

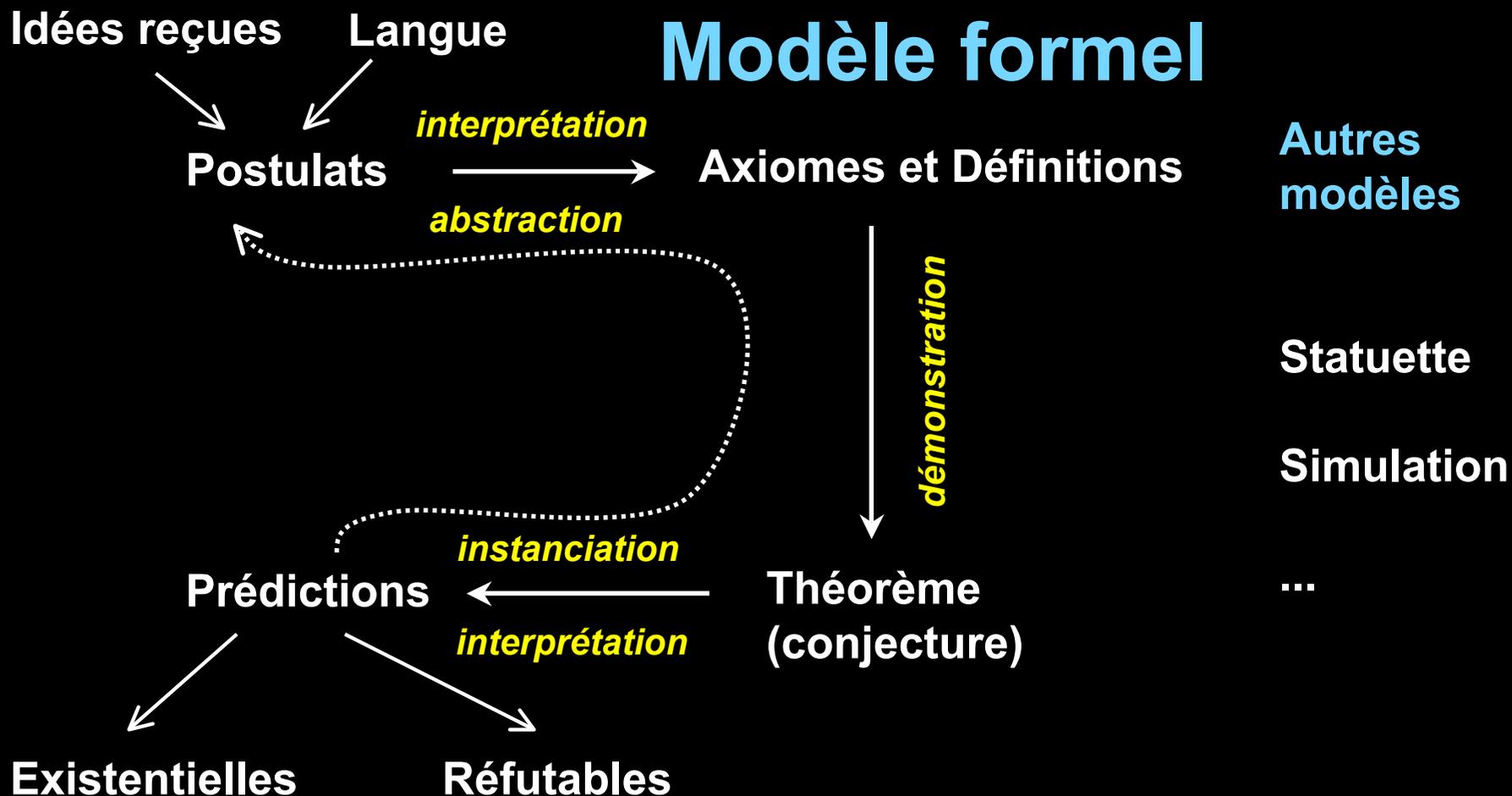
CE QUE NOUS APPREND LA STRUCTURE DES GÉNOMES BACTÉRIENS DE L'ORIGINE DE LA VIE À LA BIOLOGIE SYNTHÉTIQUE

唐善安東

Cours de Mastère I, Ecole Normale Supérieure

7 janvier 2008

LA MÉTHODE CRITIQUE GÉNÉRATIVE



--- T H É O R I E ---

→ VIE ET CALCUL

→ LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME

→ DE LA SOUPE À LA SURFACE DES PIERRES

→ LA TRANSFORMATION HOMÉOTOPIQUE

→ DES ACIDES AMINÉS AUX NUCLÉOTIDES

→ LE MONDE DES ARN

→ UNIVERSAUX DANS LES GÉNOMES

→ LE CŒUR DU GÉNOME : CE QUI PERSISTE

→ EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME

CONTEXTE : LE “PROGRAMME GÉNÉTIQUE”

- **Physique** : *matière, énergie, temps*
- **Physique statistique** : *Physique + information*
- **Biologie** : *Physique + information, codage, contrôle...*
- **Arithmétique** : *suites d'entiers, récursivité, codage...*
- **Calcul** : *Arithmétique + programmes + machine...*

La métaphore du « programme génétique » rend explicite un certain nombre de points communs entre le calcul et le fonctionnement cellulaire : **pouvons-nous pousser la métaphore jusqu'à ses conséquences ultimes ?**

CE QU'EST LA VIE

Trois entités constituent la vie :

→ **Métabolisme**

| une machine

→ **Compartimentation**

| (principe : “**substrat**”,
| reproduction

→ **Transfert d'information**

| un “programme”
| (déclaratif, non
| prescriptif, principe :
| “**matrice**”, réplication)

La cellule est l'atome de vie

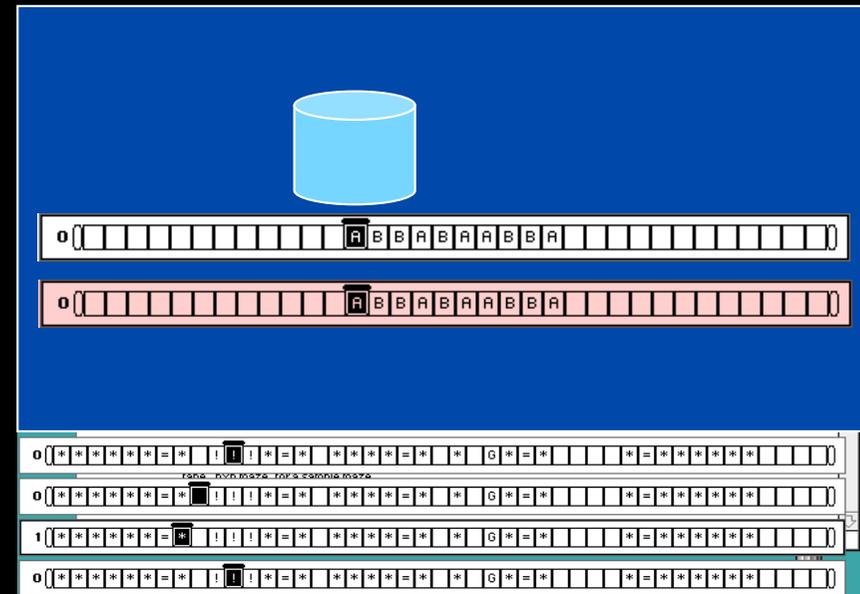
VISION PHYLOGÉNÉTIQUE DU MONDE

- Produire une descendance suppose faire du neuf avec du vieux
- Les systèmes évoluent par mutations (parfois de simples réarrangements, sans changement du « texte » génomique)
- La réplication ne peut qu'accumuler les erreurs et courir à la catastrophe (sauf échanges avec l'extérieur)
- Il faut distinguer la **reproduction** (d'un état métabolique) de la **réplication** (Freeman Dyson)
- La reproduction peut faire du neuf avec du vieux

QU'EST-CE QUE CALCULER ?

Deux entités sont nécessaires au calcul :

- Une machine capable de lire et écrire
- Un programme sur un support physique (typiquement, une bande perforée ou magnétique illustre la suite séquentielle des symboles qui forment le programme), séparé (en pratique, mais pas conceptuellement) en deux entités :
 - Programme (fournissant l'objectif)
 - Données (fournissant le contexte)



La machine est distincte du programme

CELLULES ET ORDINATEURS

La génétique repose sur la description des génomes comme **textes écrits avec un alphabet de quatre lettres** : mais **les cellules se comportent-elles comme des ordinateurs ?**

- Transfert Génétique Horizontal
- Virus
- Génie génétique
- **Transplantation d'un chromosome d'une espèce dans une autre espèce (2007)**

tout indique une séparation entre

« Machine » (l'usine cellulaire)

et

Données + programme

VUE ALGORITHMIQUE DES ACTIONS BIOLOGIQUES

Réplication, transcription, traduction : parallélisme élevé

“Début, Routine répétitive et Points de contrôle, Fin”

L'action est toujours orientée, avec un début et une fin

Les processus de contrôle temporel (check points) sont rarement pris en compte (excepté pour les processus de réplication/division), **mais leur rôle est essentiel pour permettre la coordination de multiples actions en parallèle**

TROIS NIVEAUX DE L'INFORMATION

- Ce que voit la réplication => Information à la Shannon
- Ce que voient la transcription et la traduction => complexité algorithmique
- Ce que voit l'évolution => profondeur logique

CE QUE VOIT LA RÉPLICATION

Au cours de la réplication la séquence de l'ADN est recopiée exactement. La copie ne se soucie nullement de la signification de la séquence

L'enveloppe génétique (la collection des séquences recopiées) peut être étudiée à un premier niveau de l'information qui lui correspond. Typiquement on utilise à ce stade l'« information » de Shannon

COMPLEXITÉ ALGORITHMIQUE

- L'« entropie » de Shannon ne se soucie pas de la signification (réplication)
- Kolmogorov et d'autres ont proposé de définir un autre niveau d'information, qui donne une définition de l'aléatoire, la longueur du plus petit programme définissant exactement une suite de symboles : la complexité algorithmique
- Cela nous fournit un **programme de recherche** : afin d'approcher la complexité algorithmique d'une séquence, on s'intéresse à la façon dont elle a été engendrée (au cours de l'évolution, au cours des biosynthèses, etc)

PROCARYOTES ET EUCARYOTES

→ **Procaryotes** : une seule enveloppe, plus ou moins compliquée ; génome « aléatoire »

→ **Eucaryotes** : multiplication des membranes et des peaux ; génome riche en répétitions

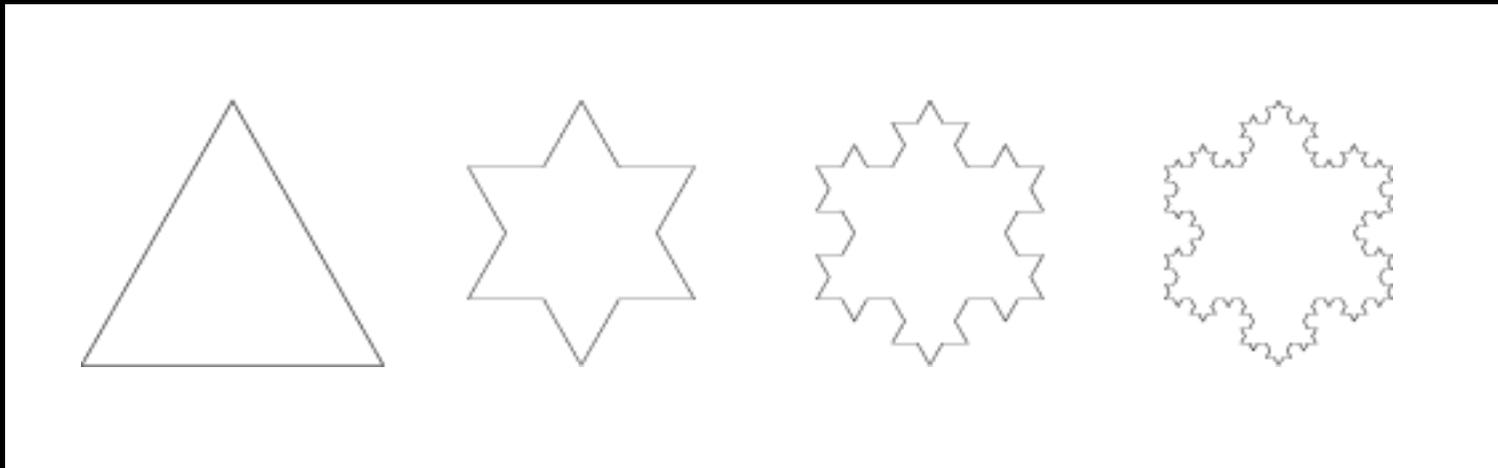
MISE EN GARDE

La métaphore alphabétique prise au premier niveau est insuffisante :
Même les répétitions ont un sens

« Les poules du couvent couvent »

LA PROFONDEUR LOGIQUE

Un programme très court (**faible complexité algorithmique**) peut décrire une simple séquence répétée, mais aussi une figure fractale comme le flocon de Koch



PROPRIÉTÉS INATTENDUES DES SUITES DE SYMBOLES

- Le texte de l'ADN peut être considéré comme une suite de symboles
- Cette suite est transposée via un code, ce qui lui permet la récursivité
- En utilisant l'arithmétique Gödel a montré que le codage permettait de créer une imprévisibilité de principe via le retour sur soi



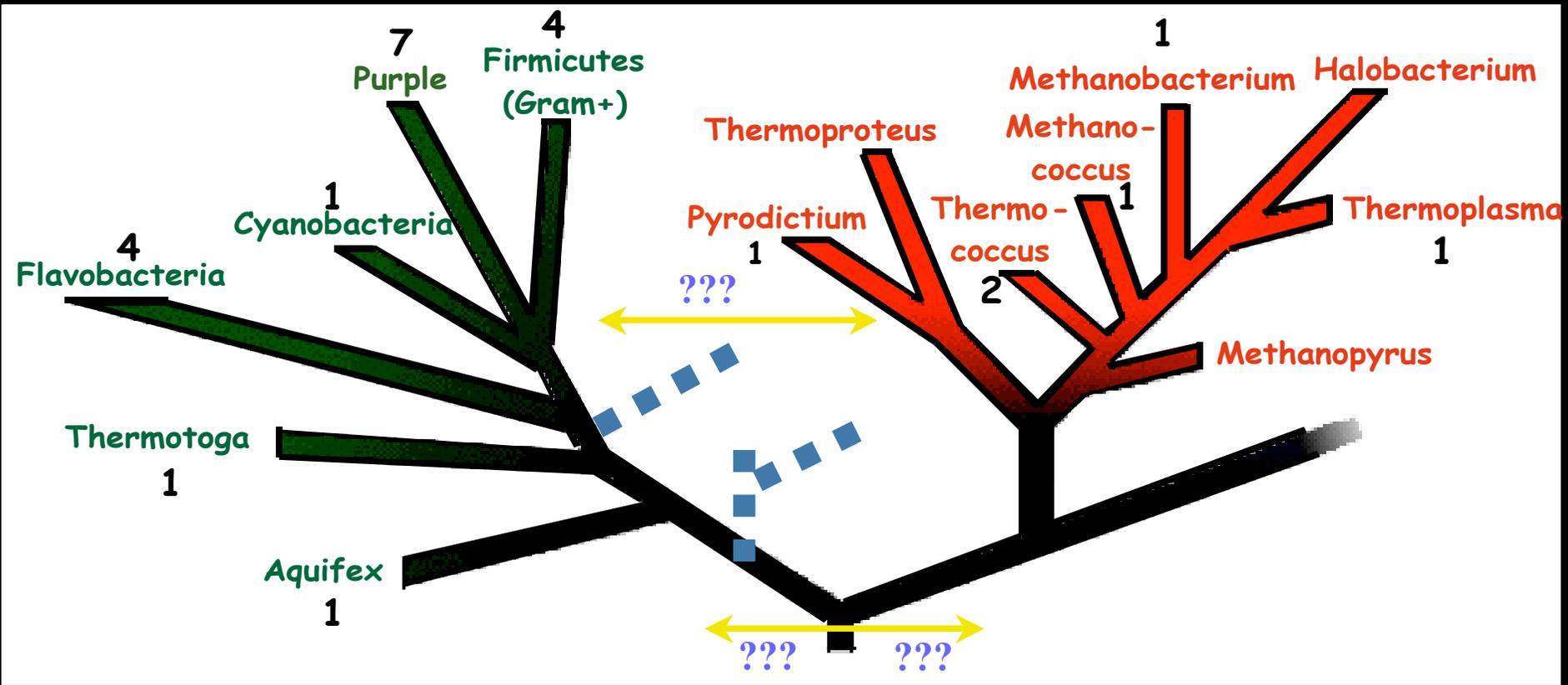
LA PROFONDEUR LOGIQUE

Une conséquence : pour un programme suffisamment compliqué (branché et récursif), il devient impossible de prédire son comportement dans un temps raisonnable. La seule façon de le faire est de **dérouler programme...**

Au cours de l'évolution l'ADN vient de l'ADN, vient de l'ADN, etc... chaque base a donc une certaine « profondeur », qui fait que les organismes vivants sont, par construction, destinés à **être imprévisibles...**

Il s'en suit que, dans la descendance, certains individus seront aptes à survivre dans un futur imprévisible (et qu'il n'y a pas d'ADN « poubelle » (junk DNA)

PHYLOGENÈSE



(Woese, 1990)

➔ **VIE ET CALCUL**

➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**

➔ **DE LA SOUPE À LA SURFACE DES PIERRES**

➔ **LA TRANSFORMATION HOMÉOTOPIQUE**

➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**

➔ **LE MONDE DES ARN**

➔ **UNIVERSAUX DANS LES GÉNOMES**

➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**

➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**

ONTOGENÈSE: Y A-T-IL UNE IMAGE DE LA CELLULE DANS LE CHROMOSOME?

