



# Génomique et séquençage d'ADN

Fidèle TIENDREBEOGO  
Ezechiel B. TIBIRI

# Plan

Rappels Historiques

Introduction à la Génomique

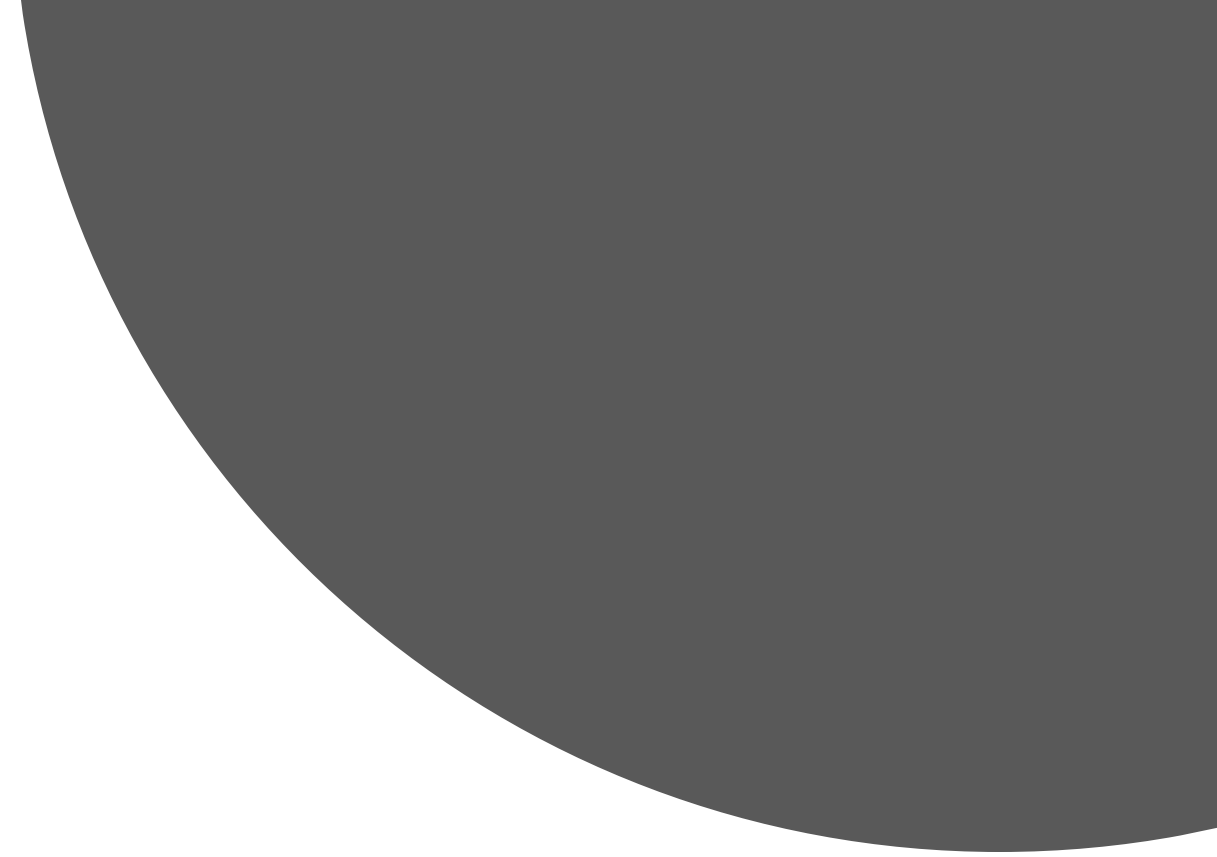
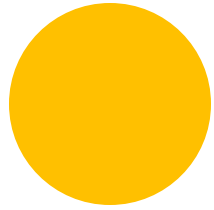
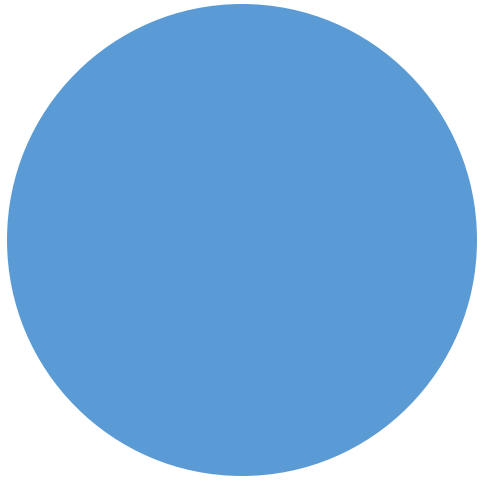
Introduction à la Métagénomique

Séquençage- Principe généraux

Next-generation sequencing

polymorphisme de séquence: Marqueur génétique

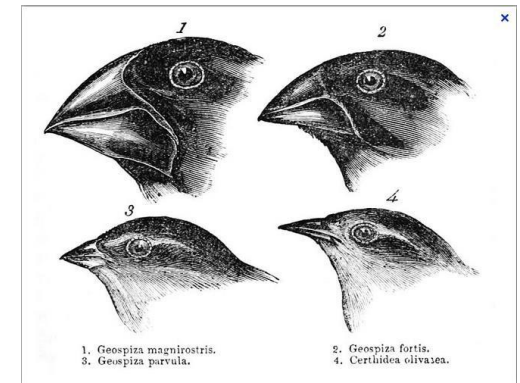
Introduction à la Bioinformatique : Analyse de séquences



# I. Rappels Historiques

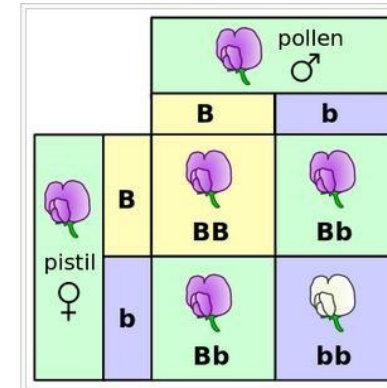
La Génétique

❖ 1859: Publication de « De l'origine des espèces » (Ch. Darwin).



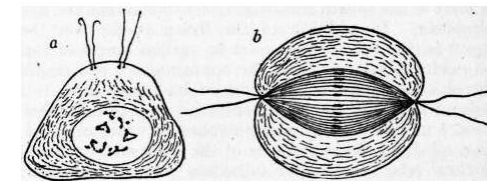
[https://fr.wikipedia.org/wiki/Pinsons\\_de\\_Darwin](https://fr.wikipedia.org/wiki/Pinsons_de_Darwin)

❖ 1865: Découverte des lois de l'hérédité (G. Mendel).



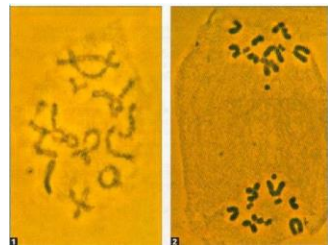
[https://fr.wikipedia.org/wiki/Mendelian\\_inheritance](https://fr.wikipedia.org/wiki/Mendelian_inheritance)

❖ 1880: Mise en évidence de la localisation de l'hérédité et le noyau dans la cellule (O. Hertwig et E. Strasburger).



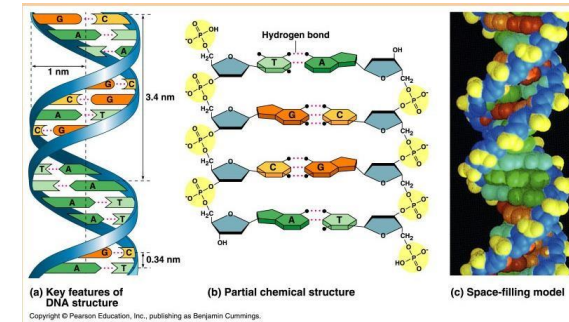
[Eb11.co.uk/index.php?p=CYTOLOGIY38709](http://Eb11.co.uk/index.php?p=CYTOLOGIY38709)

❖ 1911: Les chromosomes sont les supports de l'hérédité (W.Sutton et T.Boveri).



Slideplayer.fr/slide/1790373

❖ 1953: Découverte de la structure en double hélice de l'ADN (J.Watson, F. Crick et R. Franklin).



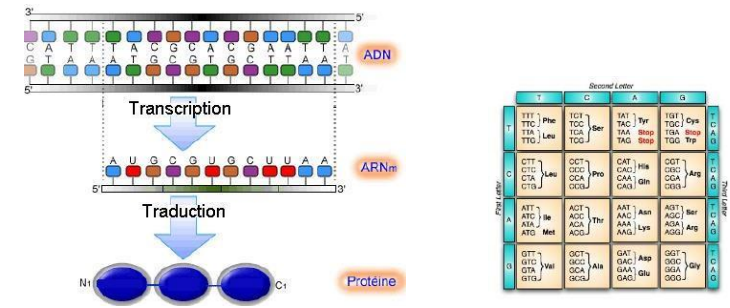
❖ 1958: premier lien entre un handicap mental et une anomalie chromosomique, la trisomie 21. (R. Turpin, J. Lejeune et M. Gauthier).



www.24matins.fr/trisomie-21  
depuis l'âge de 1 an  
22903



❖ 1961: découverte du code génétique, relation codon de trois bases et un acide aminé. (J. Monod, F. Jacob et A. Lwoff).



[www.alliot.fr/bio.shtml](http://www.alliot.fr/bio.shtml)

# La question centrale de la Génétique:

# La relation Phénotype/Génotype

- Le **phénotype** est l'ensemble des caractères observables d'un individu.
- Le **génotype** est l'ensemble de l'information génétique d'un individu qui se transmet des parents à leurs descendants
- L'expression du génotype produit dans un **environnement donné** un phénotype.
- **Phénotype = interaction {Génotype x Environnement}.**



# MARQUEURS GÉNÉTIQUES: DE 1970 À 2000

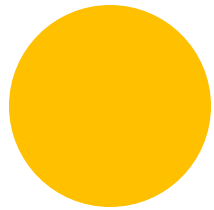
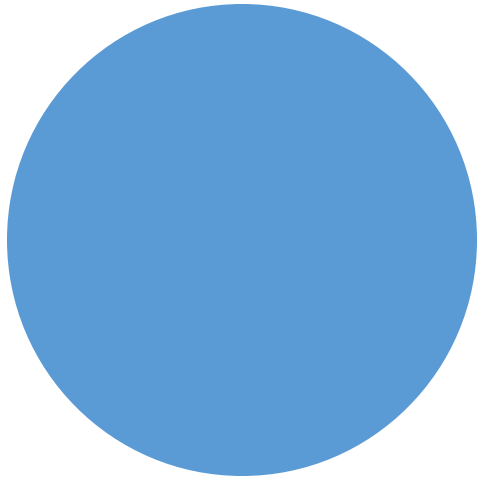
- Les **marqueurs phénotypiques** sont couleur des coquille, des pétales, des yeux,...
- Les **marqueurs moléculaires** sont des fragments d'ADN ou sa représentation moléculaire ( ARN ou protéine) qui permettent d'approcher ou mieux de connaître l'information génétique.
- Un marqueur est partagé par tous les individus d'une espèce, d'un règne, voire des êtres vivants.
- Il est transmis à chaque génération sans action de l'environnement.
- Leur nombre est très important par rapports aux marqueurs phénotypiques.
- Ils ont une forte variabilité.
- Ils sont de caractère discret ( pas de chevauchement comme les traits phénotypiques)

# LA GÉNOMIQUE : À PARTIR DE 2000

- 1977 premier séquençage d'un génome complet de virus à ARN.
- 1986: Lancement du programme de séquençage du génome humain
- 1995: Mise à disposition de la cartographie du génome humain
- 2000 Séquençage du génome humain
- 2000 Séquençage génome végétal: *Arabidopsis thaliana*

# LA GÉNOMIQUE : À PARTIR DE 2000

- Le **Génome** est l'ensemble du matériel génétique d'un individu ou d'une espèce codée dans son acide désoxyribonucléique (ADN), à l'exception des virus à ARN.
- La **Génomique** étudie la structure nucléique et le fonctionnement d'un organisme, d'un organe, d'un tissu à l'échelle du génome au lieu de la limiter à l'échelle d'un seul gène.
- Deux volets complémentaires:
  - la **génomique structurale**, qui décrit l'organisation du génome, réalise son séquençage et dresse l'inventaire des gènes,
  - la **génomique fonctionnelle**, qui étudie la fonction des gènes, leur mode de régulation et leurs interactions.



