



# **Les gènes du démon de Maxwell et la sélection naturelle**

**Antoine Danchin**

唐善 · 安東

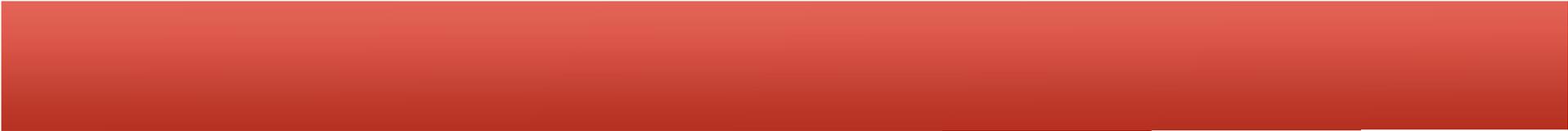
**AMAbiotics SAS & CEA / Genoscope**

**Centre Interdisciplinaire d'Étude de l'Évolution des  
Idées, des Sciences et des Techniques**

**Orsay, 2 décembre 2009**

# Evolution et Biologie Synthétique

- La SB repose sur un modèle, celui du programme génétique, où un **algorithme de construction** est **répliqué** au cours des générations
- Une **machine (« chassis »)**, qui se **reproduit** au cours des générations, exécute l'algorithme, dont elle est physiquement séparée du support, comme dans les ordinateurs
- Mais, si les ordinateurs font des ordinateurs, comment font-ils pour gérer le vieillissement ?
- L'analyse comparative des programmes génétiques met au jour des gènes (**gènes du démon de Maxwell**) nécessaires pour **faire des organismes jeunes à partir d'organismes âgés**. Elle montre que les organismes vivants sont des **pièges à information**



## Des ordinateurs faisant des ordinateurs?

***“Les organismes vivants comme pièges à information”***

<http://www.normalesup.org/~adanchin>

# Une imprimante 3D auto-reproductrice

Le projet RepRap (Replicating Rapid-prototyper, depuis 2004) vise à créer une imprimante laser 3D auto-reproductrice :

- La machine produit la plupart de ses composants (= “biobriques”)
- Ce qui manque :
  - Le programme
  - La ligne d’assemblage (gestion du temps et de l’espace, et fonctions spécifiques comme la lubrification)
- ➔ Deux types d’information
  - Information du programme
  - Information de la machine

<http://reprap.org/>



*“Les organismes vivants comme pièges à information”*

<http://www.normalesup.org/~adanchin>

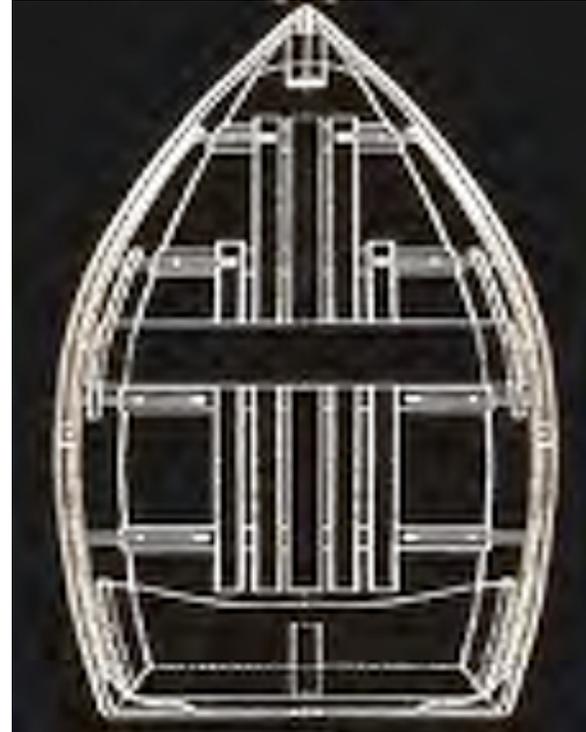
# L'information comme nouvelle catégorie du réel

- 1929 **Leo Szilard**, analysant le démon de Maxwell, a l'intuition (fausse) de la relation entre énergie et information: la création d'1 bit consomme  $1/2 kT$
- 1949 **Claude Shannon** crée une théorie de la communication
- **1961 Rolf Landauer**, prouve que le calcul est réversible (aucune énergie n'est requise pour la création d'information) ; l'énergie est requise pour effacer la mémoire du processus de création
- ~1974 **Andrey Kolmogorov**, **Gregory Chaitin**, **Ray Solomonoff** définissent la **complexité algorithmique**
- 1988 **Charles Bennett** définit la **profondeur logique** (lien entre temps et complexité algorithmique) pour prendre en compte la **valeur** d'une information et illustre concrètement le théorème de Landauer
- 1989 **Wojciech Zurek** lie la complexité algorithmique et l'énergie, en réfléchissant à ce que serait le démon de Maxwell
- 2007 **Scott Muller** définit l'information comme attribut quelconque permettant de déterminer l'état d'un système, via une **asymétrie**
- 2009 **Takahiro Sagawa** et **Masahito Ueda** reprennent le théorème de Landauer liant énergie et effacement de la mémoire

# La barque de Delphes

➔ La biologie est une science des relations entre objets

➔ C'est une information qui exprime ce qui est conservé dans la barque, pas la matière de ses planches !

