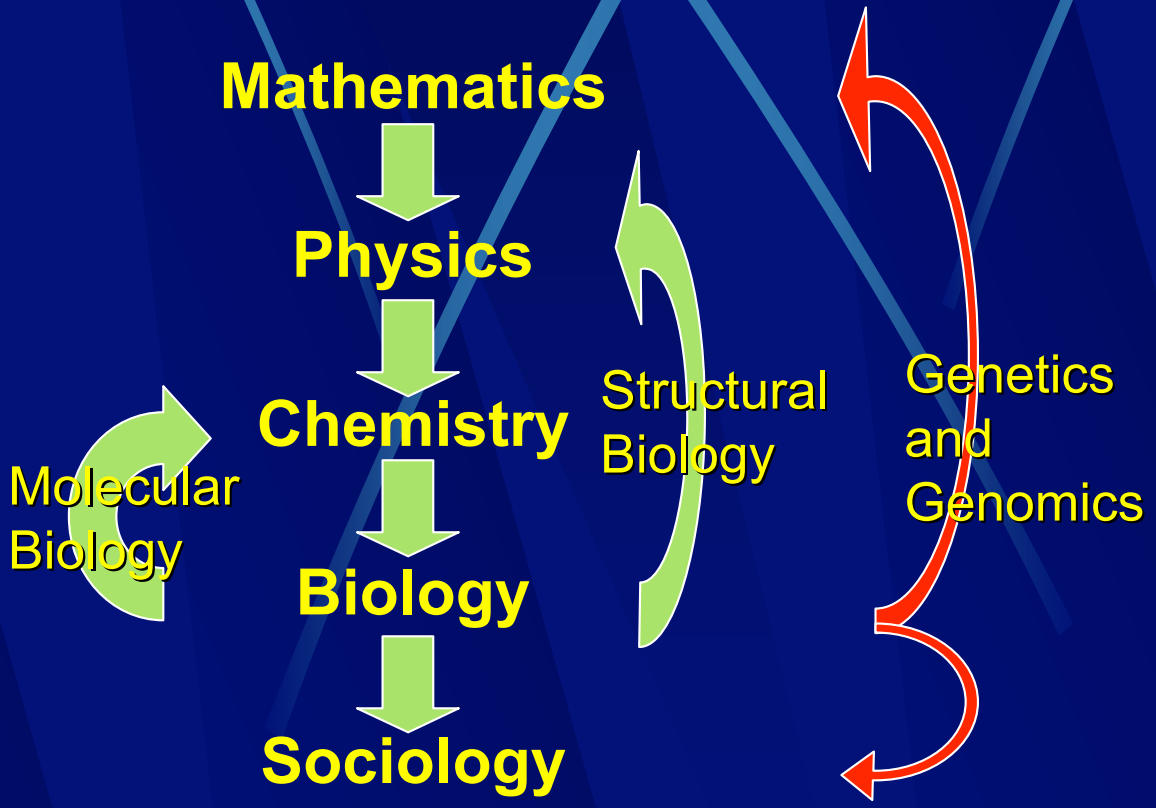
The cell as a living computer

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

adanchin@pasteur.fr

A preconceived ideology



Background

- Physics: *matter, energy, time*
- Biology: Physics + *information, coding, control...*
- Arithmetics: *strings of whole numbers, recursivity, coding…*
- Computing: Arithmetics + *program* + *machine...*

adanchin@pasteur.fr

Information Transfer

As is the case for building up a machine, one needs a book of recipe to build up a cell

This asks for changing the text of the recipe into something concrete: this transfers « information »

In a cell, information transfer is managed by the genetic program

adanchin@pasteur.fr

What is Life?

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 (Internal organisation)**
- **Compartmentalization (General structure)**

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

What is computing?

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 program), split (in practice) into two entities:
  - Program (providing the goal)
  - Data (providing the context)

The machine is distinct from the program

The cell factory

A cell behaves like a computer that would program the construction of similar computers

It has a magnetic tape, or hard disk (the « genetic program ») and reading devices which allow it to read the program and put it into action

The « cloning » of the ewe Dolly was exactly that: changing the program from a machine (an egg) to another one (an egg without a nucleus)

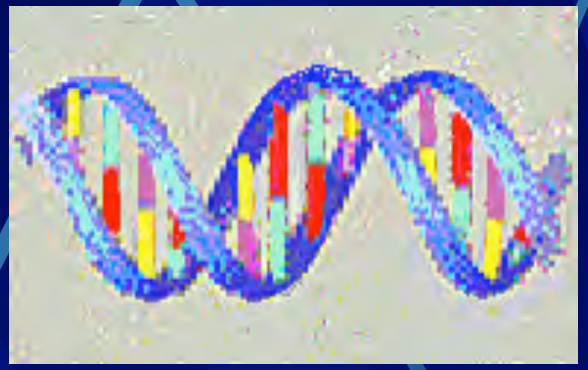
adanchin@pasteur.fr

From the recipe to the dish: from the genetic program to the cell

香港大學－巴斯德研究中心
HKU-Pasteur Research Centre

INSTITUT PASTEUR

When you read the recipe, you perform actions to make the dish. A special machinery reads the DNA and copies it into active agents, the proteins (enzymes are proteins).



DNA

protein

adanchin@pasteur.fr

Cells as computers

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

Conjecture: do cells behave as computers?

Genetic engineering

Viruses

Horizontal gene transfer

Cloning animal cells

all point to separation between

Machine

Data + Program

adanchin@pasteur.fr

Is there a map of the cell in the chromosome?

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** (J. von Neumann)

A. Danchin The Delphic Boat. What genomes tell us (2003) Harvard University Press

adanchin@pasteur.fr

Genome organisation

Is the gene order random in the chromosomes?

At first sight, despite 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

First question: how are generated and where are located repeats in the genome sequence?

adanchin@pasteur.fr

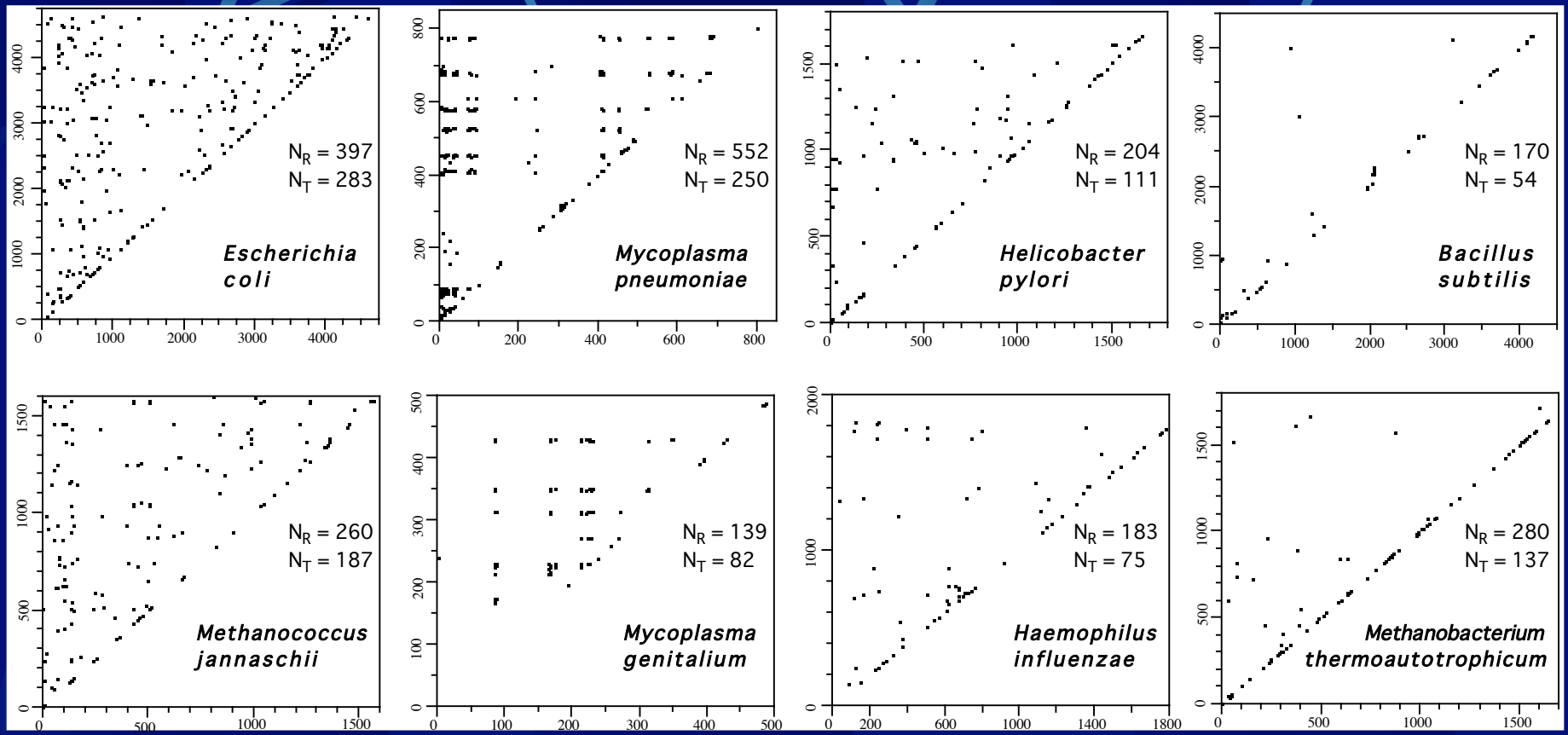
Repeats in bacteria

- Abcissa: first occurrence of the repeat
- Ordinate: second position of the repeat

- Diagonal: repeats are located near to each other

adanchin@pasteur.fr

DNA management: Repeats in genomes



adanchin@pasteur.fr

Repeats: DNA management differs according to organism

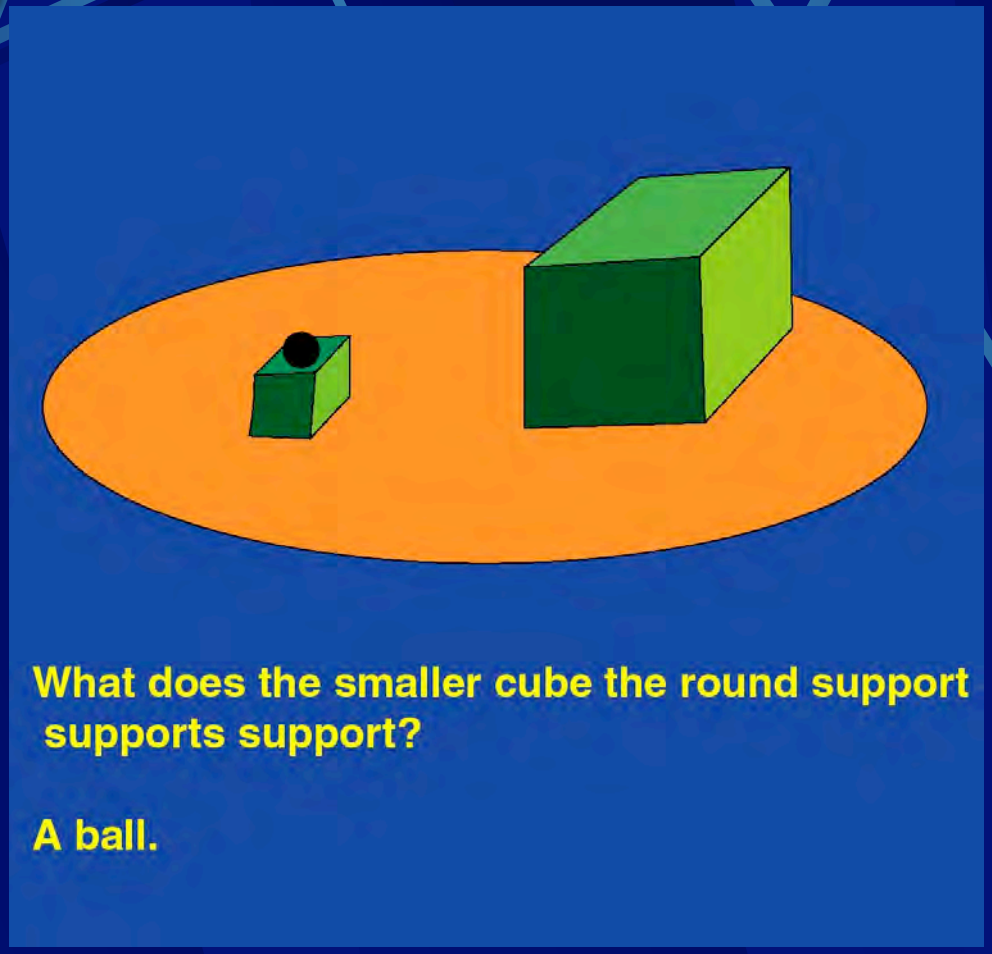
- No correlation with the length of the genome
- Non-random distribution of repeats in bacteria that can catch up DNA from the environment
- When repeats are rare, they are located not far from each other (10-15 kb)
- DNA is managed very differently in different bacteria