## II. Introduction à la génomique

- Explosion d'informations sur les séquences d'ADN ces dernières années.
- Nombreux **génomomes entiers** et ont été entièrement séquencés.
- Séquençage du premier ADN humain, en 10 ans et 1000 millions \$

Règnes	Génomomes	Nucléotides
Animaux	1560	200 960 91
Bactéries	30 418	53 432 805
Humain	1717	47 201 748
Plantes	293	32 947 944
Virus	113	4 226 787

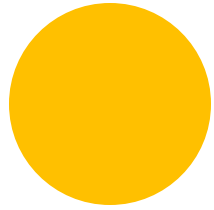
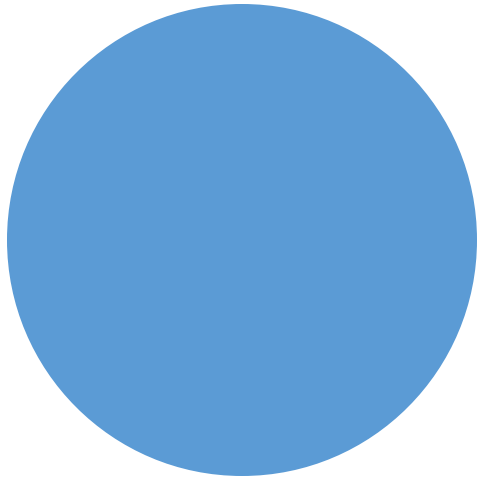
<https://www.ncbi.nlm.nih.gov/search/all/> du 20/03/19

# Génome humain

- Avant 2005, le génome de l'homme était estimé à 35 000 gènes
- Aujourd'hui estimé à 25 000 gènes avec environ 20 000 qui codent pour une protéine

# *BASES DE DONNÉES ACCESSIBLES SUR LE WEB*

- National Center for Biotechnology Information (NCBI)
- BLAST Home
  - Le logiciel BLAST recherche des régions qui présentent des similarités entre des séquences biologiques.
- BLAST “Assembled Genomes”
  - Human
  - Mouse
  - Rat
  - *Arabidopsis thaliana*
  - *Oryza sativa*
  - *Bos taurus*
  - *Danio rerio*
  - *Drosophila melanogaster*
  - *Gallus gallus*
  - *Apis mellifera*
  - Microbes, etc....



### III. Introduction à la métagénomique



# Métagénomique???

## Quelques définitions

- ❖ **Métagénome:** est l'ensemble des gènes contenus dans un environnement donné, toutes espèces mélangées
- ❖ **Microbiote:** est l'ensemble des micro-organismes (bactéries, virus, champignons, levures) vivants dans un environnement spécifique appelé **microbiome**

# Définition métagénomique

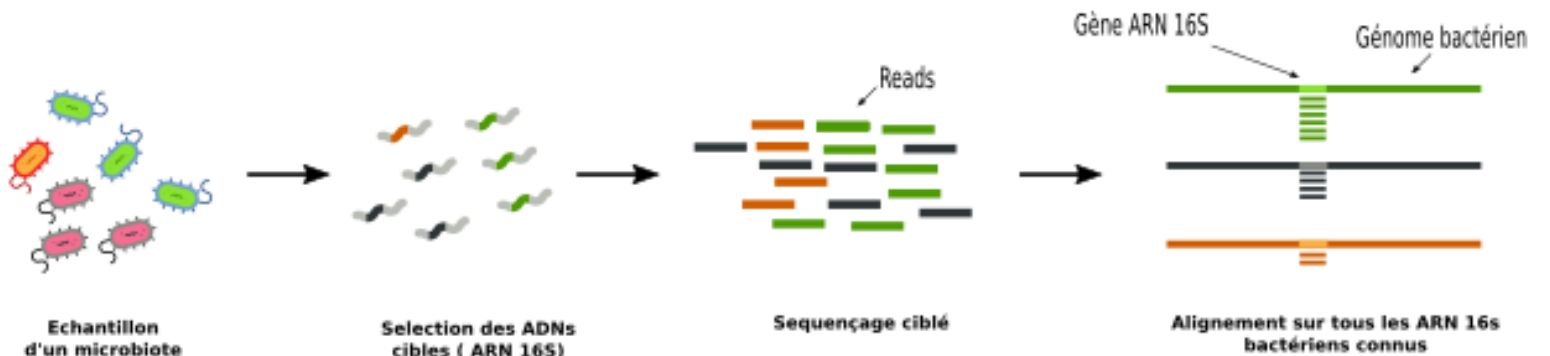
- Métagénomique = méthode d'étude du **microbiote**.
- C'est une technique de séquençage et d'analyse de l'ADN contenu dans un milieu.
- A l'inverse de la génomique qui consiste à séquencer un génome unique, la métagénomique séquence les génomes de plusieurs individus d'espèces différentes dans un milieu donné.
- Une analyse typique de métagénomique donnera la composition du microbiome (espèces, leurs abondances et leurs diversités).

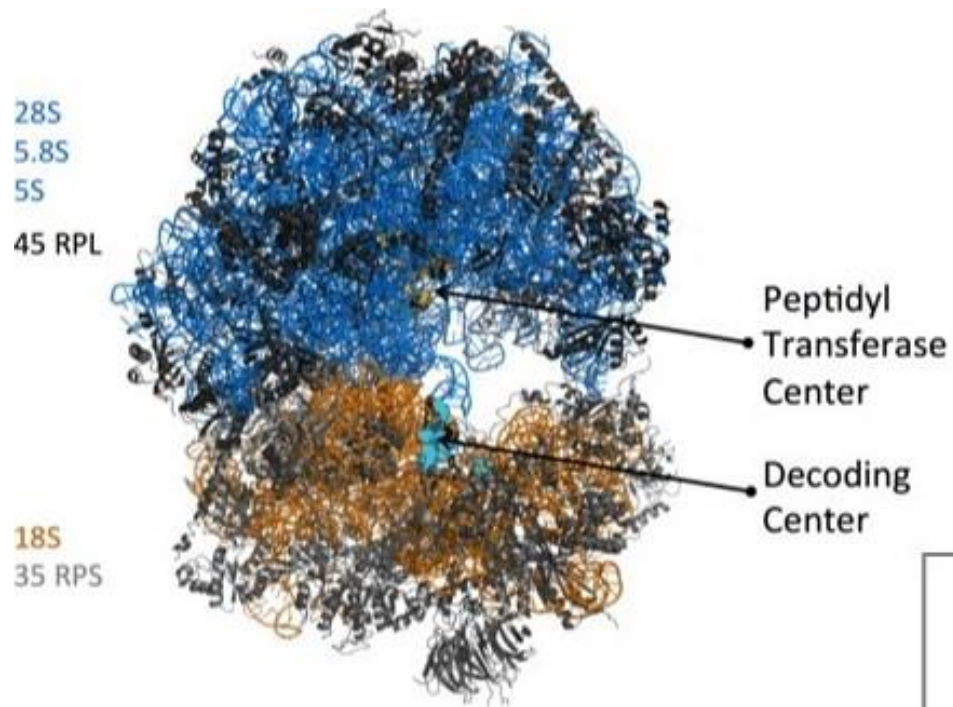
# Stratégie en métagénomique

- Deux grandes stratégies de séquençage :
  - **Métagénomique globale** : consiste à fragmenter tous les ADNs présents dans un échantillon en courts fragments et les séquencer à l'aide d'un séquenceur haut débit.

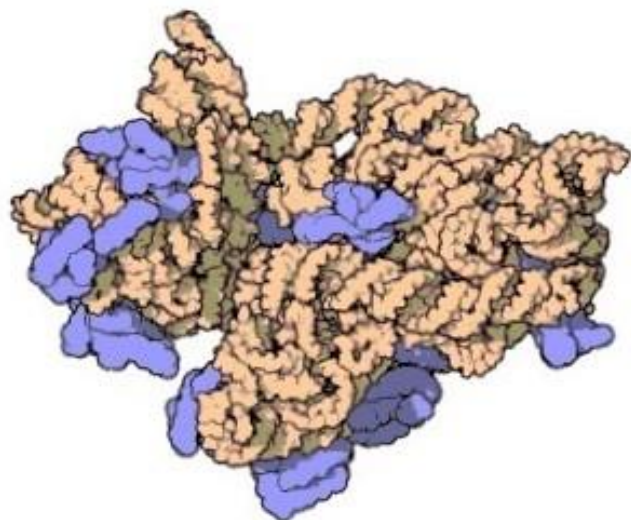


➤ **Métagénomique ciblée**: Cette stratégie consiste à séquencer un unique gène au lieu d'un génome complet (rARN 16S, commun aux bactéries).





