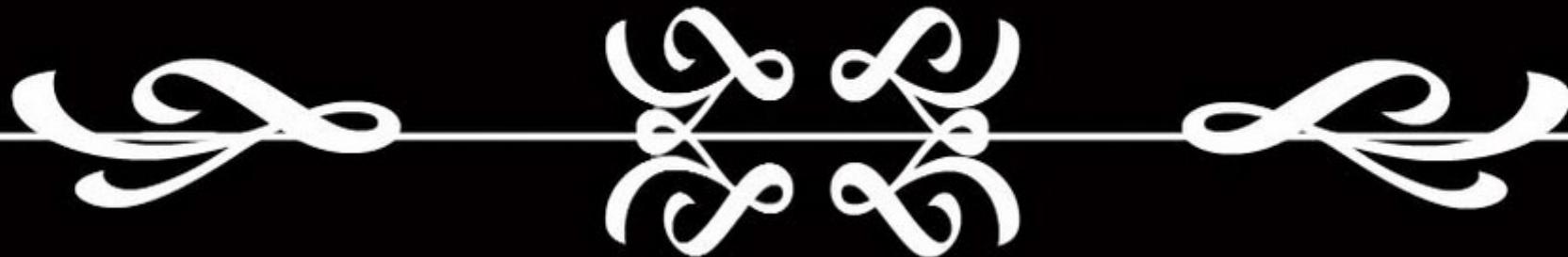


CGO

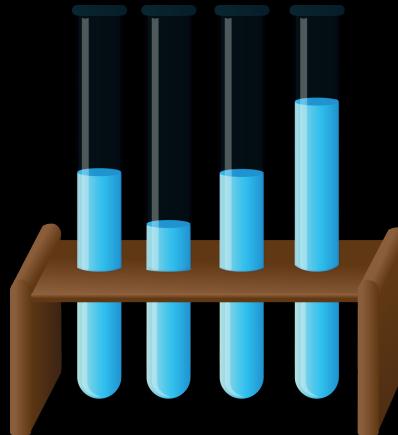
Une brève histoire  
– *de la* –  
BIOINFORMATIQUE



# **PARTIE II**

# **(1970-1990)**

# FACTEURS LIMITANTS EN RECHERCHE... ...en 2015



“Générer les données”



“Analyser les données”