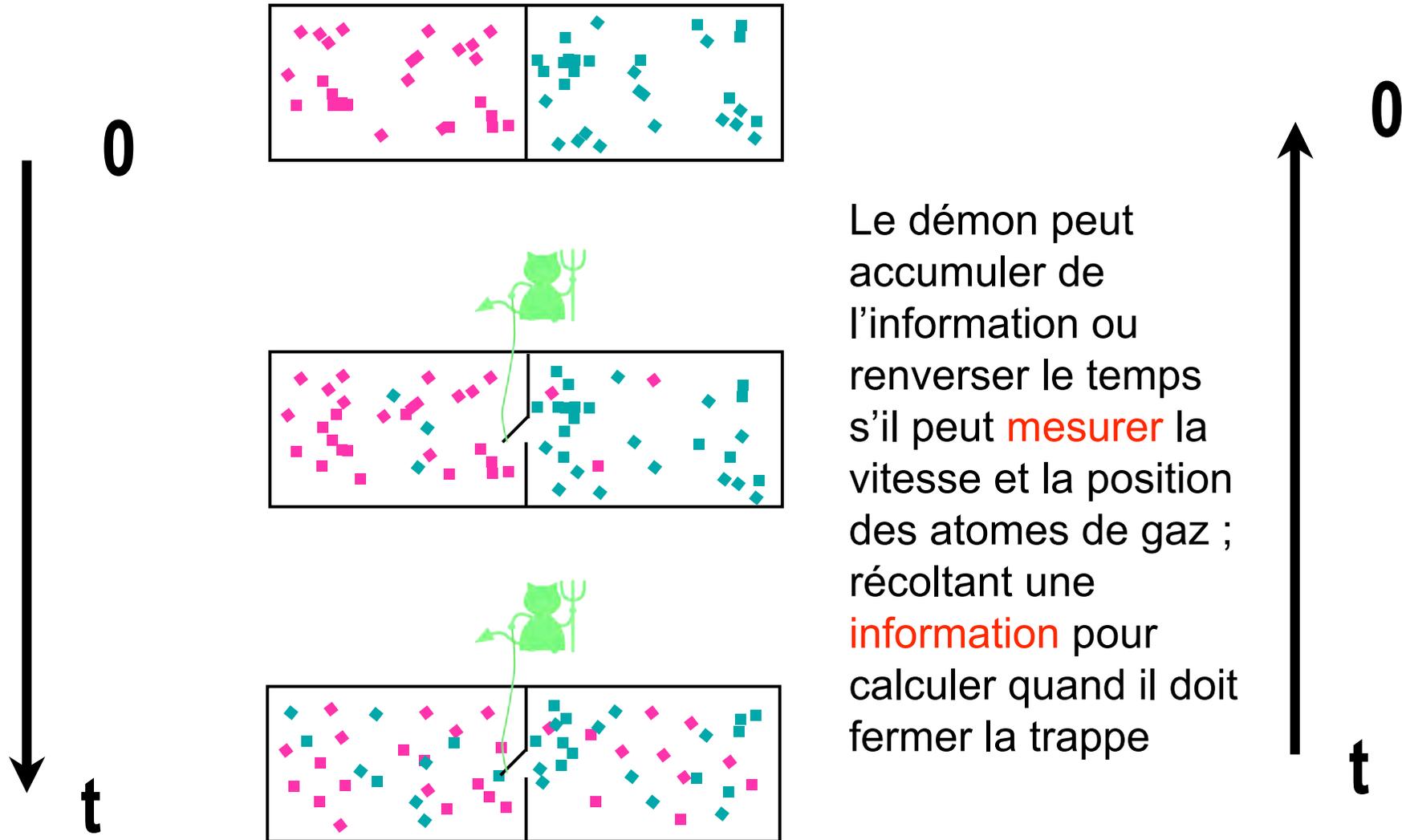


A. Danchin **The Delphic Boat, What genomes tell us.** Harvard University Press, 2003

V. de Lorenzo, A. Danchin **Synthetic Biology: discovering new worlds and new words** 9: 822-827. EMBO Reports, 2008

A. Danchin **Bacteria as computers making computers.** *FEMS Microbiology Reviews* 2009, **33**: 3-26

# Le démon de Maxwell



Le démon peut accumuler de l'information ou renverser le temps s'il peut **mesurer** la vitesse et la position des atomes de gaz ; récoltant une **information** pour calculer quand il doit fermer la trappe

# Concrétiser l'information : l'infotaxie



*Saturnia pyri*

<http://pdubois.free.fr/>

Comment une phalène peut-elle trouver un partenaire à un kilomètre de distance ?

Remonter un gradient chimique est impossible à une telle distance (turbulences, obstacles...)

Vergassola et ses collègues ont montré que maximiser la collecte d'**information** (via un simple calcul) permet à l'insecte d'atteindre cet objectif ...

'Infotaxis' as a strategy for searching without gradients

Vergassola M, Villermaux E, Shraiman BI  
*Nature* (2007) **445**: 406-409



# Pensée implicite de l'information

*Fait remarquable, la durée de l'exploration mentale varie de manière linéaire avec les distances réelles des points marqués par le sujet sur la carte, de la plage au cocotier, à la hutte, au trésor. La carte mentale contient donc la même **information** sur les distances que la carte réelle*

**Jean-Pierre Changeux (1983) Collège de France**

*Engineered biological systems have been used to **manipulate information**, construct materials, process chemicals, produce energy, provide food, and help maintain or enhance human health and our environment*

**Drew Endy (2005) MIT**

***“Living organisms as information traps”***

<http://www.normalesup.org/~adanchin>

# Heuristique: une cinquième catégorie de la réalité

## Matière / Energy / Espace / Temps

- Physique Classique
- Physique Quantique
- Chimie
- Biologie
  - Développement
  - Neurobiologie
  - Linguistique
- Mathématique

Information



Nous proposons que certaines des fonctions de la vie sont calculables

Natural selection and immortality

Danchin A

*Biogerontology*. 2009 Aug;10(4):503-16. Epub 2008 Aug 22

# Ce qu'est la vie

La vie suppose:

- ➔ **Une machine** permettant d'exécuter un programme (**elle se reproduit**)
  - ➔ **1. Métabolisme** (processus dynamique)
  - ➔ **2. Compartimentation** (chassis, définit un intérieur et un extérieur)
  
- ➔ **Un programme** (un "livre de recettes, qui **est répliqué**)
  - ➔ **3. Transfert récuratif d'information** => le codage d'un niveau à un second niveau introduit une **asymétrie** essentielle (fondamentalement différente des rétroactions)

**La cellule est l'atome de vie**

**La récursivité permet l'accumulation d'information**

# Cellules et ordinateurs

La génétique repose sur la description des génomes comme des textes écrits au moyen d'un alphabet : mais **les cellules se comportent-elles comme des ordinateurs ?**

- Transfert Génétique Horizontal
- Virus
- Génie génétique
- **Transplantation d'un génome nu dans une cellule réceptrice** suivie du changement de la machine réceptrice en une nouvelle (2007)

tout concourt à séparer

«Machine» (l'usine cellulaire) et «Données/Programme» (le génome)

**CONJECTURE: les organismes vivants sont des pièges à information, et nous pouvons identifier les processus concrets qui permettent l'accumulation de l'information**

# Virus et programme

Comme un virus informatique, un virus biologique est fait d'un ou plusieurs acides nucléiques (son génome), codant une ou plusieurs protéines permettant sa réplication, sa protection dans la cellule et l'environnement, et sa propagation.