18S  
35 RPS



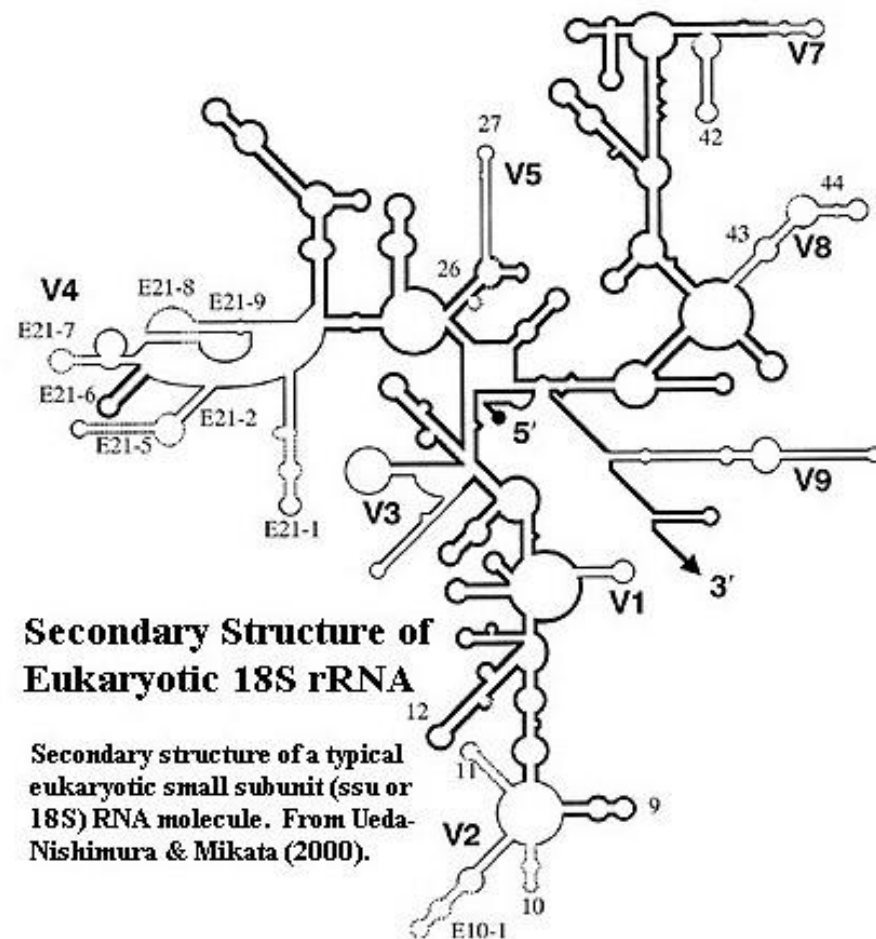
Small Sub-Unit (SSU)

16S | Bacteria  
Archaea  
Mitochondria  
Chloroplasts

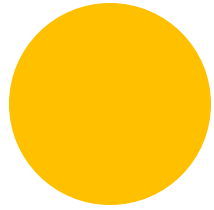
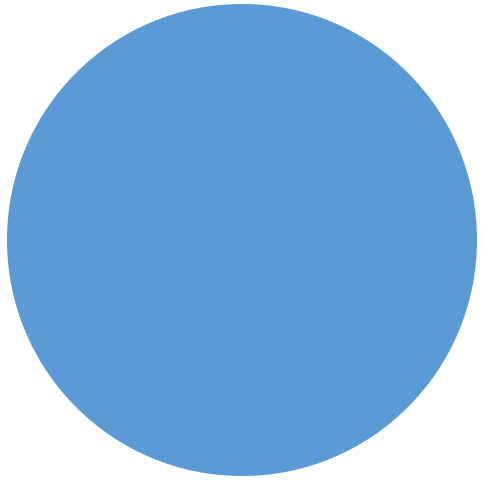
18S | Eukaryota

### Secondary Structure of Eukaryotic 18S rRNA

Secondary structure of a typical eukaryotic small subunit (ssu or 18S) RNA molecule. From Ueda-Nishimura & Mikata (2000).



other markers can be used (e.g., ITS). Requirements are: conserved distal regions for primers, variable internal regions, and available sets of reference sequences.



## IV. Séquençage- Principe généraux

# Définition

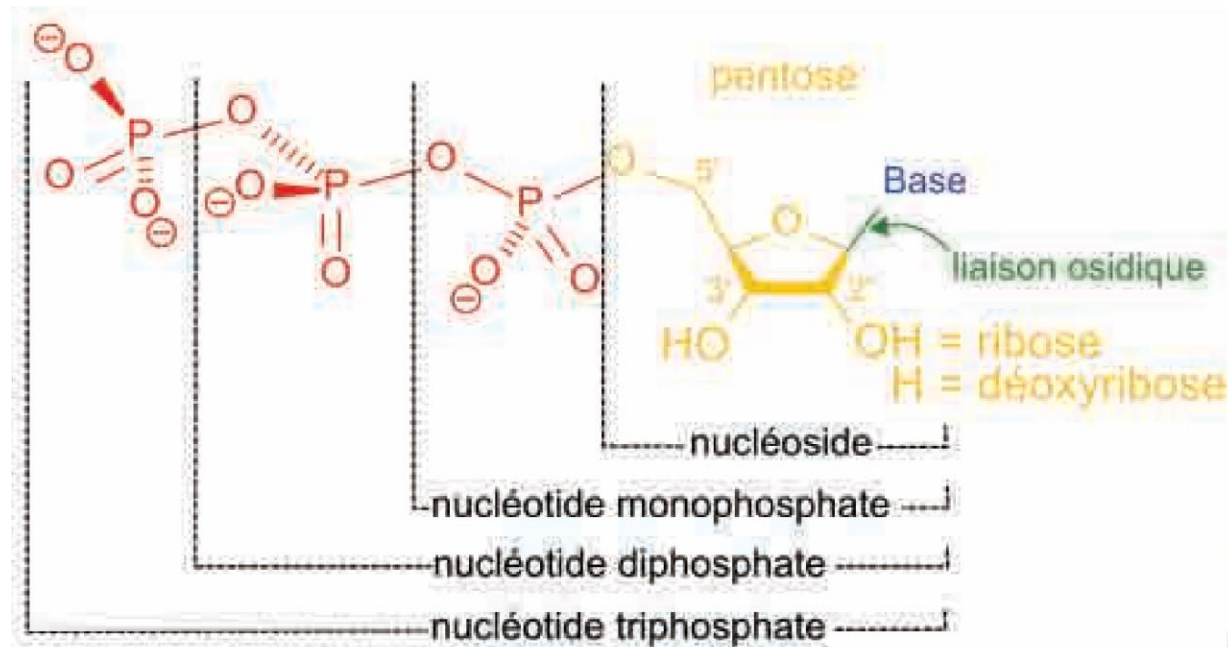
- Le séquençage de l'ADN est la détermination de la succession des nucléotides le composant.
- C'est aujourd'hui une technique de routine pour les laboratoires de biologie.
- Cette technique utilise les connaissances qui ont été acquises depuis une trentaine d'années sur les mécanismes de la réplication de l'ADN.

# Acides Nucléique

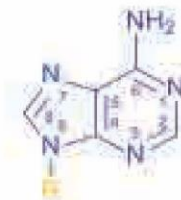
Acide nucléique = polymère de nucléotides

Nucléotides

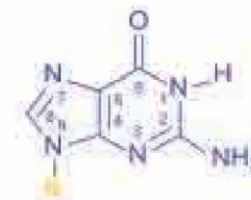
Phoebus LEVENE, 1919



## Purines

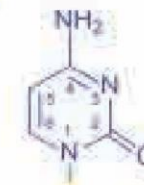


Adénine

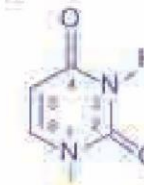


Guanine

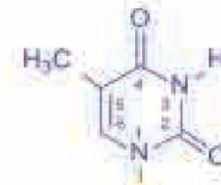
## Pyrimidines



Cytosine



Uracile



Thymine

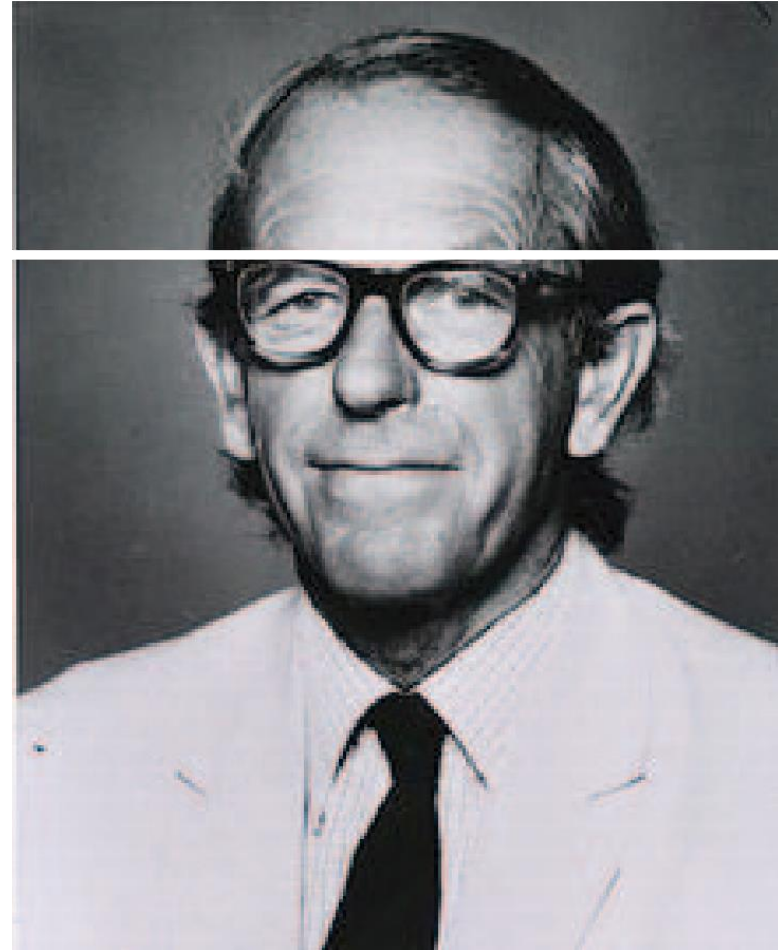
ADN et ARN ?



# SÉQUENÇAGE 1ÈRE GÉNÉRATION : MÉTHODE DE SANGER

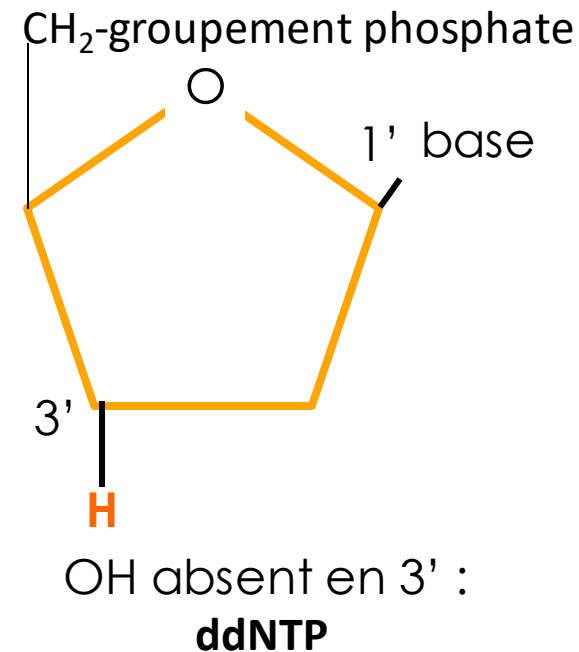
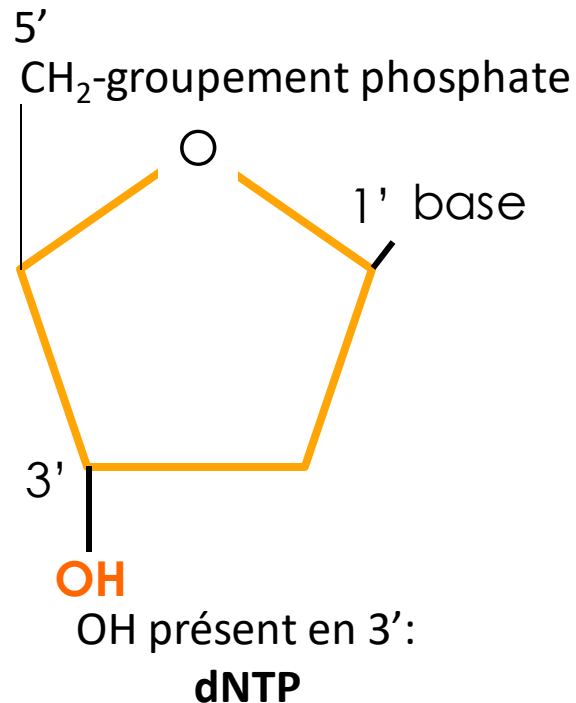
## Technique de Sanger

- Frederick SANGER (1918-2013) biochimiste anglais qui a reçu 2 prix Nobel de chimie:
  - ✓ 1958 : structure des protéines (insuline)
  - ✓ 1980 : pour le séquençage
- Les **ADN polymérases** sont capables de synthétiser un brin complémentaire d'ADN à partir d'un brin matrice.
- Pour le séquençage, des nucléotides légèrement différents sont utilisés : les **didésoxyribonucléotides (ddNTP)** au lieu des désoxyribonucléotides triphosphates (**dNTP**).