Si la machine doit non seulement se comporter comme un ordinateur mais aussi construire la machine, on doit trouver un programme géométrique quelque part dans la machine (J. von Neumann)

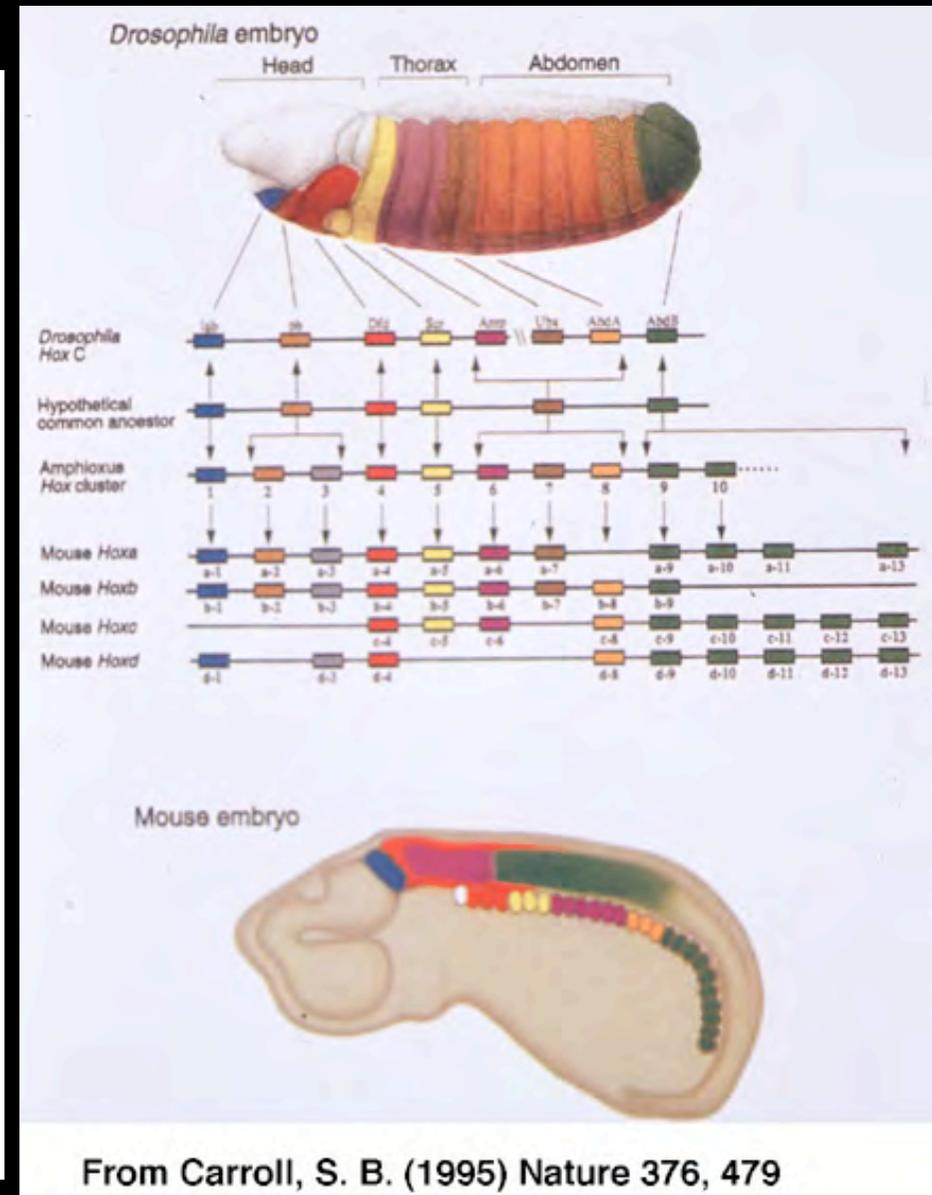
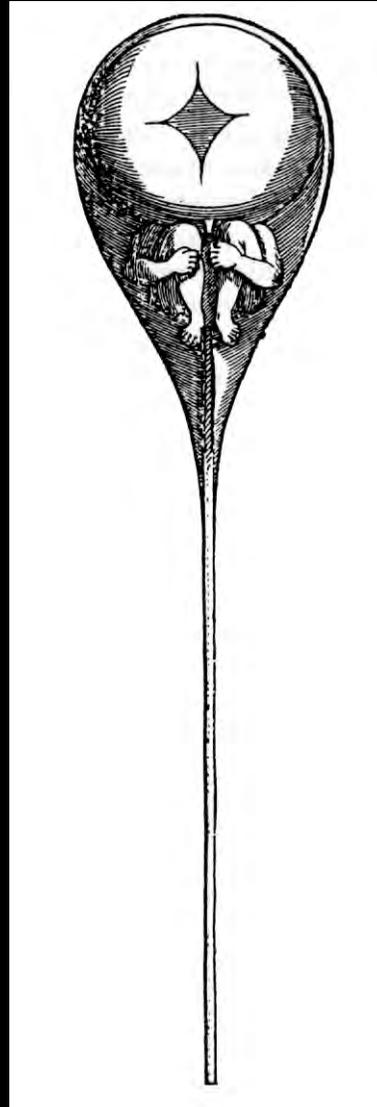
Cette réflexion, appliquée dans un contexte où elle ne peut l'être (le fonctionnement cérébral) est-elle applicable dans le cas de la cellule et du programme génétique ?

L'ordre des gènes est-il aléatoire dans les génomes?

Beaucoup d'organismes sont multicellulaires: ils passent d'un œuf à un adulte. Comment est-ce organisé si l'on retient la métaphore « alphabet - ordinateur »?

LE MYSTÈRE DE L'ORIGINE DES HOMÉOGÈNES

Drosophiloculus,
Homunculus?
Celluloculus?



FORME DE LA CELLULE

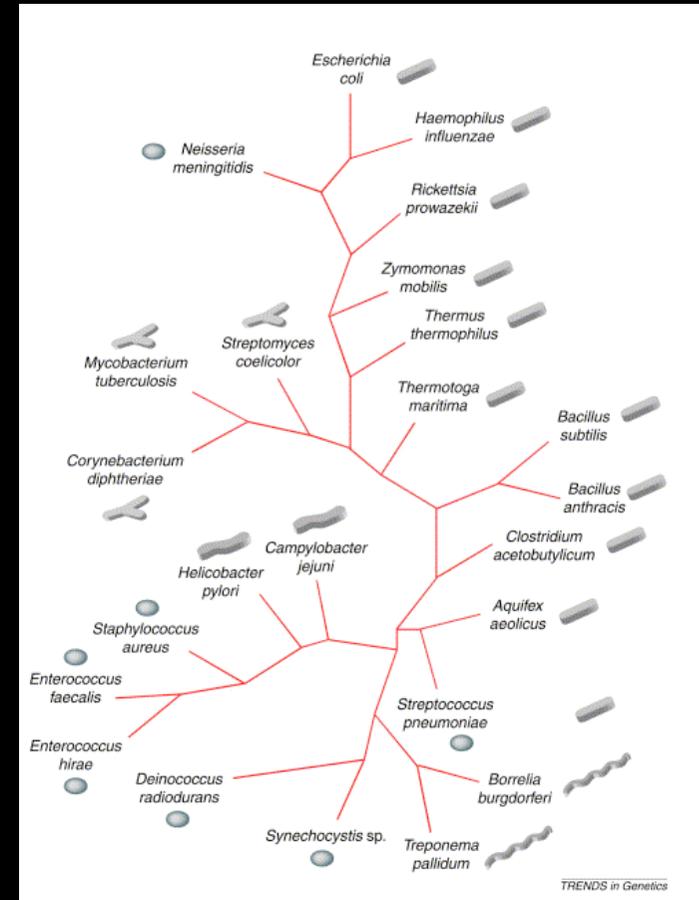
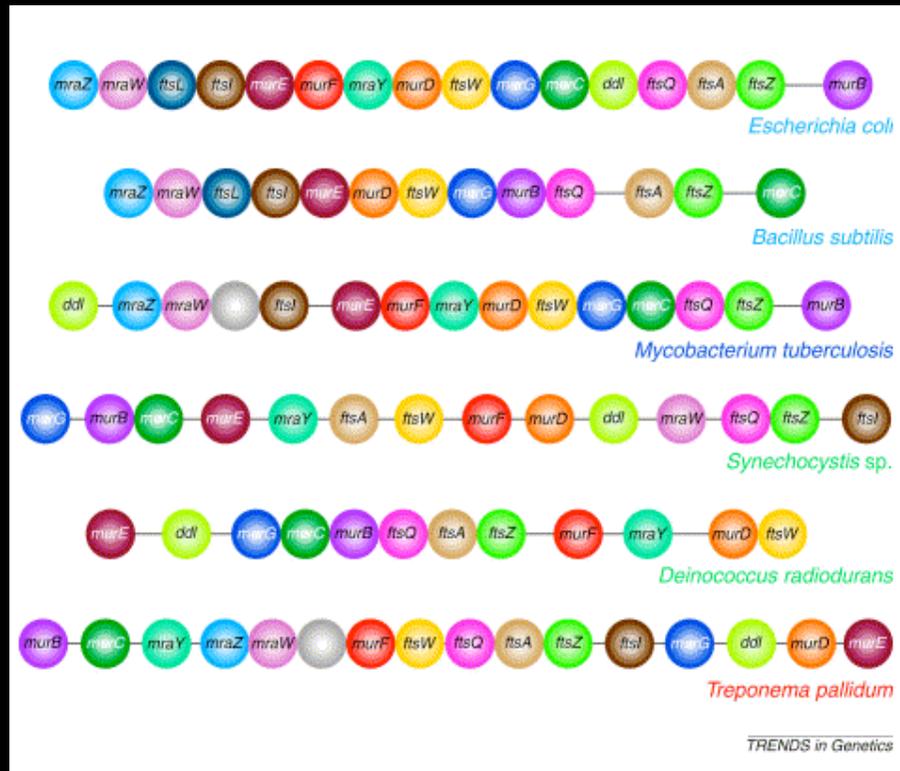
Construire une cellule synthétique doit prendre en compte cette conjecture, et respecter en pratique l'ordre des gènes dans le génomes synthétiques

Qu'apprenons-nous des génomes bactériens ?

Y trouvons-nous des principes organisateurs ?

ORDRE DES GÈNES ET FORME DES BACTÉRIES

L'îlot *mur-fts*



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

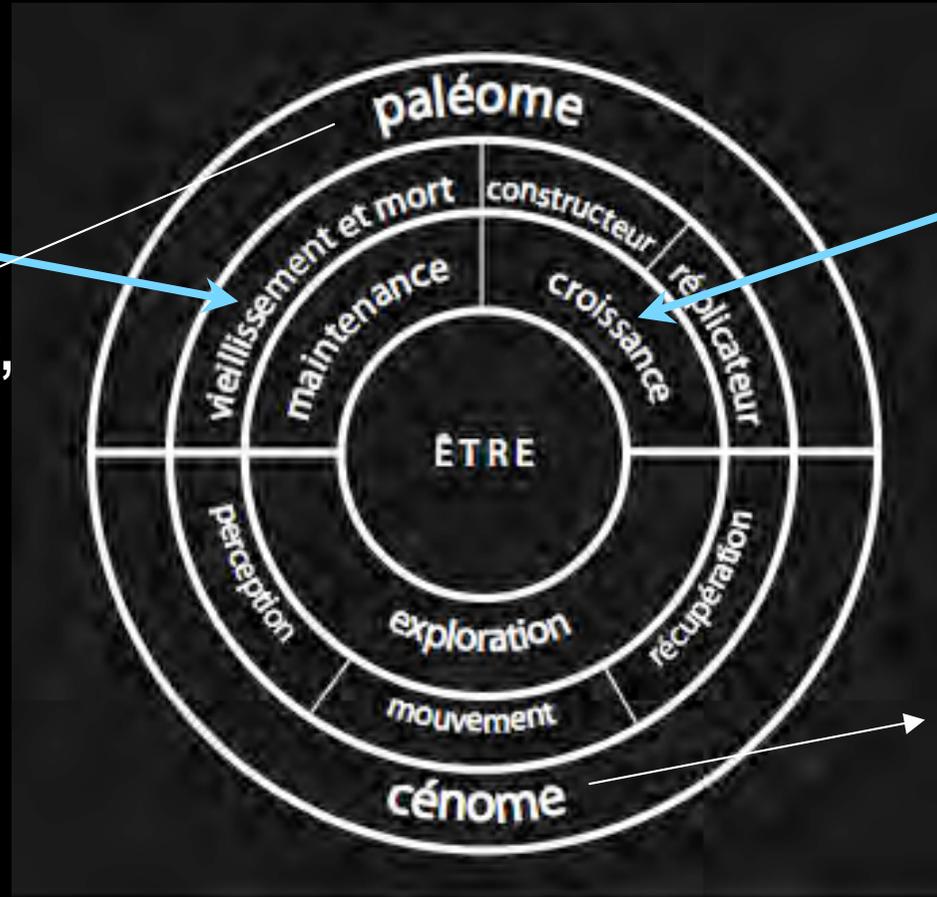
UNE MACHINE RÉCURSIVE

⇒ **Réplicateur** : l'ADN spécifie des protéines qui répliquent l'ADN

⇒ **Constructeur** : l'ADN spécifie des protéines qui forment la machine qui construit la cellule

FONCTIONS PREMIÈRES

perpétuer la vie
Du grec παλαιος,
ancien



vivre

Du grec κοινος,
commun, cf
« biocénose »

occuper le milieu environnant

- ➔ **VIE ET CALCUL**
- ➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**
- ➔ **DE LA SOUPE À LA SURFACE DES PIERRES**
- ➔ **LA TRANSFORMATION HOMÉOTOPIQUE**
- ➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**
- ➔ **LE MONDE DES ARN**
- ➔ **UNIVERSAUX DANS LES GÉNOMES**
- ➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**
- ➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**

UNE THÉORIE DE L'ORIGINE DE LA VIE DOIT EXPLIQUER :

- **COMMENT SE SONT FAITS LES COMPARTIMENTS**
- **COMMENT SE SONT CONSTRUITES LES PREMIÈRES MOLÉCULES RÉPLICATIVES**
- **COMMENT S'EST CRÉÉE LA CORRESPONDANCE ENTRE MÉMOIRE ET MANIPULATION**

**UN MÉTABOLISME PRIMITIF SUR DES SURFACES SOLIDES
DONNE UNE SOLUTION**

Origine



II N'Y A PAS DE GÉNÉRATION SPONTANÉE

J'AI LA PRÉTENTION DE DÉMONTRER AVEC RIGUEUR QUE DANS TOUTES LES EXPÉRIENCES OU L'ON A CRU RECONNAÎTRE L'EXISTENCE DE GÉNÉRATIONS SPONTANÉES, CHEZ LES ÊTRES LES PLUS INFÉRIEURS, OU LE DÉBAT SE TROUVE AUJOURD'HUI RELÉGUÉ, L'OBSERVATEUR A ÉTÉ VICTIME D'ILLUSIONS OU DE CAUSES D'ERREUR QU'IL N'A PAS APERÇUES OU QU'IL N'A PAS SU ÉVITER.

LOUIS PASTEUR



LA SOUPE PRÉBIOTIQUE IMPLIQUE LA PRODUCTION D'UN NOMBRE ÉLEVÉ DE VARIANTS D'UN PETIT ENSEMBLE DE MOLÉCULES.

OR IL FAUT METTRE EN PLACE D'ABORD UN MÉCANISME SÉLECTIF POUR COMPRENDRE L'ORIGINE DE LA CHIMIE DU VIVANT.

TROP DE MOLÉCULES STRUCTURELLEMENT VOISINES AGIT COMME LE FONT LES POISONS, EN INACTIVANT LES PROCESSUS CATALYTIQUES SÉLECTIONNÉS DANS UN PREMIER TEMPS ET EN BLOQUANT L'ÉVOLUTION.

CELA IMPLIQUE :

- UNE CONCENTRATION SUFFISANTE DES MOLÉCULES EN CAUSE**
- L'EXISTENCE D'UNE LIMITE À LA VARIÉTÉ DES MOLÉCULES PRODUITES À UN INSTANT DONNÉ**
- POUR CHAQUE TYPE, UNE MESURE DE SON ADÉQUATION**
- QUE LA DESCENDANCE DES TYPES LES PLUS ADÉQUATS S'ACCROISSE AVEC LE TEMPS**
- QUE TANT QU'UN TYPE N'EST PAS SUR LE POINT DE DISPARAÎTRE IL PRODUISE DES DESCENDANTS DONT CERTAINS SONT LÉGÈREMENT MIEUX ADAPTÉS QUE LEURS PARENTS.**

UNE EXPÉRIENCE DE MILLER

MOLÉCULES À SQUELETTE CARBONÉ

RENDEMENT

<i>Glycine</i>	2,1
Acide glycolique	1,9
Sarcosine	0,25
<i>Alanine</i>	1,7
Acide lactique	1,6
N-méthylalanine	0,07
Acide alpha-aminobutyrique	0,34
Acide alpha-aminoisobutyrique	0,007
Acide alpha-hydroxybutyrique	0,34
Beta-alanine	0,76
Acide succinique	0,27
<i>Acide aspartique</i>	0,024
<i>Acide glutamique</i>	0,051
Acide iminodiacétique	0,37
Acide iminoacétique-propionique	0,13
Acide formique	4,00
Acide acétique	0,51
Acide propionique	0,66
Urée	0,034
N-méthylurée	0,051

Origine



**ARGUMENTS THAT ATTEMPT TO EXTRAPOLATE FROM
MODERN BIOCHEMISTRY BACK TO THE ORIGIN OF LIFE
ARE FUTILE**

S. BENNER

**THE METHOD USED IS THAT OF A RECONSTRUCTION OF
PRECURSOR PATHWAYS BY RETRODICTION FROM EXTANT
PATHWAYS**

G. WÄCHTERSCHÄUSER



OUTRE LA NÉCESSITÉ D'AVOIR UNE CONCENTRATION SUFFISANTE DES MOLÉCULES EN CAUSE, UN AUTRE ASPECT DU MÉTABOLISME DOIT ÊTRE PRIS EN CONSIDÉRATION : IL DOIT FAVORISER LA SYNTHÈSE DES MACROMOLÉCULES DANS L'EAU.

À 3-DIMENSIONS, L'AUGMENTATION DE L'ENTROPIE VA CONTRE LA POLYMÉRISATION, MAIS À 2-DIMENSIONS, SI UNE MOLÉCULE D'EAU EST LIBÉRÉE PENDANT LA POLYMÉRISATION, LA FORMATION DE MACROMOLÉCULES EST FAVORISÉE SYSTÉMATIQUEMENT.

Origine



**DES MOLÉCULES CHARGÉES INTERAGISSANT AVEC DES SURFACES
PEUVENT PERMETTRE LA MISE EN PLACE D'UN PROCESSUS
SÉLECTIF.**

**DE NOMBREUX AUTEURS PAR LE PASSÉ ONT PROPOSÉ DES
HYPOTHÈSES POUR PRENDRE EN COMPTE CE FAIT :**

• 1951	BERNAL	ARGILES
• 1957	GRANICK	Fe₂O₃ + FeO + S
• 1975	CAIRNS-SMITH & HARTMAN	ARGILES
• 1988	WÄCHTERS HÄUSER	FeS (PYRITE)

Origine



PREMIÈRE ÉTAPE

MÉTABOLISME DE SURFACE PERMETTANT LA CONCENTRATION LOCALE DE
MOLÉCULES À SQUELETTE CARBONÉ, LEUR MODIFICATION ET LEUR
POLYMÉRISATION

QUELLES PROPRIÉTÉS CES MOLÉCULES DOIVENT-ELLES AVOIR ?

