

Techniques et Modélisation en Santé et en Nutrition

Jean-Michel Batto
jmbatto@jouy.inra.fr

Cours de Biologie et Bioinformatique pour Informaticiens – *Techniques et Modélisation en Santé et en Nutrition*

- Un point de départ : La science du vivant :

- Molière (*Le Malade Imaginaire*, 1673) :

- BACHELIERUS :

Mihi à docto Doctore

*Domandatur causam et rationem, quare
Opium facti dormire ?*

A quoy respondeo,

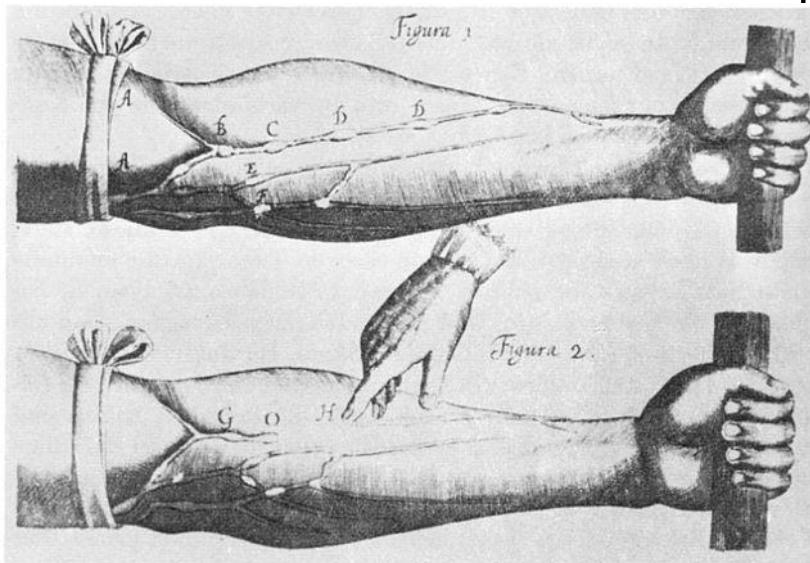
Quia est in eo

Virtus dormitiva,

Cujus est natura

Sensus assoupire.

- La circulation du sang (W.Harvey, 1628)

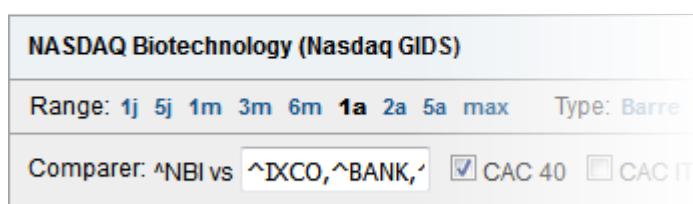


Pourquoi suivre le module TMSN ?

NASDAQ Biotechnology (^NBI) - Nasdaq GIDS

3 736,56 +38,96(1,05%) 15 sept. 23:15[Ajouter au portefeuille](#)

Graphique de base



NASDAQ Biotechnology

■ ^NBI ■ ^IXCO ■ ^BANK ■ DOW ■ CAC40



Pourquoi suivre le module TMB ?

NASDAQ Biotechnology (Nasdaq GIDS)

Range: 1j 5j 1m 3m 6m 1a 2a 5a max Type

Comparer: ^NBI vs ^IXCO,^BANK, CAC 40

Sur 5 ans écoulés, l'indice "biotech" a pris 300 %
en valeur contre 100 % pour l'informatique !

NASDAQ Biotechnology

■ ^NBI ■ ^IXCO ■ ^BANK ■ DOW ■ CAC40

15 sept, 2015

+400%

<https://fr.finance.yahoo.com>

+300%

+200%

+100%

0%

-100%

© Yahoo!

janv-11

janv-12

janv-13

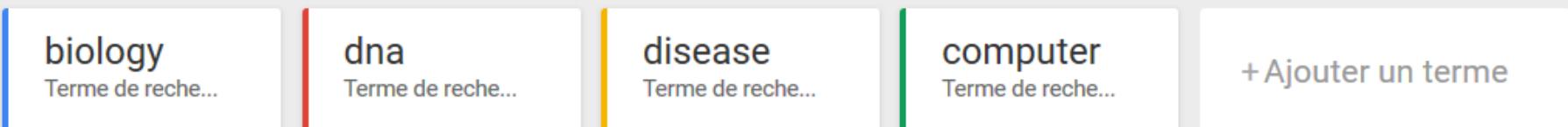
janv-14

janv-15

L'intérêt pour la biologie en 2015

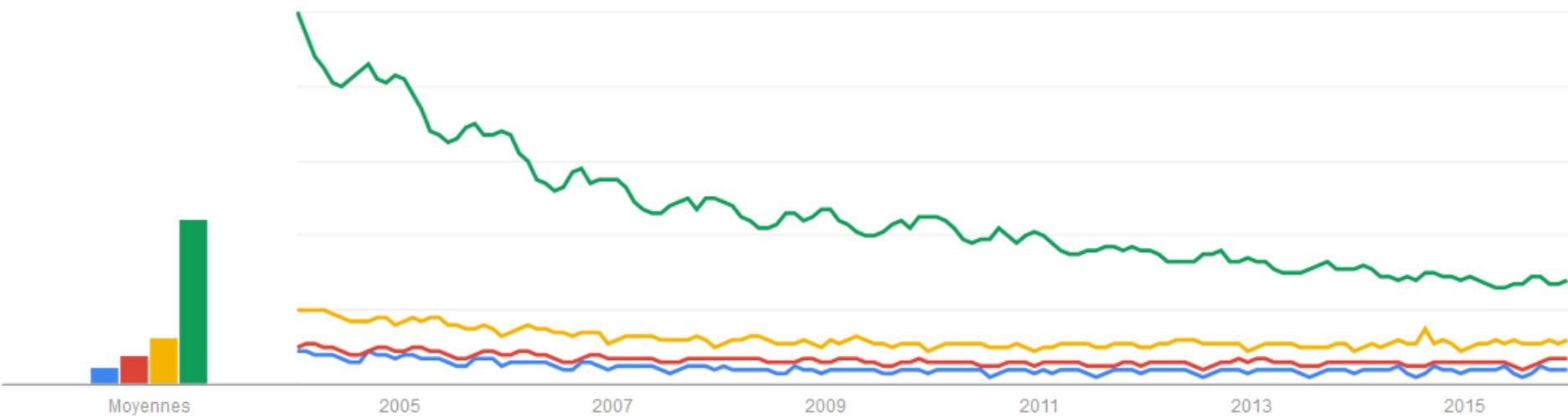
Comparer [Termes de recherche ▾](#)

<https://www.google.fr/trends/>



Évolution de l'intérêt pour cette recherche [?](#)

Titres des actualités Prévision



L'histoire du microbe par son impact sanitaire en Europe

- 1348 : La Peste : *Yersinia pestis* (victimes 100 millions?)
- 1849 : Le Cholera : *Vibrio cholerae* (victimes 40%?)
- 1918 : la grippe espagnole, H1N1 (victimes 30 millions? 20%)
- Robert Koch (1843-1910, anthrax, tuberculose, cholera)
- Louis Pasteur (1822-1895, germe, ferment, vaccin)

Le microbe et l'aliment

- Conservation des aliments
- Un exemple le Yaourt
- Un streptocoque et un lactobacille
- Biotechnologie

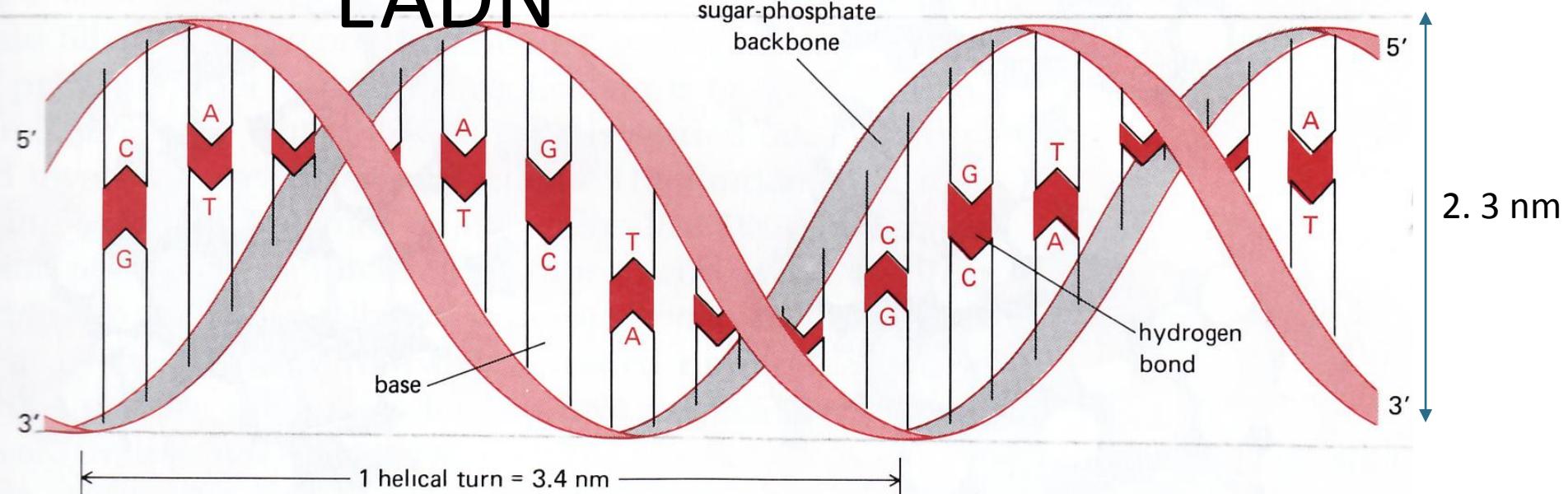
1 gramme de yaourt = 10^9 bactéries vivantes



L'ADN

- 1869 : ADN constituant « nucleus », Frederick Miescher
- 1953 : publication dans Nature de J.Watson et F.Crick
- 1966 : élucidation du code génétique
- 1 cellule humaine : 3×10^9 lettres, $20 \mu\text{m}$
(densité : 1500 To/cm^2 contre 2.5 To/cm^2 pour memristor ram)

L'ADN



ADN : 4 lettres

Adénine

Thymine

Cytosine

Guanine

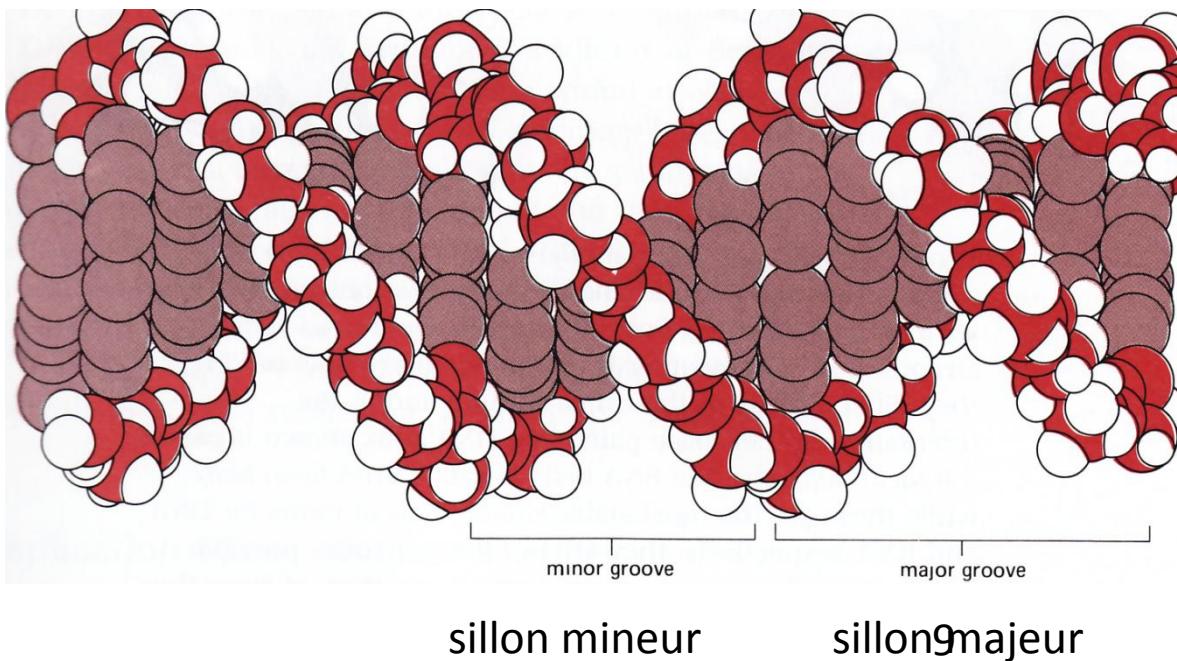
ARN : 4 lettres

Adénine

Uracile

Cytosine

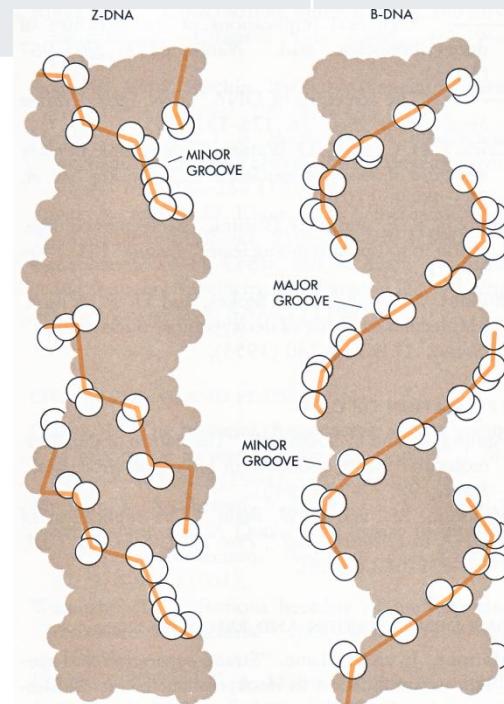
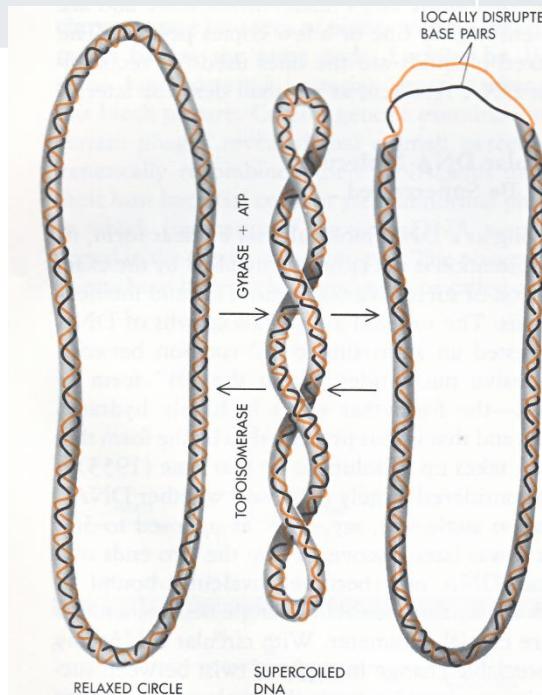
Guanine