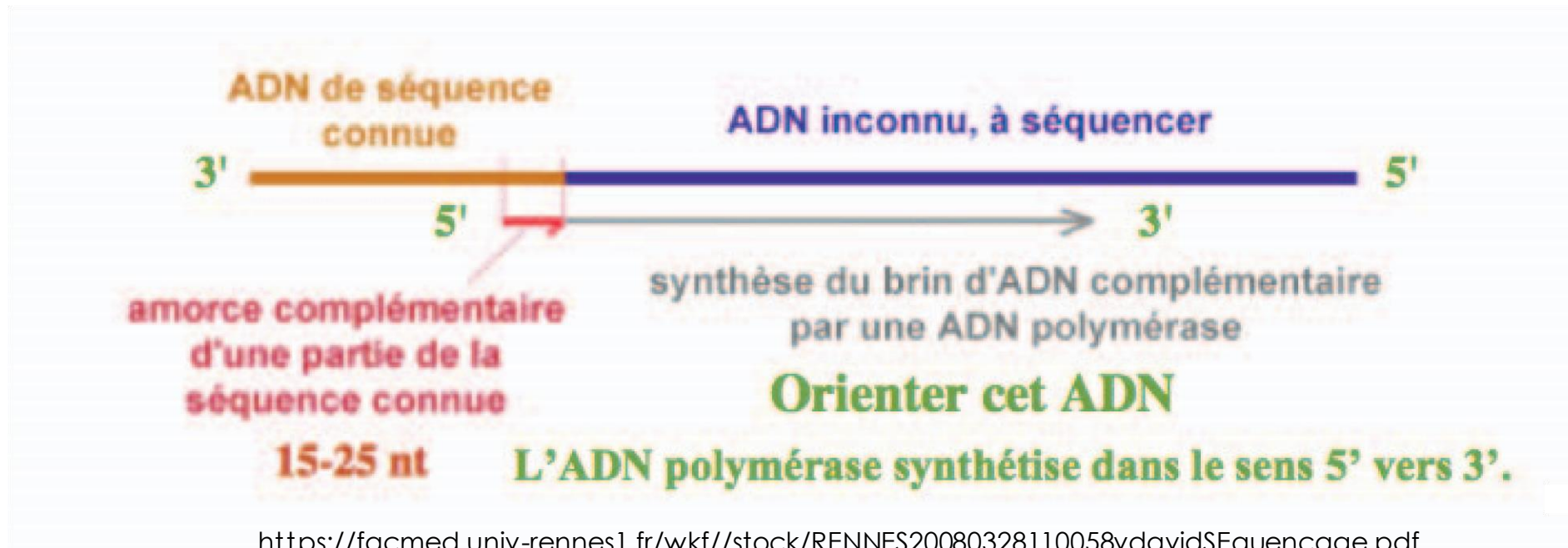
# dNTP vs ddNTP

- Les ddNTP diffèrent des dNTP par l'absence d'un groupement OH en position 3'.
- Ainsi lorsqu'une ADN polymérase utilise un ddNTP, elle n'est plus capable de rajouter le moindre nucléotide à sa suite : la synthèse du brin d'ADN s'arrête



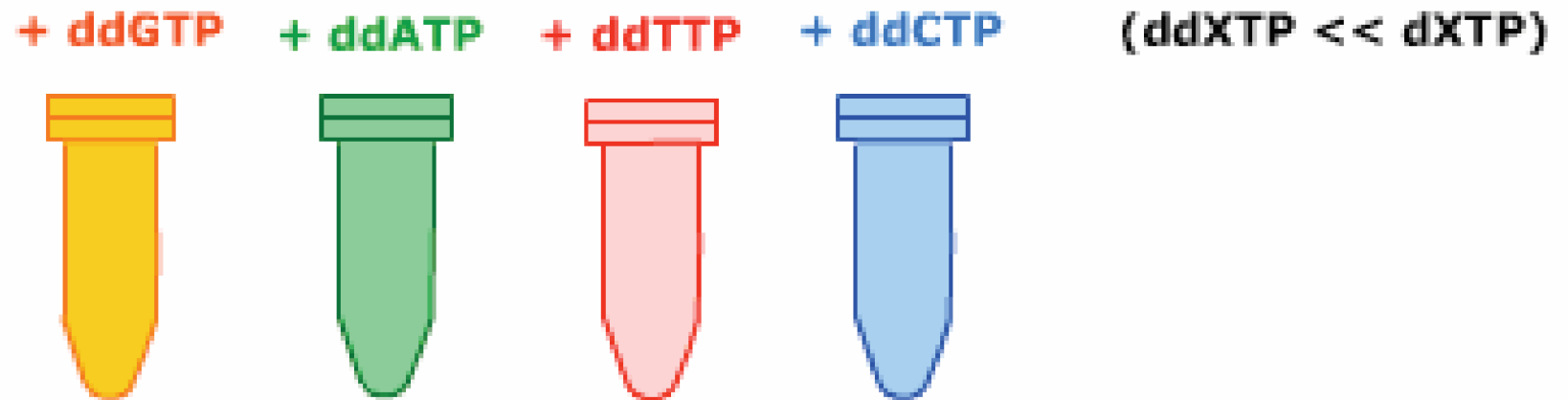
# Protocole (Sanger)

- Il faut préparer 4 solutions contenant chacune:
  - ✓ le fragment qui doit être séquencé
  - ✓ un petit morceau d'ADN dont la séquence est complémentaire à l'extrémité 3' du fragment à séquencer = amorce
  - ✓ les 4 dNTP's (dCTP, dATP, dGTP, dTTP)
  - ✓ l'ADN polymérase



# Protocole

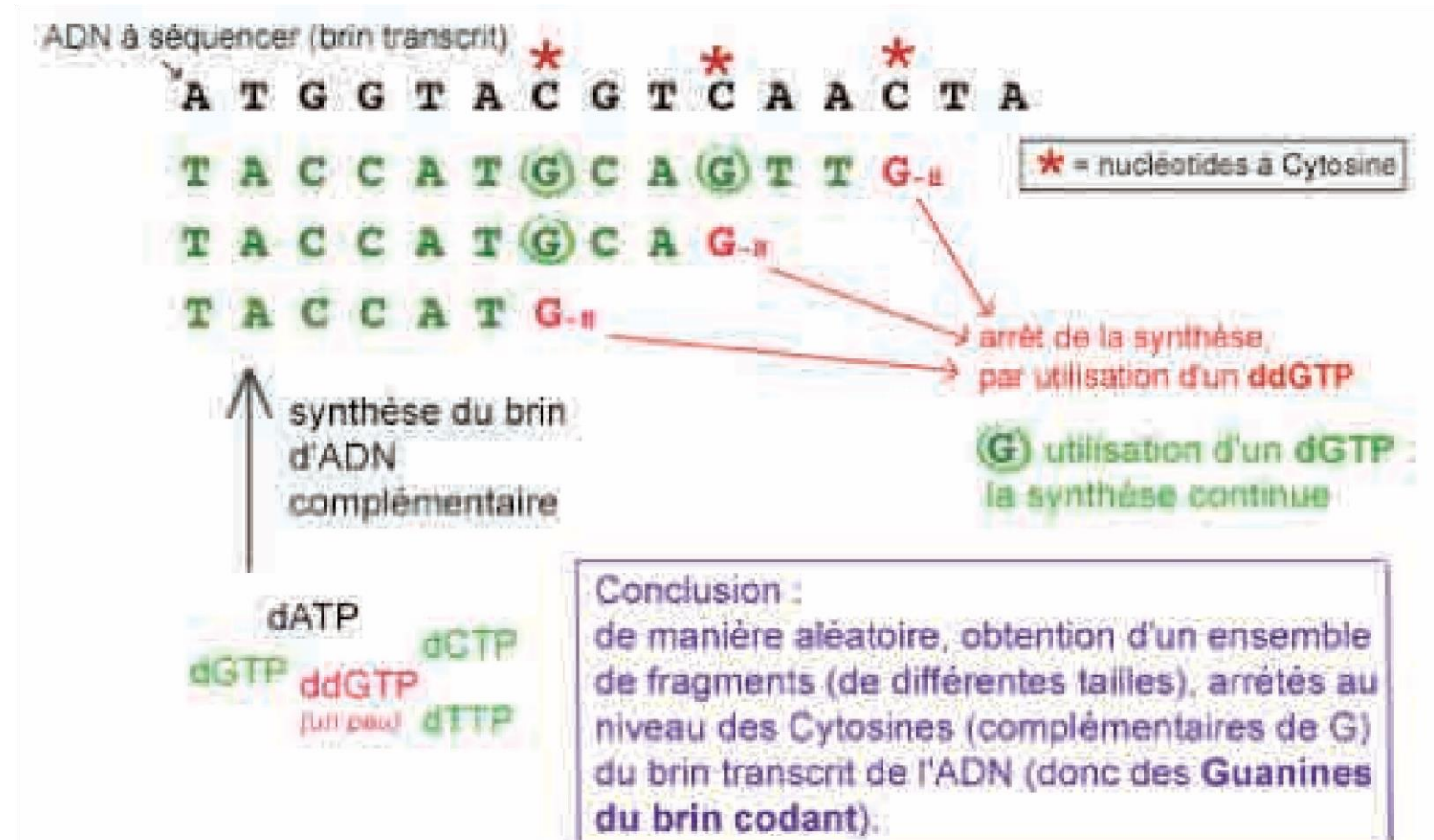
- Dans chaque tube, on met de petites quantités d'un ddNTP fluorescent ou radioactif ( $^{32}\text{P}$ )



- L'incorporation aléatoire d'un ddNTP stoppe la synthèse
  - On obtient donc à la fin des réactions un ensemble de brins d'ADN de **tailles variées**, selon l'endroit où un ddNTP se sera inséré.