# PROBLÈMES:

## 1) Notre capacité de calcul



# **PROBLÈMES:**

## **2) Surcharge de données brutes**

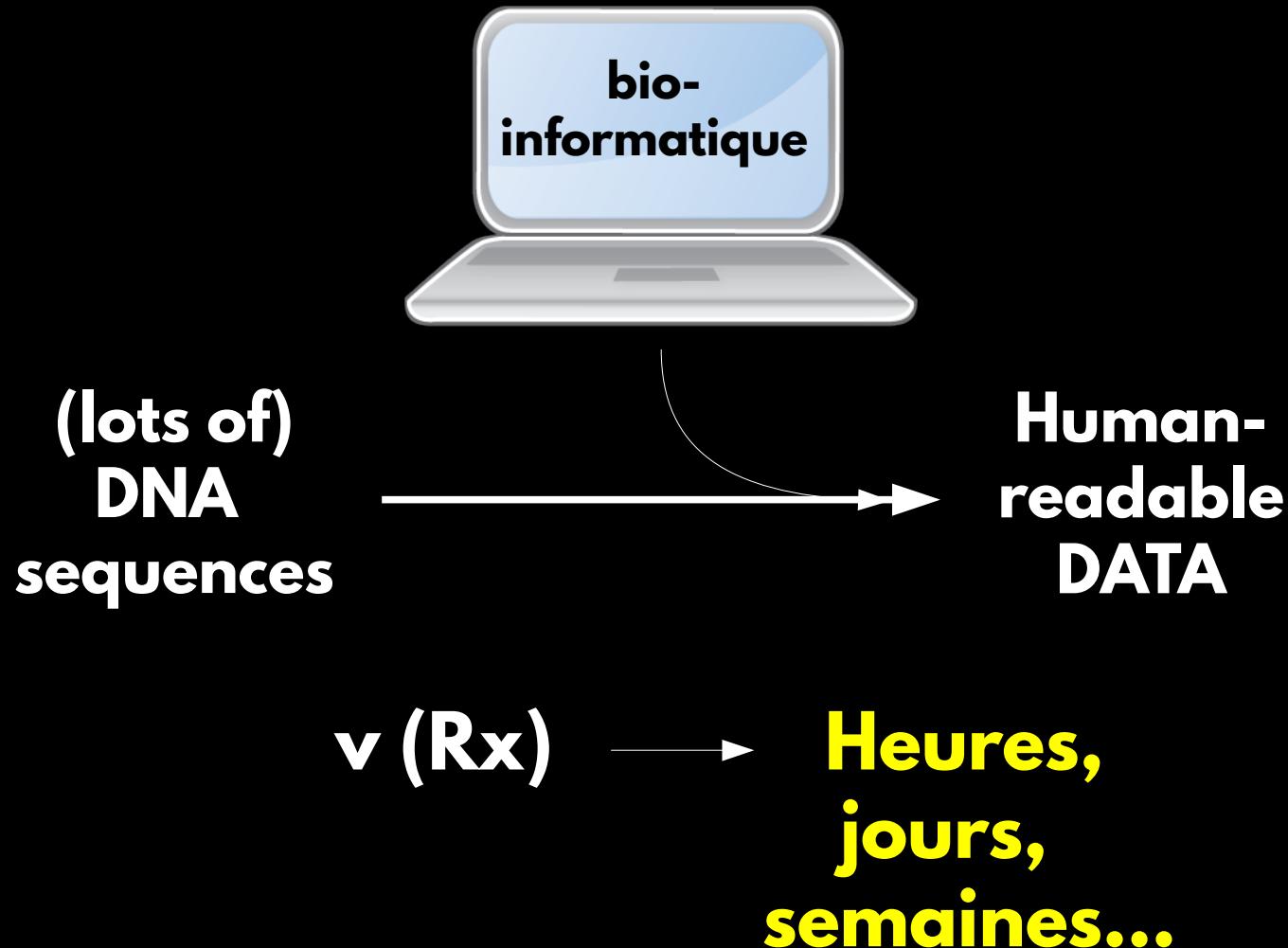


**(lots of)  
DNA  
sequences**

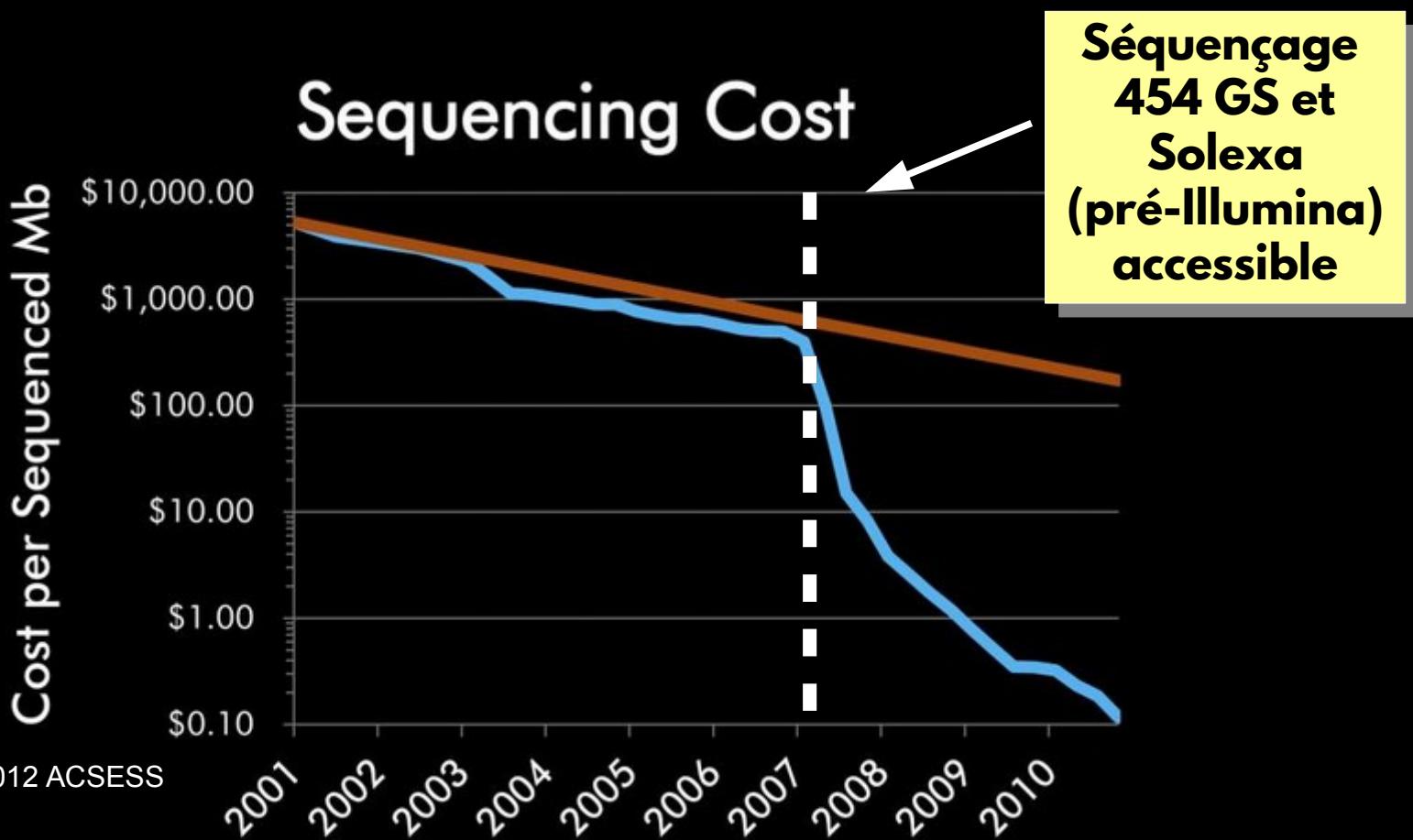


**Human-  
readable  
DATA**

$$v(Rx) \longrightarrow \infty$$



# TOUT PORTE À CROIRE QUE...



...la bioinformatique serait une discipline récente, venue prêter main forte à l'analyse de milliers de séquences d'ADN.

Est-ce vraiment le cas ?