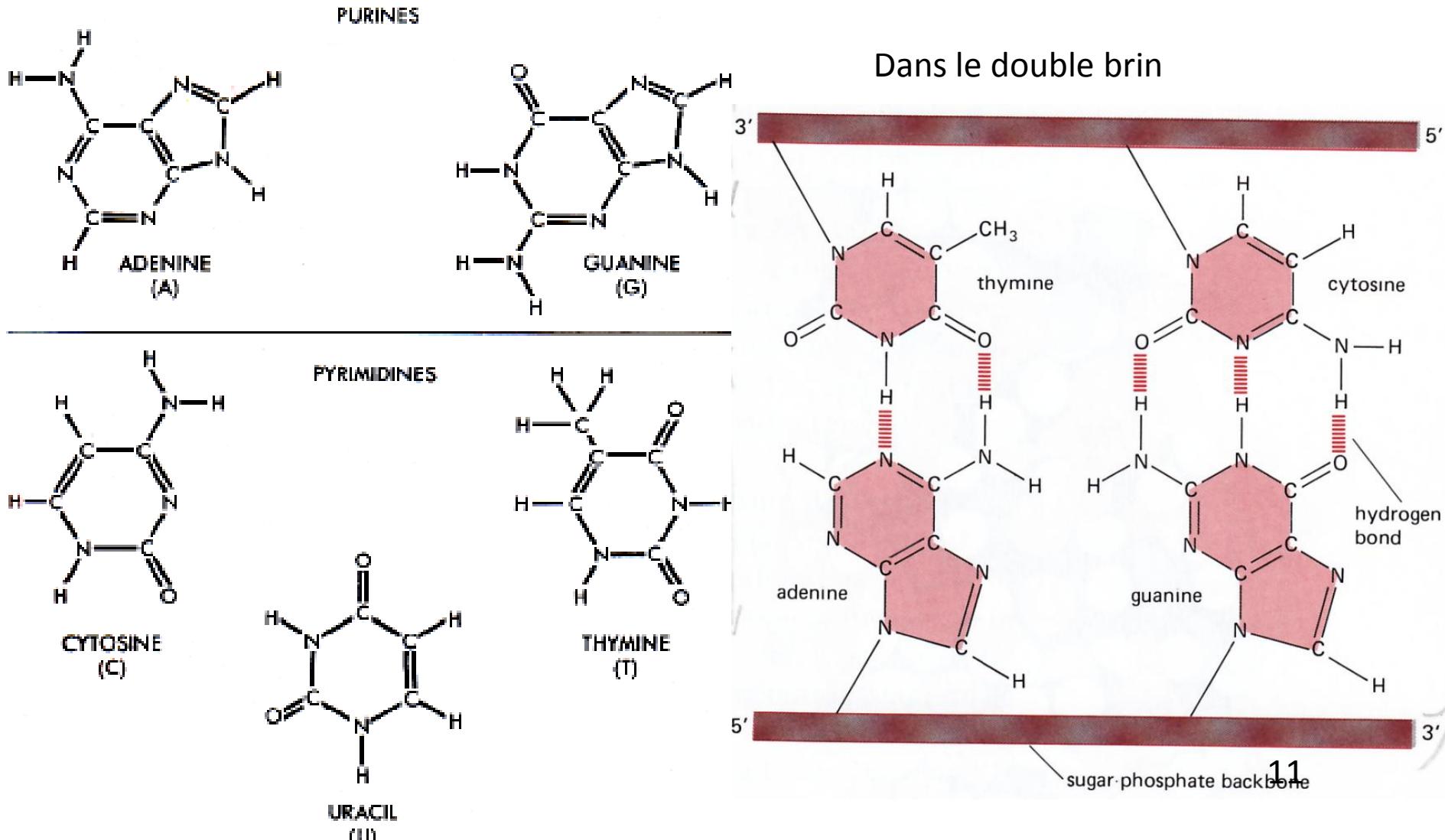
Extrait de 'molecular biology of the cell', B.Alberts et al

L'ADN est flexible

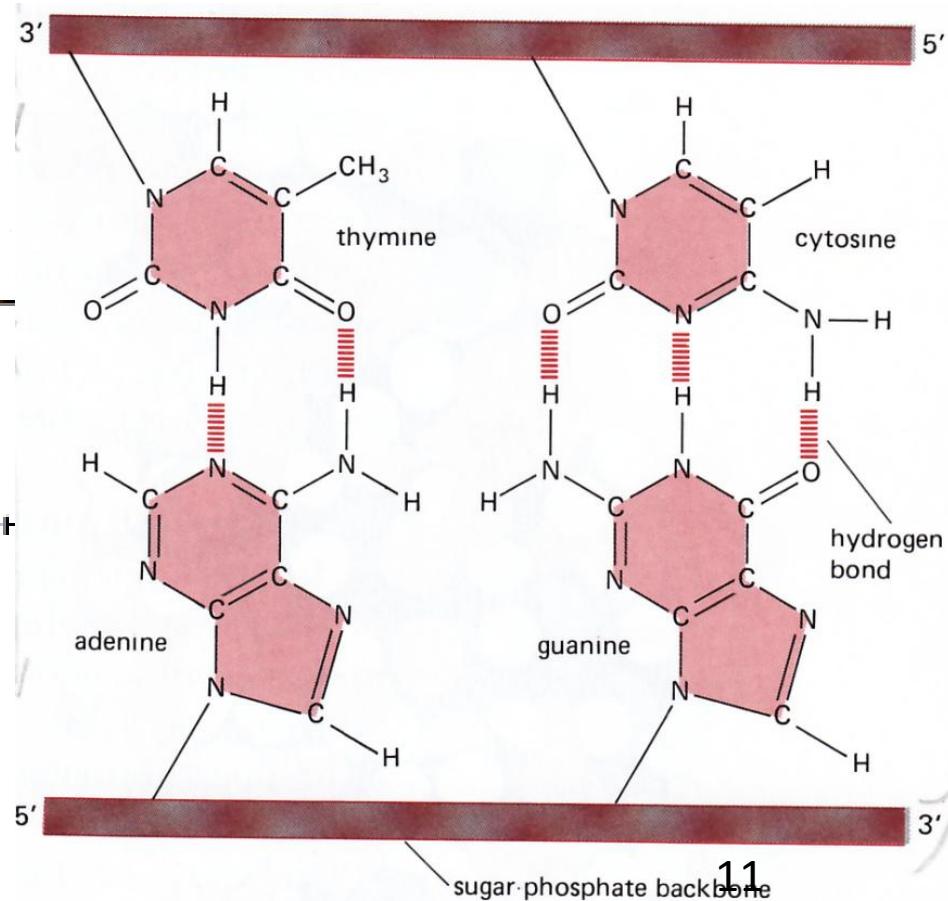
3 formes de l'ADN double brins	Forme A (déshydraté, RNA 2brins, à droite)	Forme B (standard, à droite)	Forme Z (super enroulée, à gauche)
Nbre bases/tour d'hélice	11 bases	10 bases	12 bases
Diamètre de l'hélice	3.55nm	2.37nm	1.84nm
Position de l'axe de rotation	sillon majeur	À travers les paires de base	sillon mineur
Elévation par tour d'hélice	2.46nm	3.32nm	4.56nm



L'ADN : sa chimie



Dans le double brin

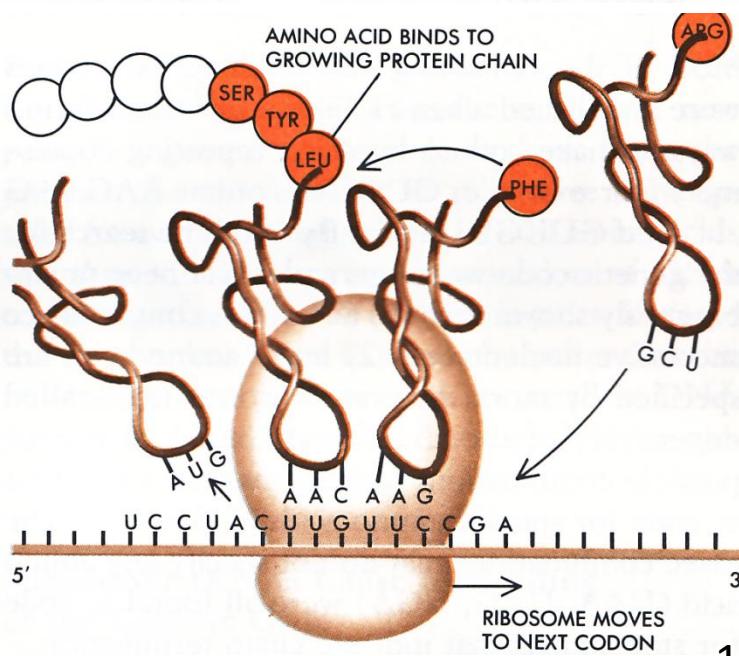
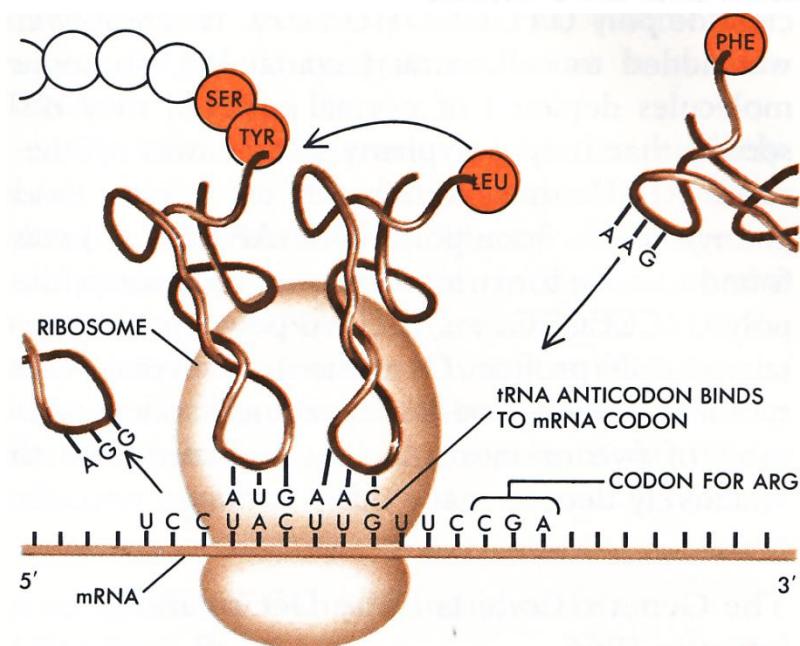
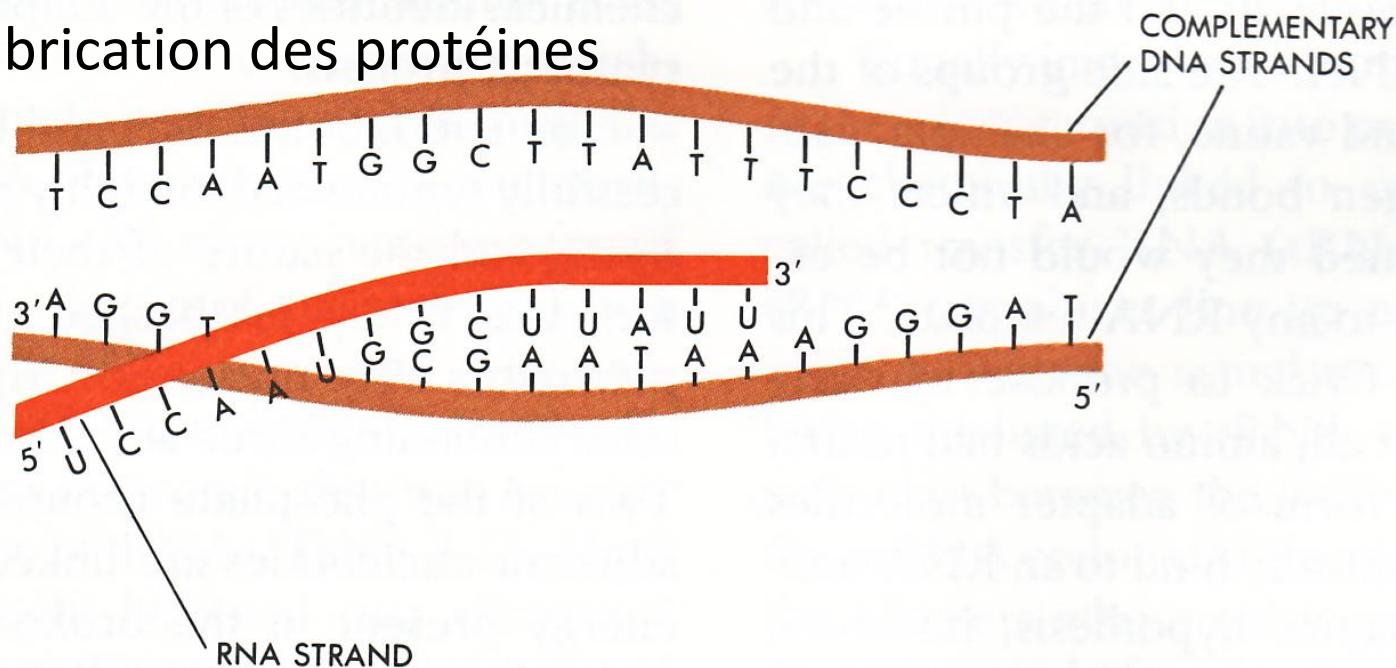
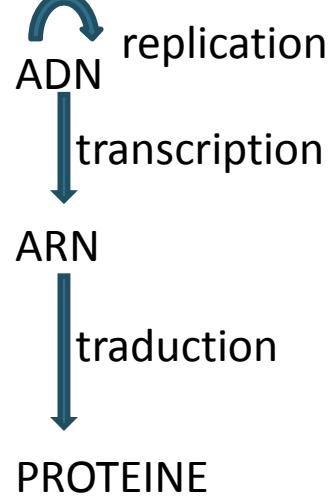


Le code génétique

20 AA
en traduction
directe

		2ème position					
		U	C	A	G		
1ère position	U	PHE	SER	TYR	CYS	U	
	U	PHE	SER	TYR	CYS	C	
	U	LEU	SER	Stop	Stop	A	
	U	LEU	SER	Stop	TRP	G	
1ère position	C	LEU	PRO	HIS	ARG	U	
	C	LEU	PRO	HIS	ARG	C	
	C	LEU	PRO	GLN	ARG	A	
	C	LEU	PRO	GLN	ARG	G	
1ère position	A	ILE	THR	ASN	SER	U	
	A	ILE	THR	ASN	SER	C	
	A	ILE	THR	LYS	ARG	A	
	A	MET	THR	LYS	ARG	G	
1ère position	G	VAL	ALA	ASP	GLY	U	
	G	VAL	ALA	ASP	GLY	C	
	G	VAL	ALA	GLU	GLY	A	12
	G	VAL	ALA	GLU	GLY	G	

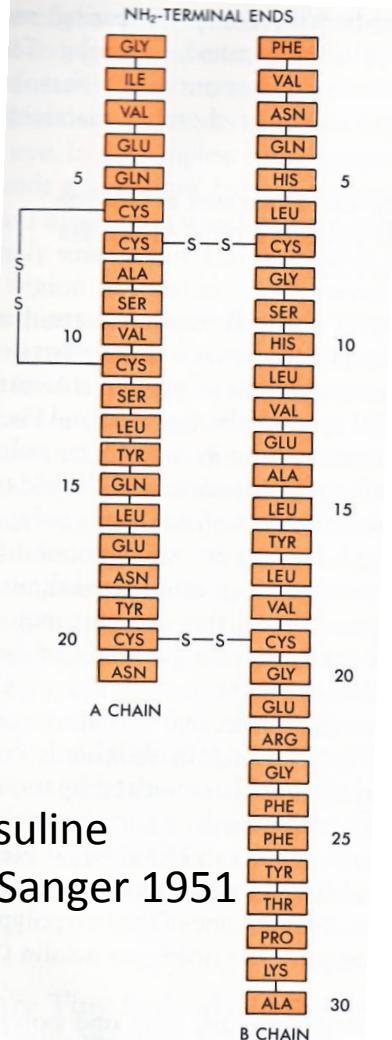
La fabrication des protéines



mRNA : 5% , tRNA 15%, rRNA 80% (ribosome)