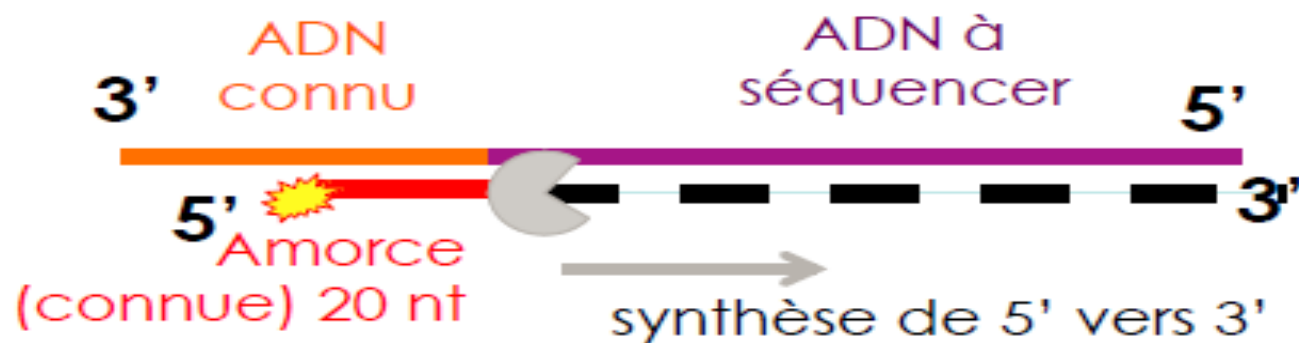
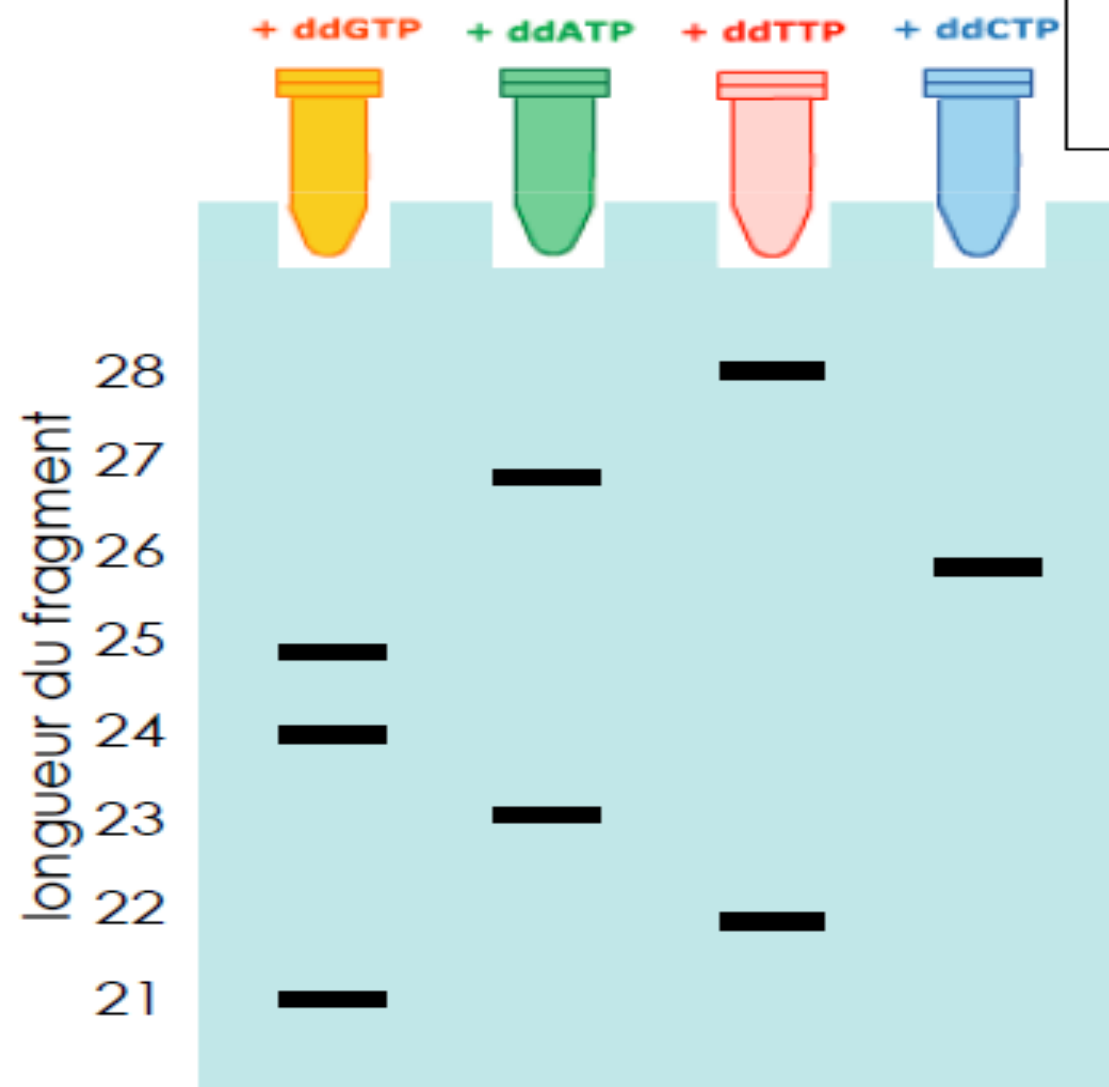
# Protocole (suite)

- Synthèse du brin complémentaire ➔ si arrêt par un ddGTP, c'est qu'il y a une Cytosine dans la séquence



# Lecture des brins

Migration électrophorétique  
(4 colonnes)



GTAGGCAT → ADN à séquencer  
5'-ATGCCTAC-3'

GTAGGCA

GTAGGC

GTAGG

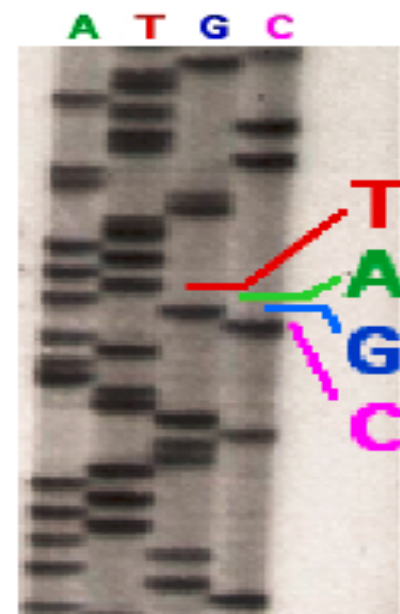
GTAG

GTA

GT

G

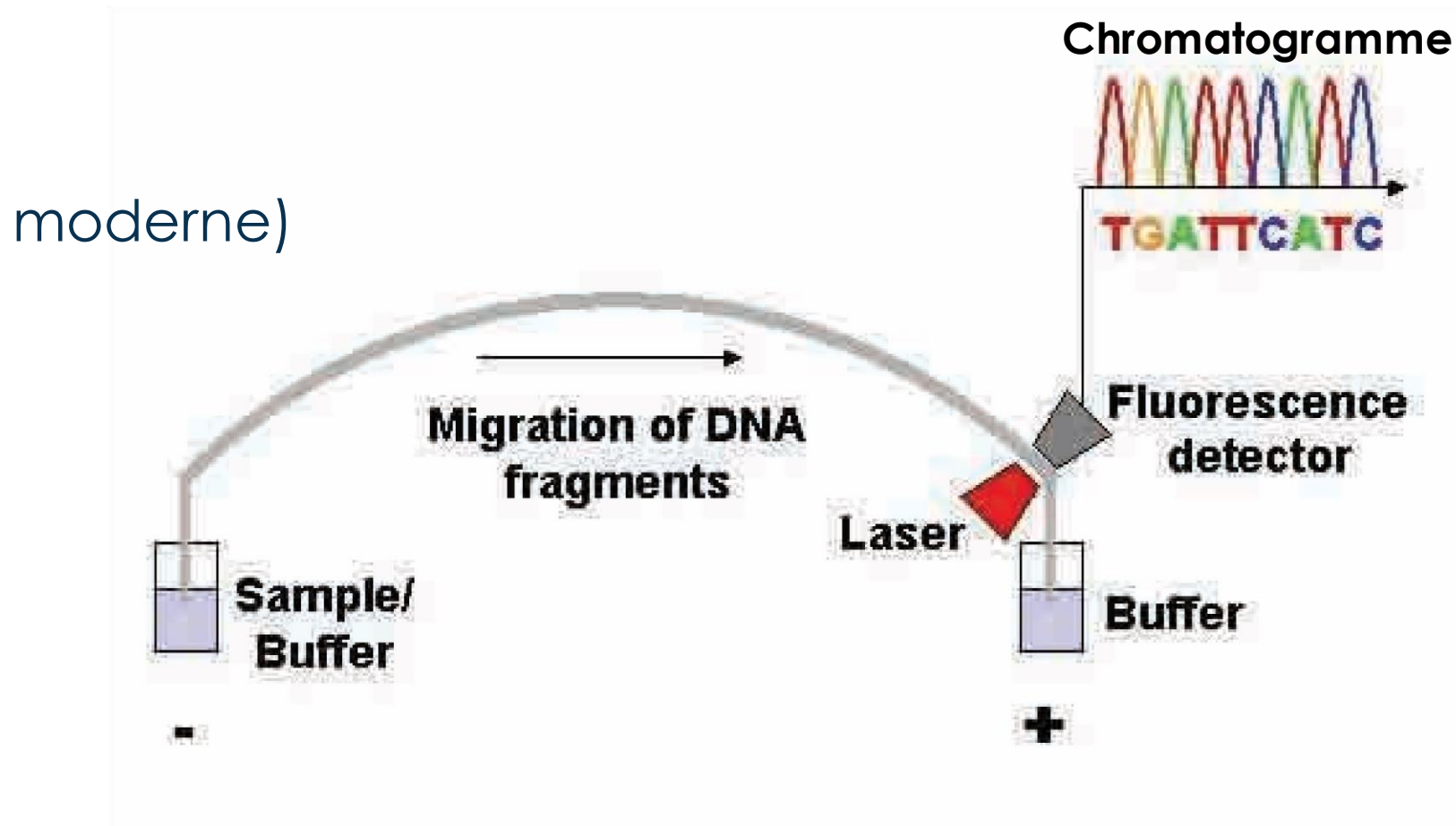
Exemple d'auto-  
radiographie  
(marquage  $^{32}\text{P}$ )  
d'un gel  
d'électrophorèse