QUELQUES RÉFÉRENCES

GRAHAM CAIRNS-SMITH

GENETIC TAKEOVER AND THE MINERAL
ORIGIN OF LIFE CAMBRIDGE UNIVERSITY PRESS 1982

ANTOINE DANCHIN

UNE AURORE DE PIERRES LE SEUIL 1990

CHRISTIAN DE DUVE

BLUEPRINT FOR A CELL PATTERSON 1990

FREEMAN DYSON

ORIGINS OF LIFE CAMBRIDGE UNIVERSITY

PRESS 1985

GUNTER WACHTHERSHAUSER

BEFORE ENZYMES AND TEMPLATES:

THEORY OF SURFACE METABOLISM. *MICROBIOL REV* 1988, 52:452-484

**AVANT DE CONSIDÉRER LE CODE GÉNÉTIQUE ET L'ADN
IL FAUT RENDRE COMPTE DES SYNTHÈSES CORRESPONDANTES
À L'INTÉRIEUR D'UN ORGANITE COMPARTIMENTÉ.**



SYNTHÈSE DES LIPIDES

SYNTHÈSE DES COENZYMES

SYNTHÈSE DES NUCLÉOTIDES

Origine



- ➔ **VIE ET CALCUL**
- ➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**
- ➔ **DE LA SOUPE À LA SURFACE DES PIERRES**
- ➔ **LA TRANSFORMATION HOMÉOTOPIQUE**
- ➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**
- ➔ **LE MONDE DES ARN**
- ➔ **UNIVERSAUX DANS LES GÉNOMES**
- ➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**
- ➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**

ANABOLISME ET CATABOLISME

LE MÉTABOLISME ACTUEL MONTRE UNE CHIMIE TRÈS DIFFÉRENTE POUR L'ANABOLISME ET LE CATABOLISME.

CE DERNIER EST SOUVENT FACILE À COMPRENDRE ALORS QUE L'ANABOLISME MET EN JEU DES GROUPEMENTS CHIMIQUES « INUTILES »

LA PLUPART DES MÉTABOLITES QUI SERVENT À LA CONSTRUCTION DES CELLULES (ANABOLISME) SONT CHARGÉS NÉGATIVEMENT, SOUVENT EN APPARENCE « INUTILEMENT »

CE QUI EST AISÉ :

**SYNTHÈSE DE QUELQUES ACIDES AMINÉS
(À CONDITION QUE L'AZOTE PUISSE ÊTRE RÉDUIT)**

**SYNTHÈSE DE MOLÉCULES À BASE D'UN SQUELETTE
DE TROIS ATOMES DE CARBONE**

Origine



CE QUI EST DIFFICILE :

SYNTHÈSE DES ACIDES AMINÉS BASIQUES

SYNTHÈSE DES COENZYMES

SYNTHÈSE DES LIPIDES

**SYNTHÈSE DES NUCLÉOTIDES (PARTICULIÈREMENT
DES PYRIMIDINES)**

Origine



CHEMINS UTILISANT DES RÉSIDUS PHOSPHATE SANS “RAISON” :

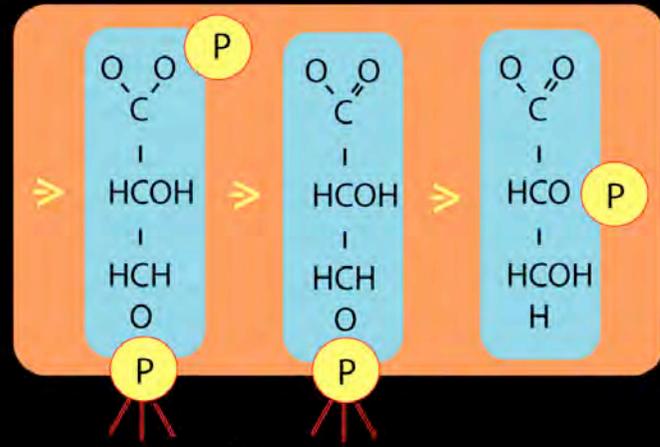
**SÉRINE (SÉRINE~PHOSPHATE)
PYRIDOXAL PHOSPHATE
ISOPRÉNYL ~ PYROPHOSPHATE
4-PHOSPHOPANTÉTHÉINE
THIAMINE ~ PYROPHOSPHATE**

CONTENANT DES NUCLÉOTIDES :

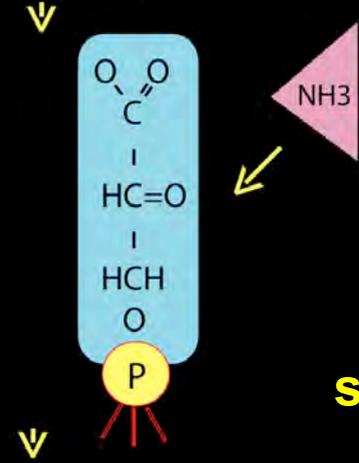
**CYTIDYLATE (LIPIDES)
ADENYLATE, GUANYLATE, URIDYLATE (SUCRES)
COENZYMES (NAD, CoA...)**



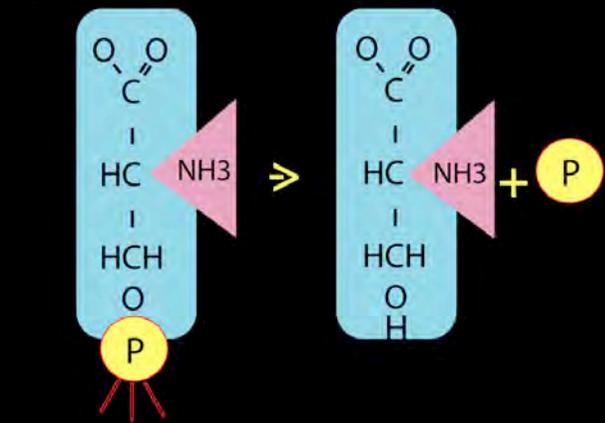
GLYCOLYSE



SYNTHÈSE DE LA SÉRINE



sérine



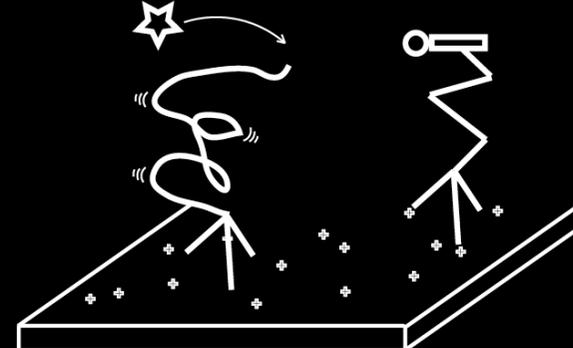
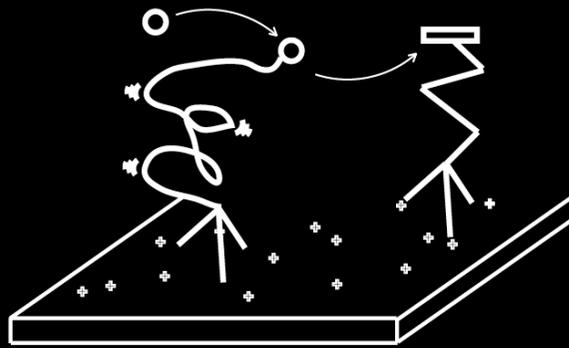
gâchis apparent !

LES MOLÉCULES À TROIS ATOMES DE CARBONE ET LES ACIDES AMINÉS SONT TRÈS FACILES À OBTENIR

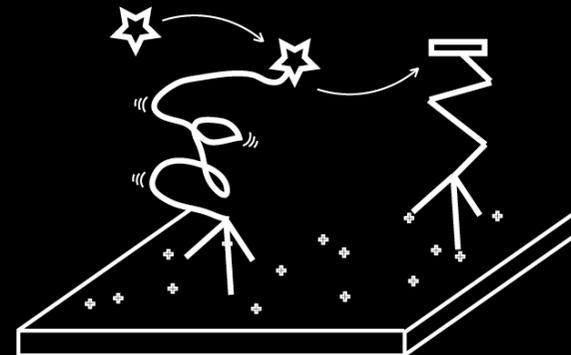
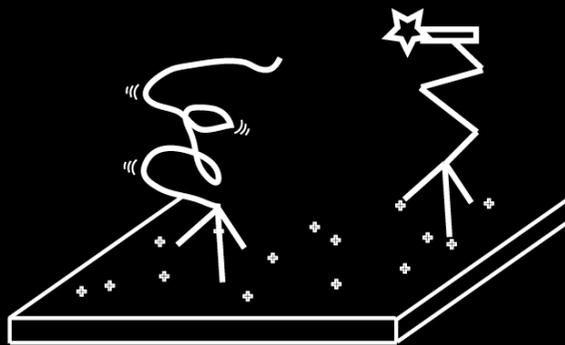
ON PEUT IMAGINER LA SYNTHÈSE SYSTÉMATIQUE DE PEPTIDES ET D'ISOPEPTIDES

LE MÉTABOLISME ACTUEL DE LA SYNTHÈSE DES NUCLÉOTIDES ET DES COENZYMES UTILISE DES ACIDES AMINÉS. IL S'AGIT DONC PEUT-ÊTRE DE LA TRACE DE CE PASSÉ PRIMITIF

SOUVENT, UNE « VARIATION SUR UN THÈME » OU UN SOCLE COMMUN RECOIT DES ACCEPTEURS DIVERS : IL S'AGIT LÀ DE « TRANSFORMATION HOMÉOTOPIQUE »



LA TRANSFORMATION HOMÉOTOPIQUE



Origine



PREMIÈRE ÉTAPE

MÉTABOLISME DE SURFACE DE MOLÉCULES À SQUELETTE CARBONÉ ET PORTEUSES DE GROUPEMENTS CHARGÉS NÉGATIVEMENT

CES MOLÉCULES COMPRENENT UN CERTAIN NOMBRE D'ACIDES AMINÉS « SIMPLÉS » (GLYCINE, ALANINE, SÉRINE, CYSTÉINE, ASPARTATE, GLUTAMATE...) ET LE CŒUR DE CE QUI CONSTITUE AUJOURD'HUI LA GLYCOLYSE

- ➔ **VIE ET CALCUL**
- ➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**
- ➔ **DE LA SOUPE À LA SURFACE DES PIERRES**
- ➔ **LA TRANSFORMATION HOMÉOTOPIQUE**
- ➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**
- ➔ **LE MONDE DES ARN**
- ➔ **UNIVERSAUX DANS LES GÉNOMES**
- ➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**
- ➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**

ACIDES AMINÉS DANS LA SYNTHÈSE DES NUCLÉOTIDES ET DES COENZYMES

PURINES Q G (S) Q D (S)

PYRIMIDINES D

GLUTATHION E C G (ISO)PEPTIDE

4-PHOSPHOPANTÉTHÉINE V A C (ISO)PEPTIDE

NICOTINAMIDE D (Q)

PYRIDOXAMINE S (Q)

PYROLLOQUINOLONE QUINONE E Y (PEPTIDE)

ETC...

Origine



**GRANICK (1957), YCAS (1974) ET JENSEN (1976) ONT PROPOSÉ
QUE LA SPÉCIFICITÉ DES ENZYMES A ÉVOLUÉ À PARTIR DU
RECRUTEMENT DE PROTÉINES PREEXISTANTES QUI CATALYSENT
DES RÉACTIONS SEMBLABLES**

CELA A ÉTÉ SOUVENT CONSTATÉ

**pabB trpE
pabA trpG guaA**

**GONCHAROFF & NICHOLS 1984
KAPLAN & NICHOLS 1983
ZALKIN et al. 1985**

**thrC dsdA ilvA trpB
metB metC
cysK trpB**

**PARSOT 1986, 1987
PARSOT et al. 1988
LÉVY & DANCHIN 1988**

**pdxB serA
pyrH thrA lysC gltK carK**

**SCHOENLEIN et al. 1989
SERINA et al. 1995**

Origine



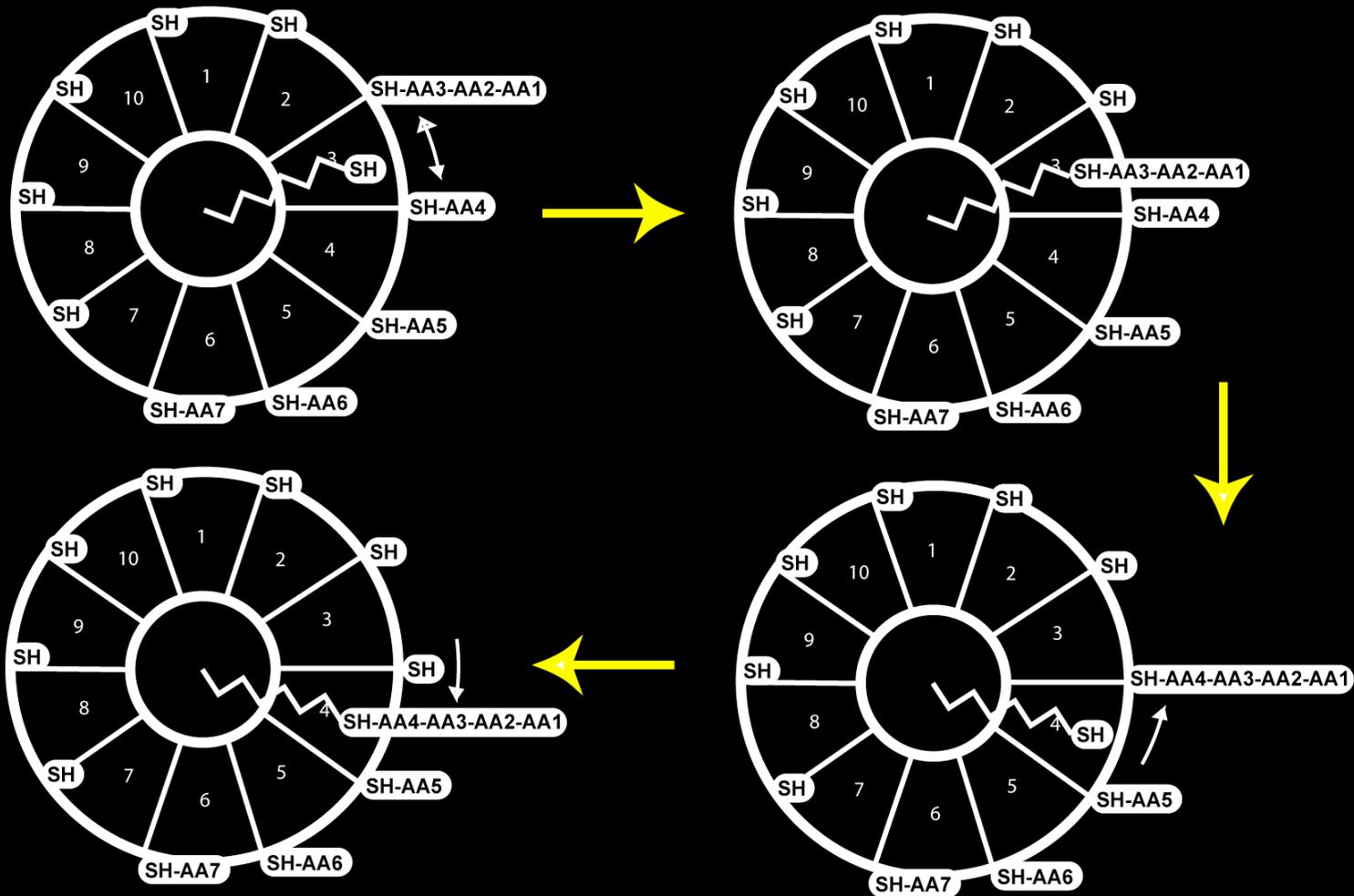
LA SYNTHÈSE « NON RIBOSOMIQUE » DES ANTIBIOTIQUES OU LA SYNTHÈSE DES ACIDES GRAS EST UNE HOMÉOTOPIE RÉCURSIVE.

LES GÈNES DE CES DEUX TYPES DE SYNTHÈSES SONT PHYLOGÉNÉTIQUEMENT APPARENTÉS.

LA FORMATION DES PEPTIDES EST CATALYSÉE PAR UN PEPTIDE PHOSPHORYLÉ (PHOSPHOPANTÉTHÉINE)

IL EST DONC PLAUSIBLE DE CONSIDÉRER DES PEPTIDES OU DES ANALOGUES DANS LA SYNTHÈSE PRÉBIOTIQUE DES COENZYMES ET DES NUCLÉOTIDES.

SYNTHÈSE PEPTIDIQUE NON RIBOSOMIQUE



Origine



IL EN RÉSULTE LA SYNTHÈSE DE MEMBRANES LIPIDIQUES, DE COENZYMES ET DE NUCLÉOTIDES.

LA POLYMÉRISATION DE CES DERNIERS, EN PRÉSENCE DE PEPTIDES, CONDUIT À DES ARN (ENCHAÎNEMENT 5'-3') ET AU MONDE ARN OU L'ANCÊTRE DE L'ARN DE TRANSFERT SE SUBSTITUE AUX SURFACES SOLIDES, EN PARTICULIER VIA LA PRÉSENCE D'ATOMES DE SOUFRE.

LA DÉCOUVERTE DU RÔLE POSSIBLE DE MATRICE PLUTÔT QUE DE SUBSTRAT VA DONNER LIEU À L'INVENTION DU CODE GÉNÉTIQUE.

DANS TOUS LES CAS SE POSE LA QUESTION DE L'ORIGINE DES NUCLÉOTIDES, CAR CES MOLÉCULES SONT TRÈS INSTABLES, CE QUI SUPPOSE UNE PRODUCTION CONTINUE.