- A side view: genomes from higher cells are much more repeated than genomes of microbes, that look highly random at first sight

adanchin@pasteur.fr

Caveat: Repeats are meaningful

….
There is no junk DNA



Remember also:

This clock has a minute minute hand

adanchin@pasteur.fr

Genome organisation

Is the genes' order random?

At first sight, perhaps because of different DNA management processes, not much is conserved, and horizontally transferred genes are distributed throughout genomes

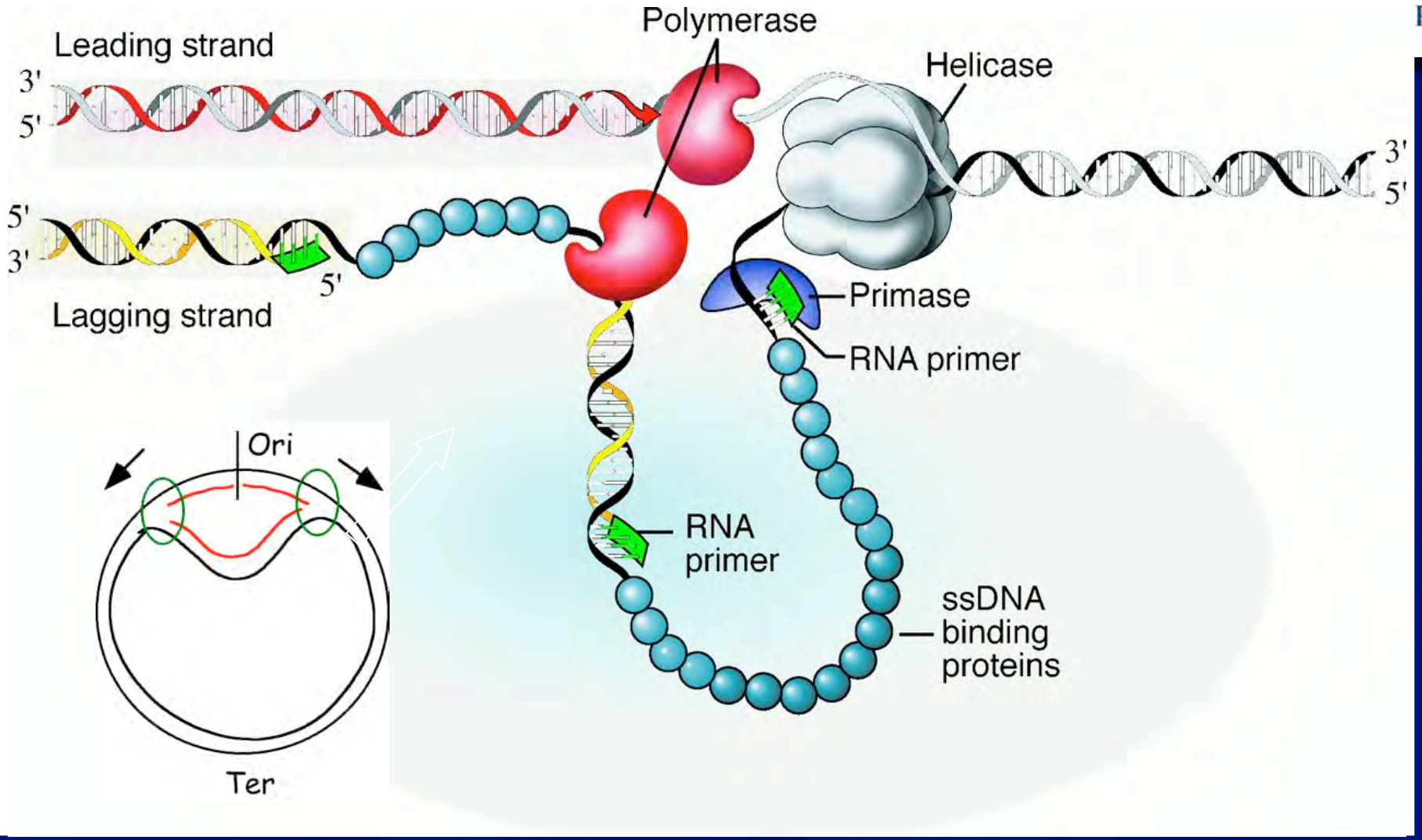
However, pathogenicity islands tend to cluster at specific places, and they code for proteins with common functions