Les briques de bases

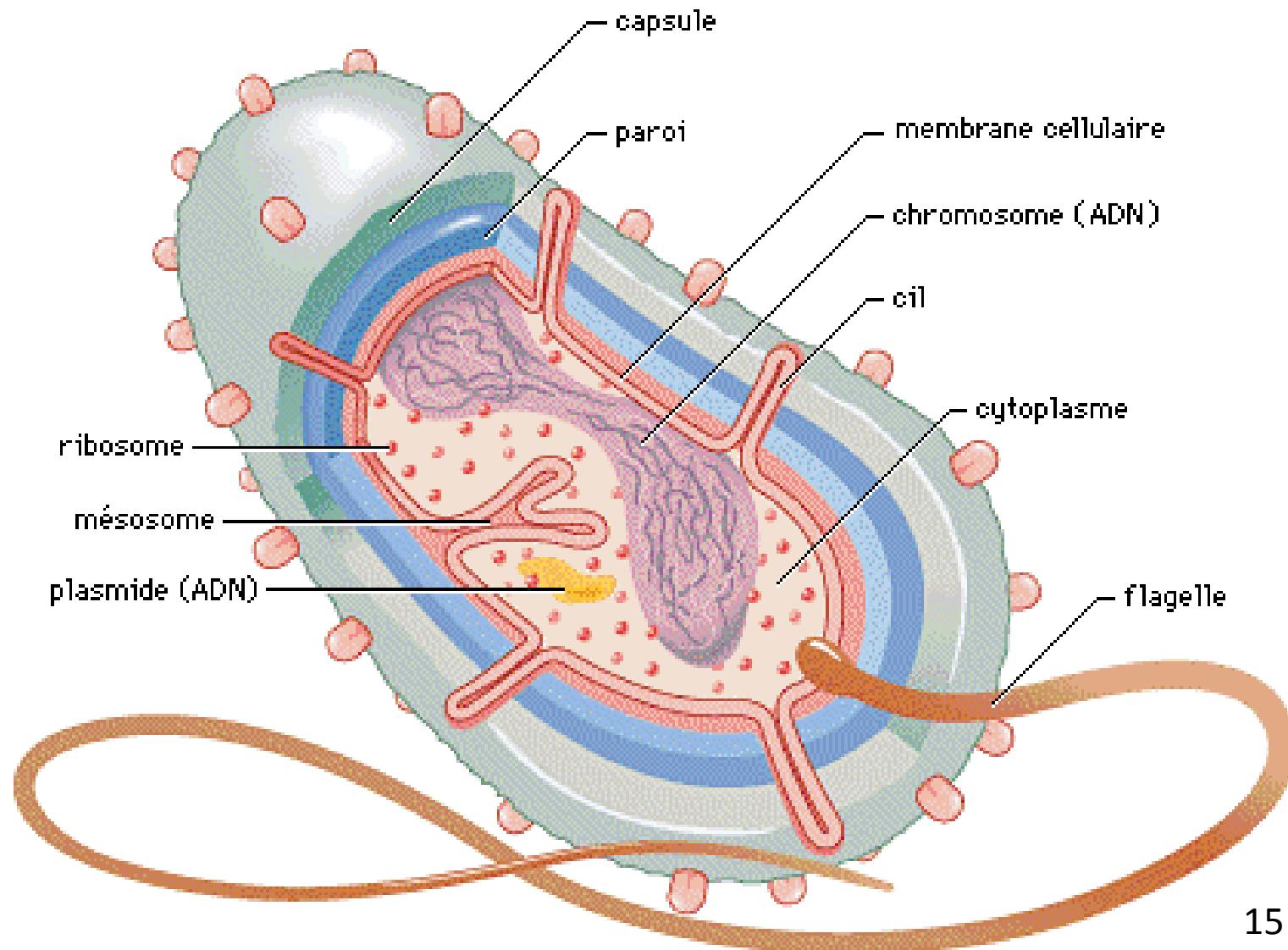
Universalité du code est fonction des tRNA disponibles



Insuline
F. Sanger 1951

Nom AA	sigle	Code IUPAC	Codon
<u>Alanine</u>	<u>Ala</u>	<u>A</u>	GCU, GCC, GCA, GCG
<u>Arginine</u>	<u>Arg</u>	<u>R</u>	CGU, CGC, CGA, CGG, AGA, AGG
<u>Asparagine</u>	<u>Asn</u>	<u>N</u>	AAU, AAC
<u>Acide aspartique</u>	<u>Asp</u>	<u>D</u>	GAU, GAC
<u>Cystéine</u>	<u>Cys</u>	<u>C</u>	UGU, UGC
<u>Glutamine</u>	<u>Gln</u>	<u>Q</u>	CAA, CAG
<u>Acide glutamique</u>	<u>Glu</u>	<u>E</u>	GAA, GAG
<u>Glycine</u>	<u>Gly</u>	<u>G</u>	GGU, GGC, GGA, GGG
<u>Histidine</u>	<u>His</u>	<u>H</u>	CAU, CAC
<u>Isoleucine</u>	<u>Ile</u>	<u>I</u>	AUU, AUC, AUA
<u>Leucine</u>	<u>Leu</u>	<u>L</u>	UUA, UUG, CUU, CUC, CUA, CUG
<u>Lysine</u>	<u>Lys</u>	<u>K</u>	AAA, AAG
<u>Méthionine</u>	<u>Met</u>	<u>M</u>	AUG
<u>Phénylalanine</u>	<u>Phe</u>	<u>F</u>	UUU, UUC
<u>Proline</u>	<u>Pro</u>	<u>P</u>	CCU, CCC, CCA, CCG
<u>Sérine</u>	<u>Ser</u>	<u>S</u>	UCU, UCC, UCA, UCG, AGU, AGC
<u>Thréonine</u>	<u>Thr</u>	<u>T</u>	ACU, ACC, ACA, ACG
<u>Tryptophane</u>	<u>Trp</u>	<u>W</u>	UGG (UGA)
<u>Tyrosine</u>	<u>Tyr</u>	<u>Y</u>	UAU, UAC
<u>Valine</u>	<u>Val</u>	<u>V</u>	GUU, GUA, GUG

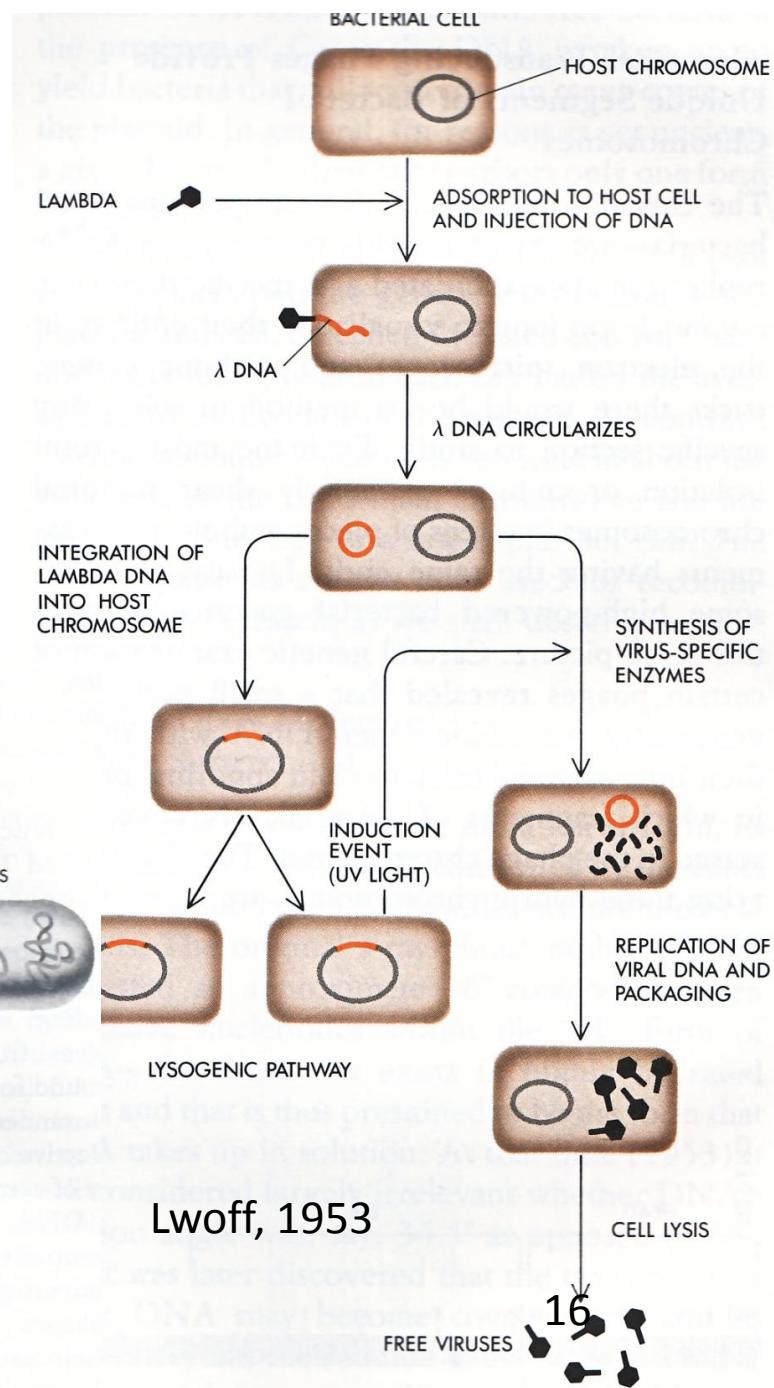
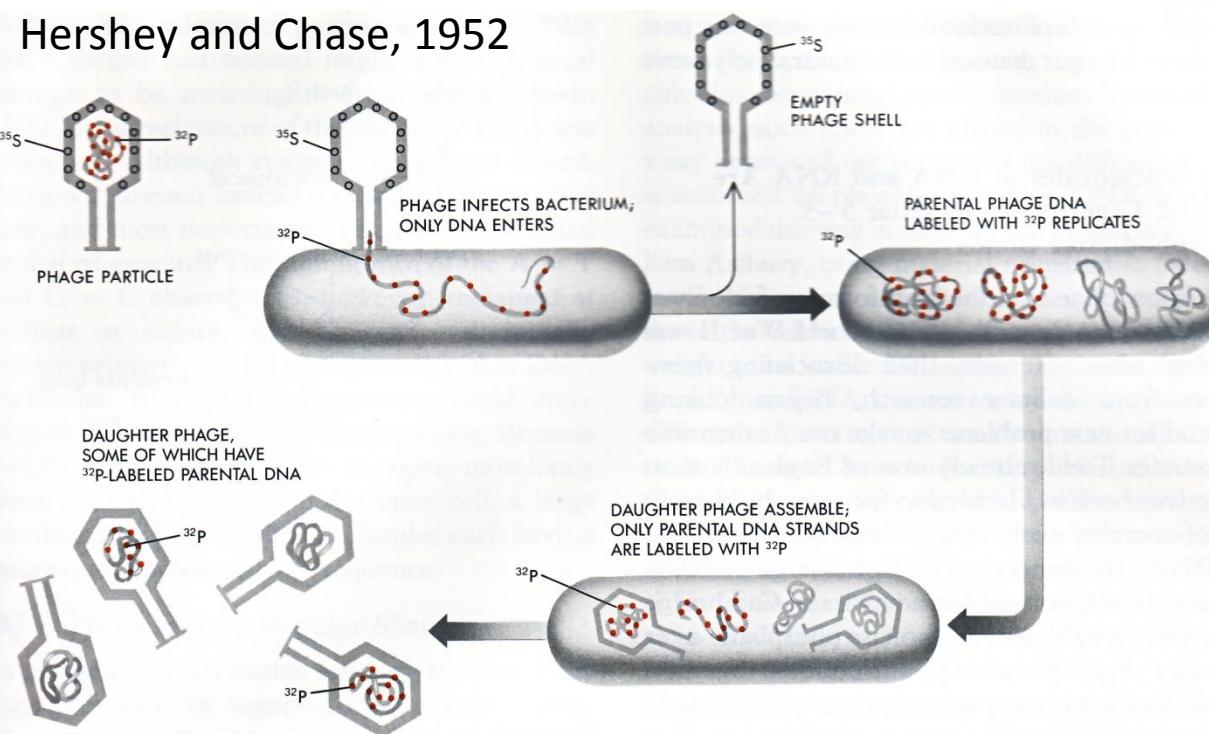
Schéma type d'une bactérie



bactérie, virus

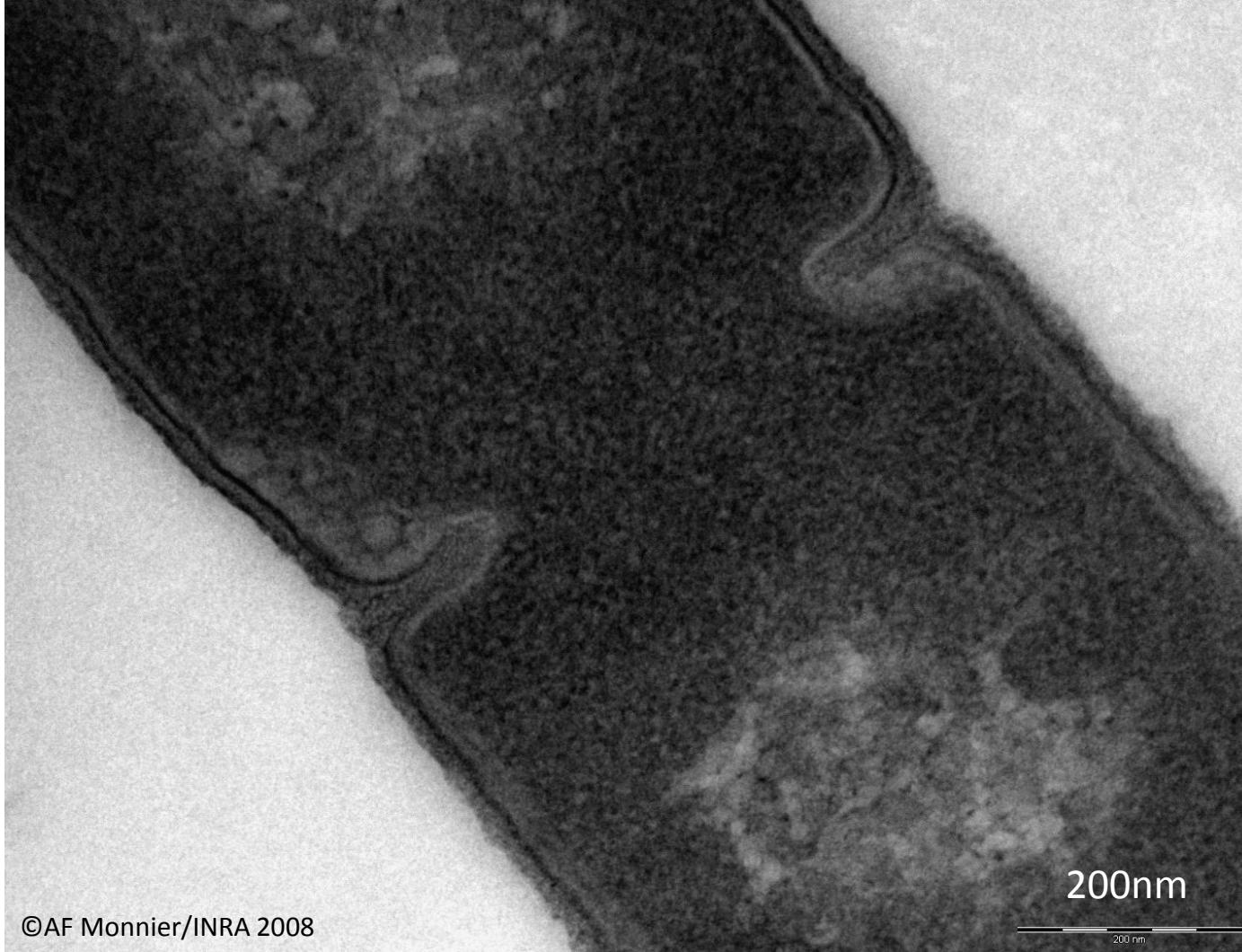
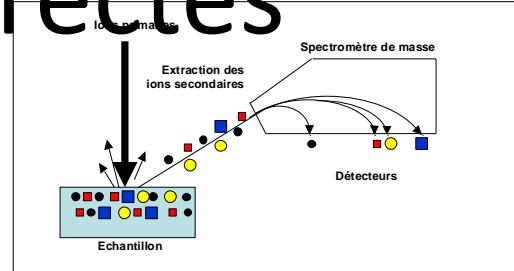
Le phage, un virus à bactérie

