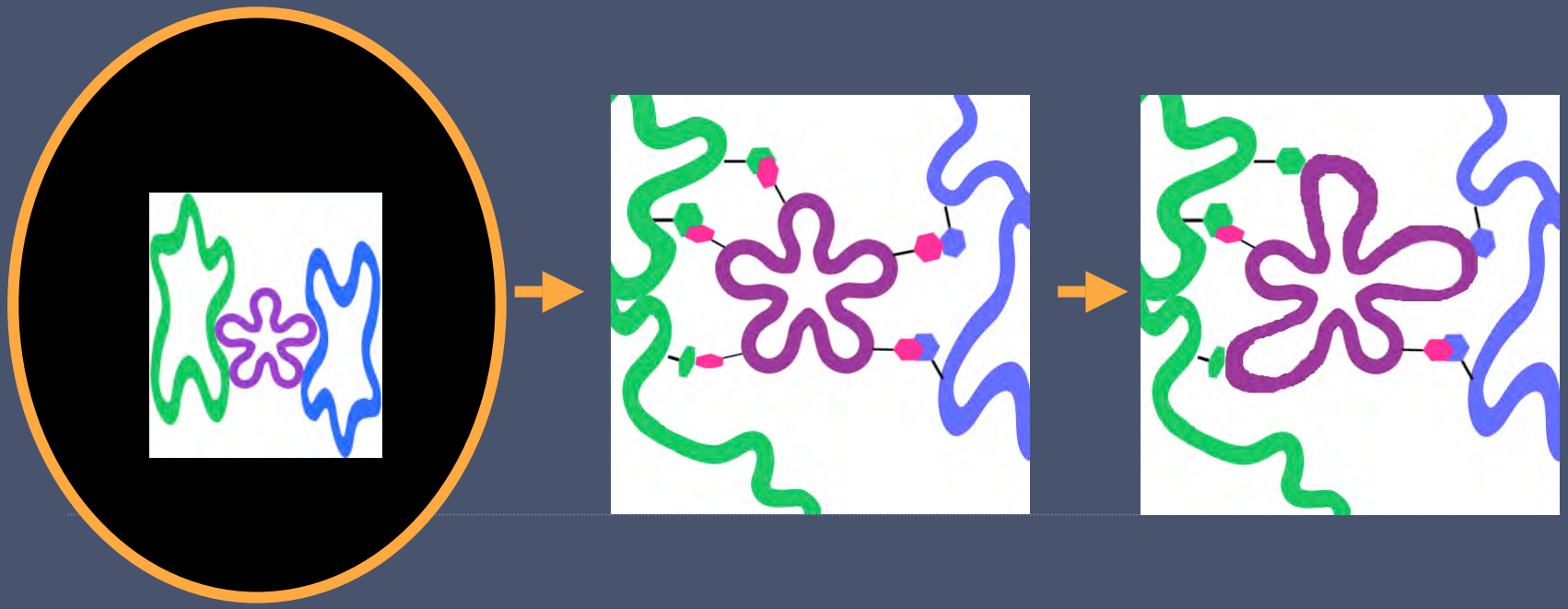
A remarkable role of aromatic amino acids creates a universal bias. Expressed orphan proteins are enriched in these residues, suggesting that they might participate in a process of gain of function during evolution. We postulate that the majority is made of proteins — **gluons** — involved in stabilising complexes, thus defining the "self" of the species.

G Pascal, C Médigue, A Danchin
Universal biases in protein composition of model prokaryotes
Proteins (2005) **60**: 27-35

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

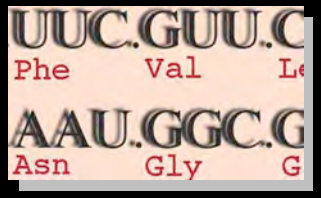
Why aromatic amino acids in orphan proteins?

From Orphans to « **Gluons** »



♣ Orphan proteins loose their status during evolution *Rocha. 2002. Pedulla. 2003*

adanchin@pasteur.fr



INSTITUT PASTEUR

Thank you

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

adanchin@pasteur.fr