adanchin@pasteur.fr

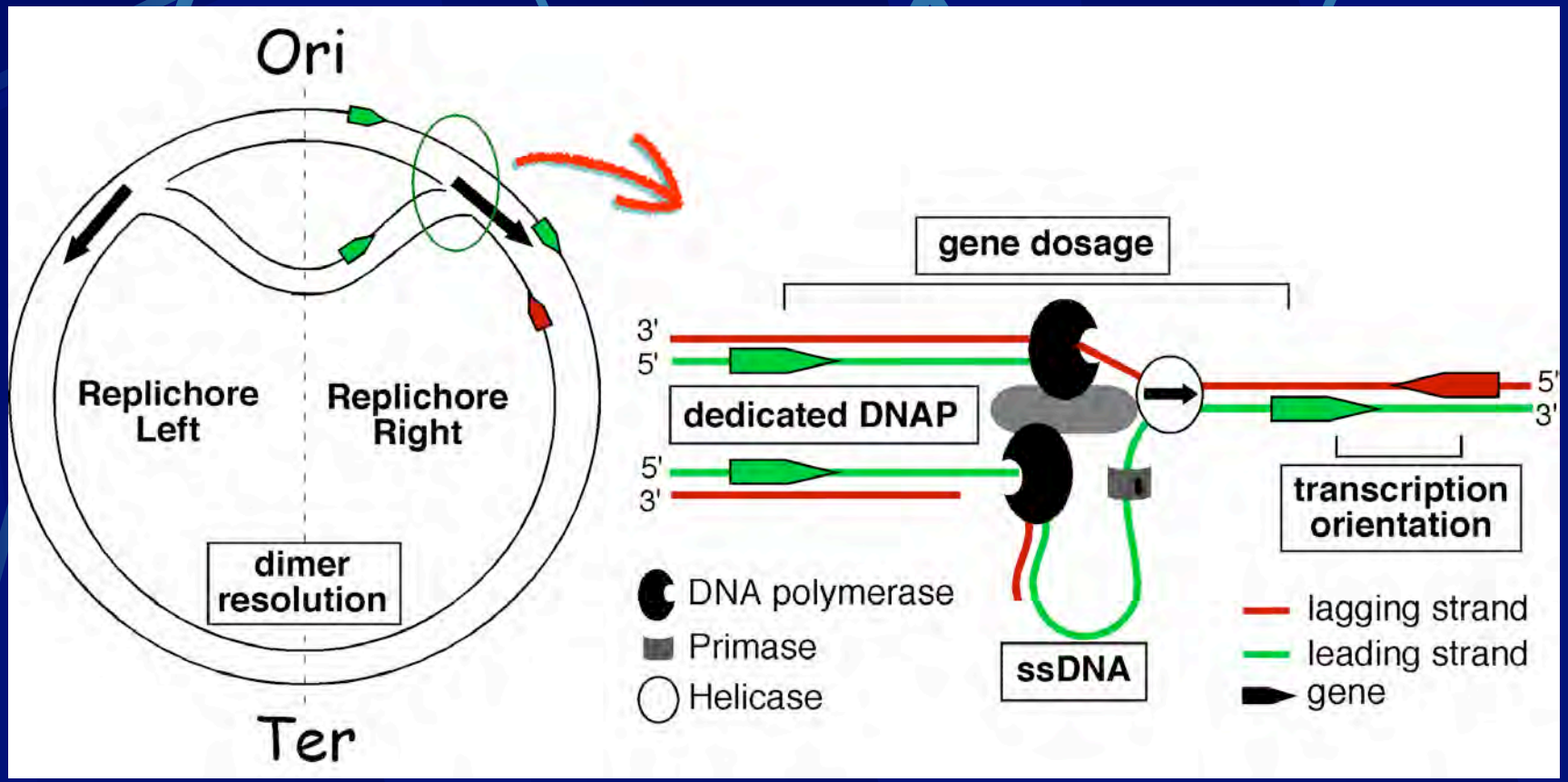
Genome organisation

The genome organisation 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

adanchin@pasteur.fr



adanchin@pasteur.fr



adanchin@pasteur.fr

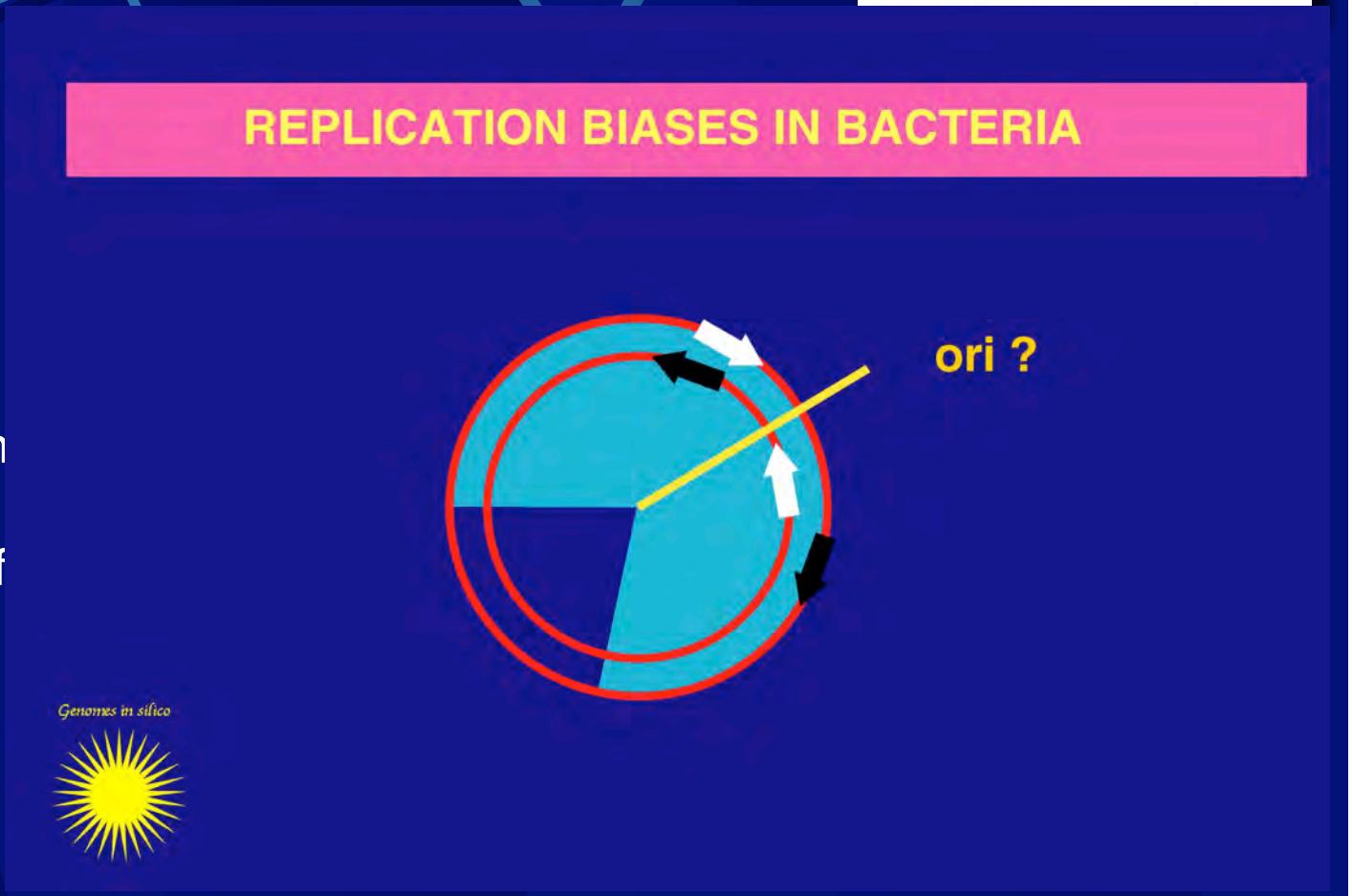
To lead or to lag...

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

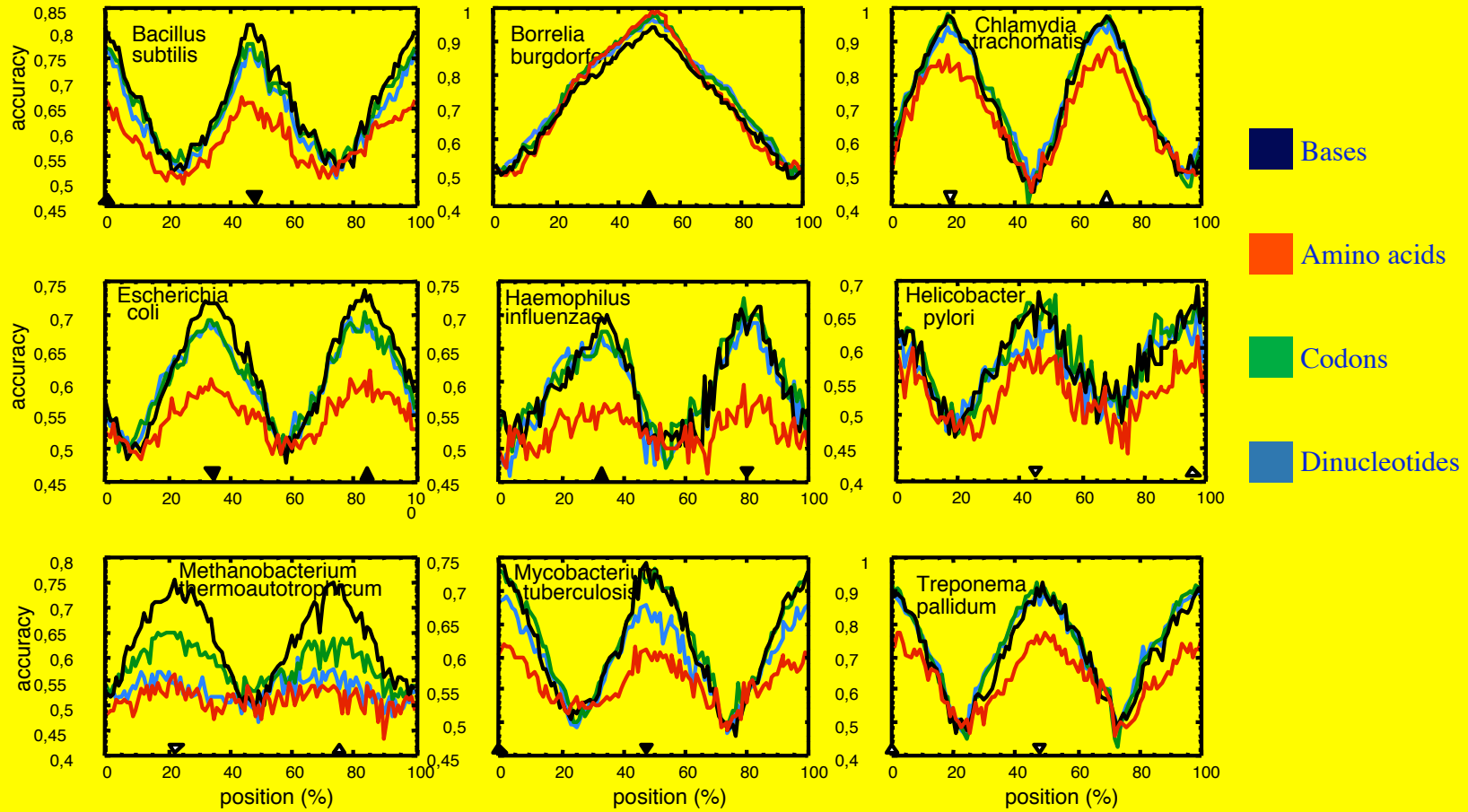
To lag or to lead…

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



adanchin@pasteur.fr

To lag or to lead, that is the question



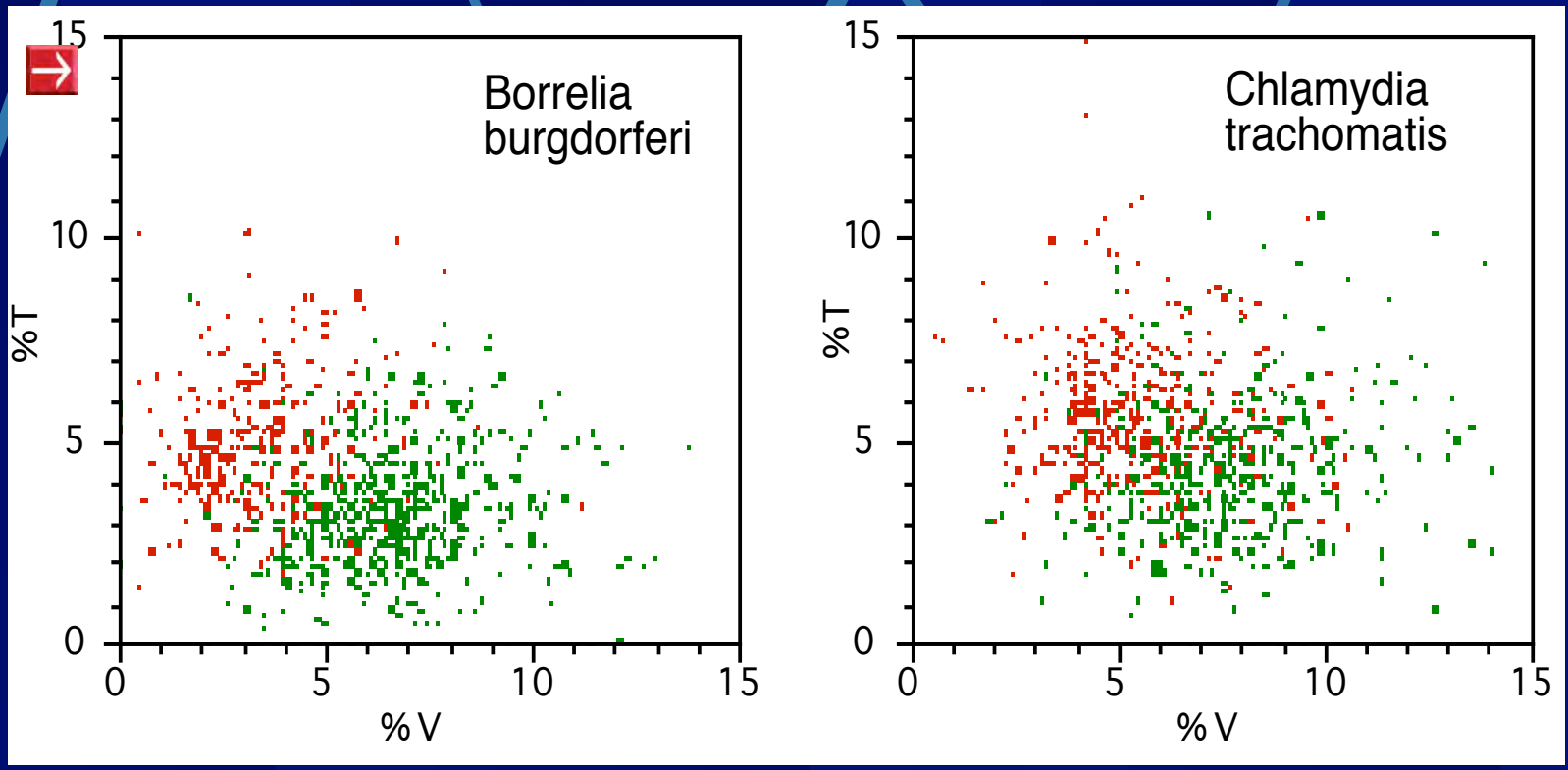
adanchin@pasteur.fr

Conclusion 1

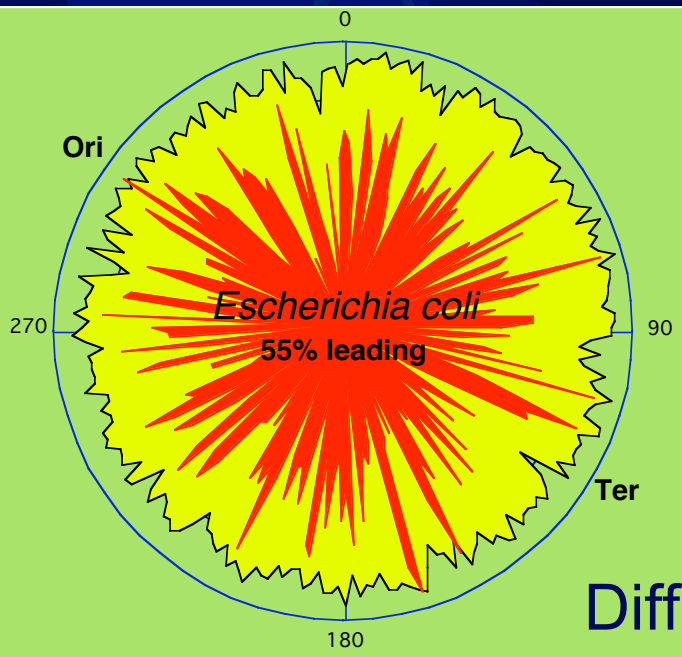
Proteins are made of 20 amino acid types, among which Valine and Threonine, and one observes that Valine-rich protein are on the leading strand while Threonine-rich proteins are on the lagging strand! Isologous proteins replace preferentially one residue for the other when their gene change strand