Hershey and Chase, 1952



ADN : les observations directes

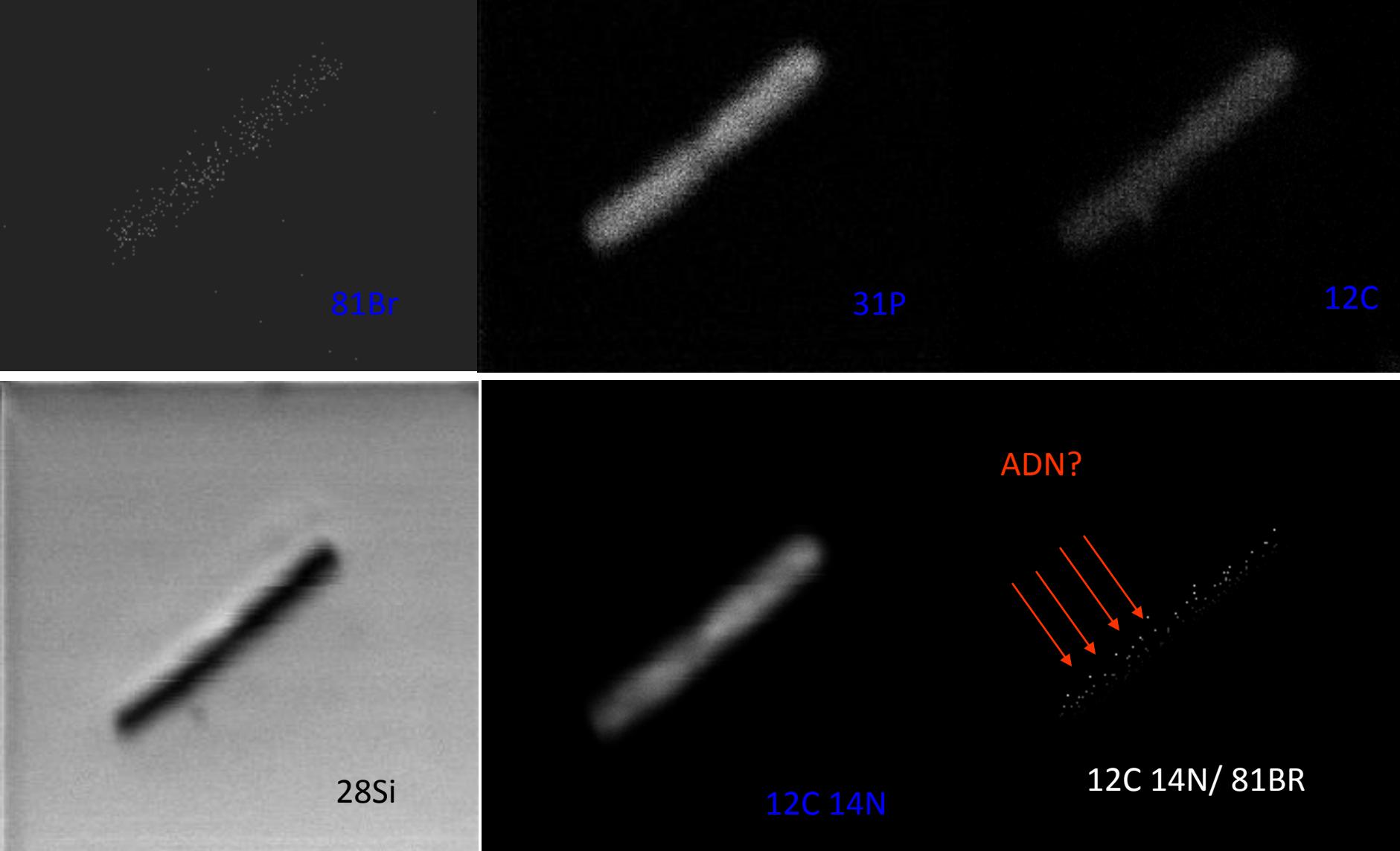
- Microscopie SIMS (Secondary Ion Mass Spectroscopie)
- Microscopie électronique (MET transmission)



200nm

17

Microscopie SIMS 1/2

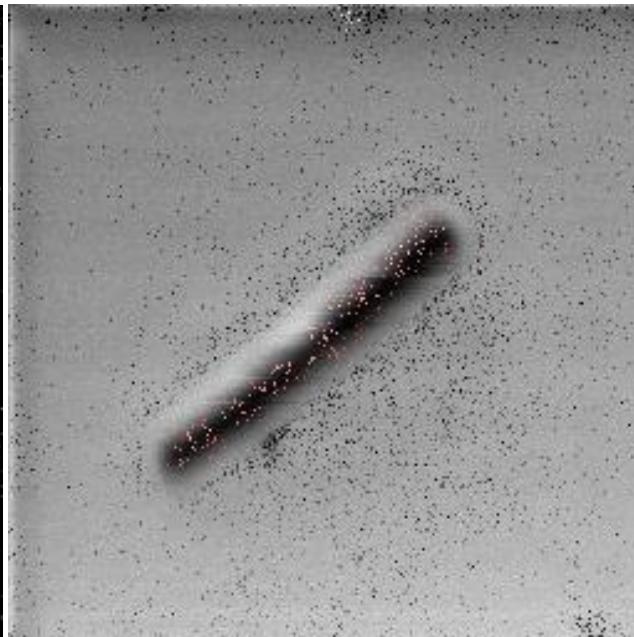


Images SIMS du Cameca Rouen, *b. subtilis* phase expo (dans le champs : 2 bactéries)
Lecture SIMS sur 5 ions : Br, P, C, Si, CN

Microscopie SIMS 2/2

31P/81Br=>dinucleotides
, trinucléotides?

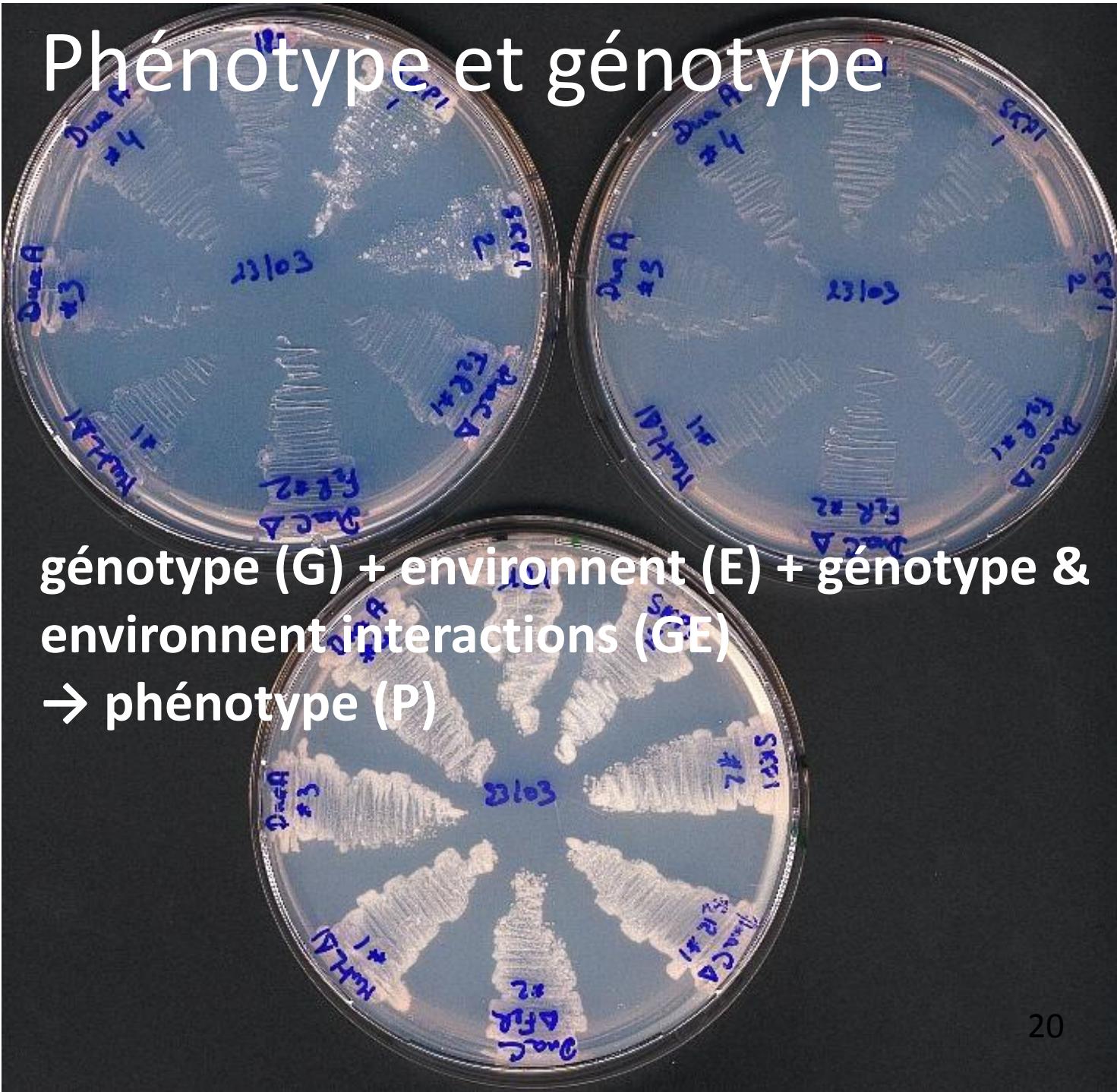
12C/12C14N => montre les sucres=>EPS?



28Si : montre le support
19

Phénotype et génotype

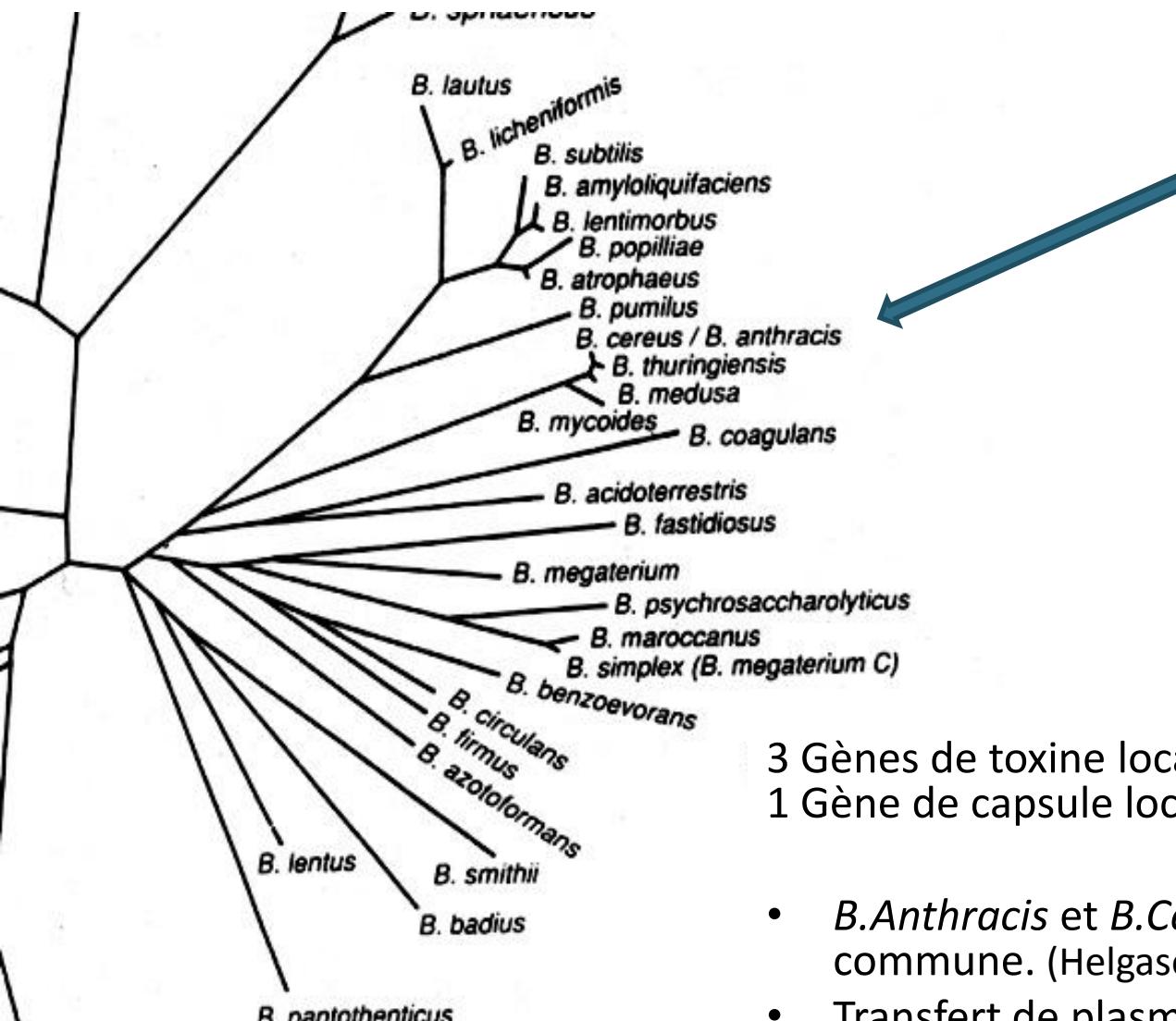
génotype (G) + environnement (E) + génotype & environnement interactions (GE)
→ phénotype (P)



Classification du vivant

- 3 domaines, 6 règnes :
 - Procaryotes :
 - Bactérie
 - Archae :
 - Archae
 - Eucaryotes :
 - Protiste (océan)
 - Plante
 - Champignon
 - Animal