# Lecture optique des brins

- Marquage de chaque ddNTP avec

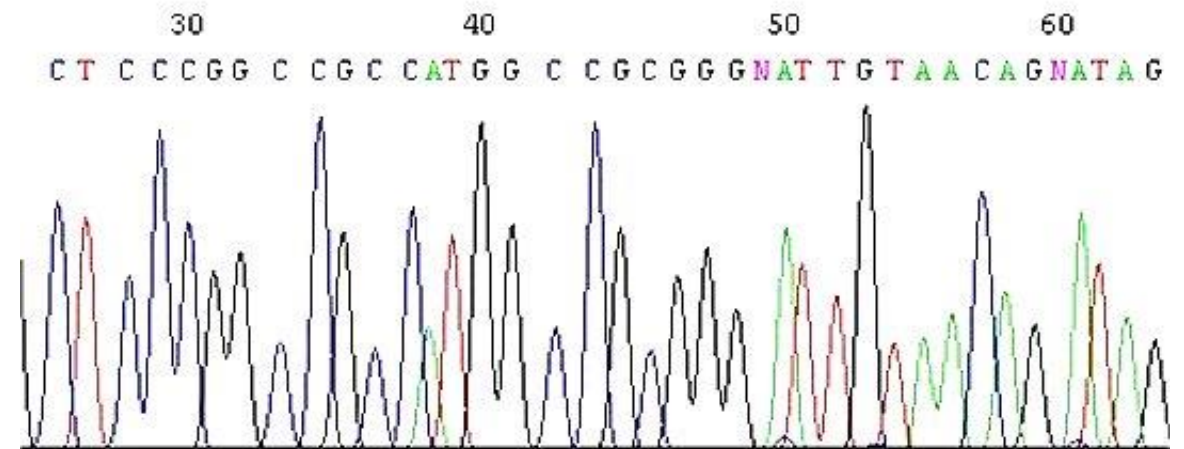
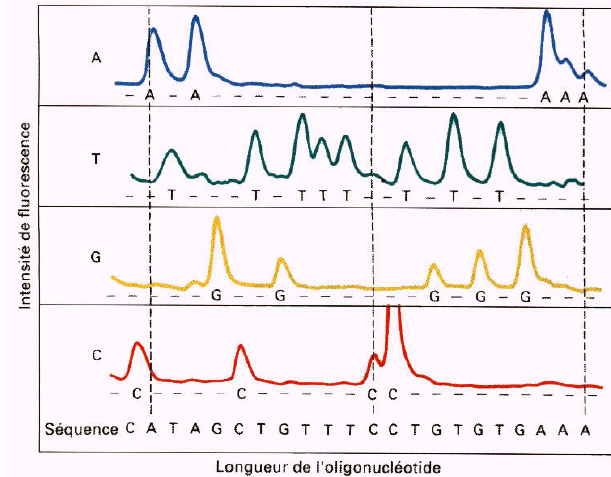


- **Avantage:** séquençage en une seule réaction au lieu de 4.

# Automatisation de la méthode

- Ajout sur les ddNTP de fluorochrome de différentes couleurs.
- 1 seul tube regroupant les ddNTP.
- Lecture par un laser des fragments qui migrent sur le gel.
- Stockage des données dans la mémoire de l'ordinateur.

- [https://agshare-my.sharepoint.com/personal/e\\_tibiri\\_agshare\\_today/Documents/Inscription%20Thèse/These\\_2019/Cours\\_TD/dna-sequencing-3d\\_0.mp4](https://agshare-my.sharepoint.com/personal/e_tibiri_agshare_today/Documents/Inscription%20Thèse/These_2019/Cours_TD/dna-sequencing-3d_0.mp4)





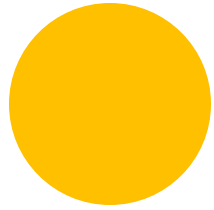
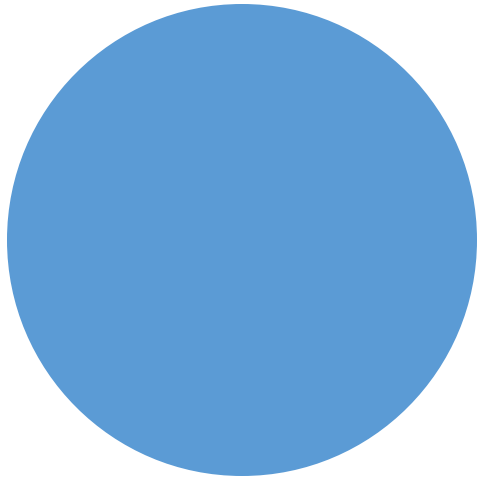
# Performances & Limitations

- **Performances des séquenceurs Sanger modernes**

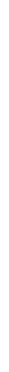
- ✓ plusieurs centaines d'échantillons simultanément et un séquençage par heure.
- ✓ Séquences de longueur 300-1000 nucléotides max

- **Limitations**

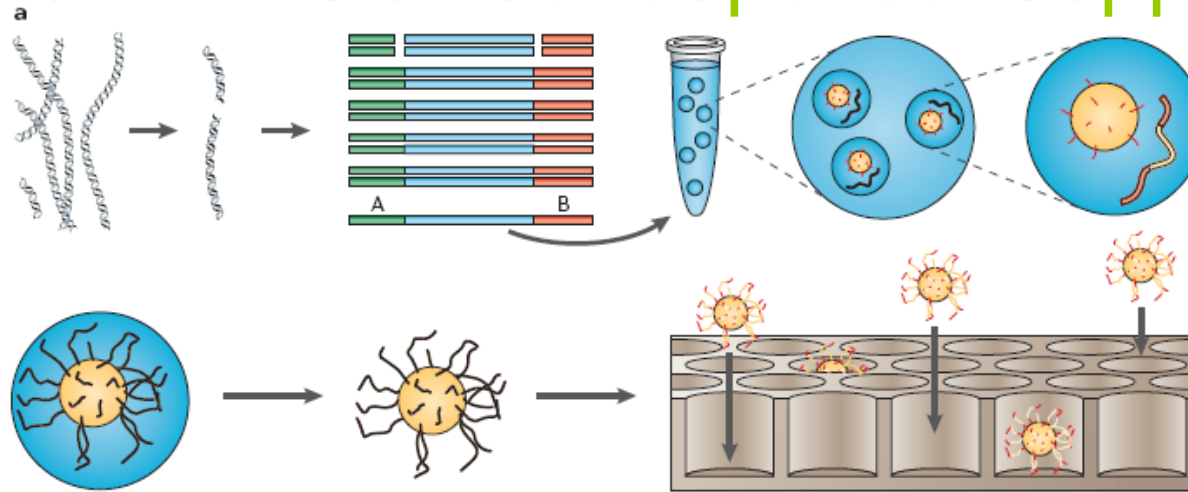
- ✓ Si amplification (PCR) avant séquençage: des parties de la séquence du vecteur d'amplification sont retrouvées dans le séquençage Sanger.
- ✓ Erreurs en début de séquence: reconnaissance imparfaite de l'amorce
- ✓ Faible pouvoir de résolution entre séquences qui ne diffèrent en longueur que d'un nucléotide



## V. Next-generation sequencing



# Un nouveau mode de préparation des échantillons et leur dépôt sur support miniaturisé

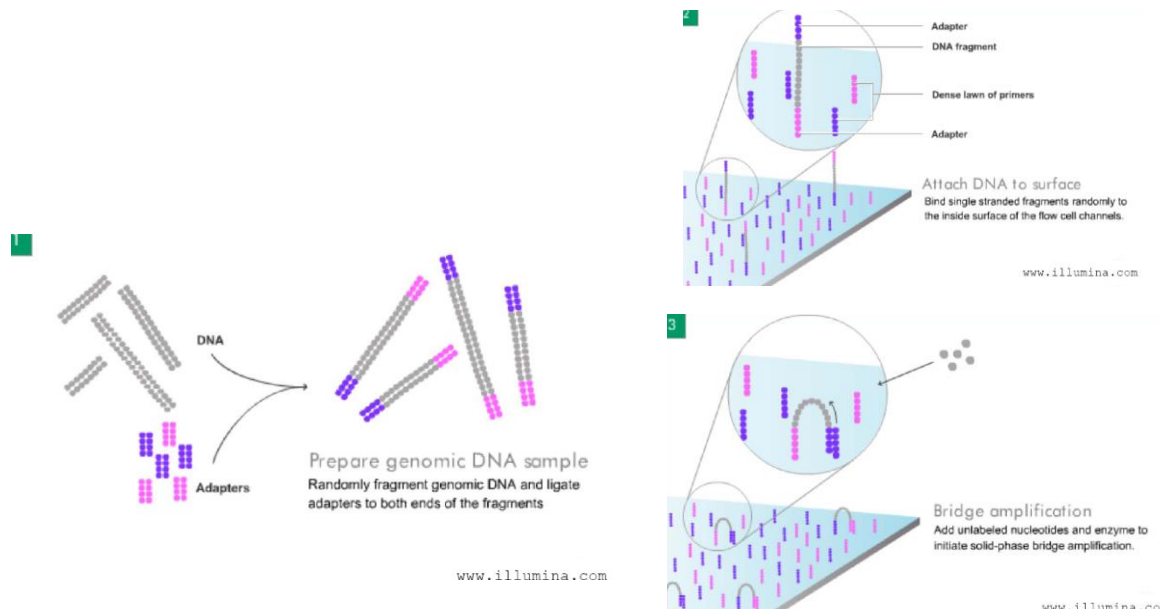


Sur une bille

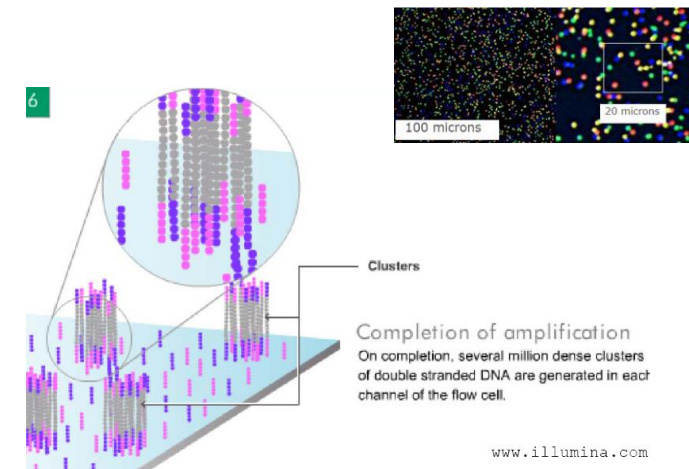
**Amplification  
simultanée d'un  
grand nombre de  
fragments d'ADN**



<http://www.454.com/>  
(Roche)