Souvent il code aussi des activités qui interfèrent avec le métabolisme de la cellule hôte, le détournant au profit de sa propre construction. Comme dans les ordinateurs il peut piéger des fonctions de la machine et les transmettre à sa descendance.

Cependant il ne code pas la totalité des objets permettant l'expression de son génome. **Il dépend entièrement de l'usine cellulaire pour se perpétuer.**

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Un ordinateur génétique

- Dans un ordinateur, la machine est séparée des données et du programme
- Données et programme jouent le même rôle (*càd* on peut les considérer comme des ‘ déclarations ’); ils peuvent être modifiés par la machine
- Une réflexion générale (due à Alan Turing et de nombreux autres penseurs) considère les actions de la machine, mais pas la façon dont elle est construite

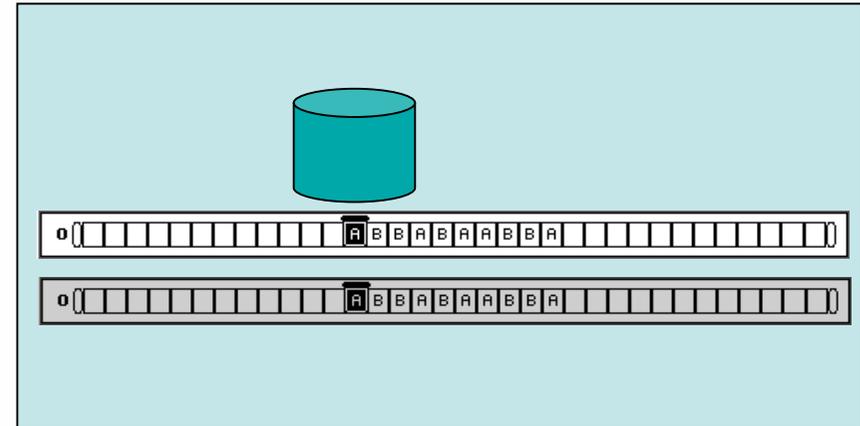
***“Living organisms as information traps”***

<http://www.normalesup.org/~adanchin>

# Ce qu'est le calcul

Deux entités permettent le calcul :

- Une machine capable de lire et d'écrire
- Un programme (sur un support physique), séparé par l'esprit humain (pas conceptuellement !) en deux entités :
  - Programme (définissant l'"objectif")
  - Données (définissant le contexte)



La machine est distincte de  
l'entité données/programme

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

## Objection au modèle de la cellule-ordinateur

**“En plus du programme génétique, la cellule comporte une quantité considérable d’information...”**

**Mais c’est exactement la même chose pour un ordinateur et son système d’exploitation !**

**Cette observation — valide — signifie simplement que notre réflexion sur le concept d’information est incomplète**

Même pour un ordinateur, n'oublions pas le support physique



**Il ne suffit pas d'avoir une molécule d'ADN avec la bonne séquence, il doit encore être correctement replié !**

Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A.* (2008) 105:20404-20409

***“Living organisms as information traps”***

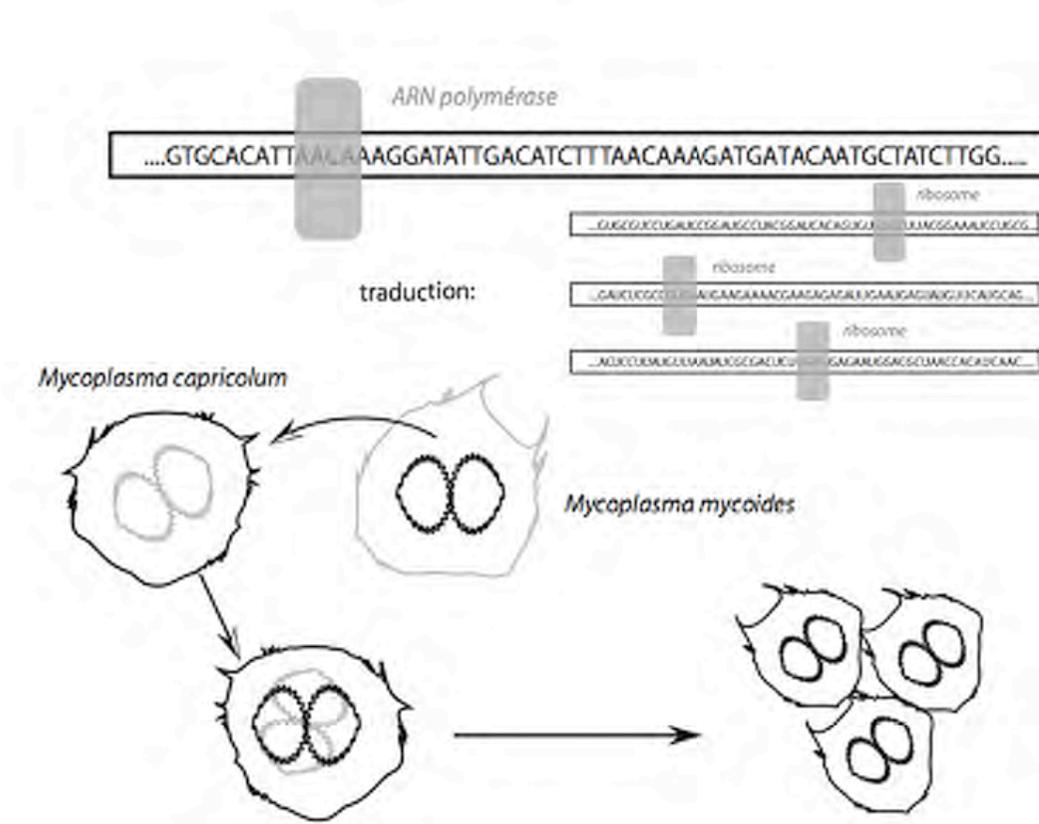
<http://www.normalesup.org/~adanchin>

# La démonstration de Lartigue-Venter

La Machine de Turing

peut être parallélisée

Transplantation d'un génome



Genome transplantation in bacteria: changing one species to another

Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC  
*Science* (2007) **317**: 632-638