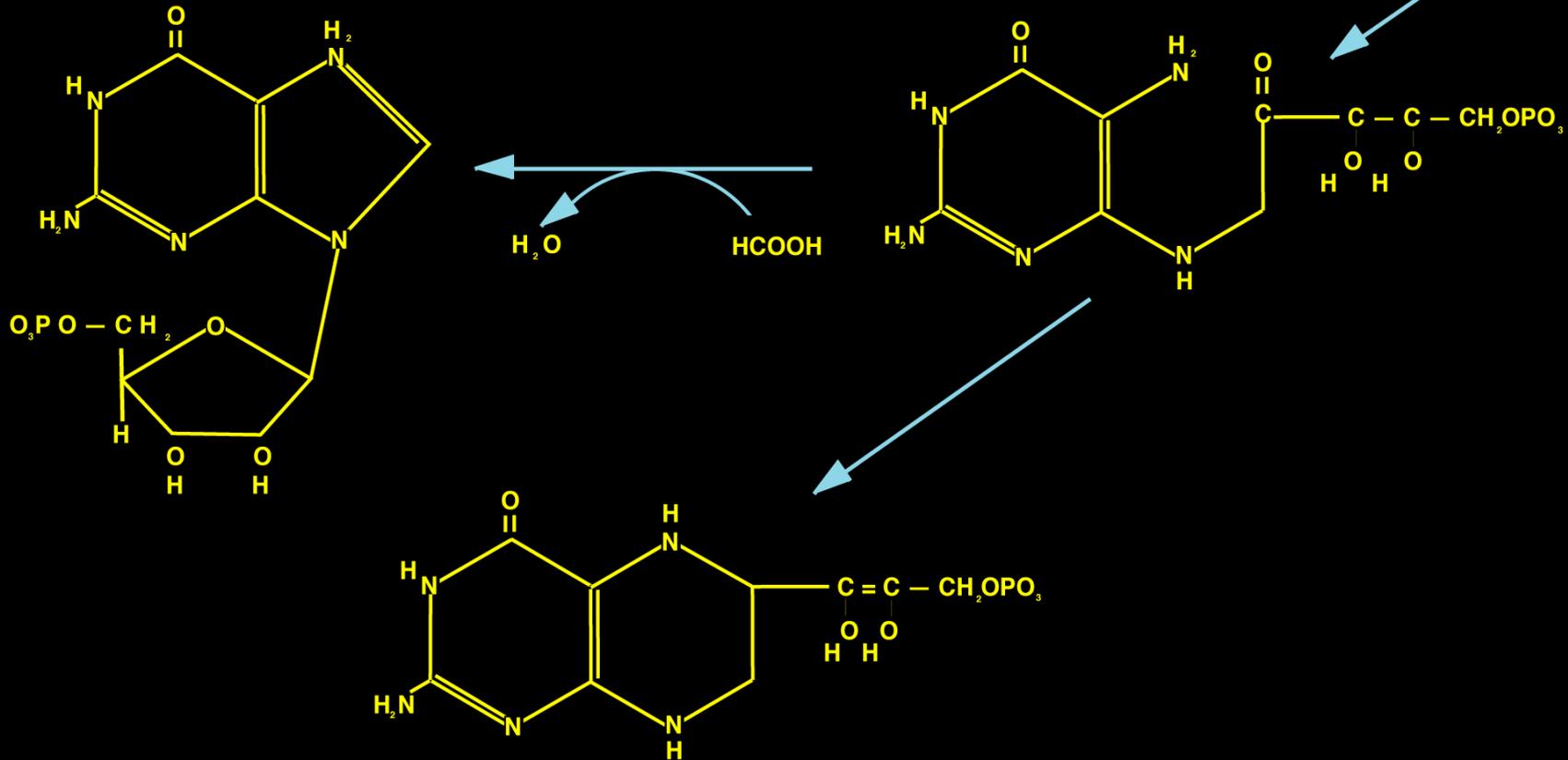
UN COENZYME SOUFRÉ (QUI SE LIE DONC FACILEMENT À DES SURFACES FER-SOUFRE), LA MOLYBOPTÉRINE, DONNE UNE IDÉE D'UN SCÉNARIO D'ORIGINE, EN INVERSANT LE MÉTABOLISME ACTUEL QUI À PARTIR DU GTP DONNE LES PTÉRIDINES.

LES PTÉRIDINES ÉTAIENT OBTENUES AU COURS D'UN PROCESSUS PRIMITIF DE FIXATION DE L'AZOTE, ET LA CYCLISATION SOUS FORME DE GMP ÉTAIT UNE « FUIITE » ACCIDENTELLE DE CE PROCESSUS.

Molybdoptérine



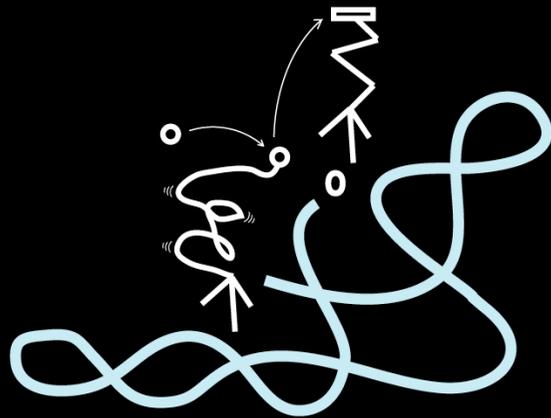
FIXATION DE L'AZOTE ?



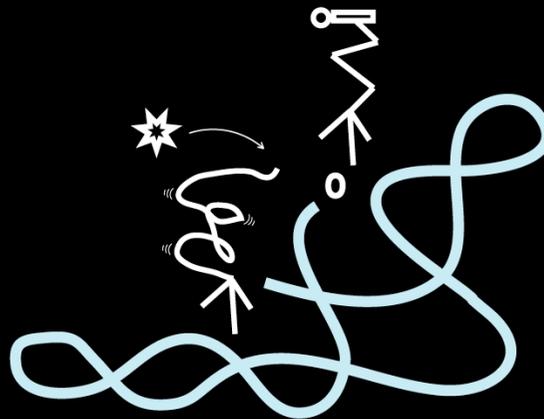
Origine



- ➔ **VIE ET CALCUL**
- ➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**
- ➔ **DE LA SOUPE À LA SURFACE DES PIERRES**
- ➔ **LA TRANSFORMATION HOMÉOTOPIQUE**
- ➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**
- ➔ **LE MONDE DES ARN**
- ➔ **UNIVERSAUX DANS LES GÉNOMES**
- ➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**
- ➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**



**L'ARN DE TRANSFERT SE SUBSTITUE
AUX SURFACES**



tRNA HORS TRADUCTION

FORMATION D'UN LIEN PEPTIDIQUE

Gly~ARNt_{Gly} muréine

Phe~ARNt_{Phe} N-terminal

Leu~ARNt_{Leu} N-terminal

Arg~ARNt_{Arg} Ubiquitine

HOMÉOTOPIE

Met~ARNt_{FMet} → FMet~ARNt_{FMet}

Glu~ARNt_{Gln} → GlnARNt_{Gln}

Ser~ARNt_{SeCys} → SeCys~ARNt_{SeCys}

DIVERS

Lys~ARNt_{Lys} Lipides

Glu~ARNt_{Glu} Aminolévulinate

AUTRES MODIFICATIONS

Modification des bases

Origine



ORIGINE DES ARNt SYNTHÉTASES

SI LA TRANSFORMATION HOMÉOTOPIQUE EST À LA BASE DU MÉTABOLISME SUR LES MOLÉCULES D'ARNt, LES ARNt SYNTHÉTASES DOIVENT PROVENIR DE LA SYNTHÈSE DES ACIDES AMINÉS :

GATTI & TZAGOLOFF ONT TROUVÉ QUE L'ASPARTYL-ARNt SYNTHÉTASE EST VOISINE DE L'ASPARAGINE SYNTHÉTASE, ET DI GIULIO (1993) A MONTRÉ QUE LA GLUTAMINE ARNt SYNTHÉTASE EST VOISINE DES GLUTAMINE AMIDO-TRANSFÉRASES

Origine



LE MONDE DES ARN

LES ARN PEUVENT AVOIR UNE ACTIVITÉ ENZYMATIQUE AUTONOME, ILS PEUVENT AUSSI LIER SPÉCIFIQUEMENT DES PETITES MOLÉCULES (APTAMÈRES), ILS PEUVENT ENFIN JOUER LE RÔLE DE MATRICE POUR L'ADAPTATION D'ARN COMPLÉMENTAIRES

LES « INTRONS » ILLUSTRENT, AUJOURD'HUI ENCORE, LA CAPACITÉ DES ARN À CATALYSER LEUR PROPRE EXCISION, AVEC L'ÉPISSAGE DE FRAGMENTS ADJACENTS

L'INVENTION DE L'ADN

LES ARN SONT TRÈS INSTABLES, EN RAISON DE LA NATURE FRAGILE DU RIBOSE.

LA POSITION 2'OH DU RIBOSE REND DIFFICILE LA POLYMÉRISATION CORRECTE (5'-3') DES ARN (EN FAIT IL FAUT L'ÉCRANTAGE DE PEPTIDES POUR QUE CE SOIT POSSIBLE).

LA RÉDUCTION DE CETTE POSITION, SOUS LA FORME DE DÉSOXYRIBONUCLÉOTIDES, A PERMIS DE FIGER LA MÉMOIRE HÉRÉDITAIRE. LES ENZYMES CORRESPONDANTES (3 CLASSES, D'UNE MÊME ORIGINE) SUPPOSENT L'ABSENCE D'OXYGÈNE ET LA PRÉSENCE INITIALE D'UN NOYAU FER-SOUFRE.

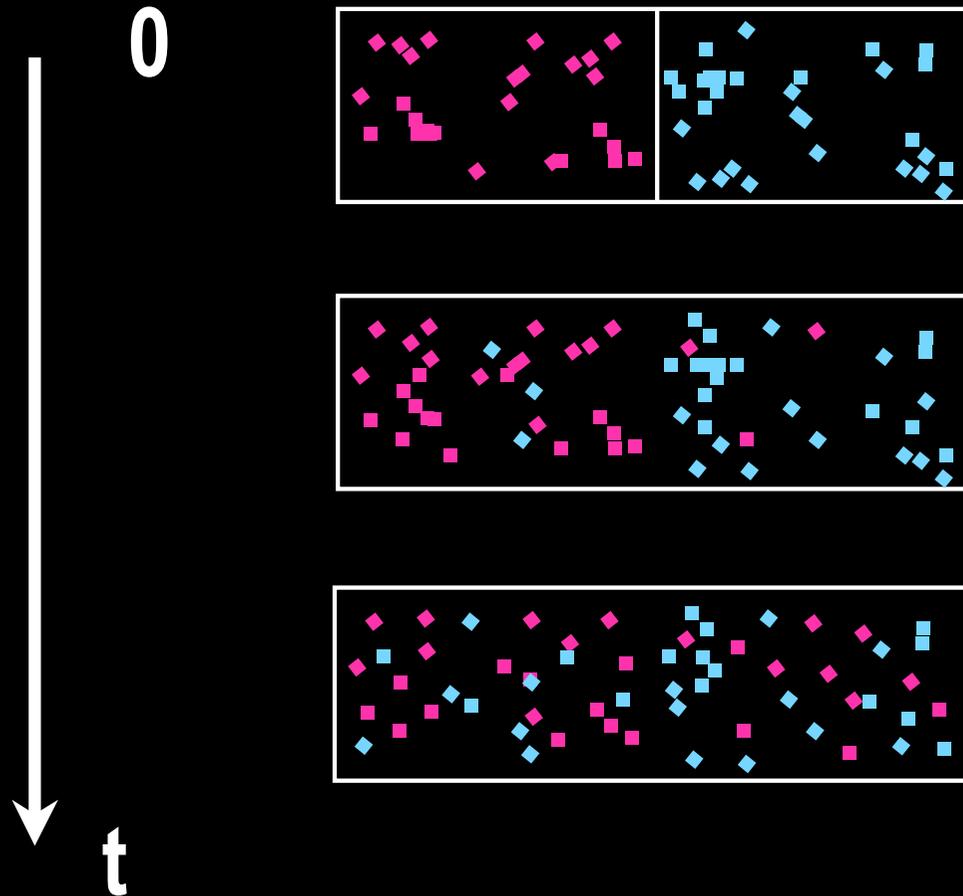
UN SCENARIO DE L'ORIGINE DE LA VIE

- LA SURFACE DE SOLIDES CHARGÉS COMME LA PYRITE DE FER (FE-S) PERMET LA SÉLECTION ET UNE COMPARTIMENTATION PRIMITIVE DE MOLÉCULES CHARGÉES ; LA POLYMÉRISATION AVEC ÉLIMINATION D'EAU EST FAVORISÉE PAR EFFET ENTROPIQUE
- UNE FOIS COMPARTIMENTÉ, LE MÉTABOLISME CRÉE DES SUBSTITUTS DES SURFACES (LE MONDE ARN)
- LES ACIDES NUCLÉIQUES SONT STABILISÉS PAR L'INVENTION DES DÉSOXYRIBONUCLÉOTIDES, AU MOMENT OU S'ÉTABLIT L'ENSEMBLE DES RÈGLES DES TRANSFERTS D'INFORMATION, D'ABORD AU SEIN D'UN MONDE ARN OU DES VÉSICULES PORTEUSES DES ANCÊTRES DES GÈNES SE SCINDENT ET FUSIONNENT TOUR À TOUR, AVANT QUE SE FORMENT LES PREMIERS GÉNOMES

A Danchin Une Aurore de Pierres. Aux origines de la vie, Le Seuil 1990

- ➔ **VIE ET CALCUL**
- ➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**
- ➔ **DE LA SOUPE À LA SURFACE DES PIERRES**
- ➔ **LA TRANSFORMATION HOMÉOTOPIQUE**
- ➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**
- ➔ **LE MONDE DES ARN**
- ➔ **UNIVERSAUX DANS LES GÉNOMES**
- ➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**
- ➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**

UNE IDÉE REÇUE (FAUSSE) ASSIMILE L'ENTROPIE AU DÉSORDRE

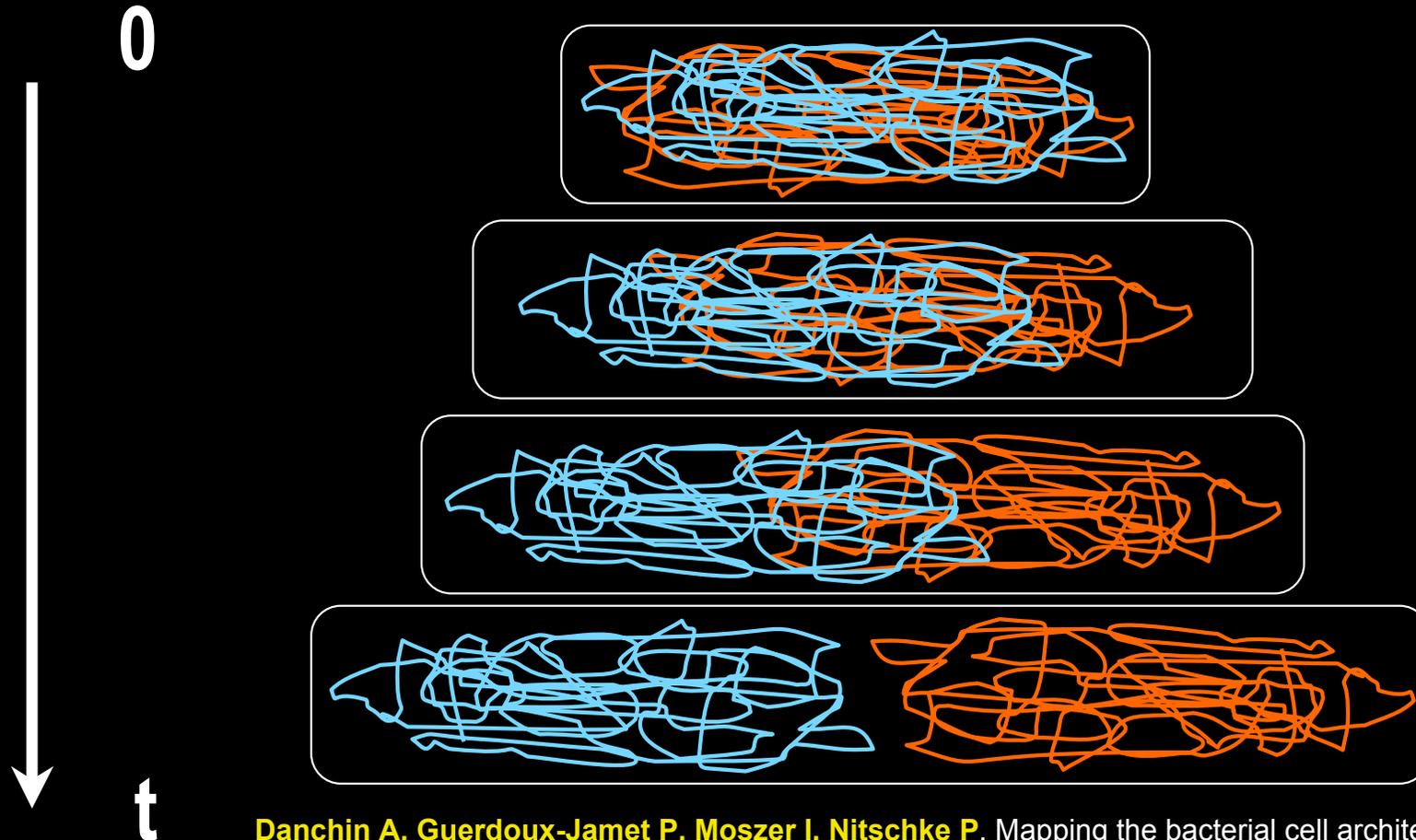


$$S = k \log \Omega$$



Benjamin Crowell, licensed under the Creative Commons Attribution-ShareAlike license

L'AUGMENTATION D'ENTROPIE SUFFIT À SÉPARER LES 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

CONCEPTS ET « RUSTINES »

Les processus formant la vie s'analysent conceptuellement. Ils doivent cependant être réalisés par des objets concrets, qui ont des propriétés idiosyncratiques. La séquence physique de l'ADN ne peut pas être une double hélice régulière, simplement en raison de la nature chimique des nucléotides ; elle est tordue, surenroulée et son axe est gauche. Elle doit aussi être reconnue par des objets spécifiques. Comment ces contraintes opposées sont-elles réalisées simultanément ?

MODÉLISATION RÉCURSIVE

→ Modèle réaliste 1 \Leftrightarrow Séquence réelle

Prédiction 1

→ Modèle réaliste 2 \Leftrightarrow Séquence réelle

Prédiction 2

→ Modèle réaliste 3 \Leftrightarrow Séquence réelle

Prédiction 3

.....