Classification des bactéries : anthrax

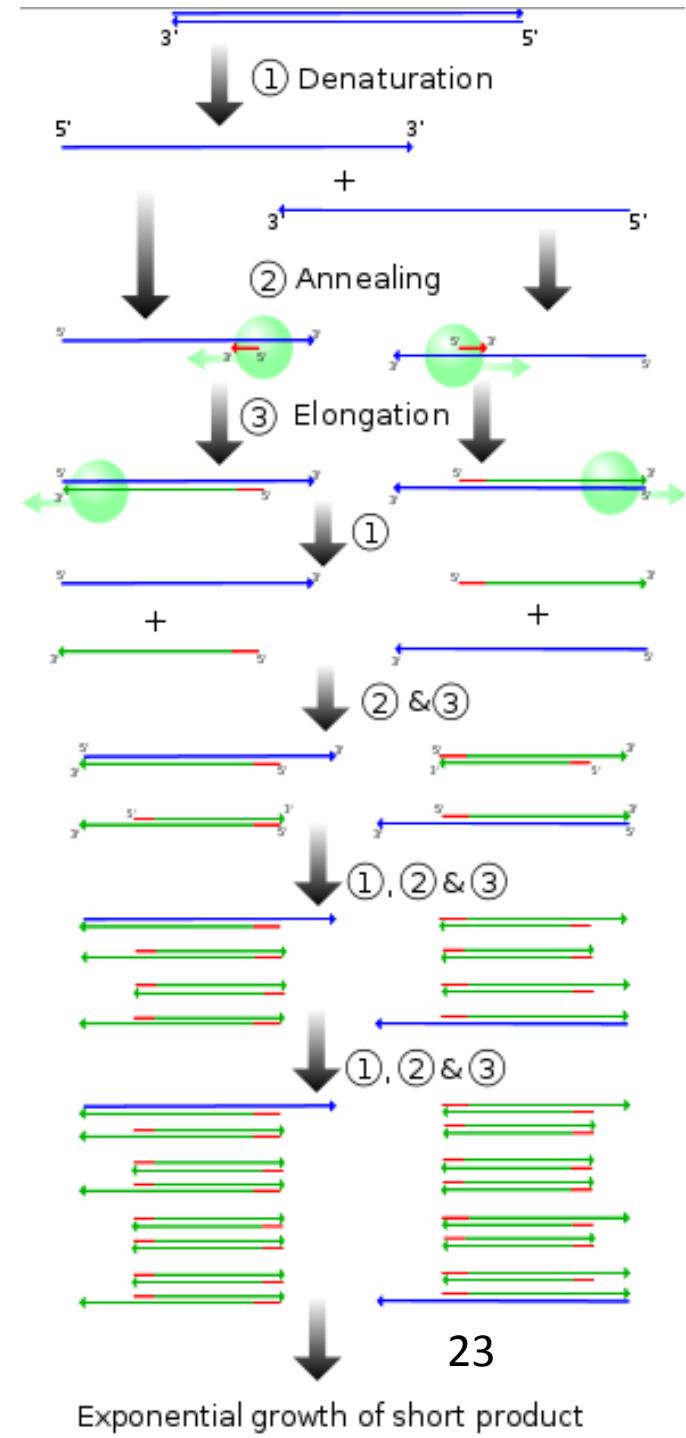
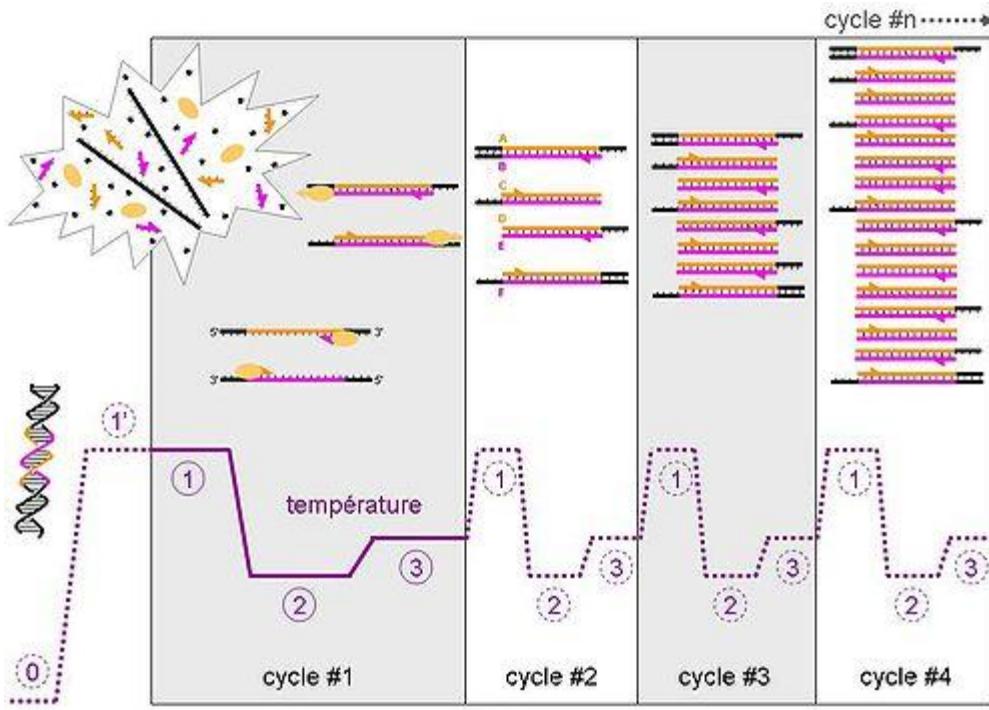


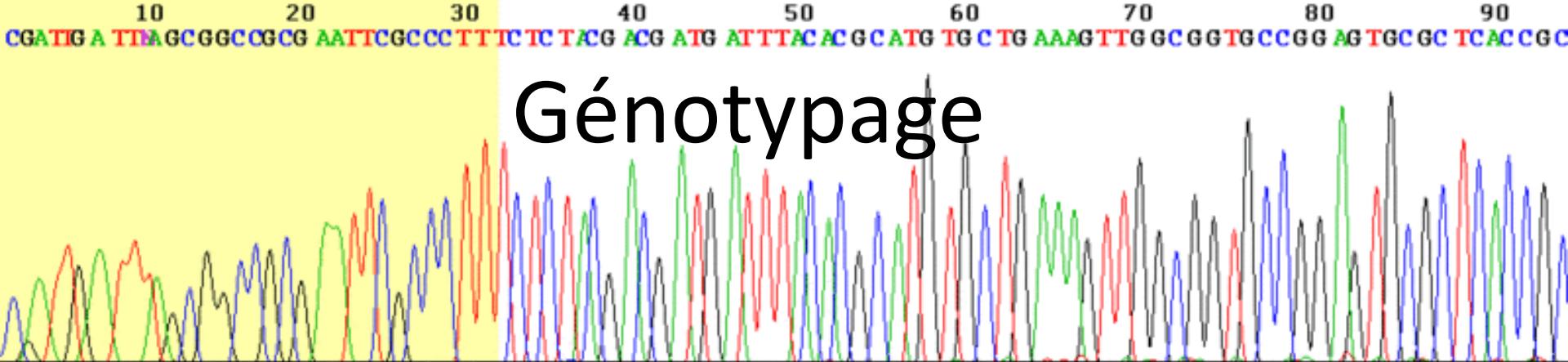
3 Gènes de toxine localisés sur PXO1 (182Kb)
1 Gène de capsule localisé sur PXO2 (96Kb)

- *B.Anthracis* et *B.Cereus* ont une origine commune. (Helgason et al, 2000)
- Transfert de plasmides entre 2 souches. (Hoffmaster et al, 2004)

PCR

- Polymerase Chain Reaction (1983)
 - réaction fondatrice de l'analyse industrielle de l'ADN
 - basée sur une enzyme haute T_p°C
 - 20 cycles multiplie par 10⁶ l'ADN





Génotypage

- Les technologies par date :
 - Chromatographie 2D : 1970
 - Sanger-ABI : 1987
 - Sanger-Capillaire eletrophorese : 2000
 - Pyroséquençage : 2005
- Les structures de données par date :
 - séquence d'un gène : 1970
 - séquence d'un plasmide : 1976 (phage 1977)
 - séquence d'un génome bactérien : 1995
 - séquence d'un génome humain : 2001
 - description d'un riboswitch (ARN) : 2005

La diversité des structures de données

- Les ordres de grandeurs
 - Gène->1..n protéines : 100 - 10000 lettres
 - ARN messager : ~100 - 10000 lettres
 - ARN antisens : ~100 lettres (=100b)
 - Opéron : ~ 3 gènes
 - Plasmide/Virus/Phage : ~100 gènes (kb)
 - Un génome bactérien : ~2000 gènes (2Mb)
 - Un génome humain : ~20000 gènes (3.2Gb)

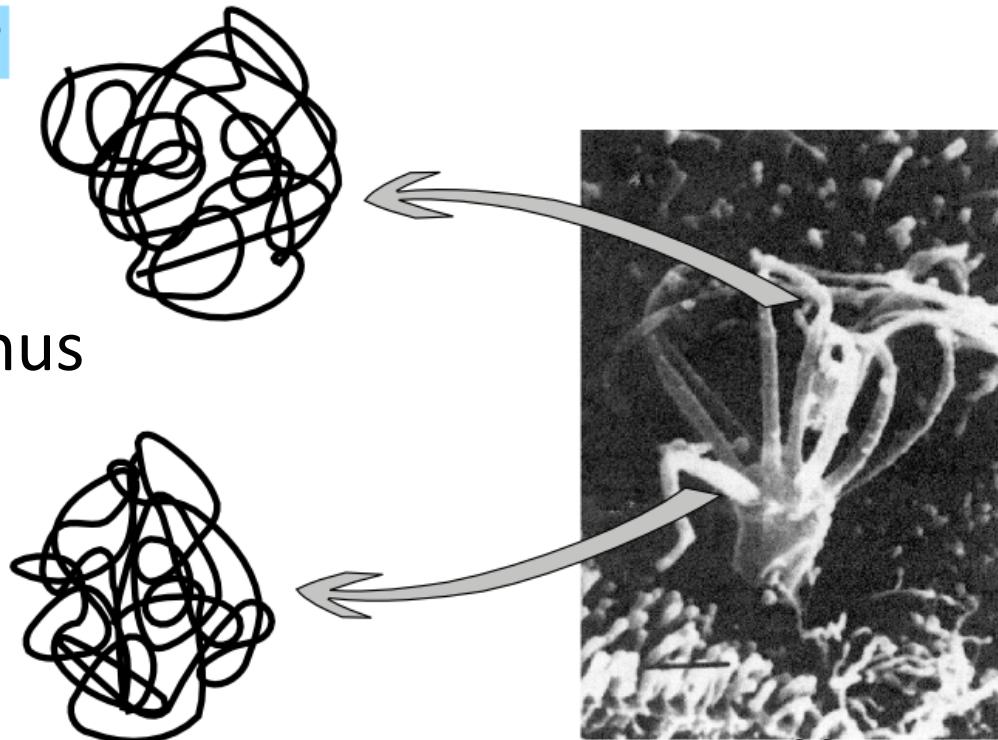
Que faire avec les données brutes ?

- Un génome est une collection d'informations
 - Des données brutes associées à des fonctions

>CfOR194 | Canis Olfactory Receptor
PMYKVILTPIMAYDRYL
KLMNIPLMNPLMSATT
TLMWNIPLMN...

- Des fonctions inconnus

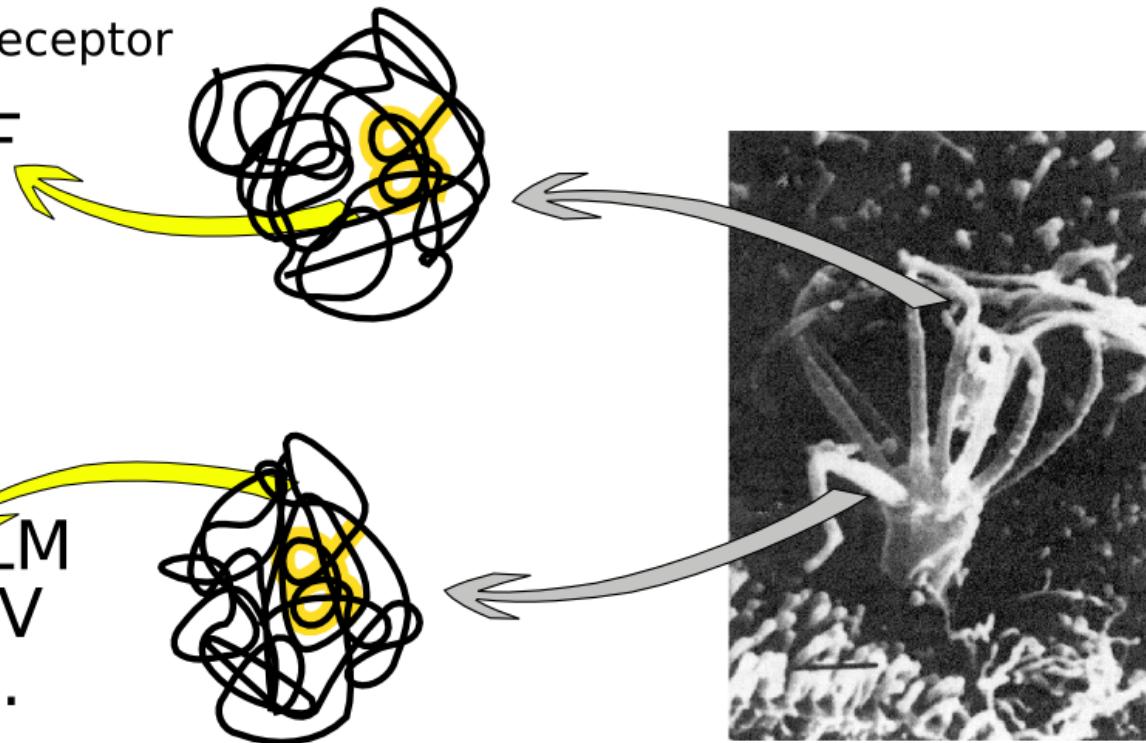
>Unknown protein
PMWFGLSMAYDRYCLM
NLMHSQWCVNPPPIYKV
TLIVKYMTAMTPCVI....



Identification par homologie

- Par homologie, prédiction des fonctions

>CfOR194 | Canis Olfactory Receptor
PMYKVILTPIMAYDRYL
KLMNIPLMNPLMSATT
TLMWNIPLMN...