**“Living organisms as information traps”**

<http://www.normalesup.org/~adanchin>

# Reproduction et réplication

Dyson a montré qu'il nous faut distinguer **reproduction** — faire une copie similaire — de **réplication** — faire une copie exacte

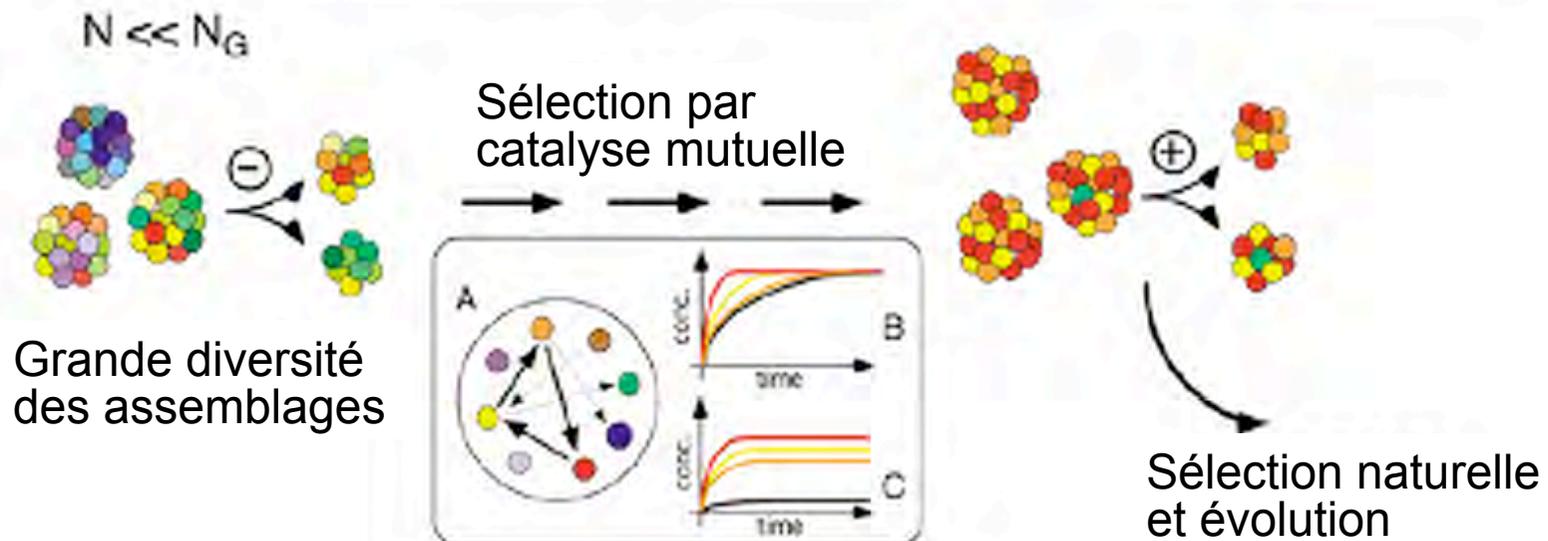
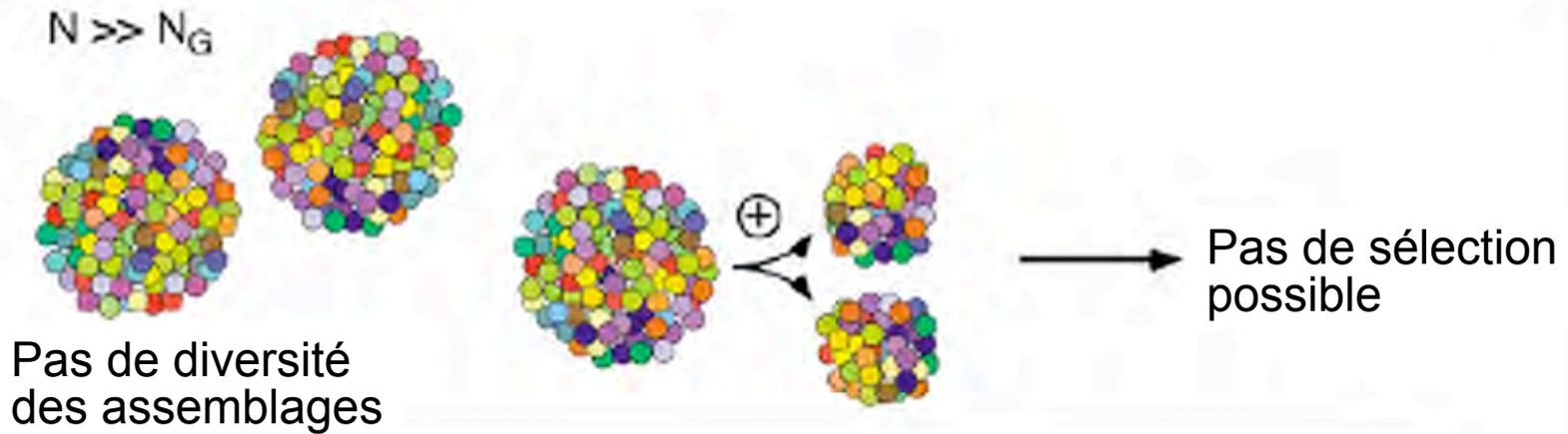
La reproduction peut s'améliorer au cours du temps. La réplication est généralement condamnée à accumuler les erreurs

Ainsi, la **reproduction est antérieure à la réplication**

La reproduction suppose un réseau d'interactions, et elle est facilement imaginable comme un type particulier de réseau métabolique compartimenté