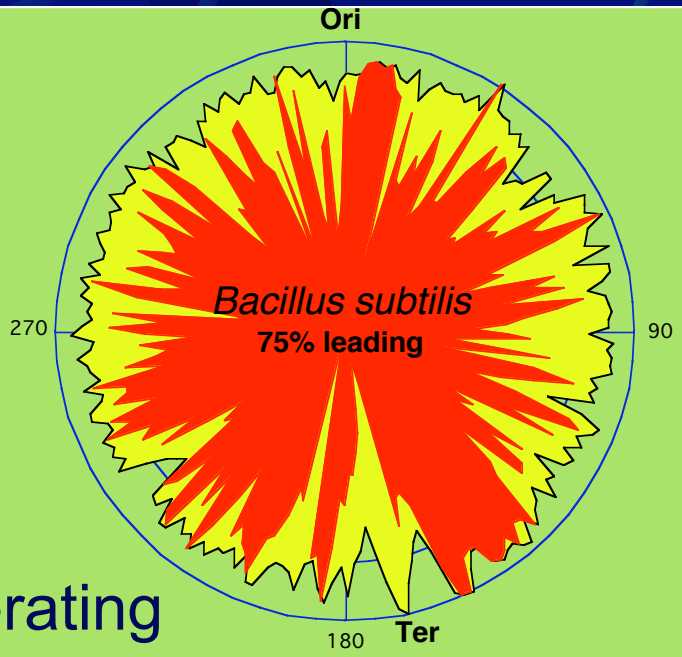
adanchin@pasteur.fr

To lag or to lead, that is the question

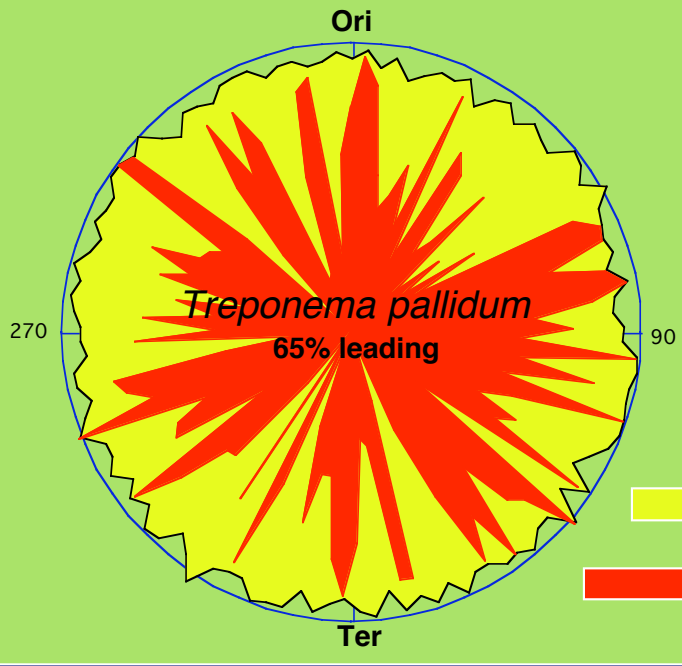


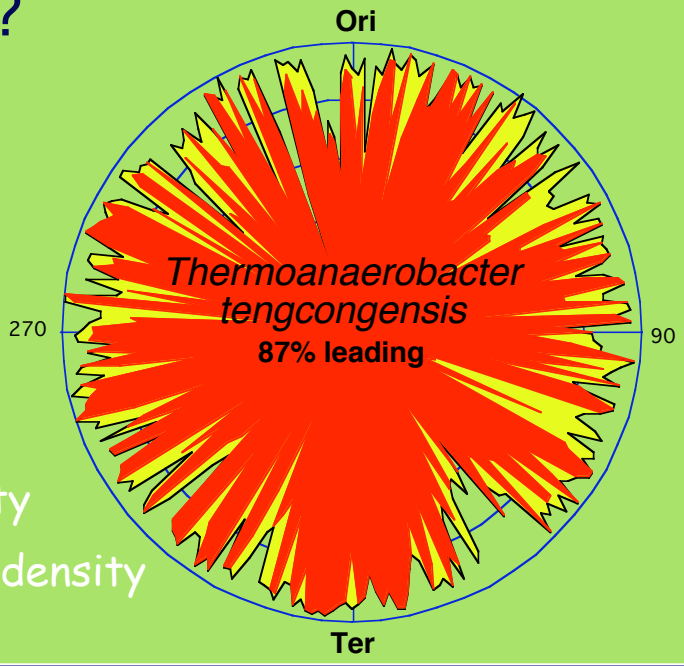
adanchin@pasteur.fr





Different "Operating Systems"?





CDS density

Leading CDS density

(updated from Kunst etal , Nature, 97)

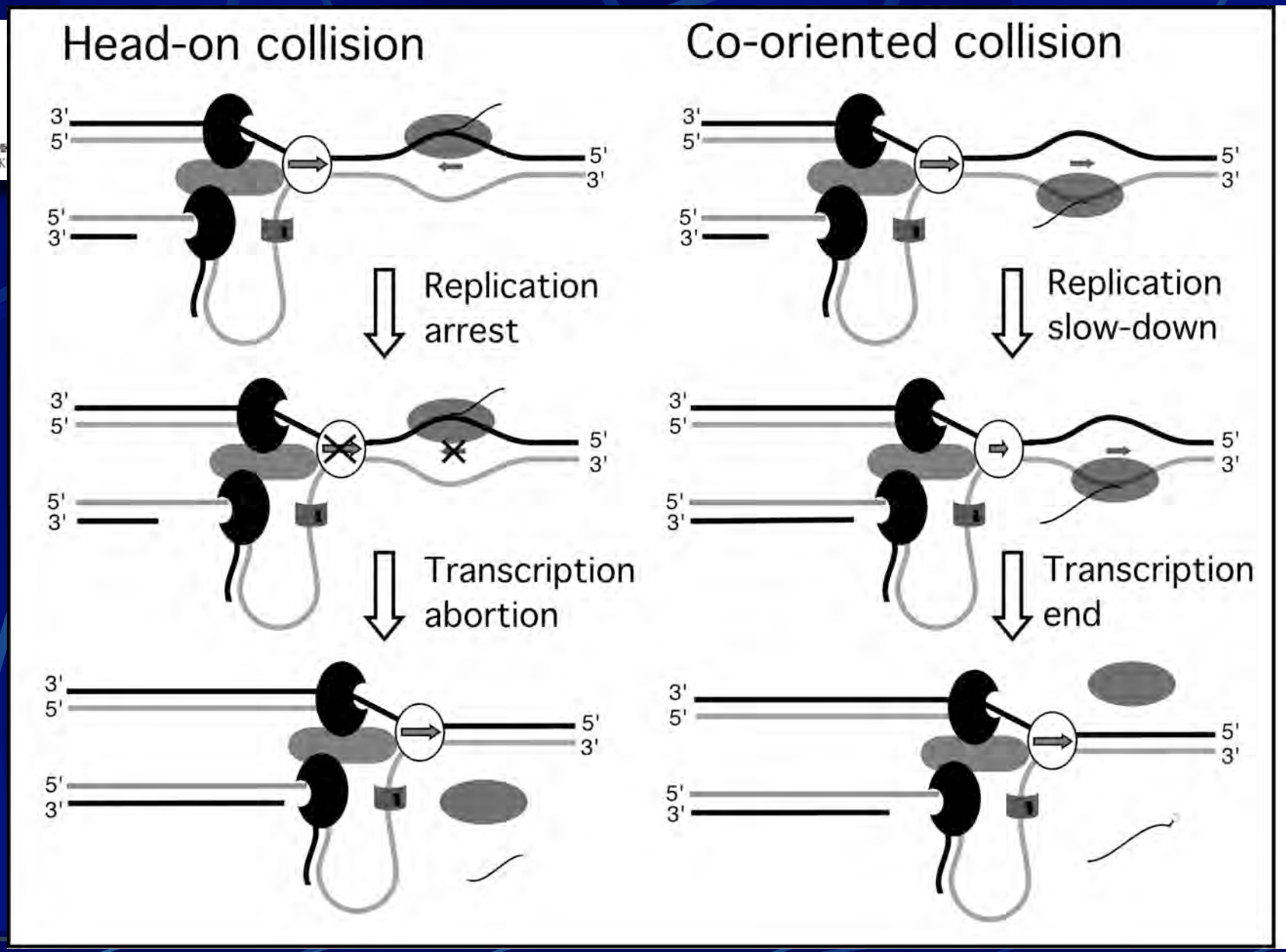
Conclusion 2 and more questions ...

The genome organisation is much more rigid than usually assumed. Some regions (such as the terminus) are rather unstable, but most of the genome structure is preserved through evolution. The distribution of genes on the leading and lagging strands is highly non-random. Is it associated to some particular function (such as the Operating System in a computer)? Where are essential genes located?

Essentiality, not expressivity dominates the strand-choice

- Most essential genes are located in the leading strand
- Many highly expressed genes are located in the lagging strand
- Essentiality organises the genome's architecture

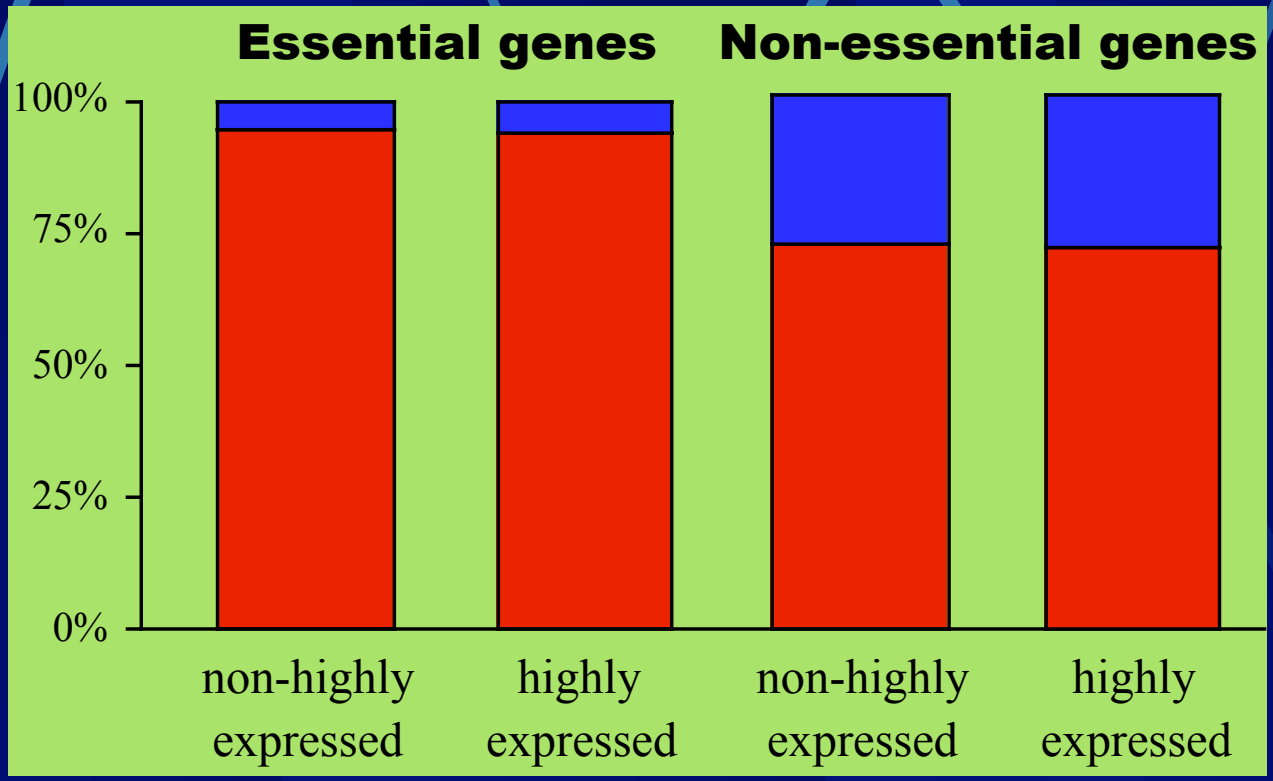
adanchin@pasteur.fr



(adapted from Brewer, Cell, 88)