Difficulté de la prédiction

Ces prédictions sont difficiles :

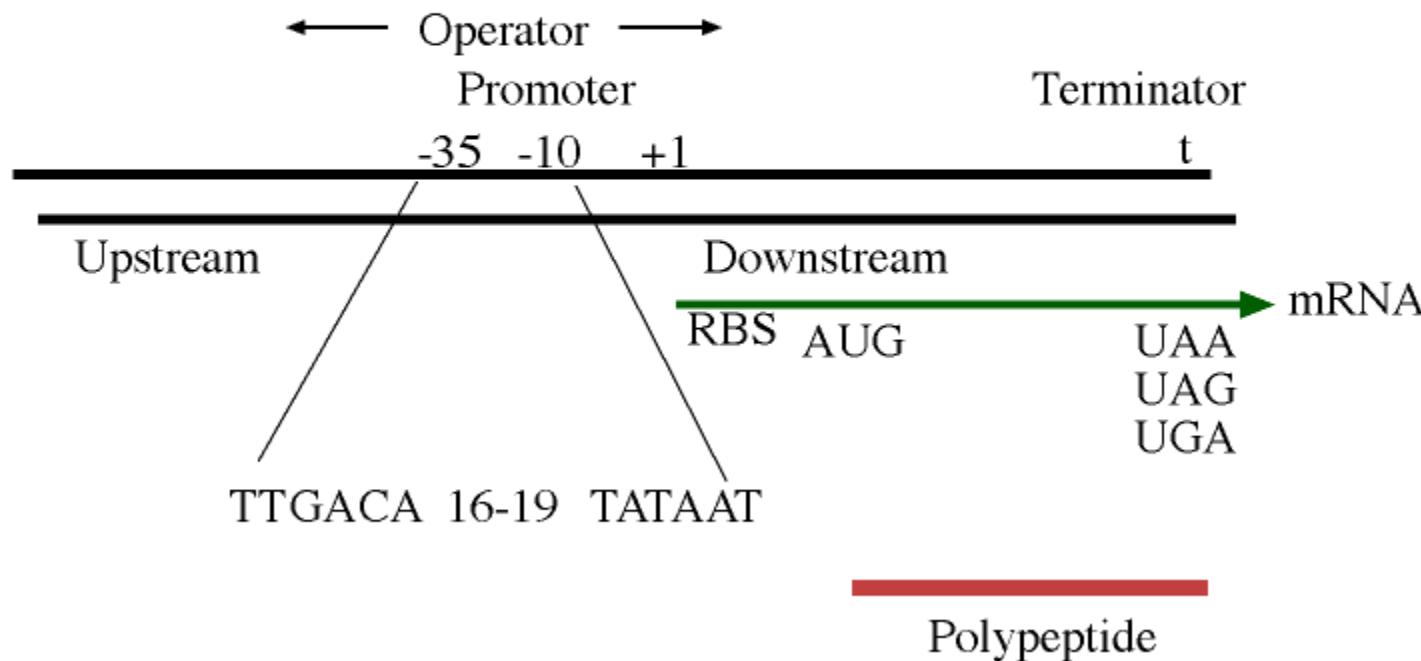
- La prédiction dépend de la qualité du corpus (jusqu'à 20% de faux positifs sur les orthologs genbank)
- Par exemple : les prophages et les IS
-> dans un génome bactérien on peut avoir 15% des gènes qui sont des séquences IS

La complexité est exponentielle :

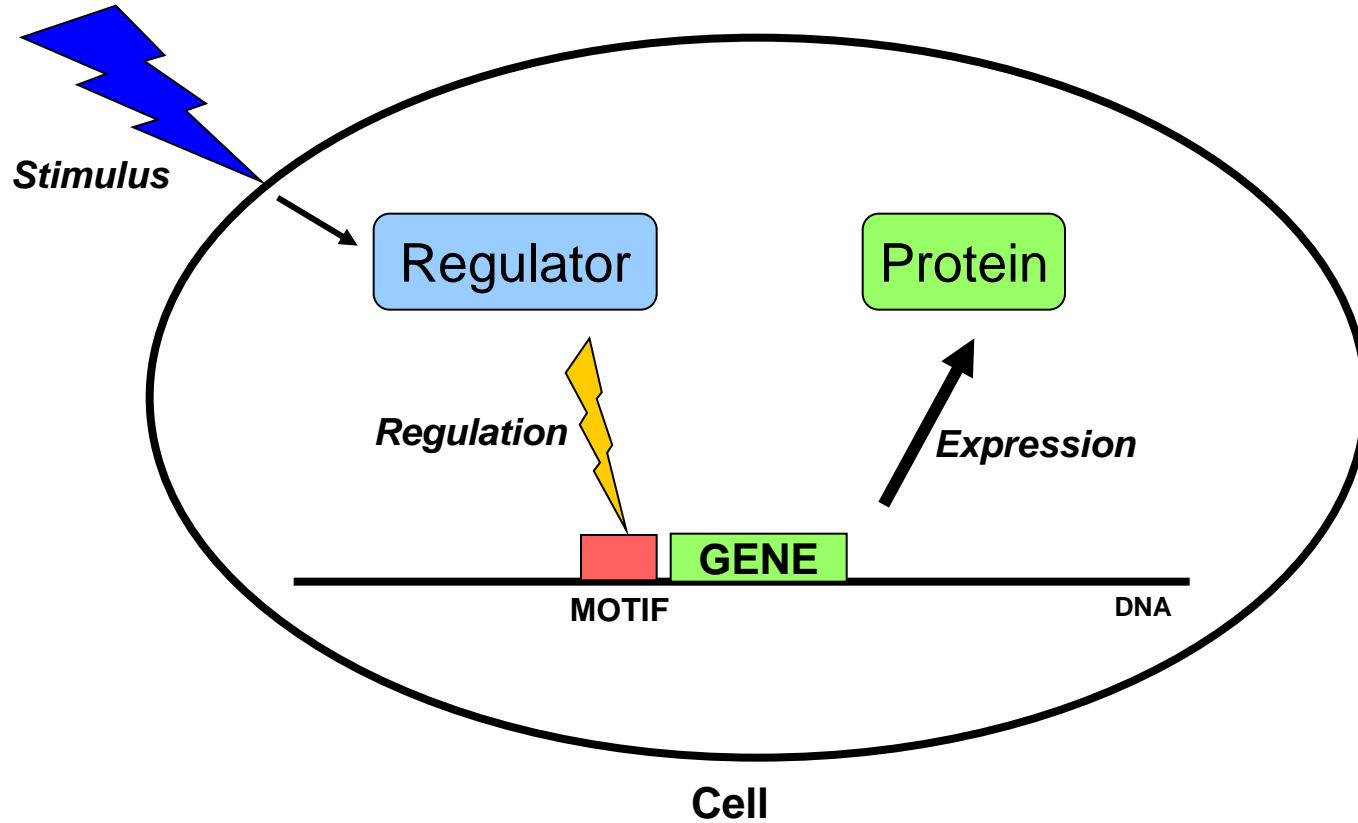
un modèle = $\prod(\text{organisation gènes}) * \prod(\text{interactions}) * \prod(\text{régulations})$

La structure d'un gène bactérien

Gene Structure

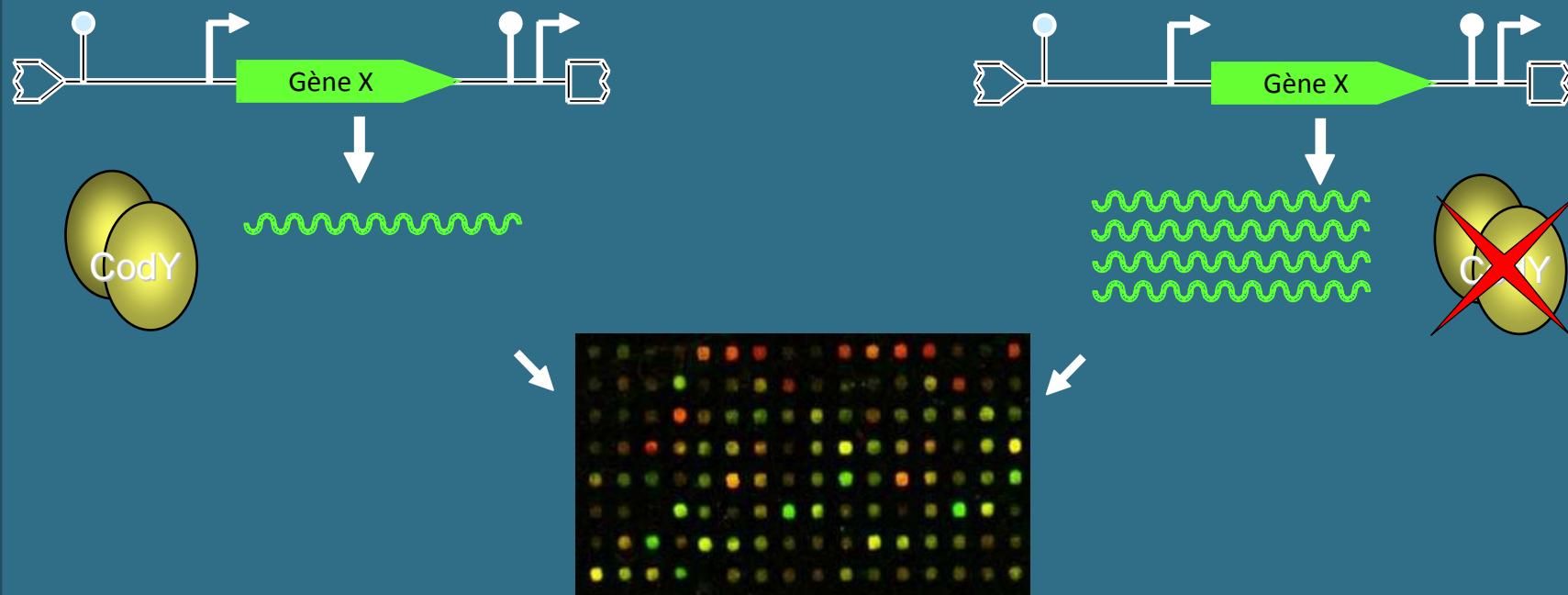


Le régulon



$$\text{REGULON} = \text{MOTIF} + \begin{matrix} \text{GENE} \\ \text{GENE} \\ \text{GENE} \\ \text{GENE} \end{matrix}$$

Exemple de recherche sur IL1403

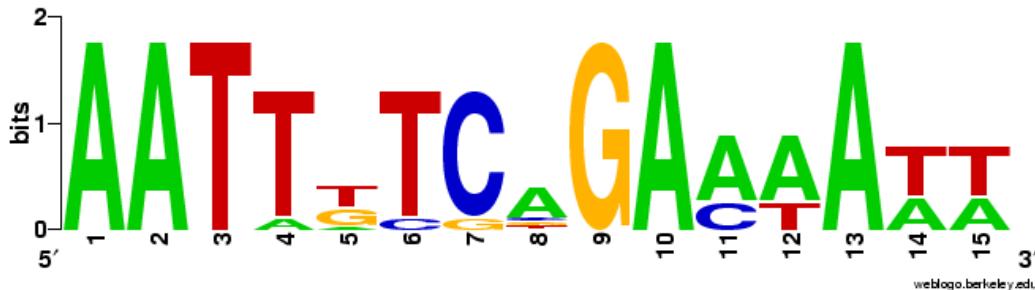


- 110 gènes différemment exprimés (5 %)
- 95 % sont réprimés

Guédon *et al.*, 2005. Microbiol.

Identification d'un motif comme site de fixation de la protéine d'intérêt.

Sa validation

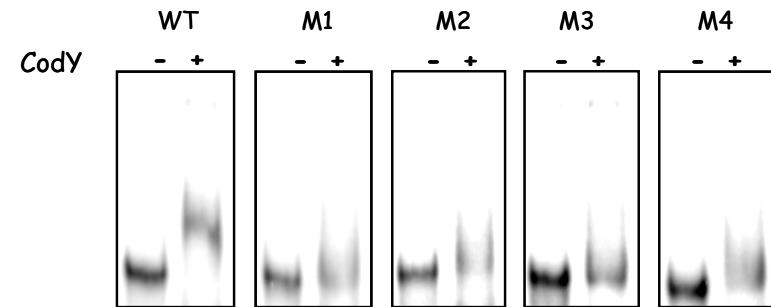


- Séquence palindromique de 15 nucléotides
- Proximité des boîtes -35/-10
- 65 % des gènes régulés par CodY chez *L. lactis*
- Présent en amont de gènes “orthologues” chez les G+

<i>L. lactis</i>	<i>bcaT</i>	AATTCATGAATATT	-77
<i>B. subtilis</i>	<i>ybgE</i>	AATATTCTGAAATT	-118
<i>B. cereus</i>	<i>BC1396</i>	AATTTTCTAAAAATT	-88
<i>S. aureus</i>	<i>ilvE</i>	AATTTTTAGAAAATT	-92
<i>S. agalactiae</i>	<i>gbs1227</i>	ATTTTTCTGAATTTT	-83
<i>S. epidermidis</i>	<i>SE0318</i>	AATATTTAGAAAATT	-90
<i>S. mutans</i>	<i>ilvE</i>	AATTTTCAGAATTTT	-83
<i>S. pneumoniae</i>	<i>ilvE</i>	AATTTTCTGAAAATT	-88
Consensus		AATTTTCTGAAAATT	

Comparaison entre mutants dans la zone d'intérêt

WT	TA	AATTTTCAGAAAATA	TA
M1	-C	GG-----	--
M2	--	-----GT-----	--
M3	--	-----TGG-----	--
M4	--	-----GC-----	--



L'apport du NCBI

(national center for biotechnology information)

Héberge PubMed, Genbank, Blast



- 1988 NCBI
- 1992 Genbank

Index de <ftp://bio-mirror.net/biomirror/ncbigenomes/Bacteria>

[/Lactococcus_lactis_II1403_uid57671/](#)

[Vers un rép. de plus haut niveau](#)

Nom	Taille	Dernière modification
NC_002662.GeneMark-2.5m.gz	97 KB	25/10/2010 00:00:00
NC_002662.GeneMarkHMM-2.6r.gz	34 KB	25/10/2010 00:00:00
NC_002662.Glimmer3.gz	32 KB	25/10/2010 00:00:00
NC_002662.Prodigal-2.50.gz	70 KB	04/12/2010 00:00:00
NC_002662.asn.gz	1325 KB	23/01/2012 02:33:00
NC_002662.faa.gz	439 KB	04/11/2011 18:00:00
NC_002662.ffn.gz	616 KB	23/01/2012 02:33:00
NC_002662.fna.gz	678 KB	22/10/2010 00:00:00
NC_002662.frn.gz	5 KB	22/10/2010 00:00:00
NC_002662.gbk.gz	1971 KB	23/01/2012 02:33:00
NC_002662.gff.gz	344 KB	23/01/2012 02:33:00
NC_002662.ptt.gz	63 KB	04/11/2011 18:00:00
NC_002662.rnt.gz	2 KB	22/10/2010 00:00:00
NC_002662.rpt.gz	1 KB	20/10/2011 14:57:00
NC_002662.val.gz	1204 KB	23/01/2012 02:33:00