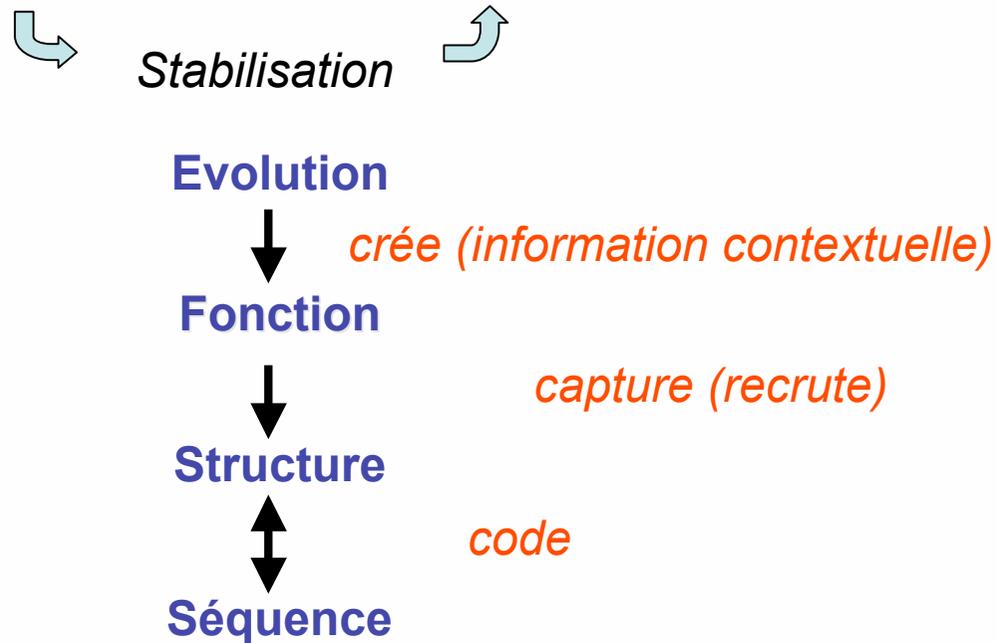
**Quelles fonctions (codée par des gènes) permettent l'accumulation de l'information?**

# Reproduction d'un métabolisme compartimenté



# A la recherche des fonctions ubiquistes

Variation / Sélection / Amplification



**ubiquité fonctionnelle n'implique pas ubiquité structurale**

# De l'ubiquité fonctionnelle à la persistance des gènes

Ubiquité fonctionnelle n'implique pas ubiquité structurale

Heureusement, les entités fonctionnelles tendent à persister à travers les générations:

➔ Rechercher la « persistance » permet d'identifier les fonctions ubiquistes

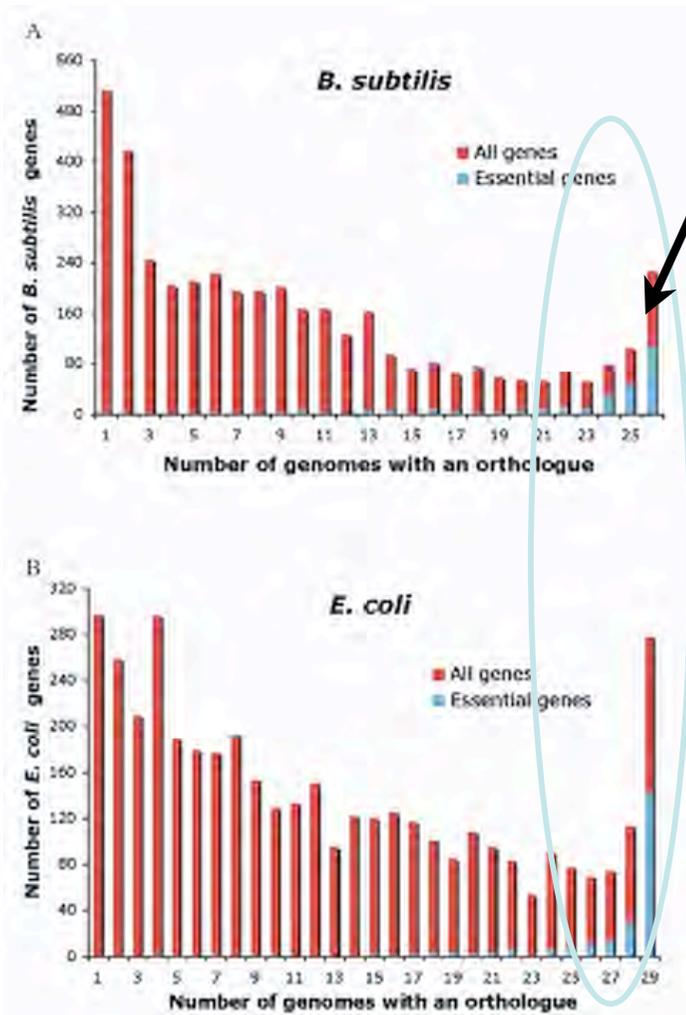
➔ « Ubiquiste » est-il synonyme d'« essentiel »? **NON**

~ 500 gènes persistent dans les génomes bactériens ; environ 250 ne peuvent être inactivés sans perte de viabilité ; **que dire de l'autre moitié ?**

*“Les organismes vivants comme pièges à information”*

<http://www.normalesup.org/~adanchin>

# Persistence : deux fois trop de gènes !



gènes persistants

Gènes essentiels et ....

Stress, maintenance et réparation

Dégradation énergivore

« Frustration » métabolique  
(chemical constraints)

Gang Fang & Eduardo Rocha

# Proximité des gènes persistants

Avec 228 génomes (>1500 gènes) annotés correctement, nous avons identifié les gènes qui tendent à rester proches dans la plupart des génomes

Deux ensembles de conservation de proximité apparaissent : les gènes persistants et les gènes rares

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Un détour par les origines de la vie

- Le physicien Freeman Dyson a montré qu'il ne peut y avoir une seule origine de la vie
- La **réplication** conduit presque inmanquablement à une avalanche d'erreurs (sauf en cas d'échanges avec l'extérieur qui permettent de revenir en arrière)
- La **reproduction**, qui concerne des relations entre objets, peut s'améliorer au cours du temps
- **La reproduction a donc dû précéder la réplication.** Elle est le fruit d'un métabolisme originel ; peut-on en imaginer le scénario primitif ?