adanchin@pasteur.fr

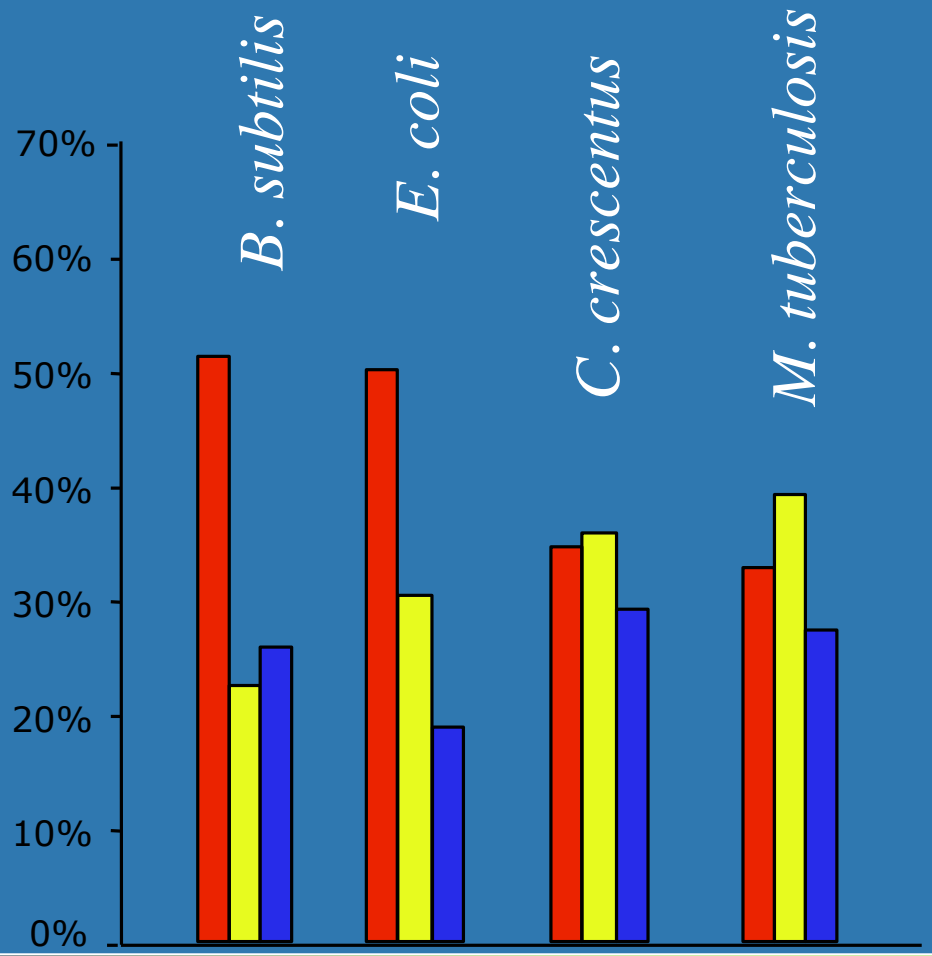
Essentiality in *B. subtilis*



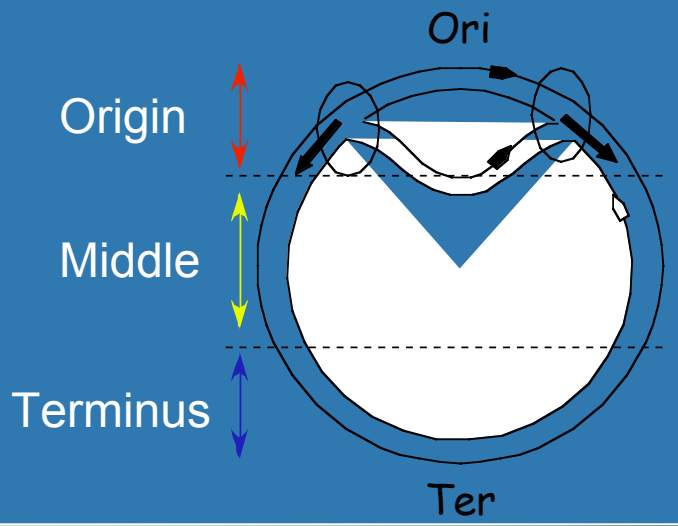
adanchin@pasteur.fr

Distribution of highly expressed genes

Fast growers | Slow growers

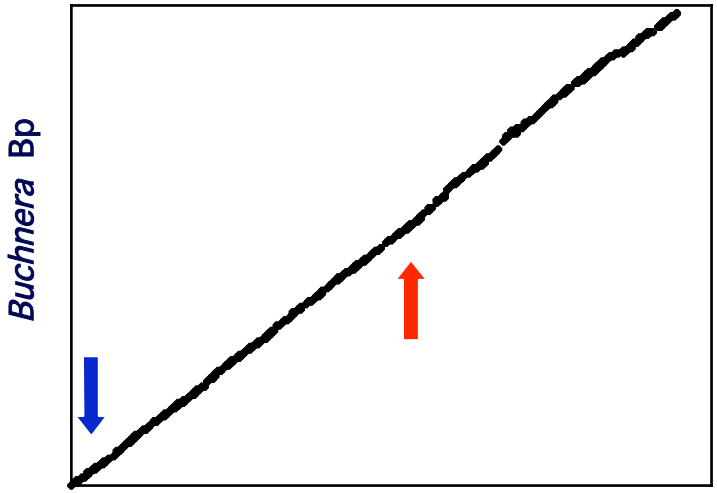


Highly expressed genes cluster near the origin in fast-growing bacteria

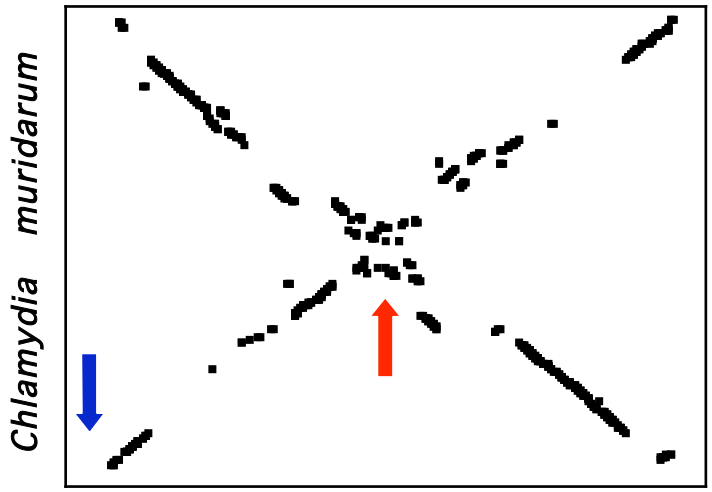


adanchin@pasteur.fr

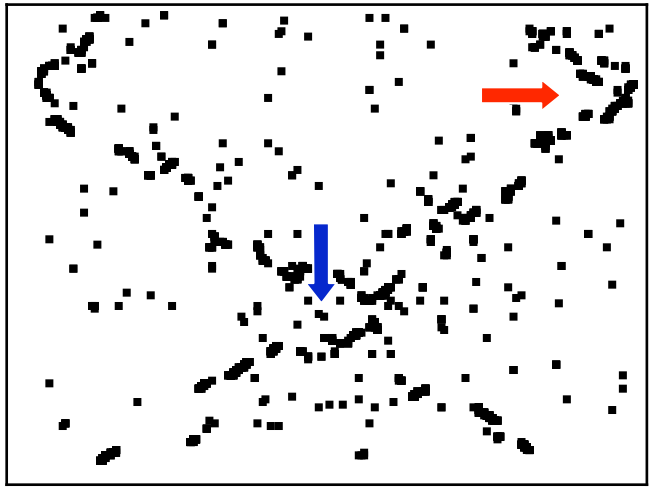
Gene order conservation



16S:.076 nt^{-1}



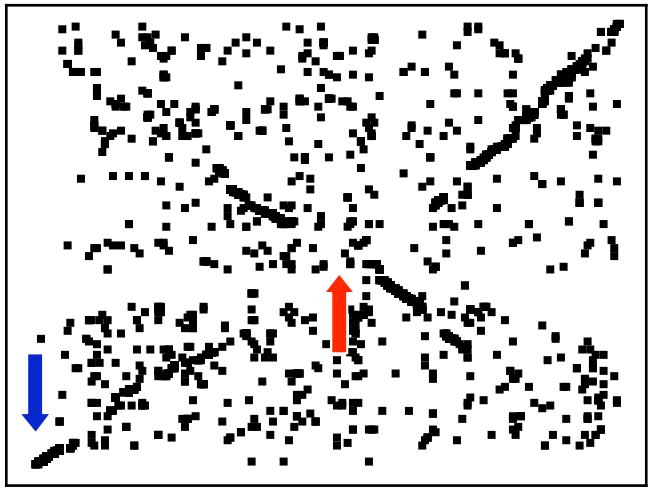
16S:.071 nt^{-1}



16S:.063 nt^{-1}

origin

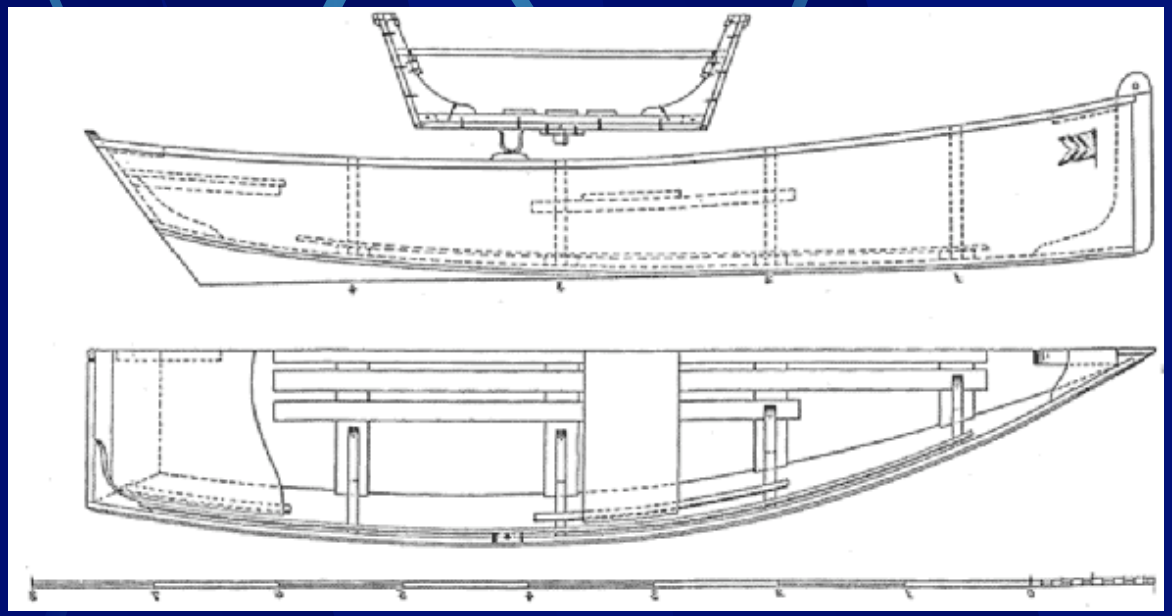
terminus



16S:.051 nt^{-1}

Exploring "neighborhoods"