L'apport du NCBI

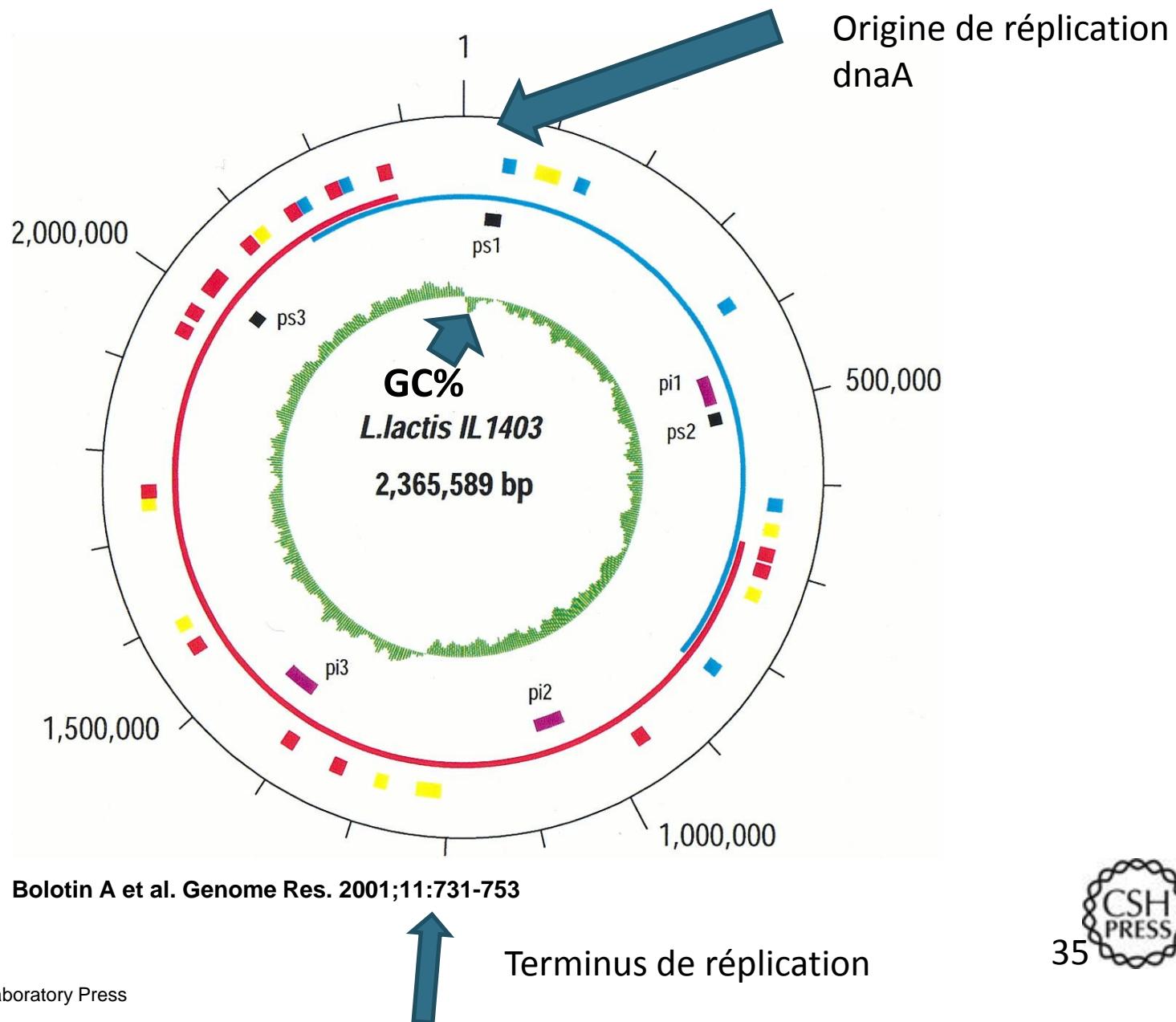
- 1992 Genbank
→format ouvert

Fichier exemple :
IL1403

LOCUS NC_002662 2365589 bp DNA circular BCT 22-JAN-2012
DEFINITION Lactococcus lactis subsp. lactis II1403, complete genome.
ACCESSION NC_002662
VERSION NC_002662.1 GI:15671982
DBLINK Project: 57671
KEYWORDS .
SOURCE Lactococcus lactis subsp. lactis II1403
ORGANISM Lactococcus lactis subsp. lactis II1403
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Lactococcus.
REFERENCE 1 (bases 1 to 2365589)
AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE The complete genome sequence of the lactic acid bacterium
Lactococcus lactis ssp. lactis IL1403
JOURNAL Genome Res. 11 (5), 731-753 (2001)
PUBMED 11337471
REFERENCE 2 (bases 1 to 2365589)
CONSRTM NCBI Genome Project
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) National Center for Biotechnology
Information, NIH, Bethesda, MD 20894, USA
REFERENCE 3 (bases 1 to 2365589)
AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de
Vilvert, Jouy en Josas 78352, France
COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The
reference sequence was derived from AE005176.
regions of prophages are represented by misc_features. Complete
genome of 5 L.lactis prophages can be found as separate proviral
genome records: 34907..49847 NC_002669 - Bacteriophage bIL310
447143..484092 NC_002668 - Bacteriophage bIL309 502595..517773
NC_002671 - Bacteriophage bIL312 1036549..1072086 NC_002666 -
Bacteriophage bIL285 1415283..1457116 NC_002667 - Bacteriophage
bIL286.
COMPLETENESS: full length.
FEATURES Location/Qualifiers
source 1..2365589
/organism="Lactococcus lactis subsp. lactis II1403"
/mol_type="genomic DNA"
/strain="IL1403"

Organisation du chromosome d'IL1403

Distribution of IS elements and prophages in the IL1403 chromosome.



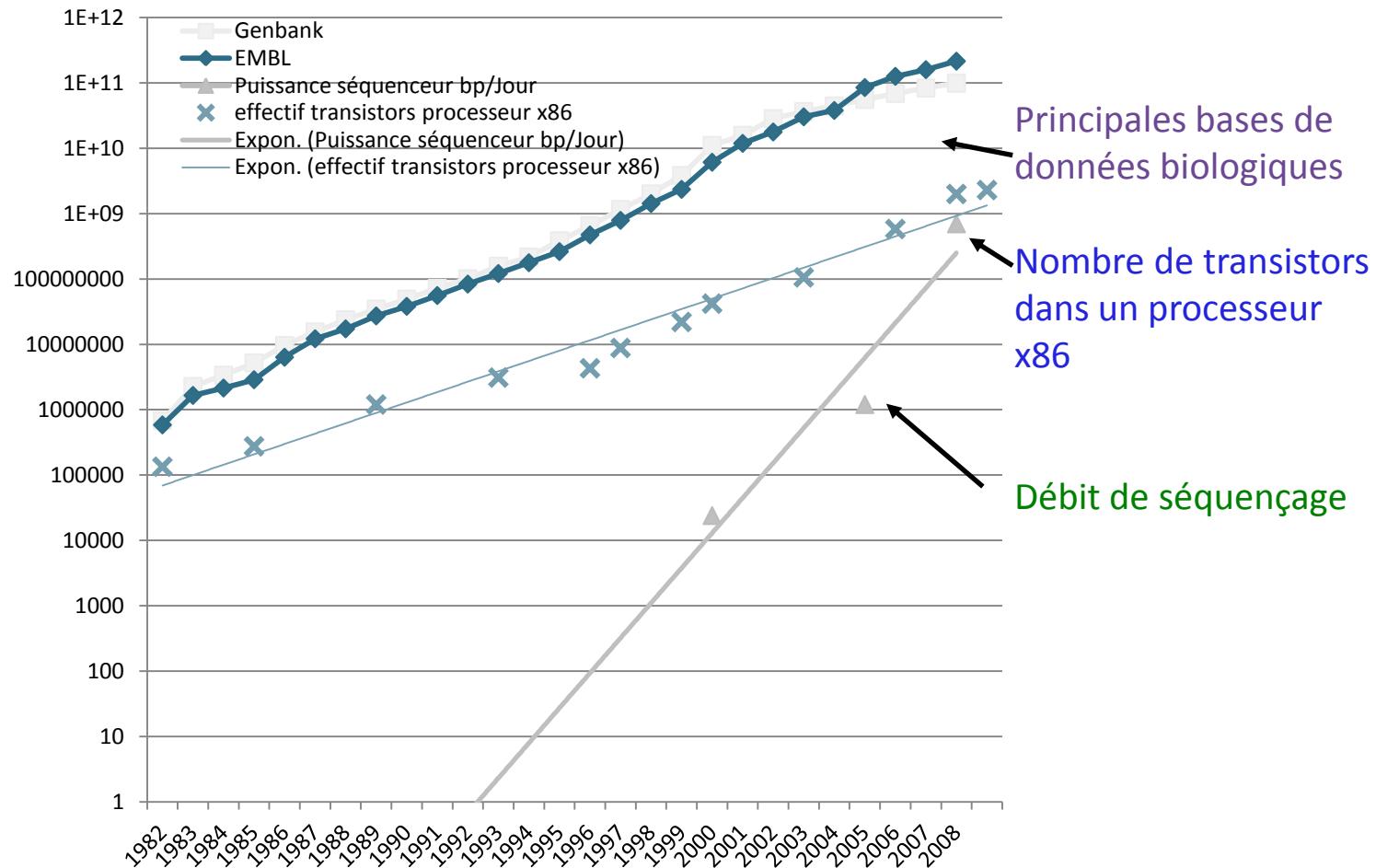
Genbank : ex d'un attribut (IL1403)

/gene="dnaA"
/locus_tag="L0274"
/db_xref="GeneID:1113599"

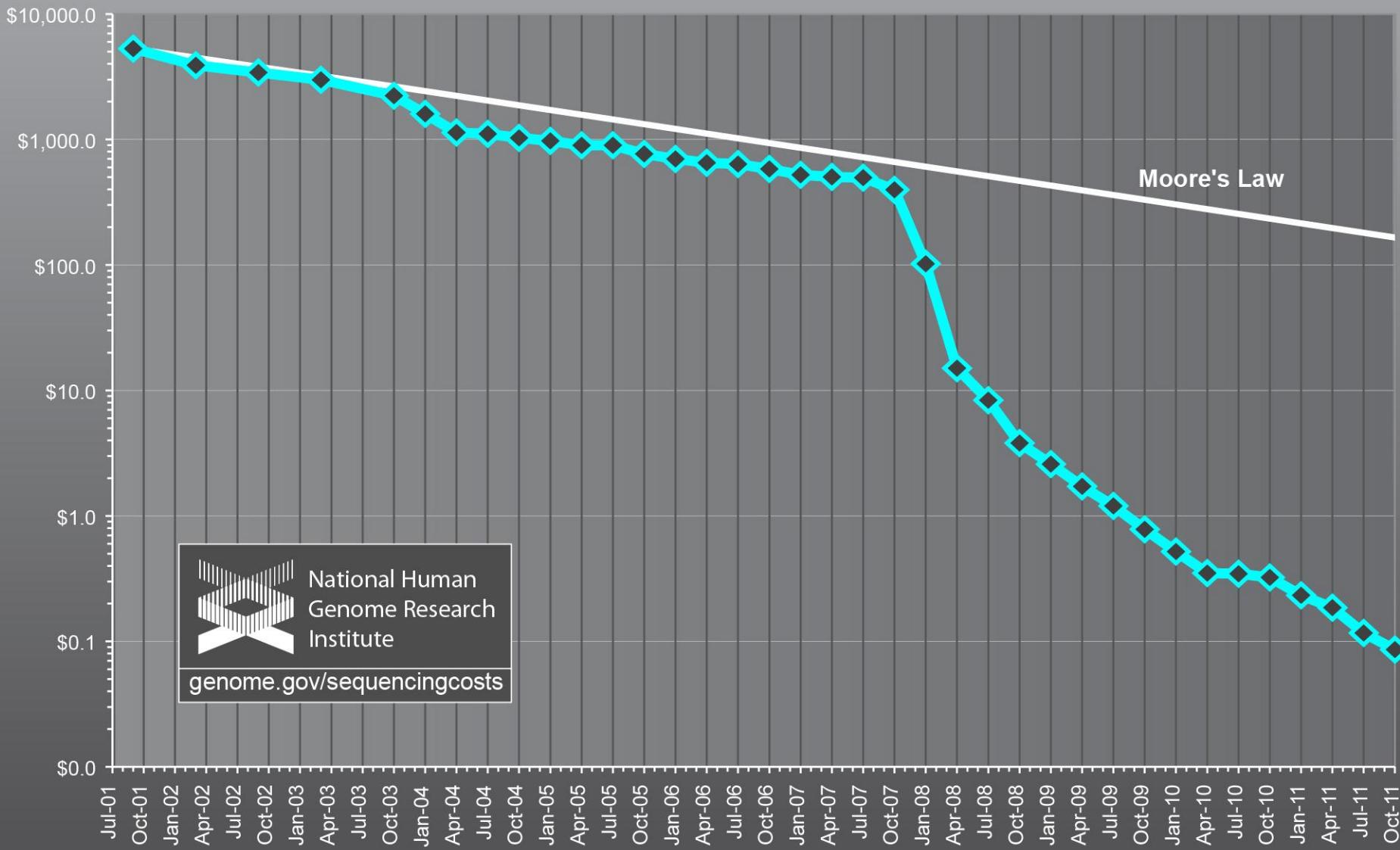
CDS 358..1725
/gene="dnaA"
/locus_tag="L0274"
/note="binds to the dnaA-box as an ATP-bound complex at the origin of replication during the initiation of chromosomal replication; can also affect transcription of multiple genes including itself."
/codon_start=1
/transl_table=11
/product="chromosomal replication initiation protein"
/protein_id="NP_266157.1"
/db_xref="GI:15671983"
/db_xref="GeneID:1113599"
/translation="MASLNENQKFWARVTELAQQSIGKQAYDFFIEPAQLMSVEQDTANILLDSGMKKDYWKKQSDLITTAGFEVFGRMIDYELYANDELTELELHRLNNQSSIEEQPRSTAKPASPLVSGLNEKYNFENFVQGPGNRWTAAAIAVADKPGDTYNPLFIYGGAGLGKTHLMHAIGNQILTNDNPTARIKYVSSENFVNNDYVNATRKNQMENFENTYRNLDLLLDDVQFFSDKEGTKNEFFNTFNALYDKGSQIVLTSDRIPQELNNLEDRLVSRFSWGLTTDITAPDYETRMAILLIKSESSRLEFPSETLSIAGQIDSNVRELEGALNRVEFVARANGIAVVDIETASQALRSLKNATQQSLSNLTICKIQDEVANYHISFSDLVGPKRKPEIAFPRQIAMYLVRRELLGTSLPAIGTAFGGRDHTTVMYAYKQISDKMKNDMDVQKDIDSIKRKF"

misc_feature 373..1719
/gene="dnaA"
/locus_tag="L0274"
/note="chromosomal replication initiation protein;
Reviewed; Region: dnaA; PRK00149"
/db_xref="CDD:178902"

Evolution de l'information et loi de Moore



Cost per Raw Megabase of DNA Sequence



La complexité de Kolmogorov

- Définition 1965
 - $K(s)$: pour une suite binaire finie s , K est la taille du plus court programme qui produit s
- Le nombre π : complexité de 1600 décimales

```
int a=10000,b,c=8400,d,e,f[8401],g;main() {for(;b-c;) f[b++]=a/5; for(;d=0,g=c*2;c-=14,printf("%.4d",e+d/a),e=d%a) for(b=c;d+=f[b]*a,f[b]=d%--g,d/=g--,--b;d*=b); }
```