*“Living organisms as information traps”*

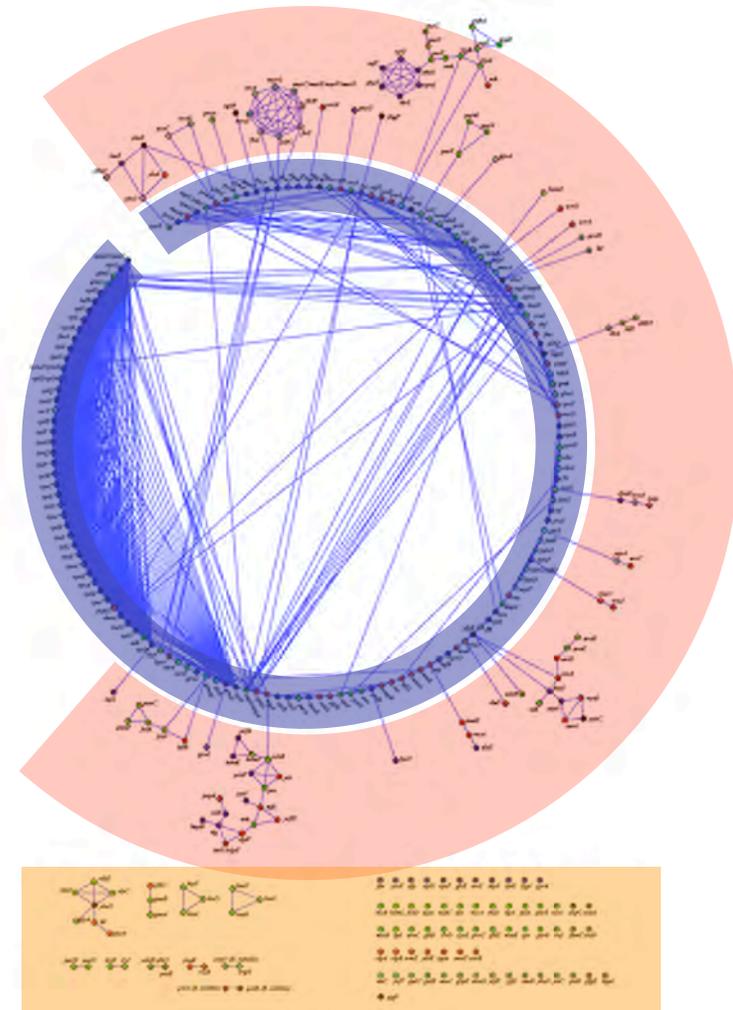
<http://www.normalesup.org/~adanchin>

# Un scénario minéral des origines de la vie

- La surface de solides électriquement chargés (p.ex. la pyrite (Fe-S)) sélectionne et compartimente des molécules chargées négativement ; cette étape forme les acides aminés, les principaux coenzymes, les acides gras et les ribonucléotides (peut-être via la fixation d'azote) ; la polymérisation avec élimination d'eau est favorisée par effet entropique
- Un métabolisme compartimenté remplace les surfaces minérales par la polymérisation de ribonucléotides en présence de peptides (le monde ARN, avec les ancêtres des ARN de transfert)
- Les ARN découvrent la loi de complémentarité, et le code génétique est inventé (des substrats aux matrices). Les règles du contrôle des transferts d'information sont découvertes, d'abord dans le monde ARN où des vésicules abritant les ancêtres des gènes se scindent et fusionnent au hasard, avant la formation des premiers génomes (le monde des génomes ARN). Puis les acides nucléiques sont stabilisés en *désoxyribonucleotides*

# Les gènes persistants récapitulent l'origine de la vie

- Le **réseau extérieur**, avec ses gènes du métabolisme intermédiaire (nucléotides et coenzymes, lipides) est très fragmenté
- Le **réseau médian** est construit autour des tRNA synthétases de class I
- Le **réseau intérieur** est organisé autour du ribosome, de la transcription et de la réplication



A Danchin, G Fang, S Noria

The extant core bacterial proteome is an archive of the origin of life  
Proteomics. (2007) 7:875-889