CONTRAINTES DE LA SÉQUENCE

L'évolution optimise la réplication, alors que l'ADN doit aussi coder les gènes

Cela est révélé par :

→ Une période 3, signature de la succession des codons dans les gènes (contrainte par la règle du code génétique)

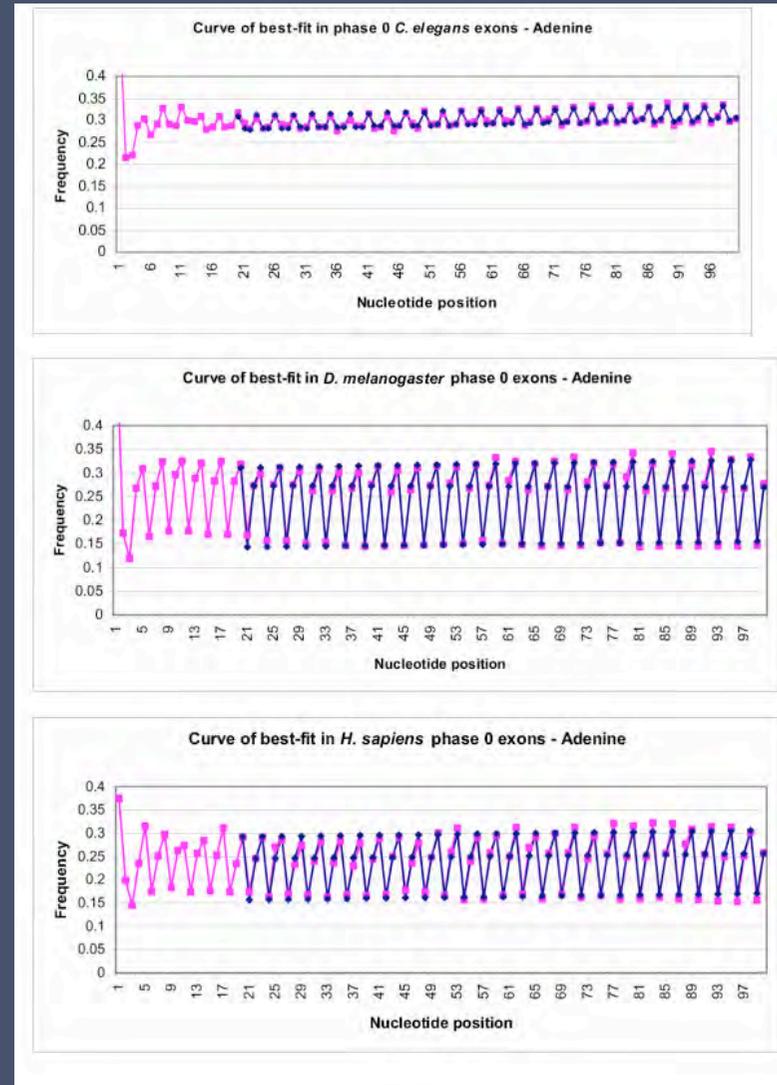
→ A période 10-11.5 de fonction inconnue ...

PÉRIODICITÉ DANS LES GÉNOMES

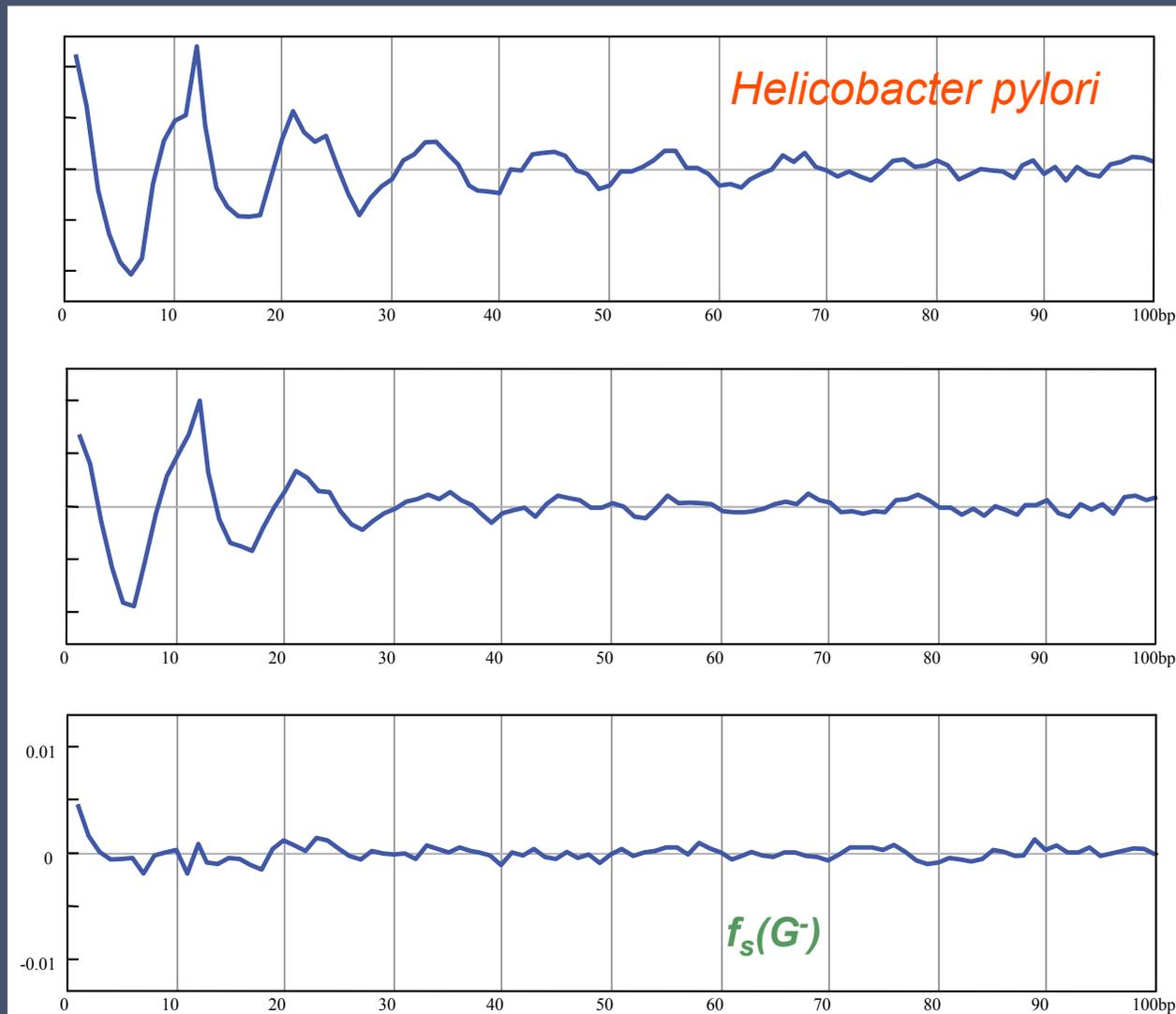
ON OBSERVE UNE LONGUE
CORRÉLATION ENTRE LES
PAIRES DE BASES
RÉPARTIES SELON UNE
PÉRIODE DE TROIS

SI L'ON DÉCONVOLUE
CETTE PÉRIODE, DUE AUX
CONTRAINTES DU CODE
GÉNÉTIQUE (MAIS MAL
IDENTIFIÉES DE FAÇON
EXPLICITE) IL RESTE UNE
PÉRIODE FLOUE DE 10 À
11,5 PAIRES DE BASES

Eskesen et coll. [BMC Molecular Biology](#) Volume 5, 12, 2004



UN CARACTÈRE UNIVERSEL DU TEXTE DU PROGRAMME : LA PÉRIODE 10-11,5



réel

modèle

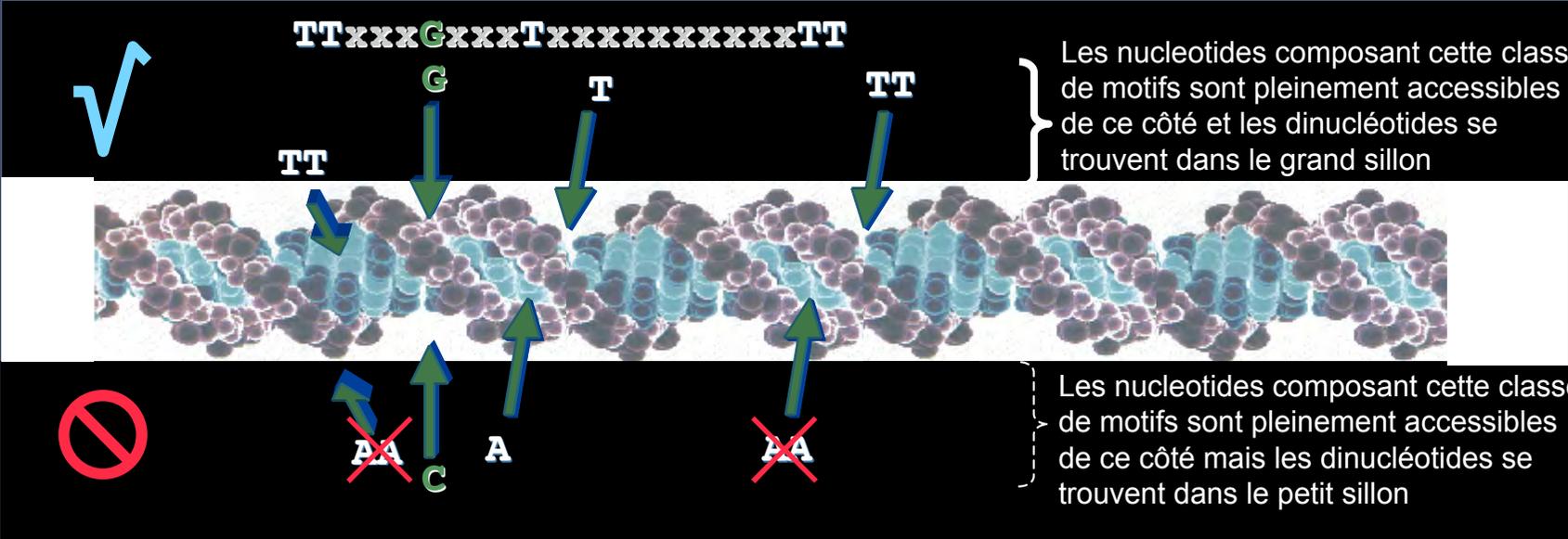
différence

Génétique des Génomes Bactériens

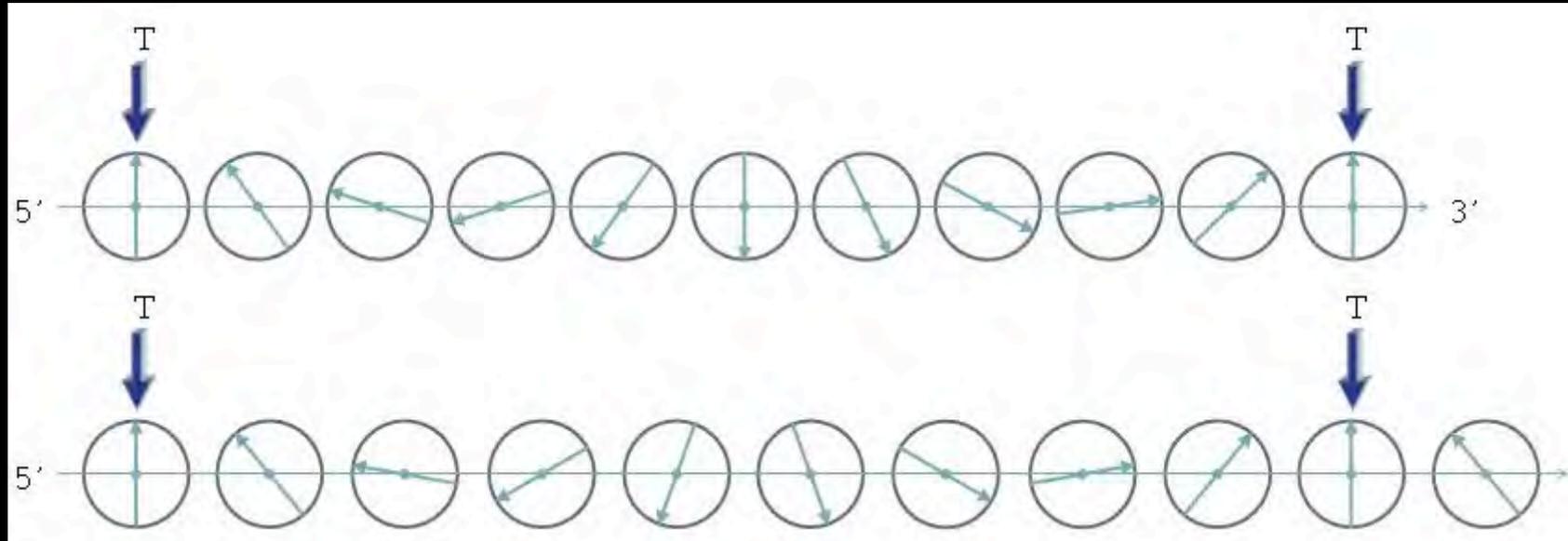
<http://www.pasteur.fr/recherche/unites/REG/>

MOTIFS FLEXIBLES DE TYPE A

- 1- $xAx\leftrightarrow x\leftrightarrow T\leftrightarrow x\leftrightarrow x\leftrightarrow Ax\leftrightarrow x\leftrightarrow TT\leftrightarrow x\leftrightarrow x\leftrightarrow Ax\leftrightarrow x\leftrightarrow T\leftrightarrow x\leftrightarrow x\leftrightarrow Ax\leftrightarrow x$: Tous règnes
- 2- $x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow G\leftrightarrow x\leftrightarrow x\leftrightarrow TT\leftrightarrow x\leftrightarrow x\leftrightarrow C\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow T\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x$: Proteobacteria
- 4- $x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow T\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow AG\leftrightarrow x\leftrightarrow x\leftrightarrow TT\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow T\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x\leftrightarrow x$: Archaea
- 5'-xxx-10xxxxxxxxxx0xxxxxxxx10xxxxxxxxbp-3'



UNE RÈGLE UNIVERSELLE : LES MOTIFS FLEXIBLES DE TYPE A

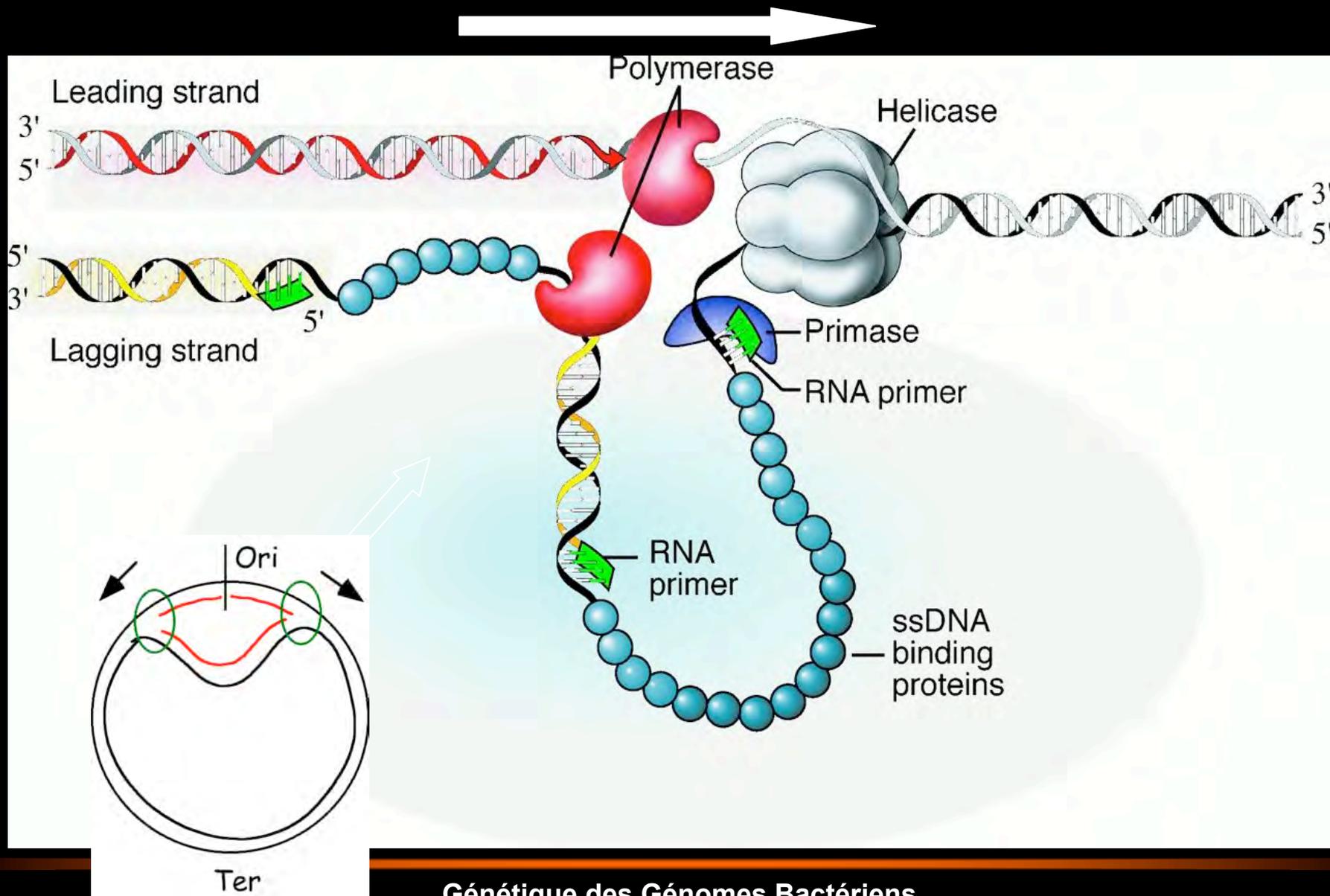


La flexibilité de ces motifs permet à l'ADN de prendre en compte les supertours et les courbures locales