- Genes do not operate in isolation
- Proteins are part of complexes, as are parts in an engine
- It is important to understand their relationships, as those in the planks which make a boat



adanchin@pasteur.fr

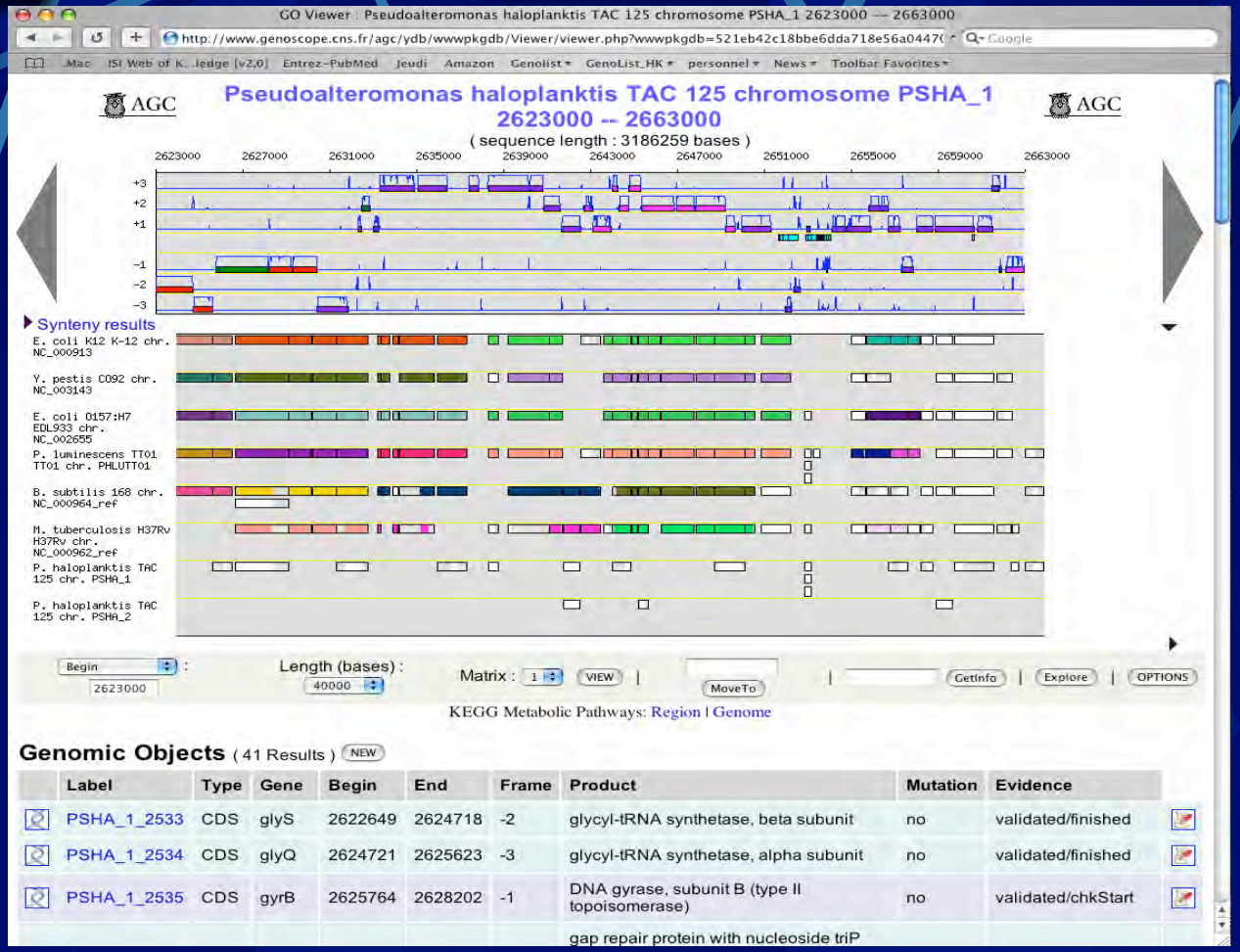
Exploring neighborhoods

To make discoveries we explore the general « neighborhoods » of genes of interest: proximity in the chromosome, in evolution, in the literature, in biochemical complexes, in metabolism etc.

Comparative genomics is essential, hence the launching of several parallel genome programs

adanchin@pasteur.fr

Gene vicinity: synteny



adanchin@pasteur.fr

What is Life?

Three processes are needed for Life:

- **Information transfer: groups of co-variant gene expression**

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

- Metabolism (Internal organisation)
- Compartmentalization (General structure)

adanchin@pasteur.fr

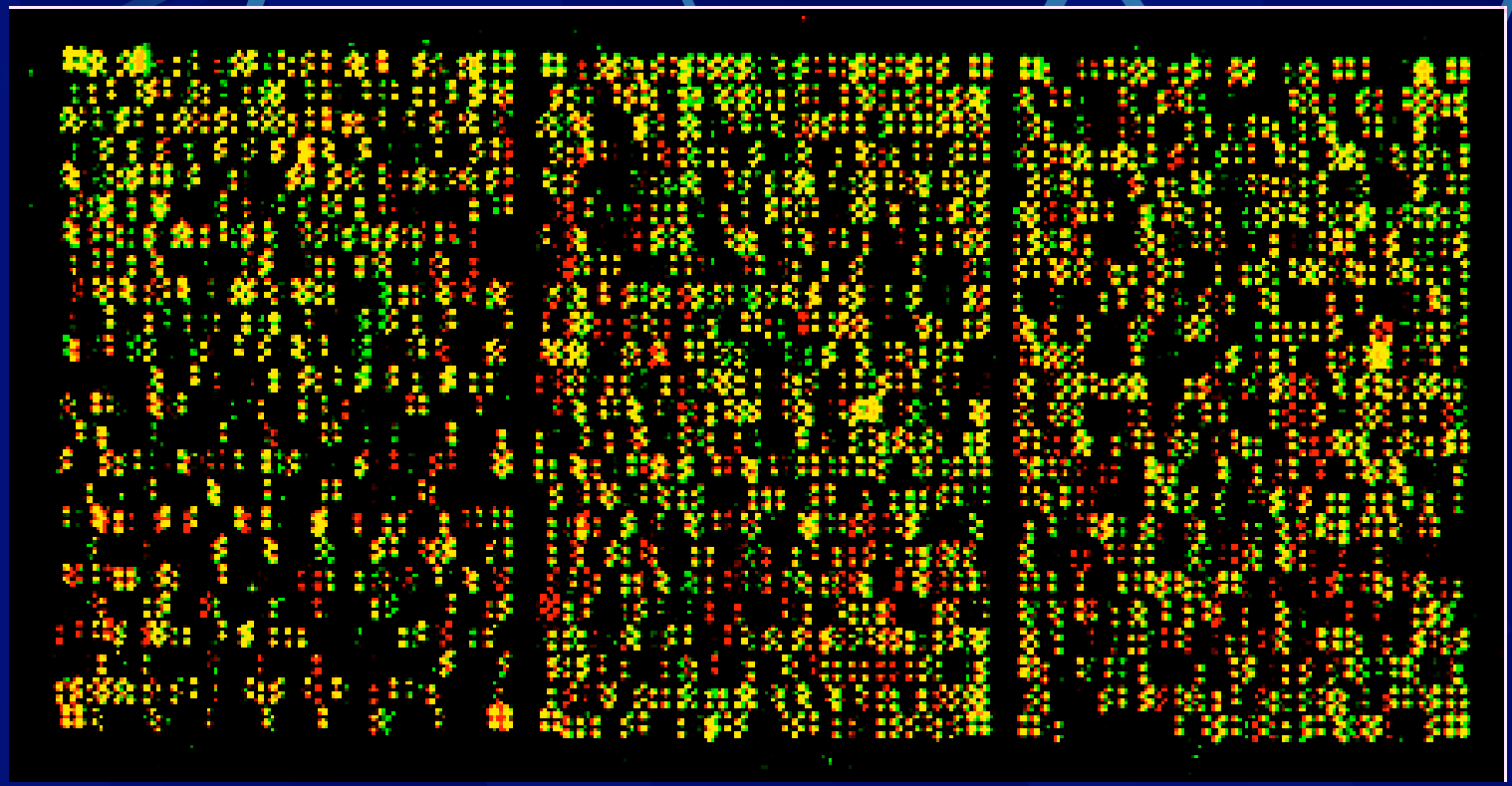
Co-variation of gene expression

Collecting data using large scale genomics techniques :

DNA chips and protein fingerprinting

adanchin@pasteur.fr

Expression neighborhood: all genes on a chip



- Green:

First condition

- Red:

Second condition

- Yellow:

no difference

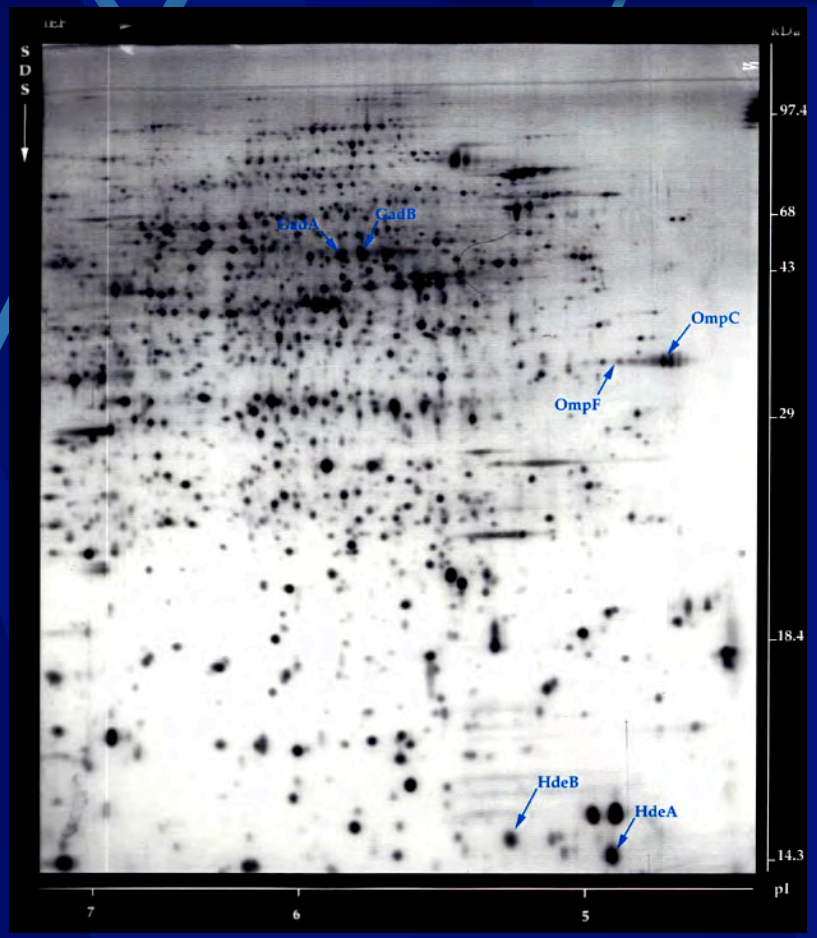
Two conditions are compared on the same chip

Protein fingerprinting

A technique named « two-dimensional gel electrophoresis » allows one to separate all the proteins of a cell, and to color them.