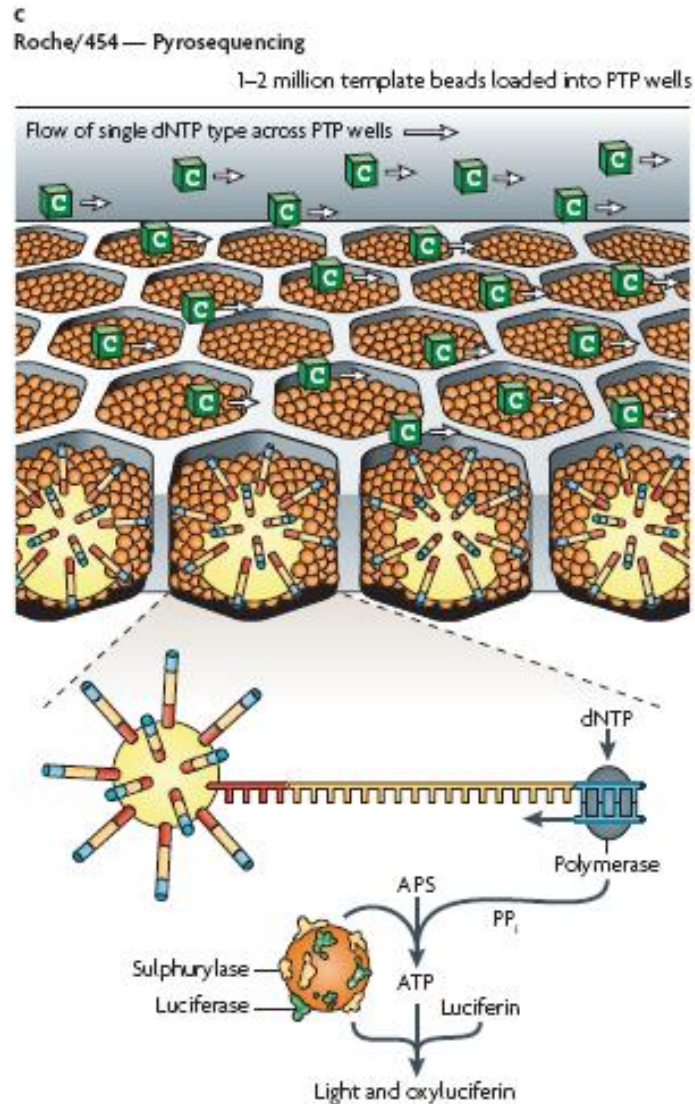


Sur une lame de verre

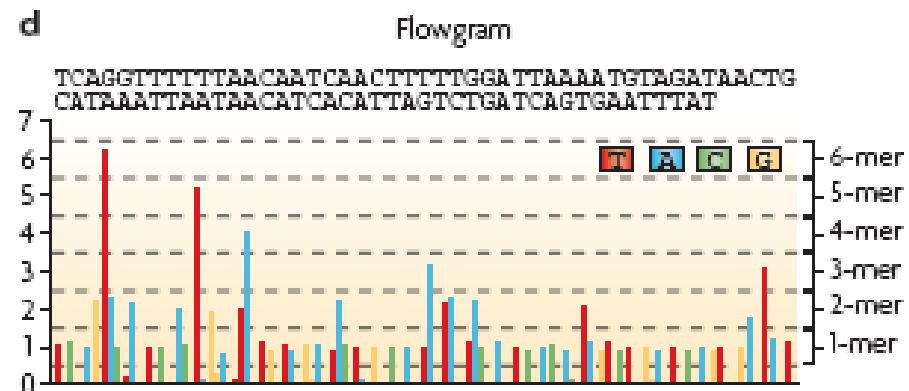


<http://www.illumina.com/>  
(Illumina)

# Pyroséquençage (454 - Roche)

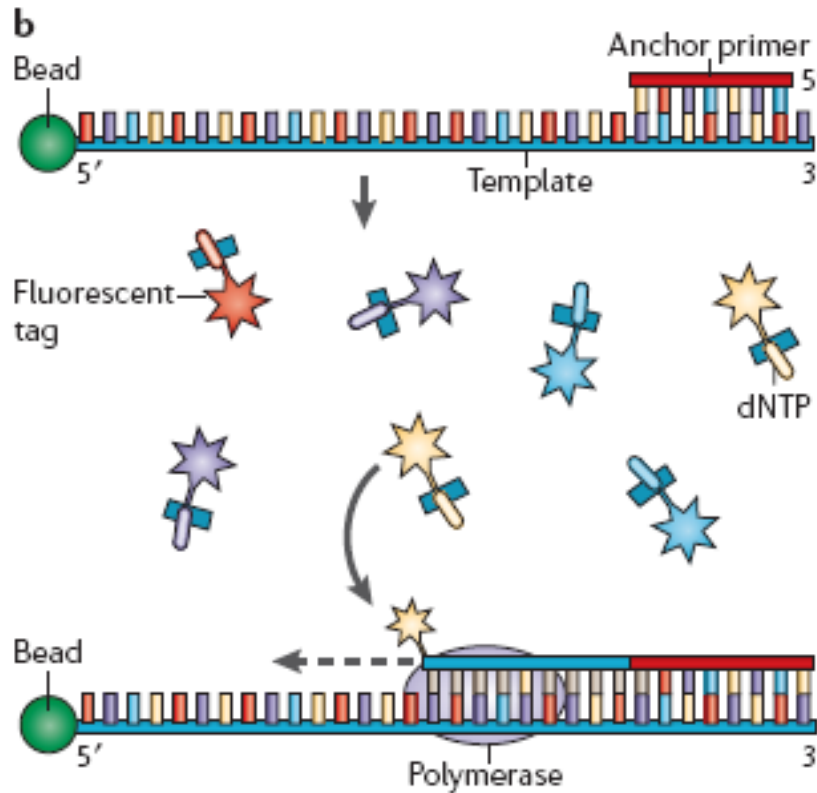


Préparation de la librairie : PCR d'ADN fixé sur bille en émulsion (1 fragment par bille)  
 Dépôt des billes dans puits (1 bille/puit)  
 Pyroséquençage : ajout d'une seule base, s'il y a incorporation alors il y a libération d'un phosphate qui va agir avec protéine qui émettra de la lumière  
 Lavage puis ajout d'une deuxième base, etc.



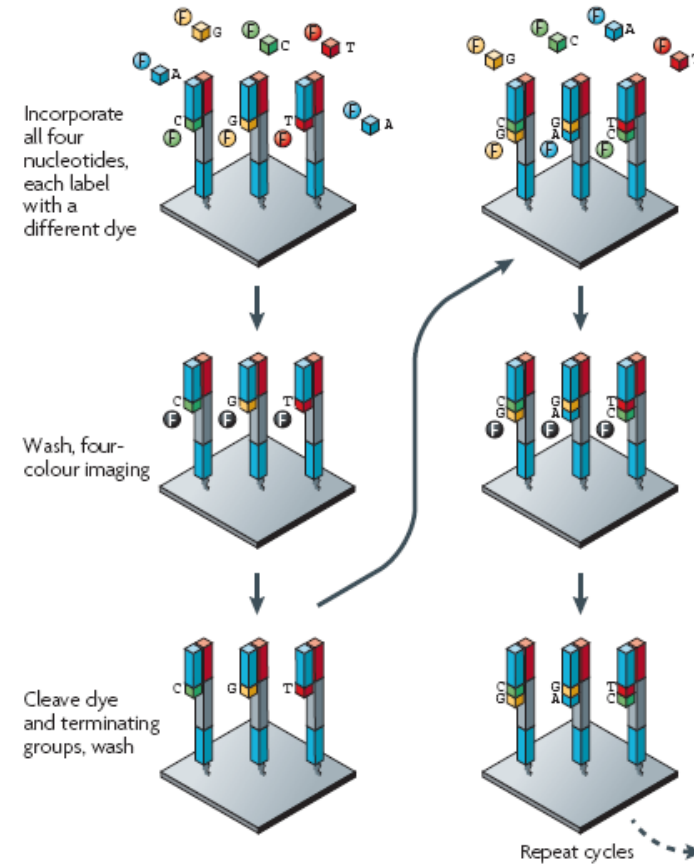
# Séquençage en temps réel

## Utilisation de terminateurs réversibles (Solexa/Illumina)



Extrémité 3' protégée - Ajoût d'un seul dNTP,  
identifié par fluorescence  
Déprotection et élimination de la fluorescence avant  
nouvel ajout de dNTPS

**a** Illumina/Solexa — Reversible terminators

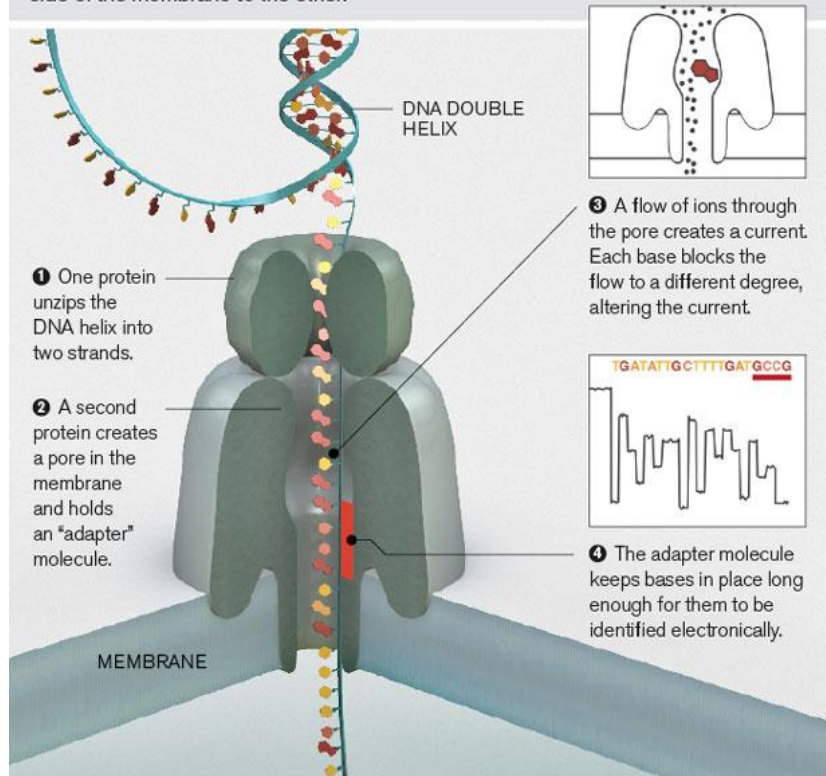


Top: CATCGT  
Bottom: CCCCCC



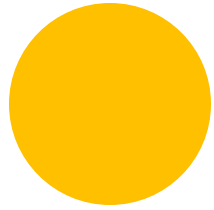
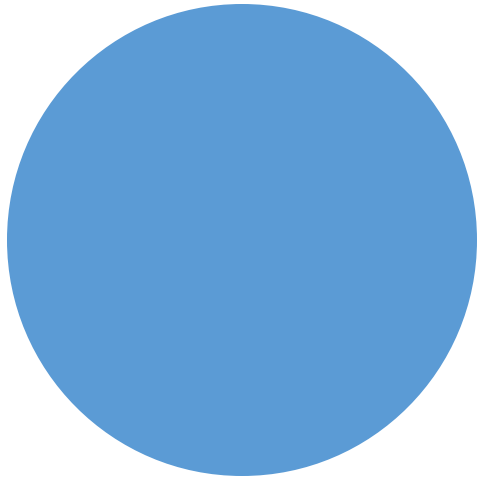
# What is Nanopore Sequencing?

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



Determine the sequence of DNA fragments by passing DNA through a protein (or other) pore in a membrane





V. Polymorphisme de séquence:  
Marqueur génétique

# Définition

- Un **polymorphisme** est simplement une différence de séquence d'ADN entre deux organismes apparentés, par exemple deux individus de la même espèce.
- Toute variation de séquence génomique entraînant l'existence d'au moins 2 formes différentes de la séquence dans la même population.
- Types de polymorphisme
  - ❖ Substitution d'un nucléotide (SNP)
  - ❖ Insertion/délétion
  - ❖ Polymorphisme de répétition (VNTR, CA repeat...)