# Organisation des génomes bactériens

## mixome: assimilation fonctionnelle

Fréquence  
de regroupement

Connu



Inconnu

**Le cénome**

(de κοινος,  
commun): le but  
de l'usine  
cellulaire

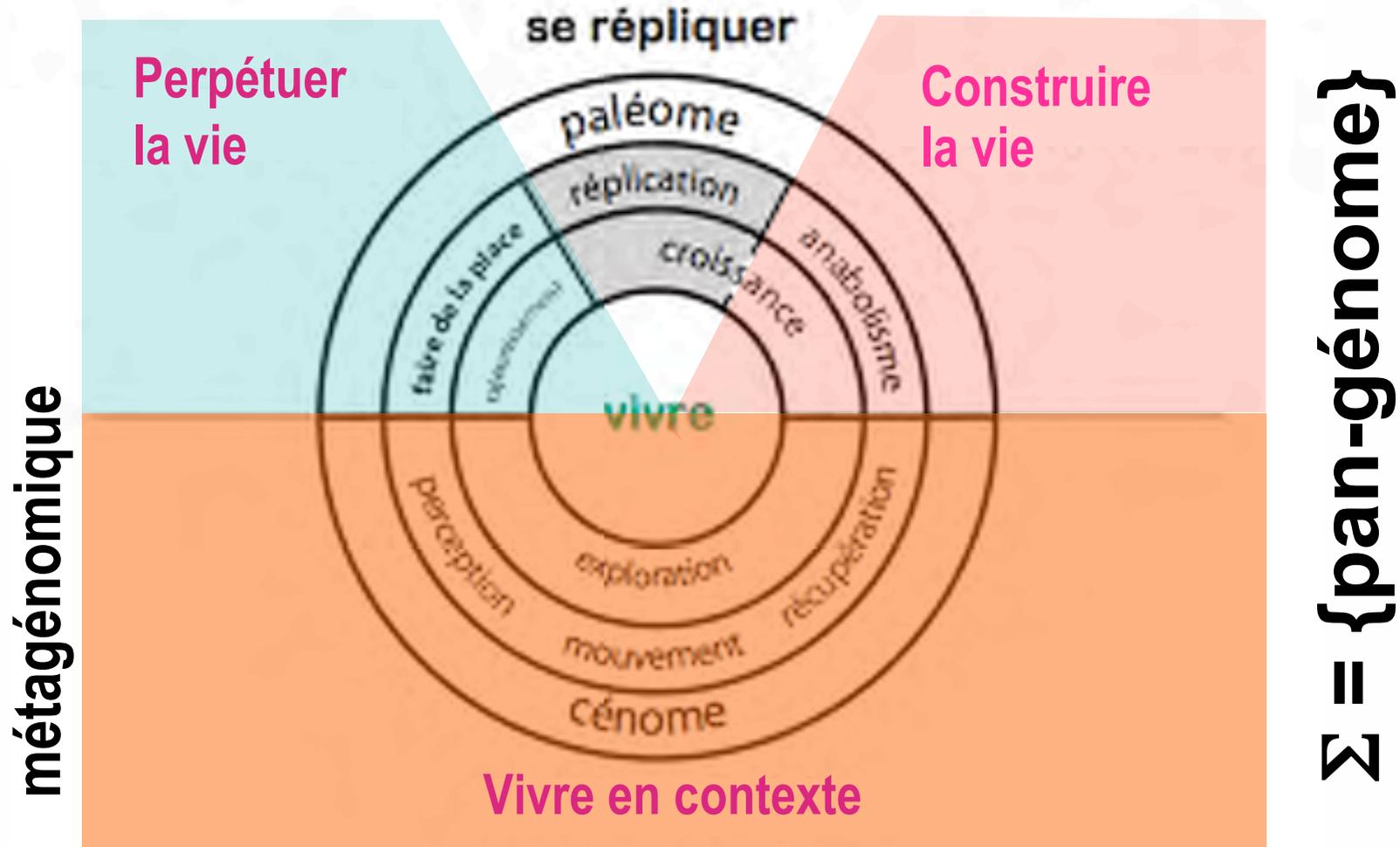
Fréquence  
dans les  
génomes

**Le paléome**

génomme conservé  
(l'usine cellulaire) 2,000 gènes

Gènes variables  
déjà > 50,000 gènes

# Histoire de deux génomes



# Le paléome et le cénome

## → La structure du paléome

→ Fonctions essentielles

→ Frustration chimique (« rustines » métaboliques : l'effet sérine)

→ **Dégradation énergivore**

## → Le cénome : du commensalisme à la virulence

→ Dégradation de composés aromatiques (*P. putida*, *P. aeruginosa*)

→ *Staphylococcus epidermidis* (Fudan University, Shanghai)

→ ColiScope: *E. coli*, *E. fergusonii* (E. Denamur et al.)

**“Les organismes vivants comme pièges à information”**

<http://www.normalesup.org/~adanchin>

# Les bébés naissent très jeunes !

- **La machine se reproduit**
  - La reproduction peut s'améliorer au cours du temps : c'est toujours un organisme âgé qui donne naissance à un jeune (cela implique création d'information)
- **Le programme se réplique**
  - La réplication accumule progressivement des erreurs

Quels gènes permettent l'accumulation d'information ?

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Le statut physique de la sélection naturelle

L'intuition dit que la création d'information requiert de l'énergie

**MAIS C'EST FAUX !**

S'efforçant de calculer les limites du calcul effectif, Landauer a démontré que la **création d'information est réversible** (càd ne consomme pas d'énergie : Landauer, 1961; Bennett, 1982, 1988) ; mais, accumuler l'information demande un **processus énergivore pour faire la place pour cette accumulation**

**Conjecture:** « Faire de la place » est nécessaire pour accumuler l'information ; comment est-ce réalisé ?  
Pouvons-nous identifier dans les génomes les gènes codant les fonctions qui mettent en œuvre ce processus ? Pouvons-nous trouver une source ubiquiste et stable d'énergie ?

# Valeur de l'information

Dans les modèles classiques de l'information on ne se soucie pas de la signification, ou de la **valeur** de l'information

L'information du programme est transmise telle quelle par la réplication, sans valeur, mais d'où vient donc l'information de la Machine ? L'idée sous-jacente est qu'il faut entièrement reconsidérer le concept dès qu'on a affaire à des objets individualisés : peut-on **imaginer les gènes d'un Démon de Maxwell** qui ferait le tri entre ce qui est fonctionnel (localement) et ce qui ne l'est pas ?

***“Living organisms as information traps”***

<http://www.normalesup.org/~adanchin>

# Trop de gènes dans le paléome

- **Paléome 1 (gènes essentiels)**
  - **Constructeur**: l'ADN spécifie les protéines qui forment la machine qui construit la cellule (reproduction)
  - **Répliqueur**: l'ADN spécifie les protéines qui répliquent l'ADN (réplication)
- **Paléome 2 (gènes persistants non-essentiels)**
  - Pérennisation de la vie, **dégradation dépendant de l'énergie**
  - Gènes de division (ftsZ)
  - Rustines métaboliques (« frustration » chimique)

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# Conjecture

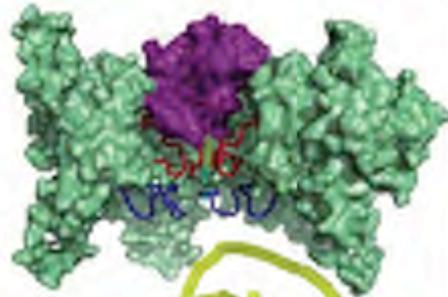
- Les processus dégradatifs dépendant de l'énergie font de la place pour des entités nouvellement synthétisées ; de l'énergie est consommée pour **prévenir** la dégradation des entités fonctionnelles
- Ce processus accumule de l'information, quelle que soit son origine, par effet de cliquet
- Comme ce processus est ubiquiste, on attend que les fonctions correspondantes soient codées dans le paléome, y compris la gestion des sources d'énergie postulées ici

***“Living organisms as information traps”***

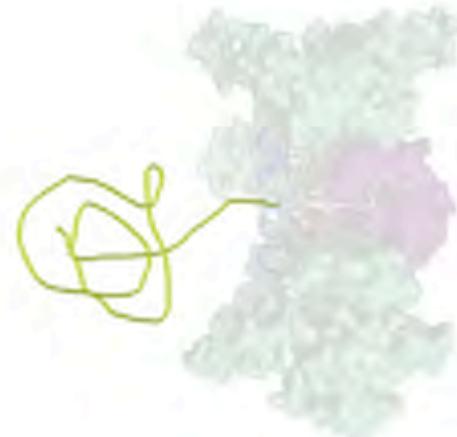
<http://www.normalesup.org/~adanchin>

# Les gènes du démon de Maxwell

ATP  
Poly(P)<sub>n</sub>



ADP + Pi  
Poly(P)<sub>n-1</sub> + Pi **<= dans le paléome**



La machinerie de dégradation utilise de l'énergie pour **rejeter** une entité fonctionnelle (ou simplement jeune)