Once colored, proteins are sorted and identified by « mass spectrometry ».

Different cells or cells in different environments have a different pattern, exactly as individuals have different fingerprints.



adanchin@pasteur.fr

What is Life?

Three processes are needed for Life:

- **Information transfer**

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

- **Metabolism (Internal organisation)**
- **Compartmentalization (General structure)**

adanchin@pasteur.fr

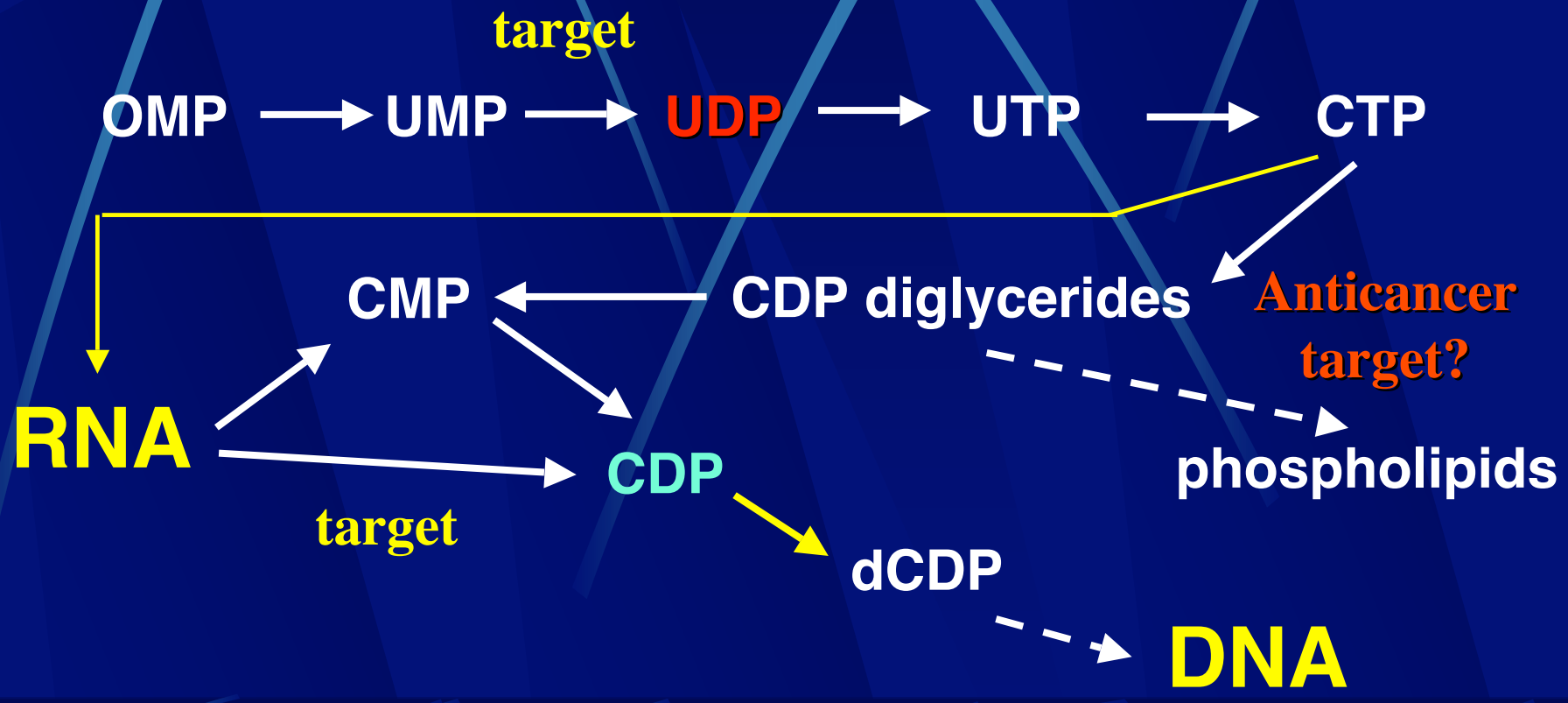
Metabolic neighborhoods: Chemical pathways

Often, drug targets are found in metabolic pathways. The idea is to mimick a normal molecule in the cell and to replace it by a similar one which kills the activity of some protein: this is exactly what poisons do! Antibiotics are poisons of a special kind, which act against microbes: to discover a new antibiotic is to discover a self-consistent pathway

adanchin@pasteur.fr

Metabolic neighborhood: unexpected selective constraints

DNA is made from NDP nucleotides, but CDP is absent:



adanchin@pasteur.fr

What is Life?

Three processes are needed for Life:

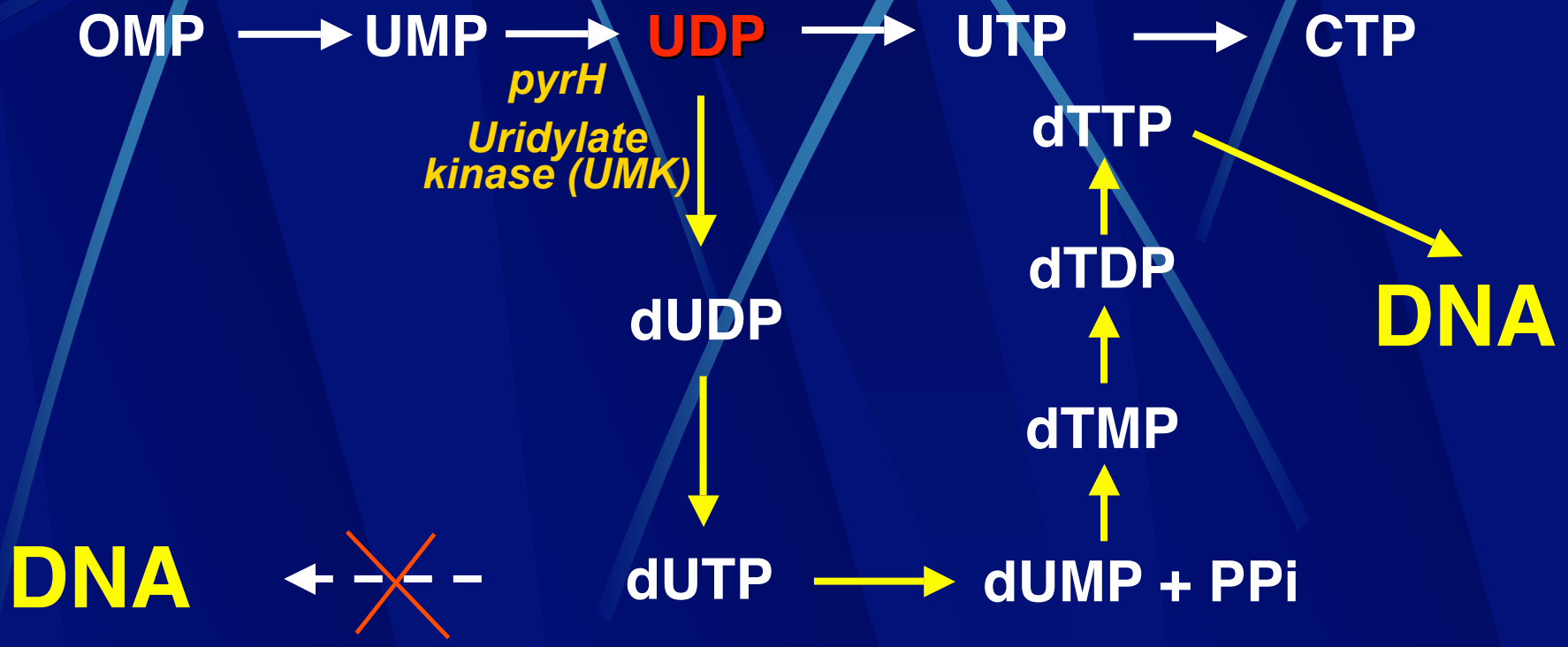
- **Information transfer**

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

- **Metabolism (Internal organisation)**
- **Compartmentalization (General structure)**

A dangerous intermediate

DNA is made from NDP nucleotides, but UDP must not get in:



adanchin@pasteur.fr

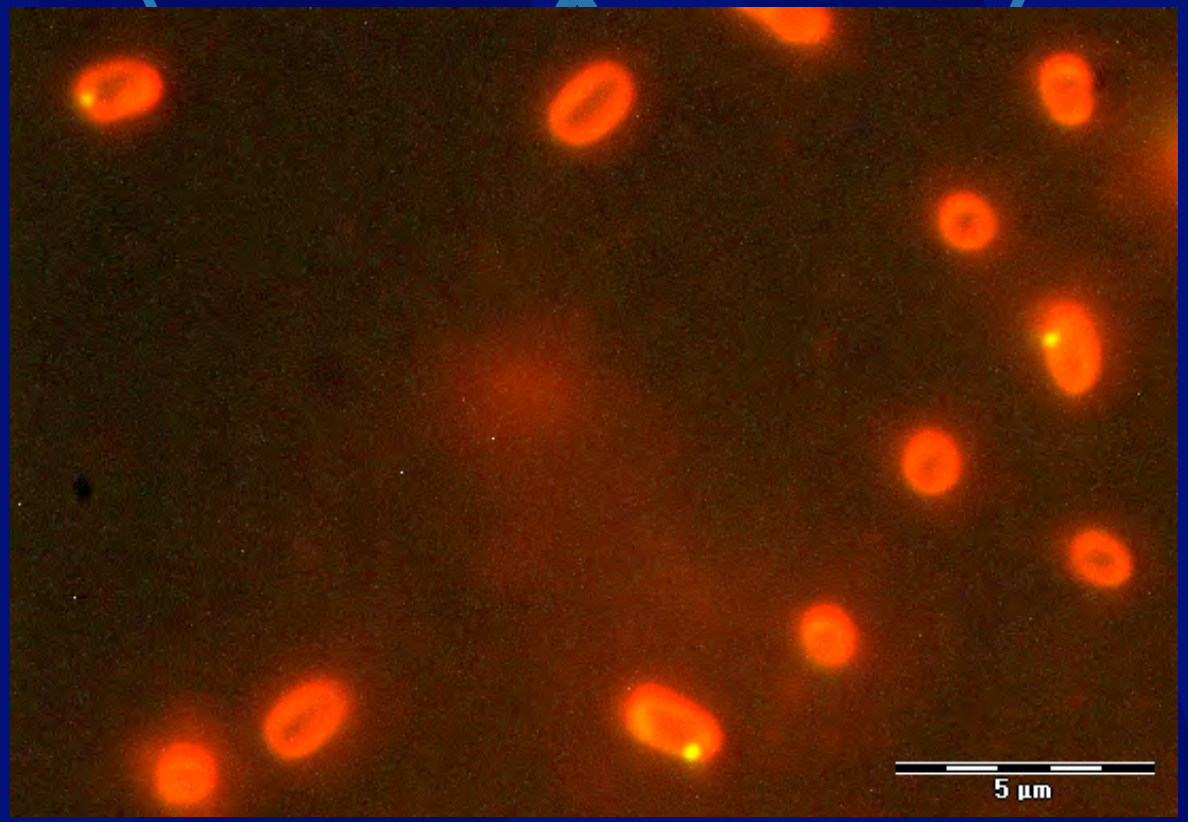
Uridylate kinase

- Essential enzyme
- Different origin in cells without a nucleus (bacteria) and in cells with a nucleus
- Conjecture: UDP must be compartmentalized to prevent U to enter DNA

adanchin@pasteur.fr

Cell compatmentalisation *in vivo*

The *pyrH* gene is fused with the reporter *gfp* gene and replaces its wild-type counterpart. One observes localisation of GFP under the membrane, and at foci….



adanchin@pasteur.fr

Conclusion

The genome organisation is much more rigid than usually assumed.