158 caractères

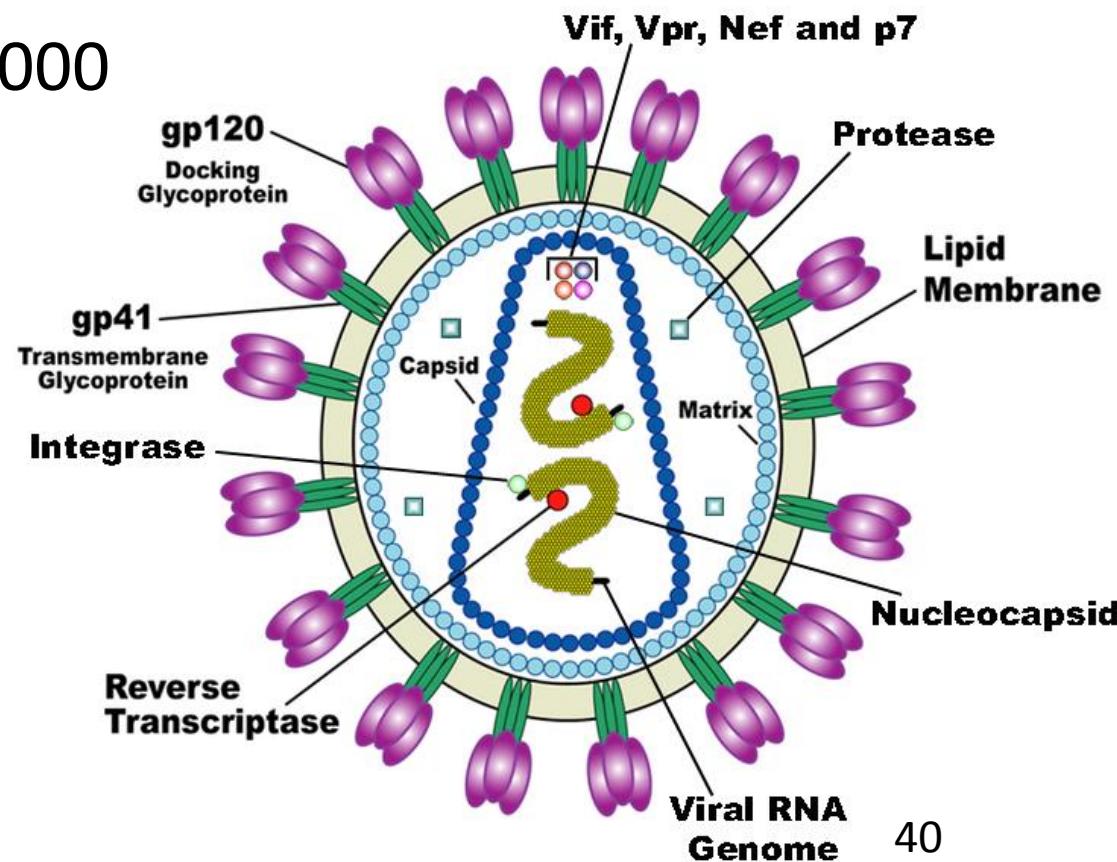
$$K(\pi) < 5000$$

Combien de bits pour un virus

- Le virus HIV : ~10 800 lettres ATGC

Compressé 7zip : 3090 octets

Complexité K(HIV)<25000



Combien de bits pour un humain

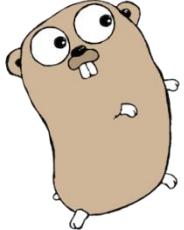
- Genome humain : $3*10^9$ (Gigabase) , $6*10^9$ bits
Compressé zip : 770 Mo
 $K(H) < 7*10^9$
- Une cellule humaine : 11k protéines, longueur moy. 500 AA, (pour décrire 20AA → 6 bits)
 $11*10^3*5*10^2*6=3*10^7$
 $K(\text{cell}) < 3*10^7$
- Effectif en cellules humaines : $50*10^{18}$
- Variabilité DNA humain : ~6 million SNP et ~1.8 million insertions, deletions et inversions → on peut la décrire avec ~20 Mo

Utilité de la complexité K

- Permet de construire une métrique sans connaître la biologie
 - Modélisation? non
 - Ne permet que de décrire
- Donc il faut connaître la biologie

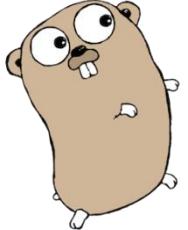
En bref,

- L'étude du vivant est une science expérimentale
- Elle nécessite des observations
- Connaitre l'ADN permet de faire des prédictions et de classifier le vivant
- L'informatique a révolutionné l'étude de l'ADN

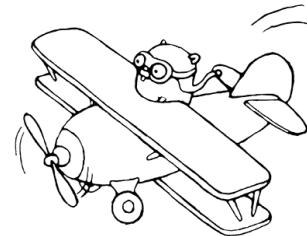


Le langage Go

- Le langage et son utilité
 - Abstraire le matériel
 - Décrire l'espace du problème
 - Favoriser les bonnes pratiques (versions, réemploi, validation, contrat)
 - Rendre le programmeur heureux (joies simples : formatage, joies complexes : suprématie)
 - Rendre le formateur heureux (effet perroquet, effet ha-ha)



Le langage Go



- La création des langages est une activité continue
50 nouveaux langages tous 10 ans

(source https://en.wikipedia.org/wiki/Timeline_of_programming_languages)

Un langage est une représentation de l'espace du problème

Un langage n'est pas universel (tour de Babel) → idée de spécialisation (en 1996, 500 langages spécialisés)

(source <http://www.cs.bsu.edu/homepages/dmz/cs697/langtbl.htm>)

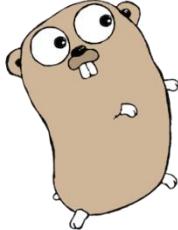
Il faut environ 7 ans pour qu'un langage soit populaire

PHP4.2 (2002) → 2009

Python2.3 (2003) → 2010

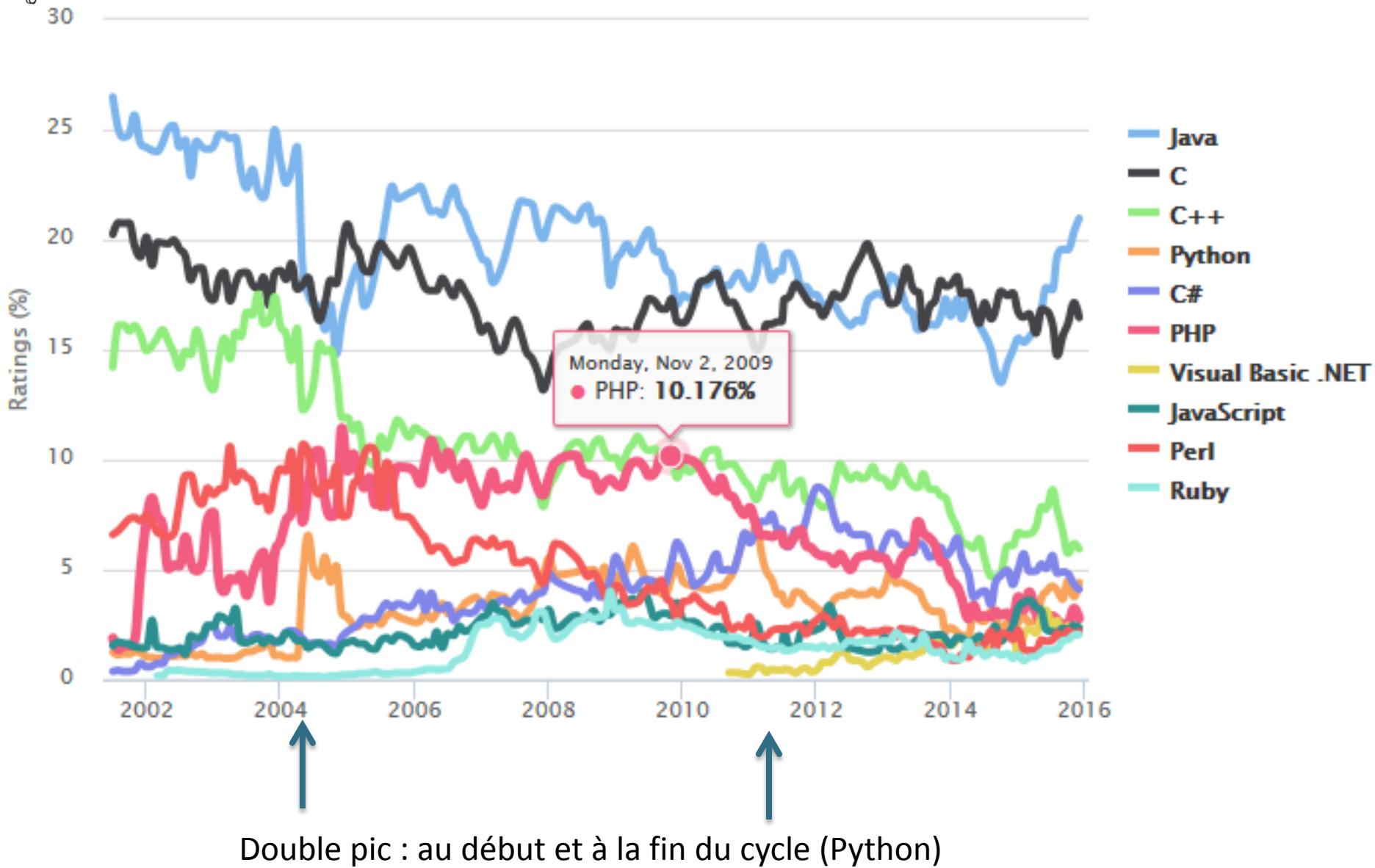
Go1.0(2009) → 2016 ??

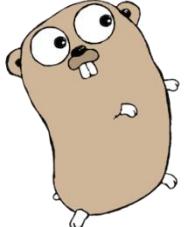
Un langage "à la mode apparaît" en fonction des pbs à "la mode"



TIOBE Programming Community Index

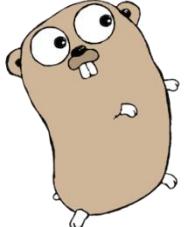
Source: www.tiobe.com





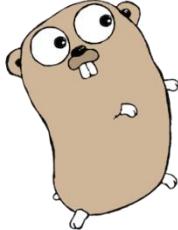
Les problèmes de 2015

- Many-cores
 - Knights landing à 72 coeurs
- SIMD et vectorisation
 - GPU
- Enveloppe thermique <5W
 - IoT
- Cycle de développement en mode "superhéros"
(=pas de cycle en V)
 - Agile, Git



Typologie d'un langage

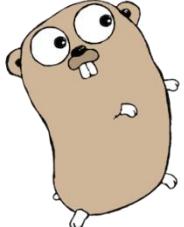
- Les paradigmes sous jacents au delà du pure langage
 - Tout est liste (lisp, ruby, smalltalk) et rien d'autre
 - Smalltalk conceptualise l'objet à travers une vision organisée
 - Tout est objet (C++) y compris le clavier
 - C++ reprend le C et branche des objets sur tout (header, opérateur, template, ...)
 - Tout est règle (SQL, Prolog)
 - L'abstraction dans l'abstraction :
 - mettre une VM
 - mettre un ramasse miette
 - mettre des références à la place des pointeurs
- Tout est relatif!



La productivité du codeur

- COCOMO

Nom du langage	Date de création	Cible	Typologie	Cocomo (lignes / jours)	Expressivité
Smalltalk	1980	bytecode	Objet Typ.dynamique	15	6* le C
C++	1985	Compilateur	Objet hybride Typ.statique	25	2* le C
VHDL	1987	FPGA	Procédural Typ.statique	2 (\rightarrow le FPGA)	1/6 du C
Delphi Pascal	1995	Compilateur	Objet simple Typ.statique	30	5* le C
Java	1995	bytecode	Objet simple Typ.statique	40	4* le C
Python	2003	Interpréteur	Tous les paradigmes Typ.dynamique	40	4* le C

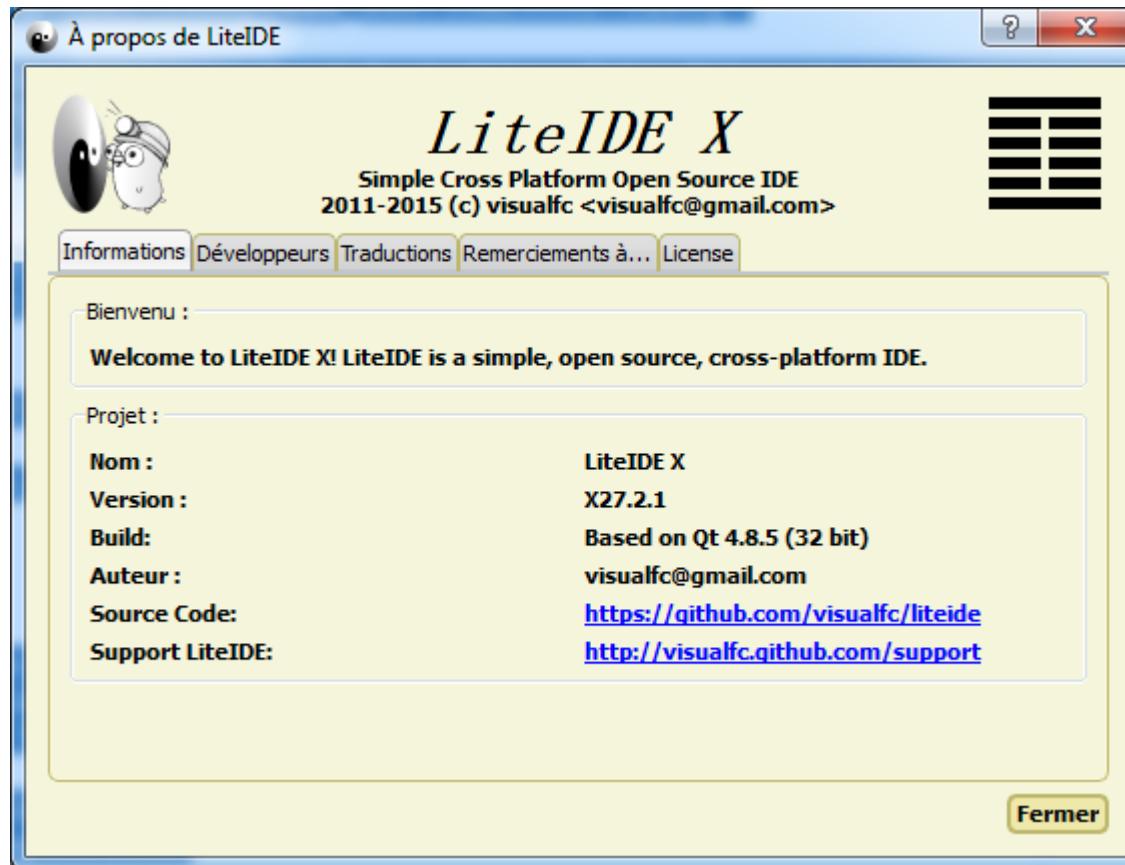


Le langage Go

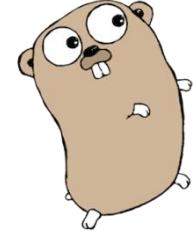
- Pas d'héritage mais de la composition
- Contrat par interface (ça se comporte comme)
→ mieux que l'héritage (idée du smalltalk)
- Dispatch simple (récepteur du message)
- Closure → contexte
- CSP (communicating sequential processes)
- Bonnes pratiques
→ un retour en arrière?
→ un usage de type "No Limit!"

Les outils
liteide

www.golang.org



package main



```
import (
    "encoding/csv"
    "flag"
    "fmt"
    "github.com/gonum/plot/plotter"
    "os"
    "path/filepath"
    "runtime"
    "strconv"
    "strings"
    "time")
var g_sizeKmer int
var xmin, xmax, ymin, ymax float64
//compute duration
func elapsedTime(s string, startTime time.Time) time.Duration {
    return time.Since(startTime)}
//replace decimal point according to local
func formatLocal(s string) string {
    return strings.Replace(s, ".", ",", -1)}
//get the local (string)à representation of float64
func formatLocalF64(val64 float64) string {
    return (formatLocal(strconv.FormatFloat(val64, 'f', 10, 64))))}
```



```
func main() { var defaultfile string
fmt.Println("MaxParralelism %d %d \n", runtime.GOMAXPROCS(0),
    runtime.NumCPU())
startTime := time.Now()
defaultfile = `E:\golang-test\R_2014_12_17_16_38_01_user_PRO-30-
    ControlMix1.mix1_17.IonXpress_017.fastq`
var imputFile = flag.String("filename", defaultfile, "the filepath to process - fastq
    format")
var kmerSize = flag.Int("kmersize", 4, "k-mer size (>2 and <12)")
var shortTreatment = flag.Bool("shortprocess", false, "shortprocess (boolean) will treat
    only the first 10000 lines")
flag.Parse()
if (flag.NArg() > 0) || (*kmerSize > 12) || (*kmerSize < 3) {
fmt.Fprintf(os.Stderr, "Usage: %s [OPTION]...\n", os.Args[0])
os.Exit(1)
}, err := os.Stat(*imputFile)
if err != nil {
fmt.Fprintf(os.Stderr, "the file %s doesn't exist! \n", *imputFile)
os.Exit(1)
durationTime := elapsedTime("prog", startTime)    fmt.Printf("time elapsed %s \n",
durationTime) }
```