Les entités non fonctionnelles sont reconnues et dégradées

# Un test pour identifier les gènes du démon de Maxwell

- L'accumulation d'information qui dépend de l'énergie est aveugle ; elle ignore d'où l'information proviendra
- L'information peut juste venir d'une mémoire; elle peut aussi être créée de novo

**Prédiction:** Les mutations adaptatives sont des créations d'information neuve; Elles devraient donc dependre des gènes impliqués dans l'accumulation d'information

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

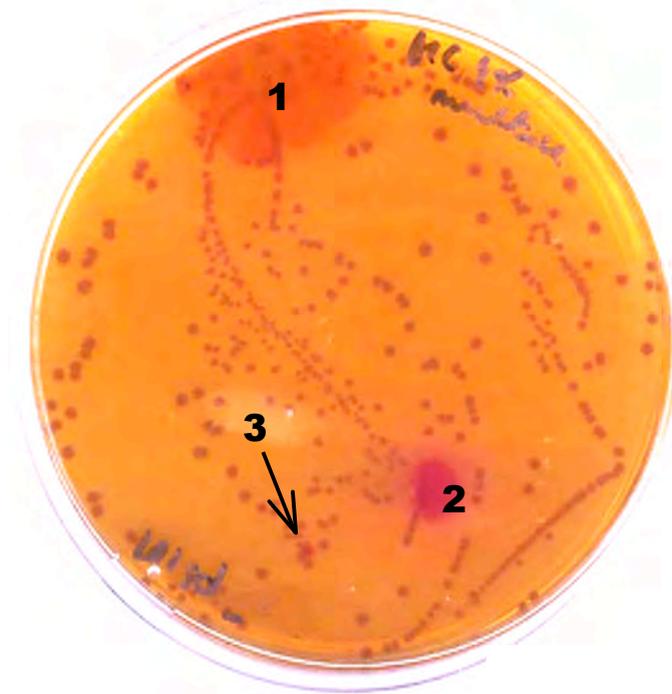
# Mise à l'épreuve : Mutations adaptatives

## Construction de bactéries "intelligentes"

Mises en croissance sur un milieu limitant en éléments nutritifs; Forment des colonies de  $10^7$  bactéries environ. Le milieu contient aussi des nutriments qu'elles ne savent pas utiliser

Au bout de quelques semaines apparaissent des papilles qui se mettent à croître et qui utilisent les éléments ajoutés. Ce sont des **mutations adaptatives**

Elles ne préexistaient pas et supposent donc la **création d'information**



Agnieszka Sekowska  
Evelyne Turlin  
Andrew Martens

***“Les organismes vivants comme pièges à information”***

<http://www.normalesup.org/~adanchin>

# Des mutations apparaissent au cours du vieillissement

**Séquençage de 14 génomes + 30 PCRs**

**Le nombre total des mutations croît avec l'âge**

**Les mutations sont réparties dans le chromosome, et majoritaires dans gène => PCR de nombreuses colonies**

**Dans ce gène on trouve différentes mutations dans les différentes papilles, 2 mutations dans 30% des cas**

**Dans certains cas l'une des mutations est silencieuse**

**Avec une source de carbone particulière on observe au moins un autre gène comme cause de la croissance**



A. Sekowska

# Une source d'énergie inattendue : les polyphosphates

- La synthèse et la dégradation des poly-P se retrouve codée dans les gènes non-essentiels persistants ; ce processus est encore peu connu et associé à la dégradation de l'ARN
- Le poly-P est un **minéral**, donc extrêmement stable ; il est présent de façon ubiquiste dans les cellules
- Les NTPs peuvent être régénérés à partir des NMP et de poly-P ; la protéase Lon peut utiliser le poly-P au lieu d'ATP; le NADP (anabolisme) peut être engendré à partir de NAD et de poly-P...

*“Living organisms as information traps”*

<http://www.normalesup.org/~adanchin>

# La sélection naturelle piège l'information

- Sélection naturelle: processus de dégradation qui fait de la place pour des entités neuves; elle utilise de l'énergie pour **prévenir** la dégradation des entités fonctionnelles
- Ce processus accumule l'information, quelle que soit son origine, avec effet de cliquet
- Parce que ce processus est ubiquiste, on attend que les fonctions correspondantes soient codées dans le paléome, y compris la gestion des sources d'énergie nécessaires à son fonctionnement
- Ce processus est **myope** : il ne peut donc avoir de grand dessein, d'où l'aspect "bricolé" de l'évolution de la vie

# Une cellule synthétique ?

- La vue d'ingénieur de la BS exclut que les cellules artificielles soient innovantes
- Il est possible d'**exclure les gènes permettant l'accumulation d'information**
- La conséquence est que, comme les usines, l'usine cellulaire vieillira et devra être systématiquement reconstruite
- Cela a l'avantage social que les risques sont minimisés
- En revanche cela pose problème pour l'utilisation à grande échelle de cellules synthétiques

# Prédictions générales

→ La persistance bactérienne dans un hôte dépend de gènes persistants non-essentiels de la famille décrite

→ L'initiation du cancer provient de cellules (souches) qui ont découvert des mutations adaptatives leur permettant de créer une descendance immortelle

→ L'accumulation d'information dans le cerveau (mémoire et apprentissage) implique des processus pour faire de la place tout en préservant les connexions fonctionnelles, d'une façon qui doit dépendre de l'énergie

# REMERCIEMENTS

## In silico

Gang Fang, Eduardo Rocha, TingZhang Wang

## In vivo

Agnieszka Sekowska, Evelyne Turlin, Andrew Martens

## Collaborations

### GenoScope

Beijing Genome Institute, Fudan University, The University of Hong Kong  
Hong Kong University of Science and Technology

