

# Méthodes pour l'Analyse de Séquences en Bioinformatique

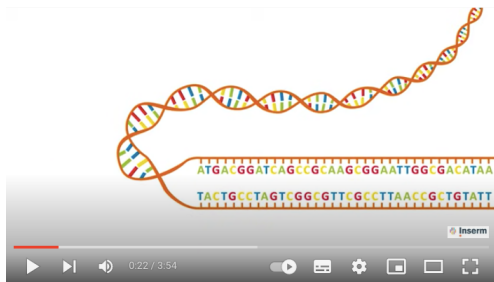
Hélène Touzet

`helene.touzet@univ-lille.fr`

CNRS, Bonsai, CRIStAL

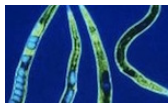
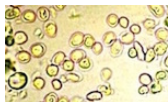
# Quelques repères historiques

## ▷ Le séquençage du génome



# 1977 : Séquençage de Sanger

## Première génération



1978 : séquence du phage phiX174 (5386 bp)

1987 : séquenceur automatique (Applied Biosystems)

1990 : lancement du programme international de séquençage du génome humain

1995 : *Haemophilus influenzae*

1996 : génome de la levure

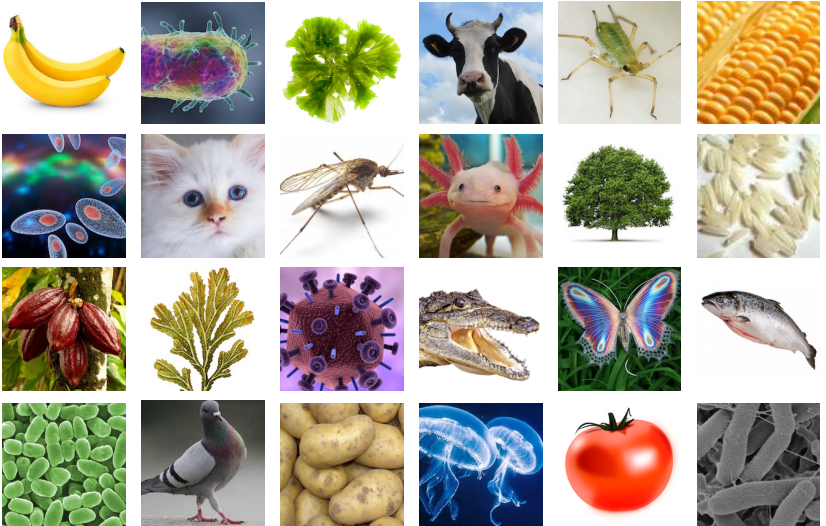
1997 : *Bacillus subtilis*, *Escherichia coli*

1998 : *Caenorhabditis elegans*  
premier organisme pluri-cellulaire

2000 : publication du brouillon du génome humain

# 2004 : Séquençage à haut débit

Deuxième génération : ABI Solid, Ion Torrent, Illumina



- coût diminué, débit augmenté → multiplication des projets de séquençage
- analyses individuelles → diversité intra-spécifique
- meilleure résolution → transcrits peu abondants, single cell
- sans culture préalable → communautés microbiennes
- matériel en faible quantité ou dégradé → ADN ancien

Get the best sale of the year. Up to **50% OFF**. Offer ends Dec 2.



The advertisement features a light gray background with several white snowflake ornaments hanging from thin lines. On the left, the text "THE BEST DNA GIFT" is rendered in large, white, 3D block letters. To the right of this text is a large magenta circle containing the text "UP TO 50% OFF" in white. Further right is a white box representing a 23andMe saliva collection kit, tied with a large magenta ribbon bow. The box has the 23andMe logo at the top, followed by the text "Welcome to you" and "saliva collection kit" below it. At the bottom of the box is a row of colorful, vertical test tubes. Below the 3D text and the discount circle is a magenta button with the text "Shop now" in white.