## Notion de marqueurs en génétique : Le génotype

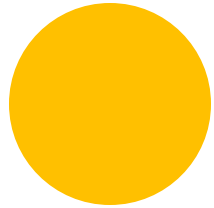
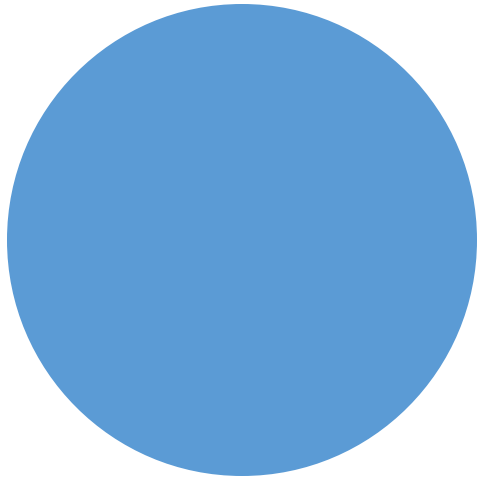
- Le marqueur génétique sert à repérer ou identifier un chromosome ou un allèle transmis
- Définissent le **génotype** (constitution génétique d'un individu) qui inclut les allèles d'un seul locus ou de l'ensemble des locus, selon les cas
- Marqueurs génotypiques :
  - variations identifiables et connues de la séquence du génome
  - de localisation connue
  - d'informativité connue
  - codominants : les 2 allèles sont détectables chez les hétérozygotes

# Les différentes types de marqueurs

- [RFLP](#), Polymorphisme de longueur des fragments de restriction,
- [AFLP](#), Polymorphisme de longueur des fragments amplifiés
- [SSLP](#), Simple sequence length polymorphism
- [RAPD](#), Amplification aléatoire d'ADN polymorphe
- [SNP](#), Polymorphisme nucléotidique,

# Les différentes types de marqueurs

- [SSCP](#), Polymorphisme de conformation des simples brins
- [EST](#), Marqueur de séquence exprimée,
- [CNV](#), Variabilité du nombre de copies,
- [VNTR](#), Séquence répétée
  - [SSR](#) (simple sequence repeat) ou polymorphisme de [Microsatellite](#),
  - [minisatellite](#),
  - [STS](#), Sequence-tagged site

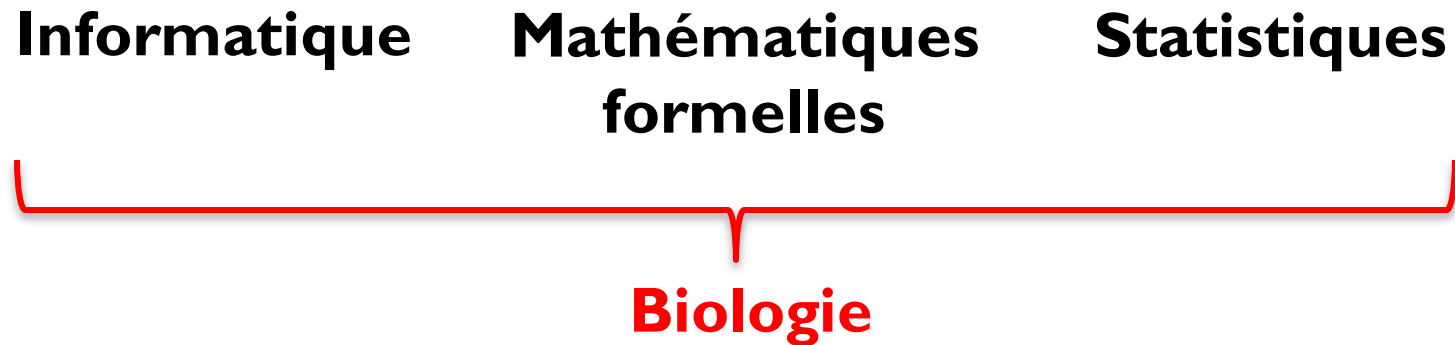


## VI. Introduction à la Bioinformatique :

Analyse de séquences

***Qu'est-ce que la bioinformatique ?***

- Ensemble de méthodes, de logiciels et d'applications en ligne qui permettent de gérer, manipuler, et analyser des données biologiques.
- La bioinformatique met en jeu plusieurs champs disciplinaires :

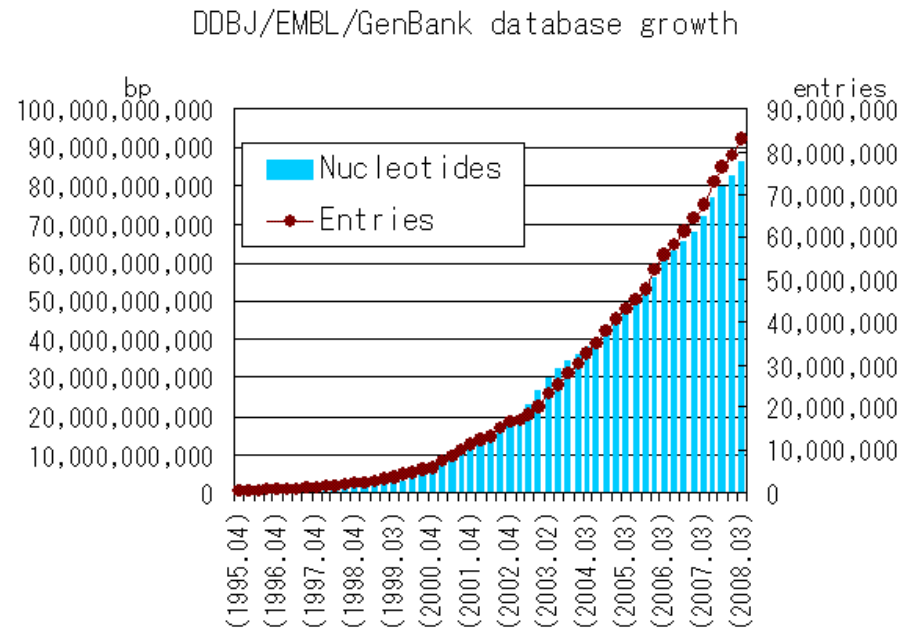


***La bioinformatique – pour quoi faire ?***

# ***La bioinformatique. Pourquoi ?***

La bioinformatique a différents objectifs et différentes applications :

**I-Collecter et stocker des informations dans des bases de données, accessibles en ligne.**



**Explosion de la  
quantité de données  
biologiques  
nécessitant des outils  
de stockage adaptés**



# ***La bioinformatique. Pourquoi ?***

La bioinformatique a différents objectifs et différentes applications :

## **2-Fournir des outils de comparaison de séquences (protéiques ou nucléotidiques).**

Séquence de  
référence



Séquence à  
analyser

*Identification ? Points communs ?*

### **Objectifs :**

- identifier une séquence par rapport à une base de données**
- déterminer le degré de similitudes entre deux séquences (intérêt en taxonomie)**
- repérer des motifs structuraux :**
  - gènes, promoteurs, etc. pour un nucléotide.**
  - zone de repliement, site actif, etc. pour un polypeptide.**

# ***La bioinformatique. Pourquoi ?***

La bioinformatique a différents objectifs et différentes applications :

## **3-Fournir des outils de traduction de séquences.**



### **Objectifs :**

- simplifier les taches de traduction**
- proposer plusieurs possibilités de protéines pour une même séquence**
- repérer exons / introns**

# ***La bioinformatique. Pourquoi ?***

La bioinformatique a différents objectifs et différentes applications :

## **4-Fournir des outils de prédiction**

**Prédiction  
physiologique et  
fonctionnelle**

### **Objectifs :**

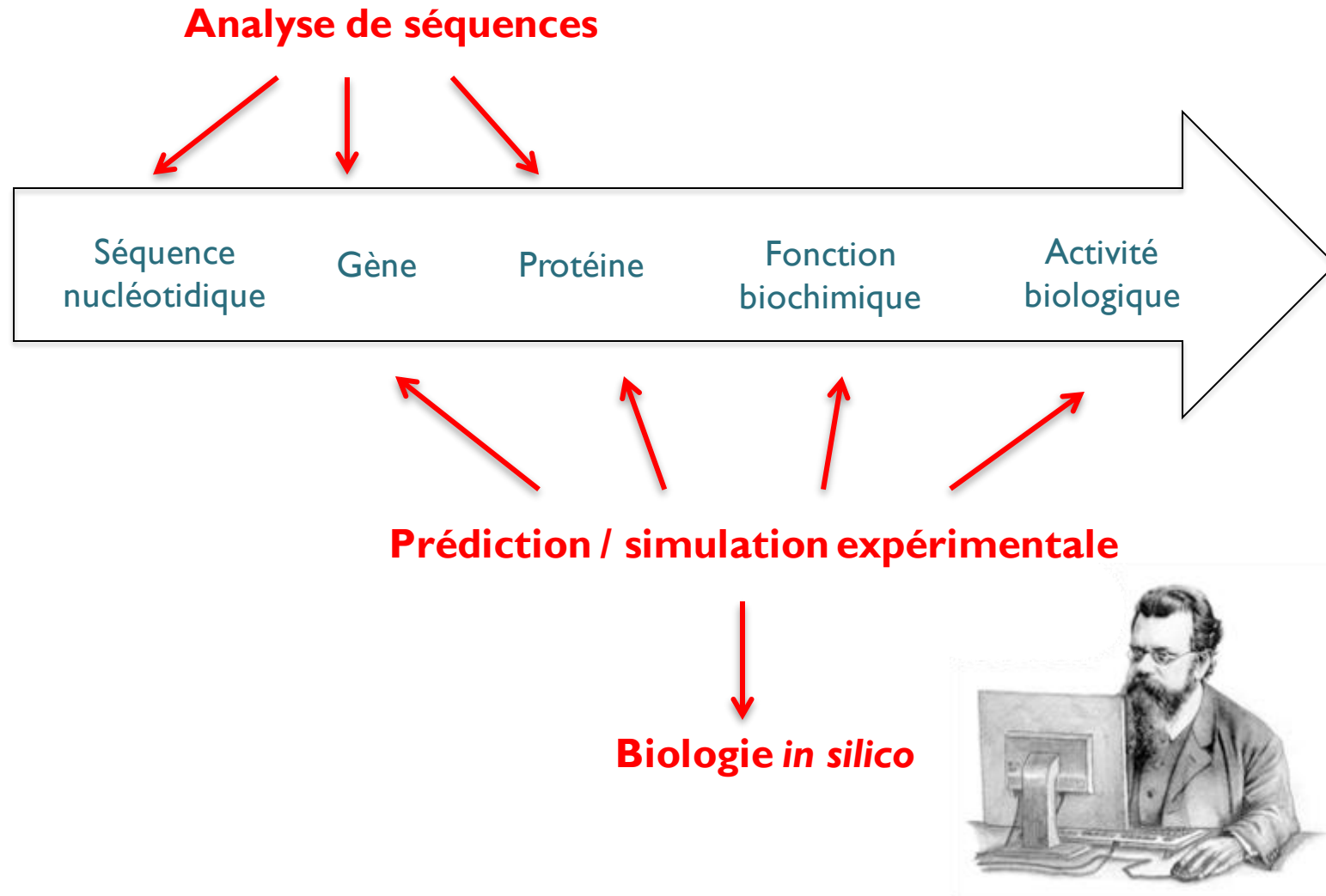
- repérer un opéron**
- repérer un gène ou une protéine anormale**
- prévoir la structure 3D d'une protéine**
- repérer des mutations**
- prédire une pathologie...**

**Prédiction  
expérimentale**

### **Objectifs :**

- repérer des sites de restriction**
- prévoir la digestion d'un nucléotide**
- prévoir / simuler la migration de fragments nucléotidiques ou protéiques lors d'une électrophorèse...**

# ***La bioinformatique. Pourquoi ?***



# Pratique: Analyse des séquences

- Du séquençage à la phylogénie

Merci pour votre attention!!!