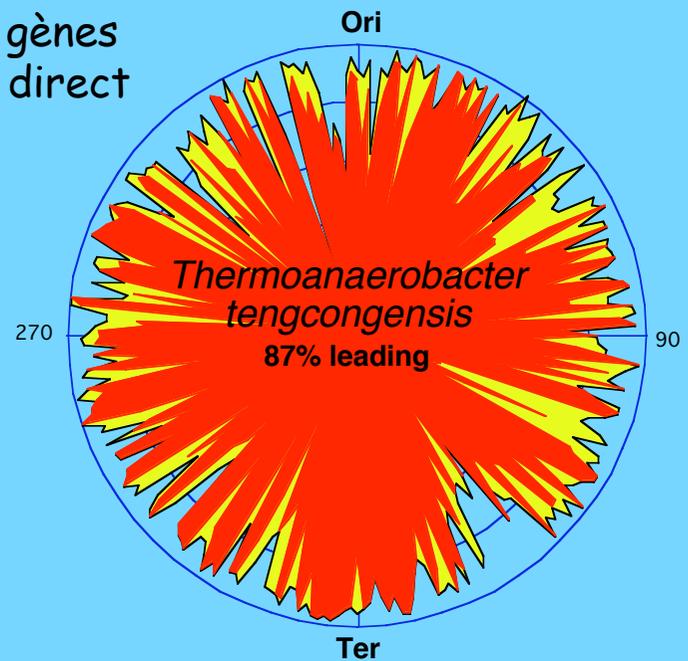
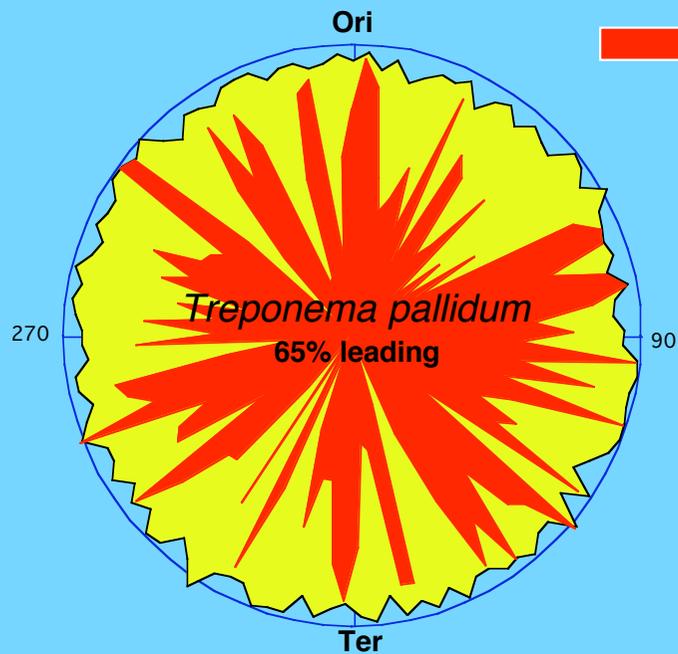
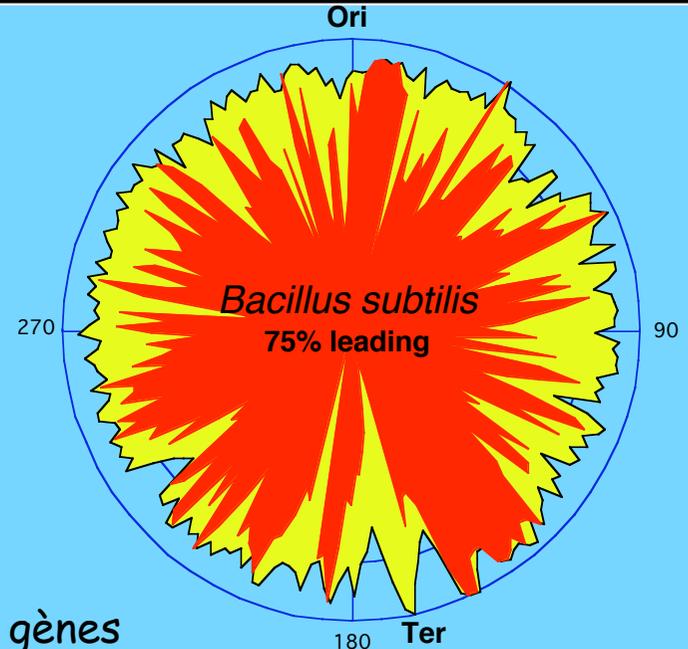
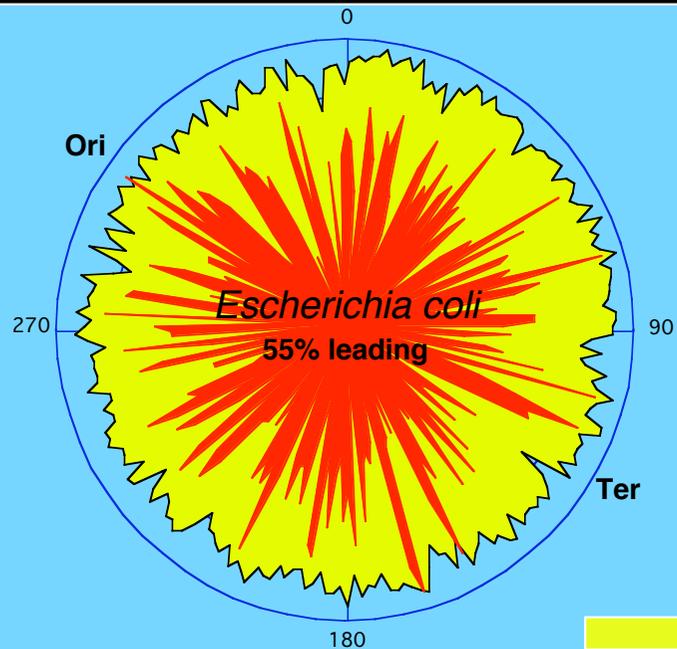
Larsabal E, Danchin A.

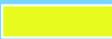
Genomes are covered with ubiquitous 11 bp periodic patterns, the "class A flexible patterns »
BMC Bioinformatics. 2005 6:206

EXISTE-T-IL DES BIAIS DUS À LA RÉPLICATION ?



Les gènes ont tendance à préférer le brin précoce de la réplication chez les Bactéries. On observe cependant une très grande variation, qui dépend de l'organisme : les bactéries à coloration de Gram positive et dont l'ADN est riche en A+T ont un biais particulièrement important



 densité des gènes
 densité des gènes dans le brin direct (précoce)

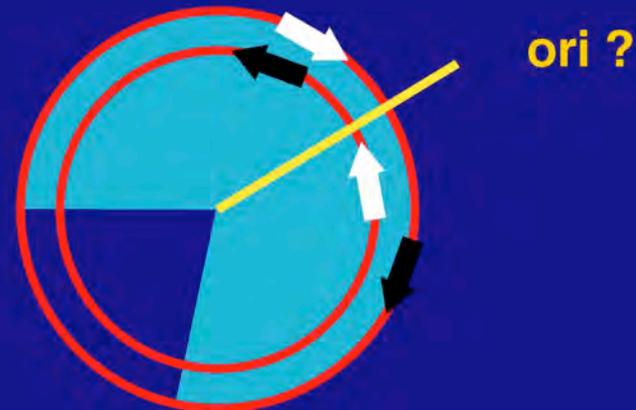
TÔT OU TARD...

Est-il possible de voir s'il y a une différence dans la composition en nucléotides, entre le brin précoce et le brin retardé ? Est-ce que cela influence le biais d'usage du code ? Est-ce que cela a une influence sur la composition en acides aminés des protéines ?

TÔT OU TARD ...

Prenant une origine arbitraire pour la réplication et une propriété du brin (composition en bases, en codons, composition en acides aminés de la protéine codée...) l'analyse discriminante linéaire permet de découvrir s'il y a une origine, et s'il y a un biais entre les brins

REPLICATION BIASES IN BACTERIA

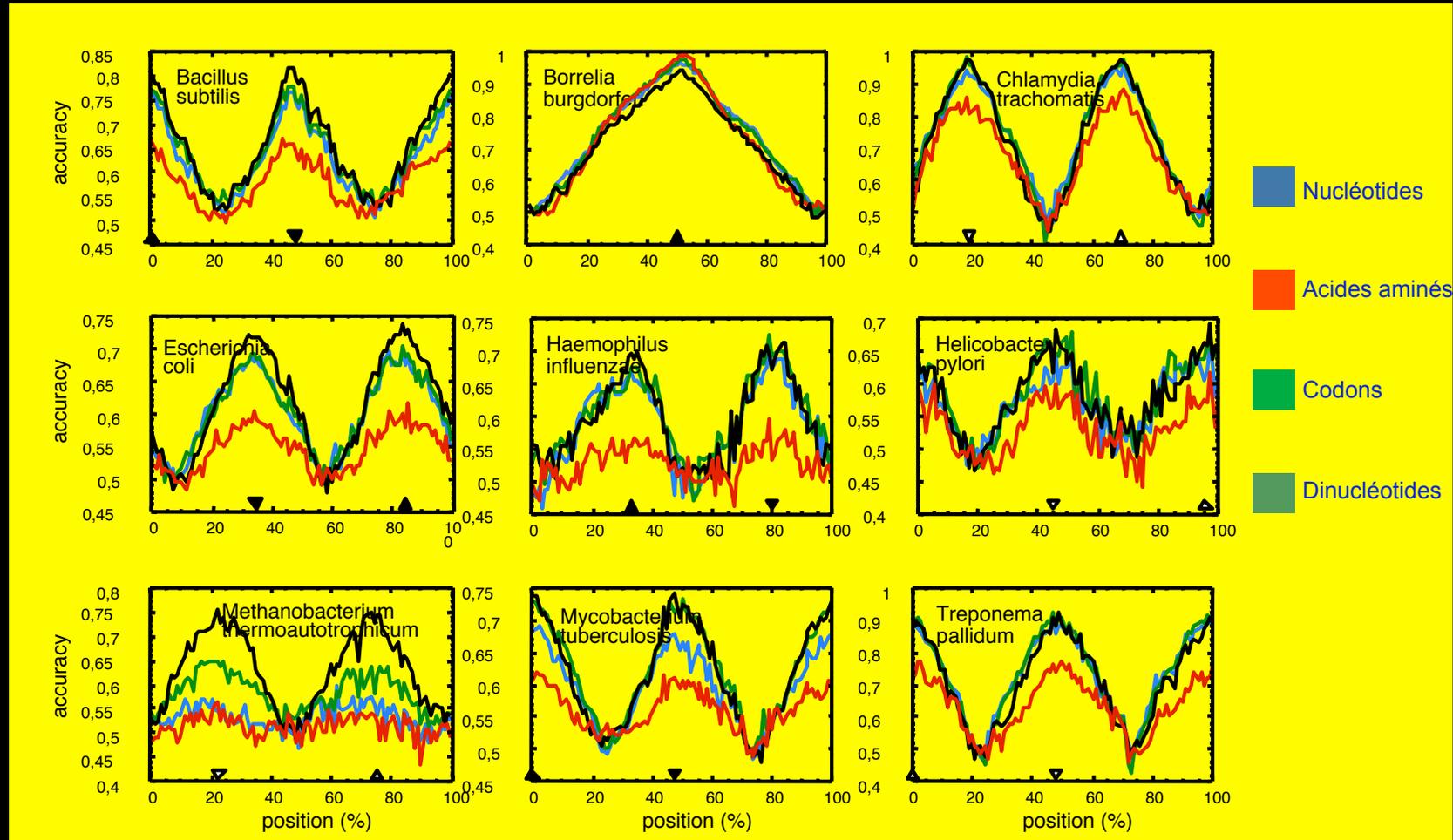


Genomes in silico



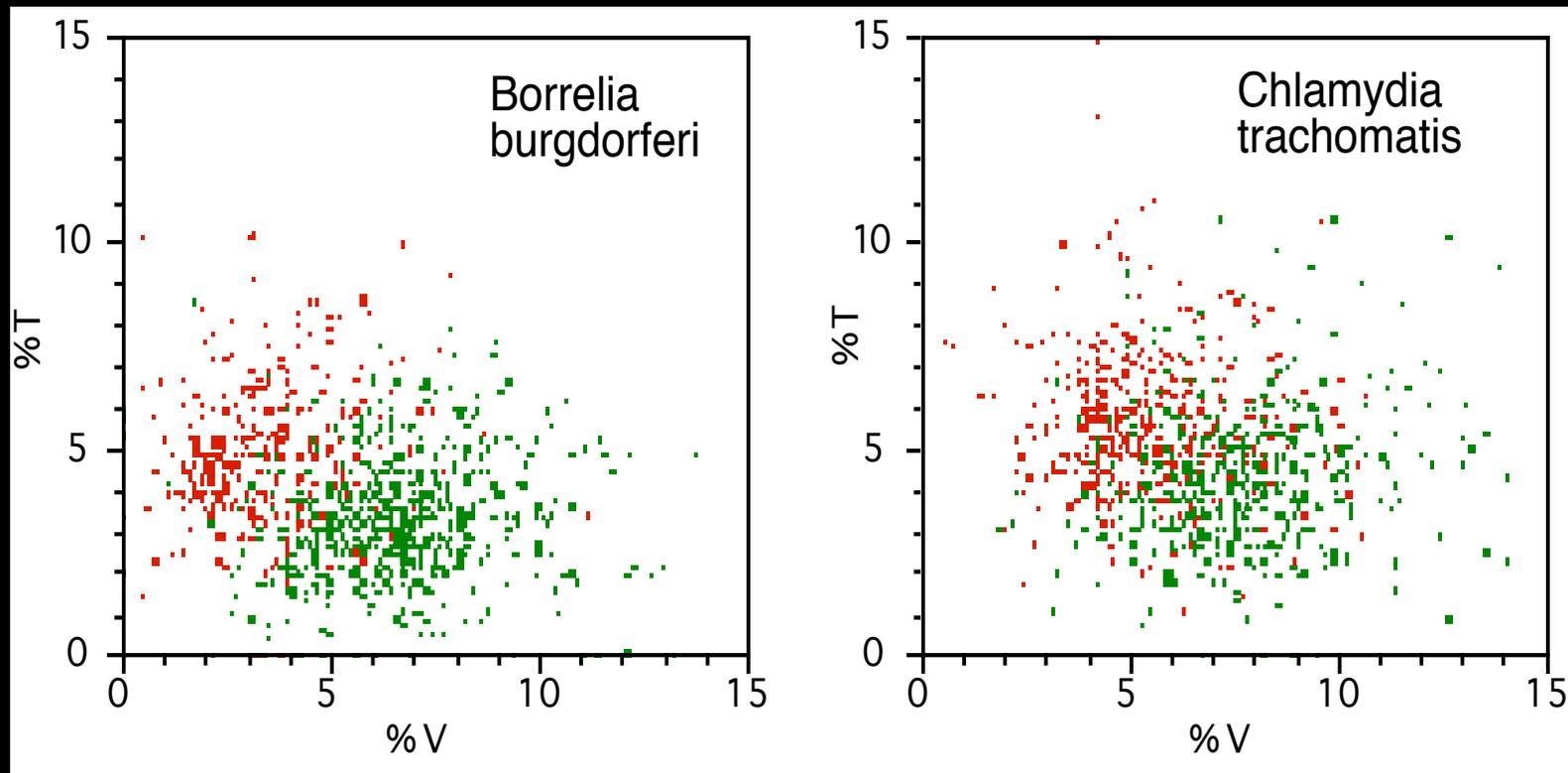
E. Rocha, A. Danchin & A. Viari Universal replication biases in bacteria. *Mol. Microbiol.* (1999) 32: 11-16

C'EST LA QUESTION...

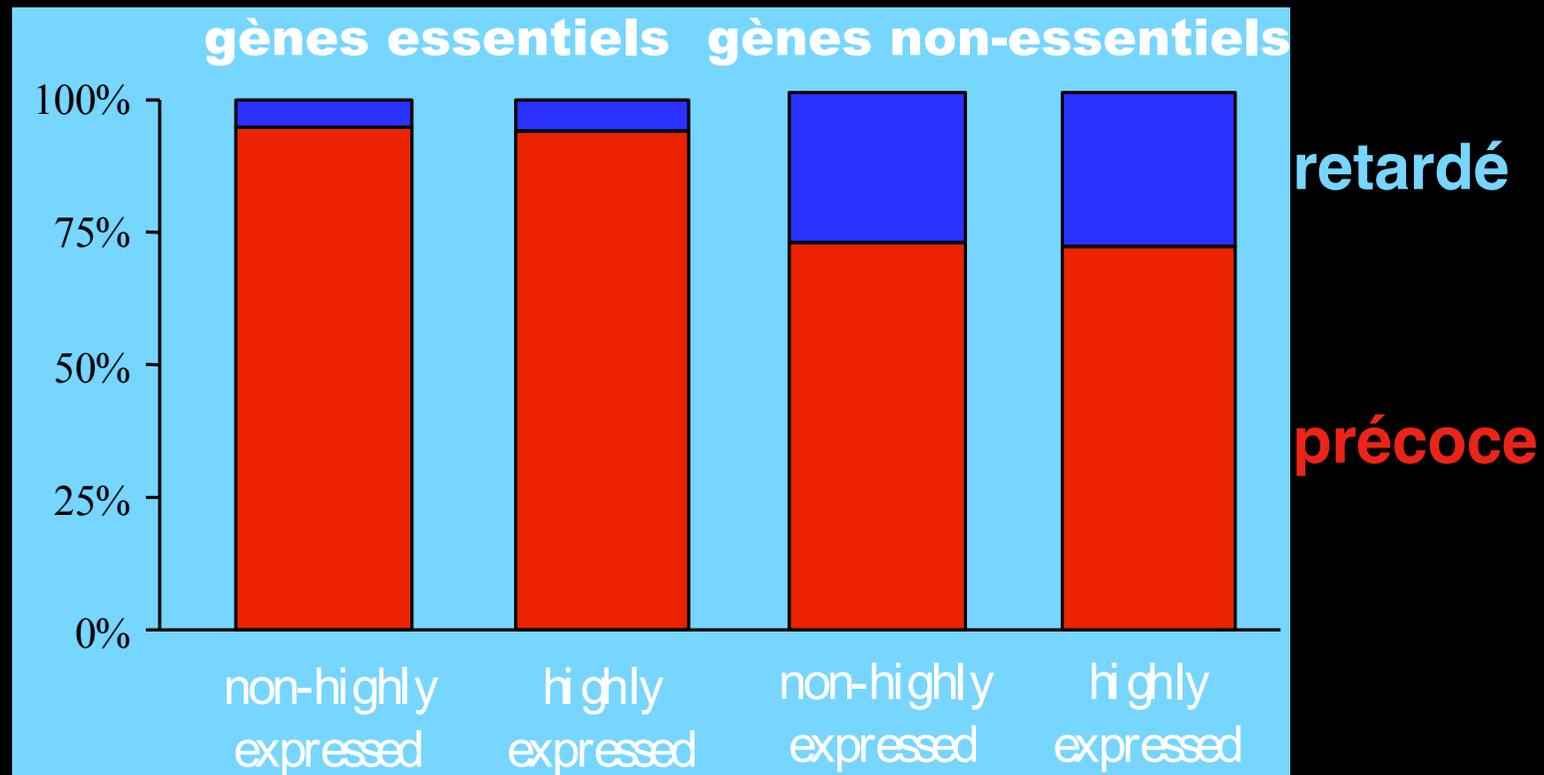


VISIBLE DANS LES PROTÉINES...