THE BEST DNA GIFT

UP TO 50% OFF

23andMe

Welcome to you

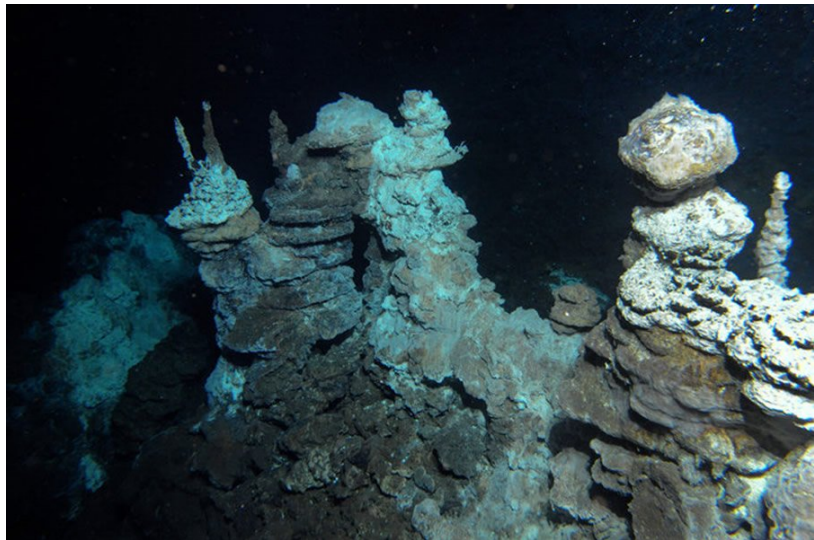
saliva collection kit

Shop now







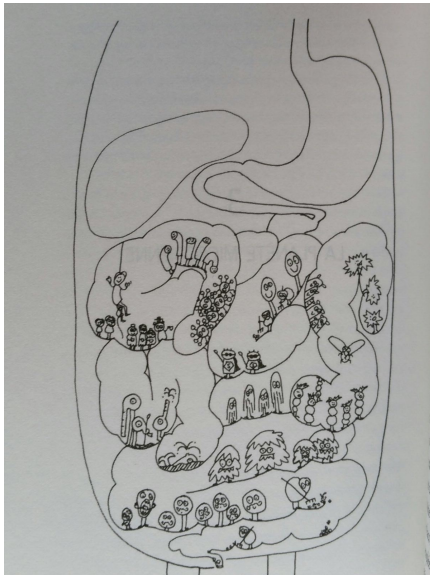


**Giulia  
Enders**

# LE CHARME DISCRET DE L'INTESTIN

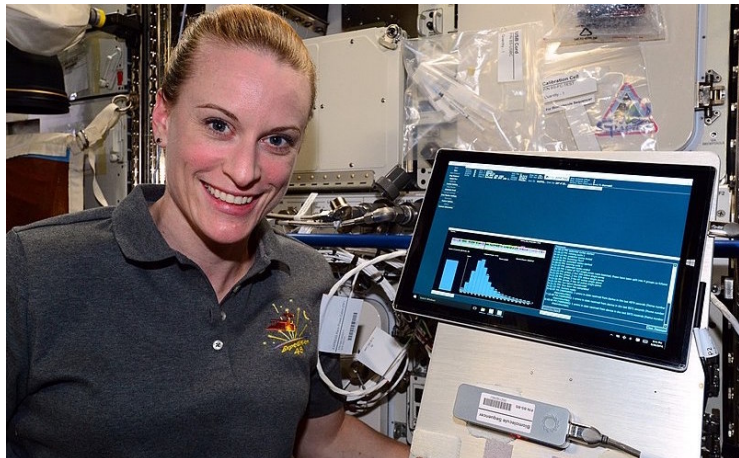
TOUT  
SUR UN ORGANE  
MAL AIMÉ...

ACTES SUD



# 2014 : Séquençage à haut débit avec longues lectures

Troisième génération : PacBio, Oxford Nanopore





- longueurs des lecture : milliers, voire dizaine de milliers de nucléotides
- résolution des longues répétitions
- réarrangements structuraux
- épissage alternatif
- phasing des SNP dans les génomes polyploides

The rate limiting step in the process of nucleic acid sequencing is now shifting from data acquisition towards the organization and analysis of that data.

Gingeras & Roberts, Science – 1980



◀  $10^3 \times 10^3$  – comparaison de gènes

- 1970 : alignement global, Needleman & Wunsch
- 1981 : alignement local, Smith & Waterman



◀  $10^3 \times 10^3$  – comparaison de gènes

- 1970 : alignement global, Needleman & Wunsch
- 1981 : alignement local, Smith & Waterman

◀  $10^3 \times 10^6 \dots 10^9$  – recherche d'homologie

- 1990 : Blast





◀  $10^3 \times 10^3$  – comparaison de gènes

- 1970 : alignement global, Needleman & Wunsch
- 1981 : alignement local, Smith & Waterman

◀  $10^3 \times 10^6 \dots 10^9$  – recherche d'homologie

- 1990 : Blast

◀  $10^6 \dots 10^9 \times 10^6 \dots 10^9$  – comparaisons de génome

- 1999 : Mummer
- 2004 : Mauve



◀  $10^3 \times 10^3$  – comparaison de gènes

- 1970 : alignement global, Needleman & Wunsch
- 1981 : alignement local, Smith & Waterman

◀  $10^3 \times 10^6 \dots 10^9$  – recherche d'homologie

- 1990 : Blast

◀  $10^6 \dots 10^9 \times 10^6 \dots 10^9$  – comparaisons de génome

- 1999 : Mummer
- 2004 : Mauve

◀  $10^8 \dots 10^{11} \times 10^6 \dots 10^9$  – lectures de séquençage

- 2009 : BWA
- 2009 : Bowtie



# MASB, organisation du module

- Planning
  - lundi de 13h30 à 16h45
  - vendredi de 9h à 12h20
- Evaluation
  - contrôle continu
  - projets de programmation

# MASB, organisation du module

- Les fondamentaux : à l'échelle des gènes et des génomes
  - Programmation dynamique et algorithmes pour l'alignement deux à deux, graines et filtrage, alignement multiple, modèles de Markov, application à la modélisation de motifs, à la prédiction de gènes
  - Mathieu Genete → projet en python, à partir du 24 janvier
  - Hélène Touzet
- Séquençage à haut débit
  - Mapping, assemblage
  - Mikaël Salson
  - Antoine Limasset