Some regions (such as the terminus) are rather unstable, but most of the genome structure is preserved through evolution

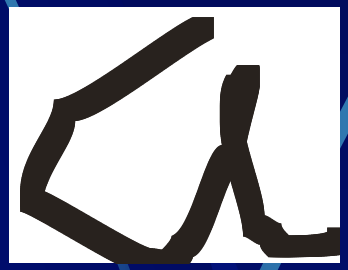
adanchin@pasteur.fr

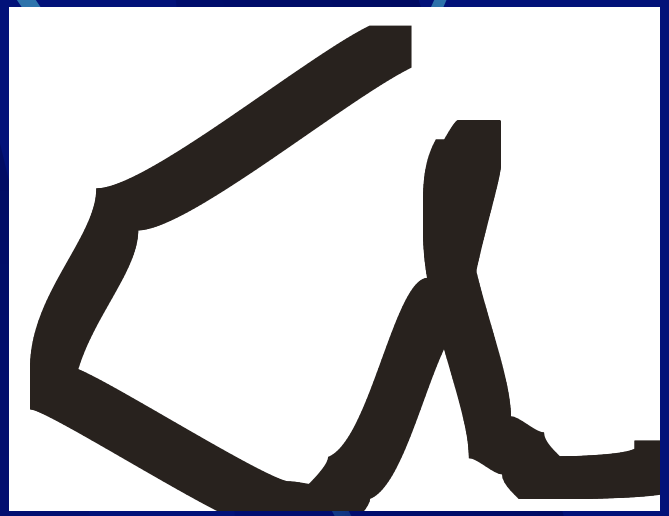
Why a computer? The power of algorithms

- The analogy between the cell is a computer goes beyong the separation between the program and the machine
- The structure of the program itself is subject to architectural constraints, that must derive from some sort of selection pressure, building up an image of the cell in the program
- Computing allows one to reinvestigate the idea of preformism: the cell is not preformed, but its algorithm of construction is

adanchin@pasteur.fr

Gestalt and Algorithms





« Start from the top, middle

Go downfrom right to left

Accelerate

Turn right

etc.

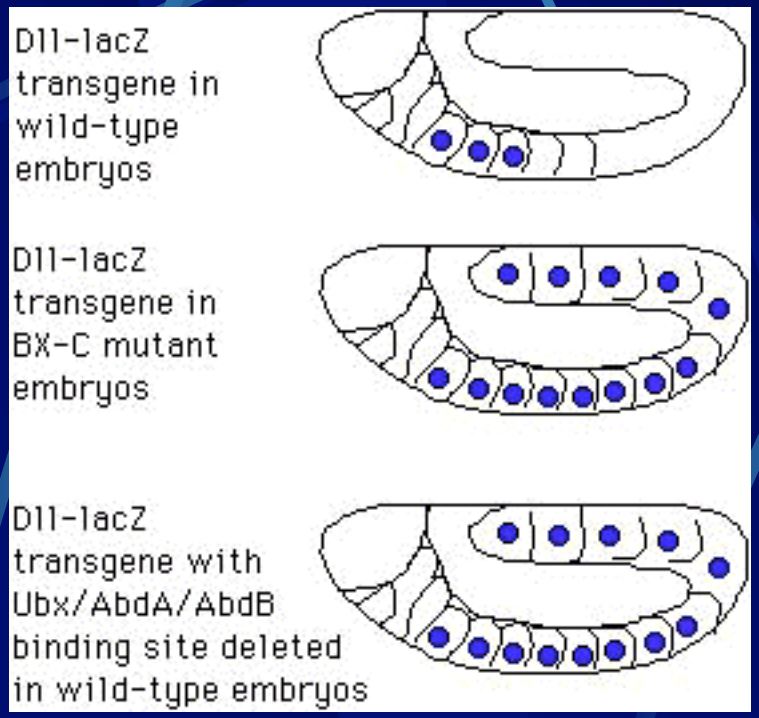
From the cell to the animal?

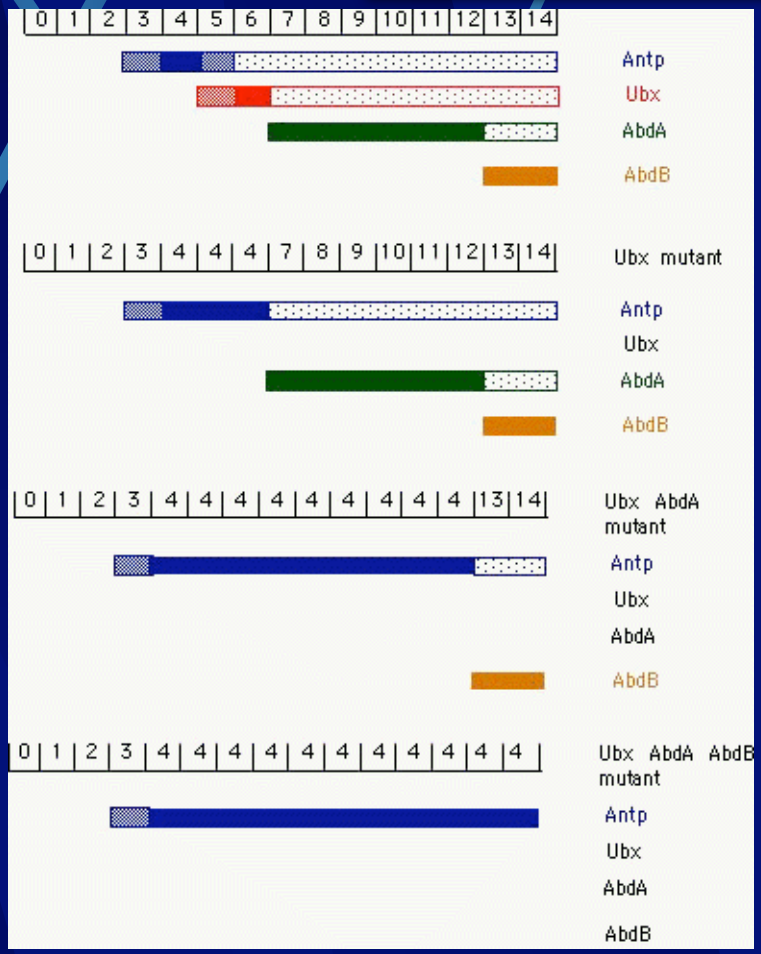
The analogy between the cell is a computer goes beyong the separation between the program and the machine

Can this analogy be extended to organisms as whole entities?

Homeogene control the development of animals (and plants): the order of these genes along chromosomes follow the plan of the body of the animal

Drosophiloculus, Homunculus ?



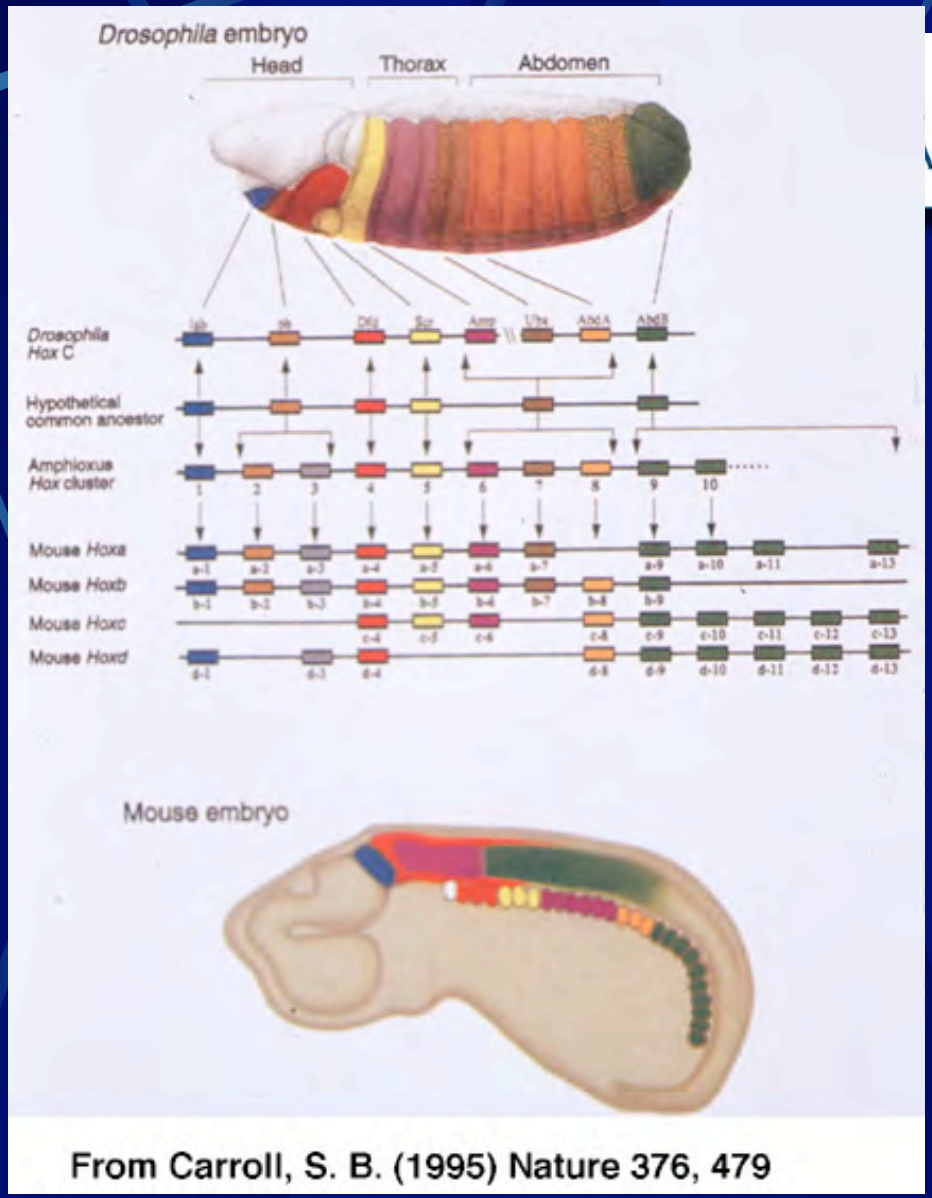


adanchin@pasteur.fr

香港大學－巴斯德研究中心
HKU-Pasteur Research Centre

Drosophiloculus,

Homunculus ?



From Carroll, S. B. (1995) Nature 376, 479

adanchin@pasteur.fr

From the genetic program to a research programme?

Our working hypothesis was that the cell would behave as a computer. This conjectures that an architectural program exists in the chromosome. This may be wrong, but science goes to India… and may find America !

A. Danchin The Delphic Boat. What genomes tell us (2003) Harvard University Press

« Bombardment of th Embassy in Belgrade »