GT dans le brin précoce, CA dans le brin retardé...



LES GÈNES ESSENTIELS SONT DANS LE BRIN PRÉCOCE



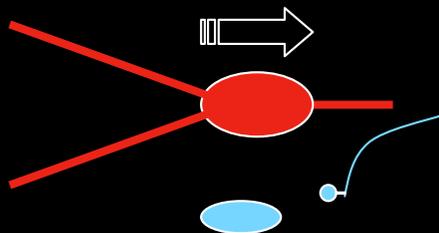
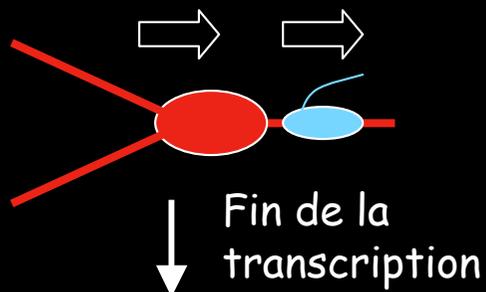
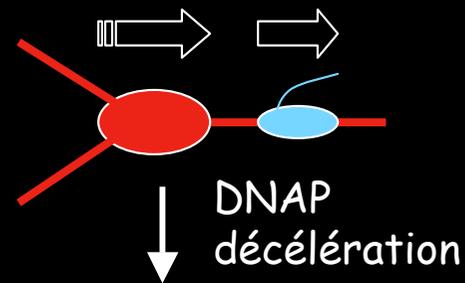
Rocha EP, Danchin A.

Essentiality, not expressiveness, drives gene-strand bias in bacteria

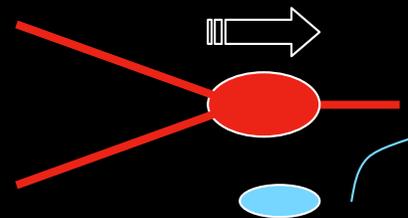
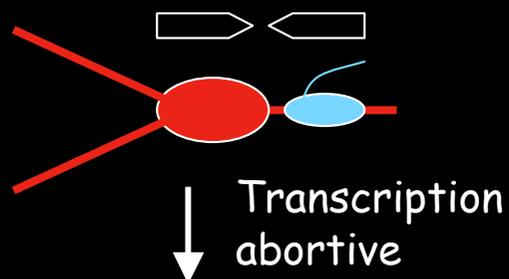
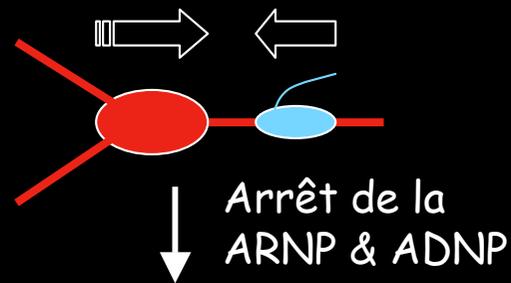
Nature Genetics. 2003 34:377-378.

COLLISION DES POLYMÉRASES

Co-orientées



Frontale



Conséquences:

1. Ralentissement de la réplication
2. Perte des transcrits

Conséquences:

1. Transcrits interrompus
2. Protéines essentielles tronquées

- ➔ **VIE ET CALCUL**
- ➔ **LE PLAN DE LA CELLULE EST DANS LE CHROMOSOME**
- ➔ **DE LA SOUPE À LA SURFACE DES PIERRES**
- ➔ **LA TRANSFORMATION HOMÉOTOPIQUE**
- ➔ **DES ACIDES AMINÉS AUX NUCLÉOTIDES**
- ➔ **LE MONDE DES ARN**
- ➔ **UNIVERSAUX DANS LES GÉNOMES**
- ➔ **LE CŒUR DU GÉNOME : CE QUI PERSISTE**
- ➔ **EXPLORATION DE L'ENVIRONNEMENT : LE CÉNOME**

UNE CONTRAINTE UNIVERSELLE, L'ÉVOLUTION

VARIATION / SÉLECTION / AMPLIFICATION
↳ STABILISATION ↻

ÉVOLUTION



CRÉE

FONCTION



CAPTURE (RECRUTE)

STRUCTURE



CODE

SÉQUENCE

LE LIVRE « PRESSE-PAPIERS » cf A DANCHIN La Barque de Delphes, Odile Jacob 1998

ANALYSES MULTIFACTORIELLES

A l'opposé de la génétique habituelle, la génomique analyse de grandes collections de gènes et de produits de gènes

Les analyses multivariées cherchent à extraire l'information en réduisant le plus possible les descripteurs des objets étudiés

L'Analyse en Composantes Principales qui utilise la valeur centrée réduite des caractères étudiés est une méthode de base

L'Analyse Factorielle des Correspondances est de la même famille, mais utilise la mesure du χ^2 comme distance. Cela permet non seulement d'étudier des objets dont les descripteurs sont hétérogènes, mais aussi de travailler sur l'espace des descripteurs et sur l'espace des objets simultanément

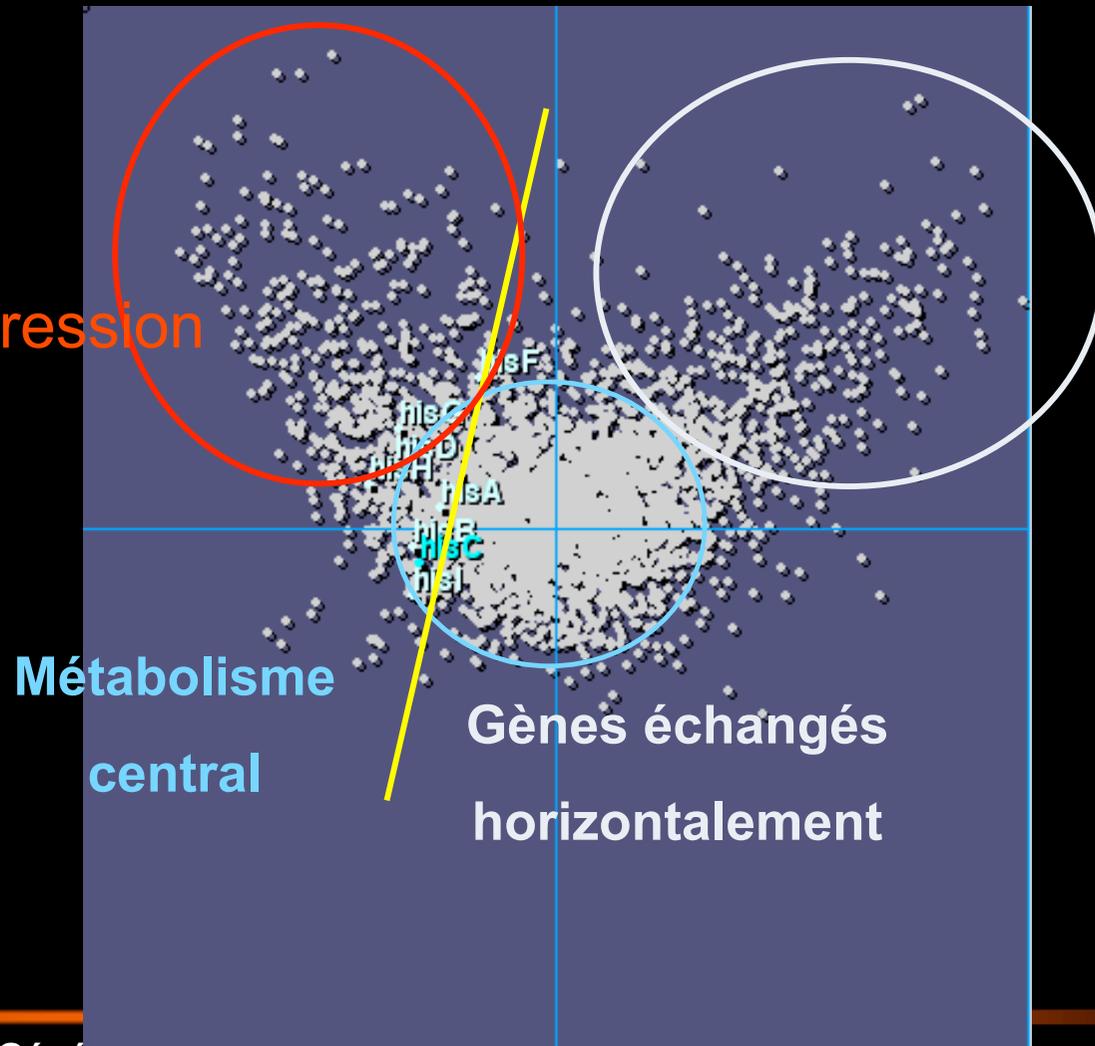
L'Analyse en Composantes Indépendantes utilise le caractère non gaussien de la répartition des valeurs, etc

LES BIAIS DE L'USAGE DU CODE SONT LIÉS AUX FONCTIONS

Classe I :
métabolisme
central

Classe II : forte expression
en croissance
exponentielle

Classe III :
transfert
horizontal



BIAIS LOCAUX DE L'USAGE DU CODE

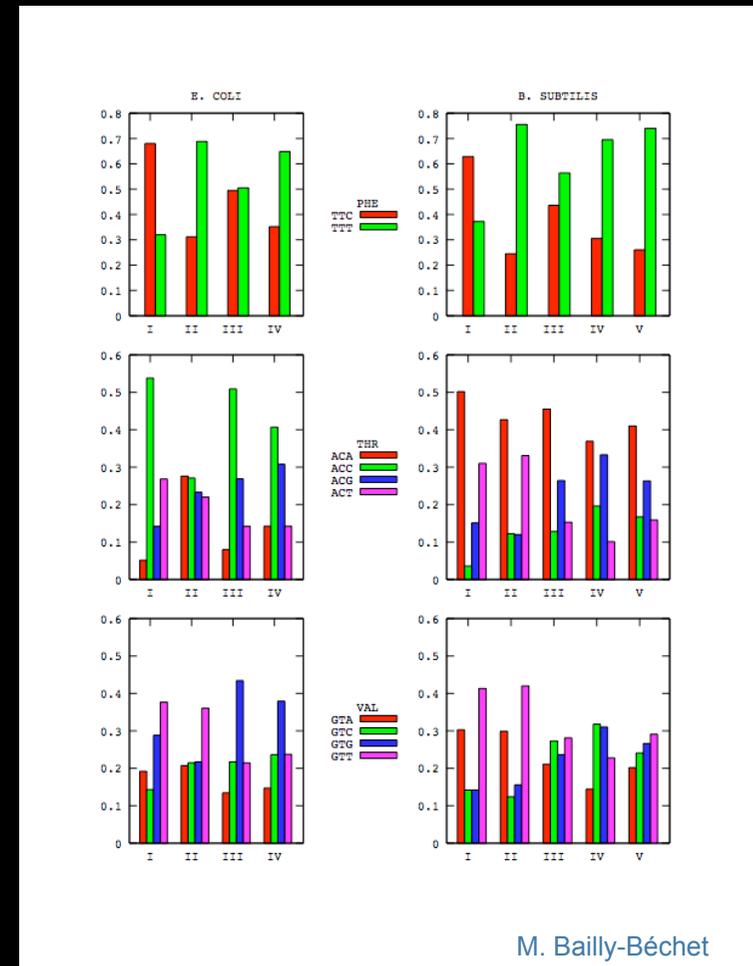
L'Analyse Factorielle des Correspondances montre que les gènes qui ont un biais similaire dans l'usage des codons sont apparentés fonctionnellement. Comment cela se répartit-il dans le chromosome ?

Une méthode de groupage fondée sur l'analyse des biais d'**usage des codons** au moyen d'une théorie de l'information regroupe les gènes en familles homogènes, qui ne se répartissent pas au hasard dans le chromosome. La méthode permet à la fois d'identifier des biais cohérents et de trouver le nombre pertinent de classes à considérer (4 pour *E. coli* et 5 pour *B. subtilis*)

ÎLOTS GÉNOMIQUES

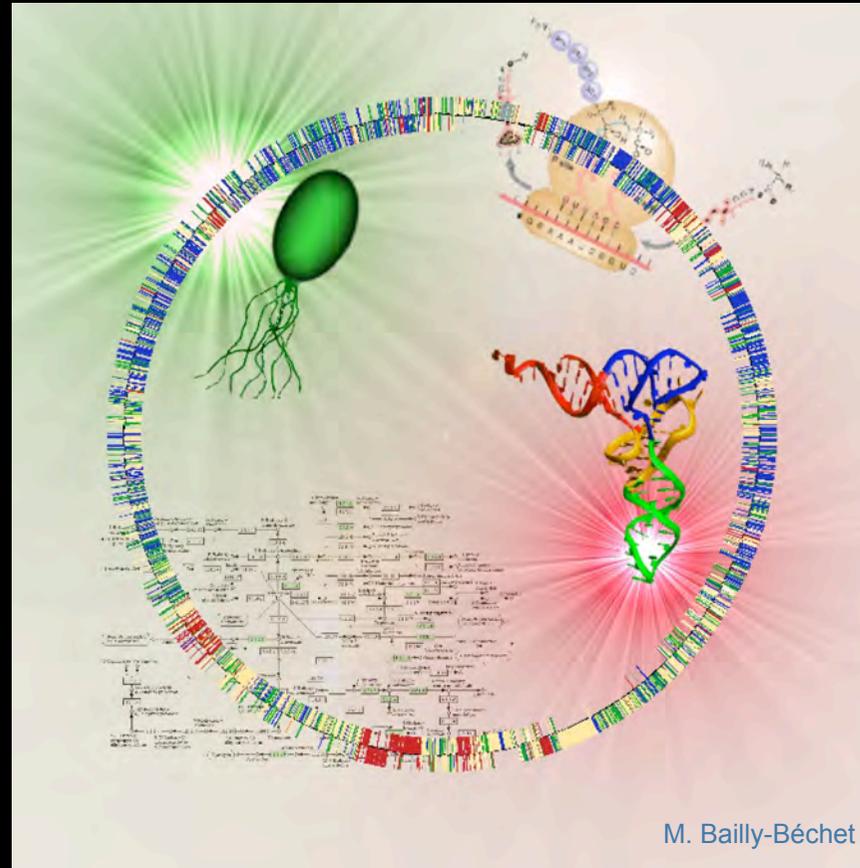
Les gènes qui ont un biais similaire sont organisés en régions plus étendues que les opérons, ce qui démontre un rôle de la traduction dans la structuration du chromosome bactérien

Une part importante de la contribution à cette effet vient du recyclage des ARN de transfert rares



ÎLOTS GÉNOMIQUES

Un groupe correspond au niveau d'expression (bleu). Les autres groupes sont fonctionnellement cohérents : gènes du transfert horizontal (rouge), motilité (jaune) et métabolisme intermédiaire (vert)



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

LA PREMIÈRE DÉCOUVERTE DE LA GÉNOMIQUE

En 1991, à la réunion de l'UE sur les programmes génomes à Elounda, en Grèce, la présentation du chromosome III de la levure et des premiers 100 kb du génome de *Bacillus subtilis* a révélé que, au contraire de ce qu'on pensait jusqu'alors, **au moins la moitié des gènes découverts étaient totalement inconnus, que ce soit par leur fonction ou la structure de leur produit**

Plusieurs raisons rendent compte de ce fait, toujours vrai aujourd'hui ; d'une part notre connaissance du métabolisme est très imparfaite, d'autre part nous ne savons pas comment se créent de nouveaux gènes ; enfin, l'évolution procède par la **sélection de fonctions**, en **recrutant des structures** dont l'adaptation va s'accroître en parallèle avec l'adéquation de l'individu à son environnement (**évolution acquisitive**)

S'il y a tant de nouveaux gènes, existe-t-il cependant un socle commun ?

GÈNES PERSISTANTS

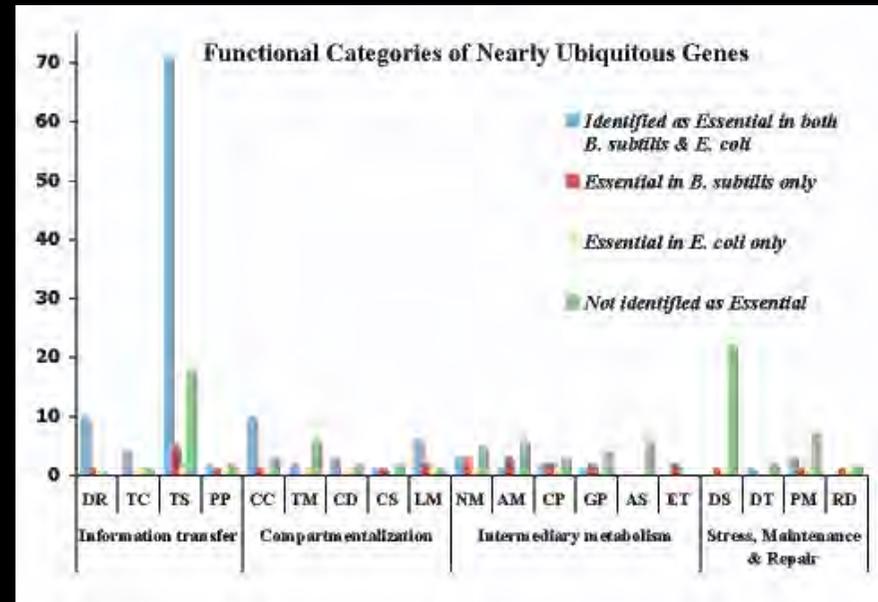
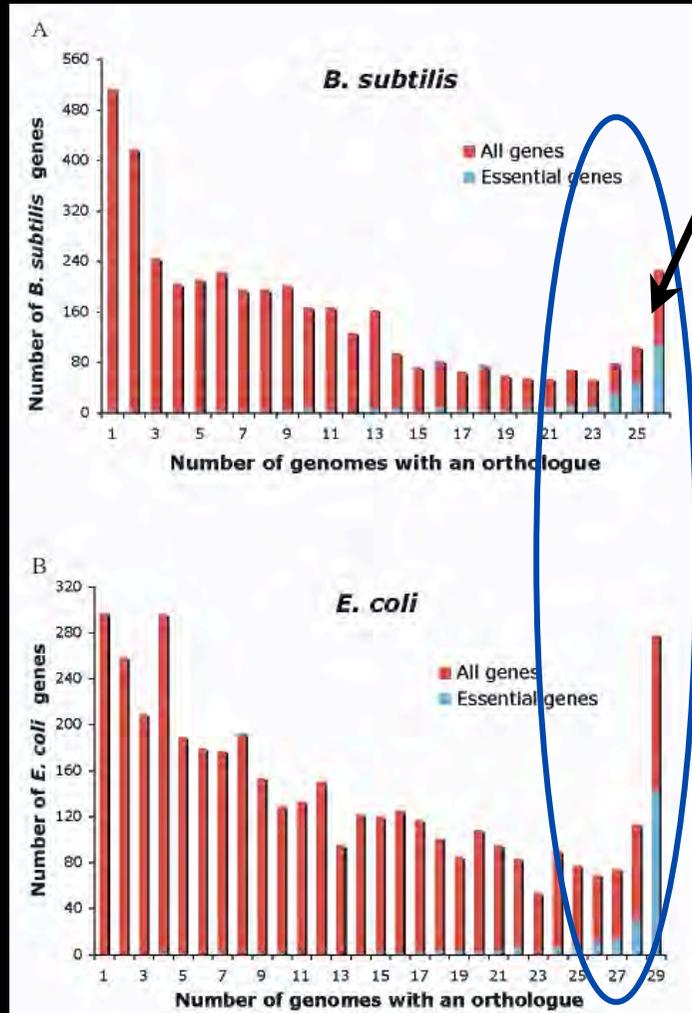
Les gènes essentiels en laboratoire sont situés dans le brin précoce. Ils sont aussi conservés dans une majorité de génomes

A l'inverse les gènes qui sont conservés et se trouvent dans le brin précoce forment une catégorie supplémentaire, qui double celle des gènes essentiels

Ces gènes forment des **universaux** ; 400-500 gènes persistent dans un grand nombre de génomes ; ils sont impliqués non seulement dans les trois processus nécessaires à la vie, mais dans la **maintenance** et dans l'**adaptation aux phénomènes transitoires** ; une fraction gère l'**évolution** de l'organisme

PERSISTANCE DES GÈNES

Gènes persistants



→ Stress, Maintenance et Réparation

PHYLOGÉNIE DE LA PERSISTANCE

Quelques uns des gènes essentiels absents de la liste ont évolué particulièrement rapidement

Pour mesurer la contribution de cet effet, on la met en parallèle avec l'évolution de l'ARN ribosomique 16S

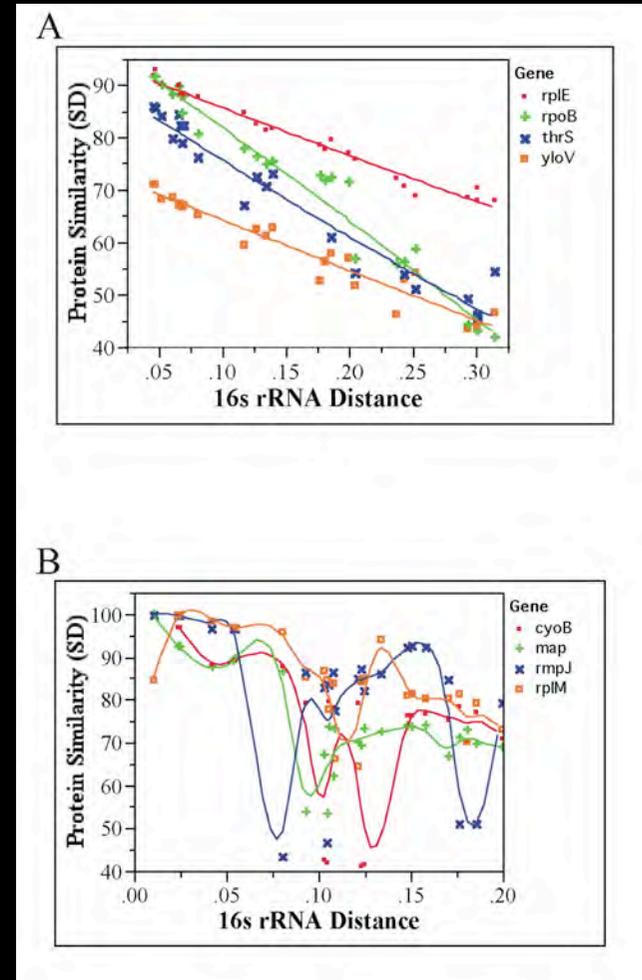
Deux scénarios distincts apparaissent : dans la moitié des cas une évolution linéaire, et dans l'autre moitié une évolution erratique

PERSISTANCE DES GÈNES

La moitié (A) des gènes persistants de *B. subtilis* (resp. *E. coli*) ont un coefficient de corrélation >0.9 entre la similitude de séquence des paires d'orthologues et l'ARN 16S

Certains gènes (B) évoluent de façon erratique. Ce peut être dû au transfert génétique horizontal, à des adaptations particulières, ou tout simplement à des erreurs dans l'attribution de l'orthologie

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

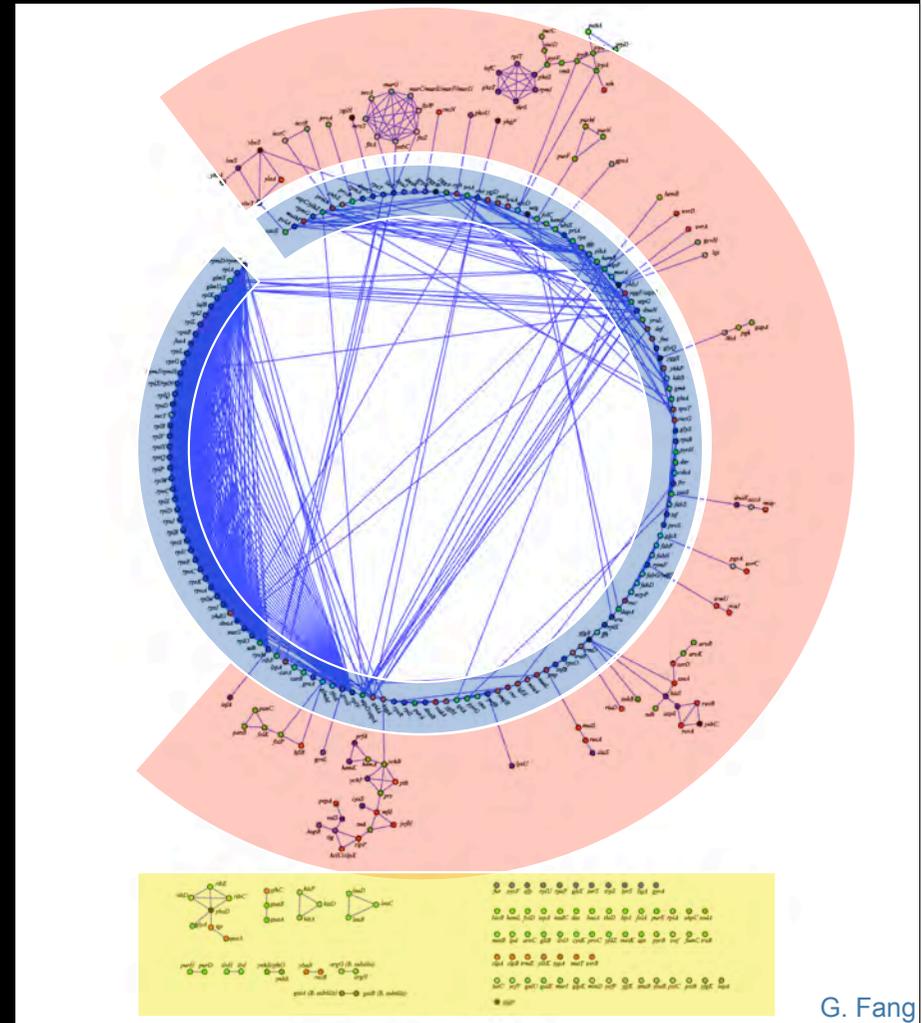


LES GÈNES PERSISTANTS SONT REGROUPÉS

Avec l'analyse de 228 génomes codant chacun plus de 1500 gènes et des annotations « correctes », on identifie les gènes persistants ; ils tendent à rester à proximité les uns des autres ; cette « attraction mutuelle » constitue un réseau remarquable fait de trois cercles concentriques qui regroupent les fonctions identifiées dans ce scénario de l'origine de la vie : un métabolisme compartimenté (à la surface des pierres) suivi de l'invention d'un substitut des surfaces, l'ARN, puis de l'usage des acides nucléiques pour gérer **les interactions avec l'environnement au travers de l'information correspondante**

LES GÈNES PERSISTANTS RÉCAPITULENT L'ORIGINE DE LA VIE

Le **réseau externe** (gènes du métabolisme intermédiaire, nucléotides et coenzymes, lipides), est très fragmenté ; le **réseau médian** est formé autour des ARNt synthétases de classe I, et le **réseau intérieur**, presque continu, est organisé autour du ribosome, de la transcription et de la réplication gère les transferts d'information

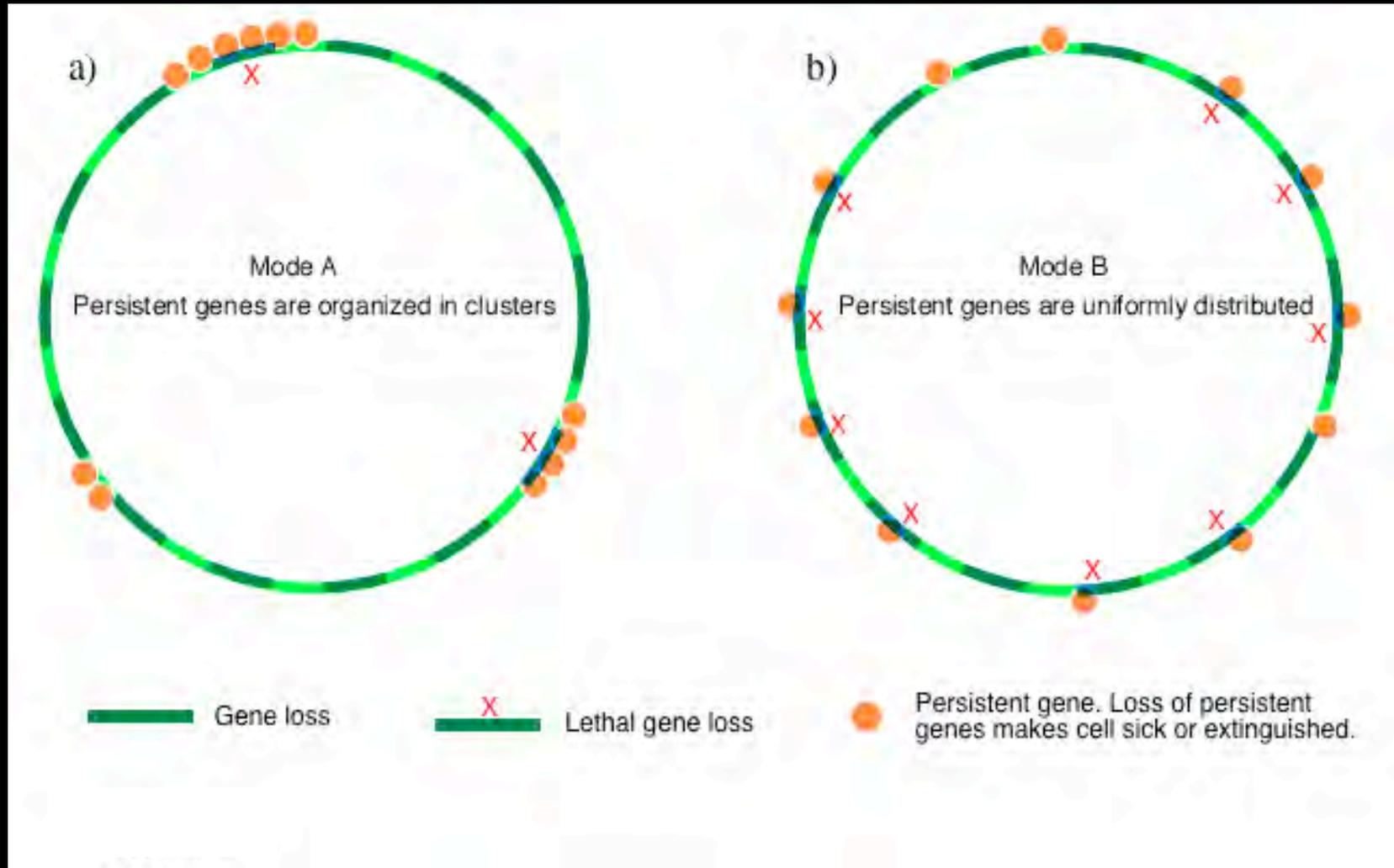


« ETRE » IMPLIQUE LE REGROUPEMENT

Pourquoi les gènes persistants sont-ils regroupés ? Un modèle simple montre que si, pour compenser le transfert horizontal des gènes, les gènes tendent à disparaître en groupes, alors tout gène contribuant à l'adéquation de l'organisme à son environnement assez fréquemment tendra à se regrouper avec les gènes ayant la même propriété. Cela explique le regroupement des gènes essentiels, mais peut-être aussi celui des gènes de résistance aux antibiotiques...

En conséquence le regroupement **précède et ne résulte pas de la** co-transcription ou des interactions protéine-protéine !

L'EXISTENCE IMPLIQUE LE REGROUPEMENT



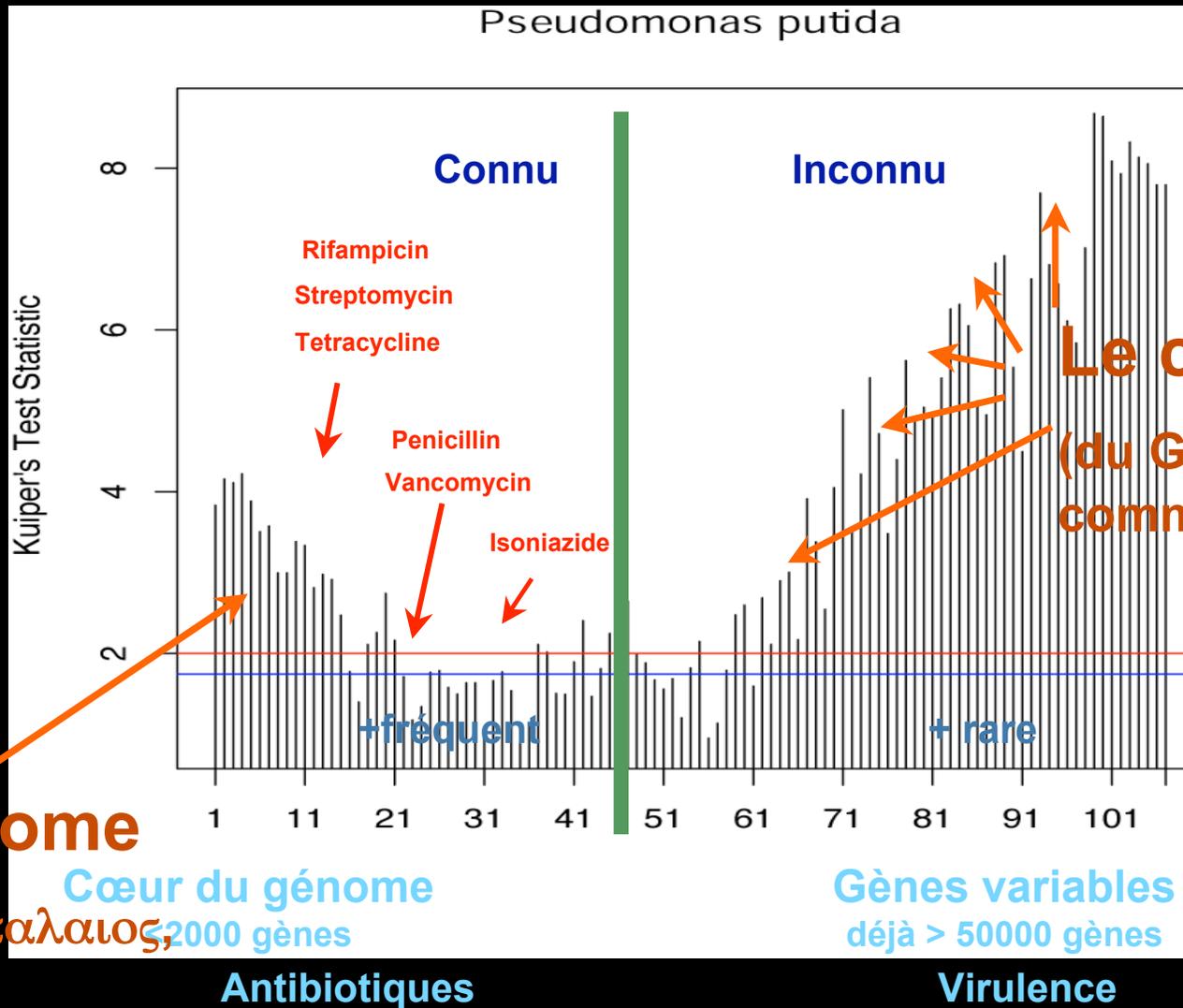
G Fang, EPC Rocha, A Danchin Persistence drives gene clustering BMC Genomics (2008) **sous presse**

LE GÉNOME MOSAÏQUE

- Une séparation conceptuelle suppose l'existence de **deux parties du génome**, codant la machine et l'”objectif” de la machine. Cela implique la distinction entre le **replicateur/constructeur** et des fonctions secondaires
- Les génomes réels doivent donc coder des fonctions ubiquistes (pas des gènes ubiquistes !) qui correspondent au premier génome (appelé ici le **paléome**) et des fonctions spécifiques à l'environnement de l'organisme (constituant ici le **cénome** — comme dans “biocénose” — pour exprimer le fait qu'il s'agit de fonctions spécifiques à l'occupation d'une niche particulière)

LE CÉNOME

Fréquence de regroupement



Le cénome
(du Grec κοινος, commun)

Fréquence dans les génomes

Le paléome
(du Grec παλαιος, ancien)

FAIRE UNE CELLULE : DU PALÉOME AU CÉNOME

La vie se manifeste par la croissance et la lutte contre le vieillissement: le génome correspondant est le **paléome**. L'exploration de l'environnement est une conséquence inévitable de l'existence, elle résulte du flux constant d'entrée et de sortie des gènes qui forment le **cénome**.