**Dans l'épisode  
 précédent...**

**PARTIE I  
(1950-1970)**

**1952**

**ON SAIT QUE... PEU DE CHOSES SUR LA STRUCTURE DE L'ADN.**

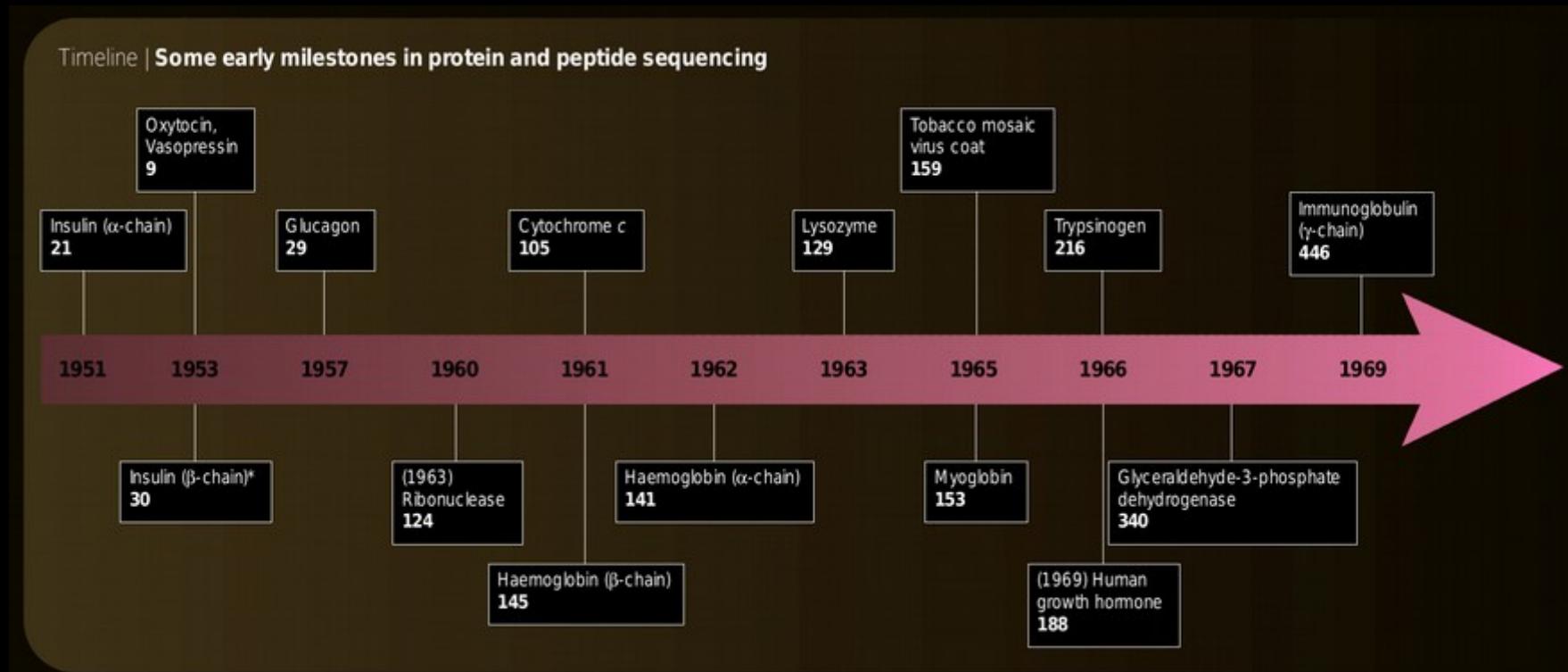
**IL FAUDRA ATTENDRE :**

- **1 an avant de connaître la structure de l'ADN;**  
(Watson et Crick 1953)
- **13 ans avant d'en déchiffrer le code;**  
(Nirenberg, Leder et al. 1965)
- **25 ans pour séquencer une molécule d'ADN.**  
(Gilbert et al. 1976 ; Sanger et al. 1977)

# 1960-1965

## OPTIMISATION DU SÉQUENÇAGE DES PROTÉINES.

### Dégradation d'Edman + Automatisation =



### PUBLICATION MASSIVE DE SÉQUENCES DE PROTÉINES.

**1960-1965**

**GÉNÈSE DE LA “ BIOLOGIE COMPUTATIONNELLE ”**

**MARGARET OAKLEY DAYHOFF**  
**(1925-1983)**

*“ ... the mother and father of bioinformatics”*  
- David Lipman (NCBI)



# 1960-1965

## GÉNÈSE DE LA “ BIOLOGIE COMPUTATIONNELLE ”

### PRINCIPALES CONTRIBUTIONS DE M.O. DAYHOFF

#### 1- Le premier “ outil bioinformatique ” (1962)

##### COMPROTEIN: A COMPUTER PROGRAM TO AID PRIMARY PROTEIN STRUCTURE DETERMINATION\*

*Margaret Oakley Dayhoff and Robert S. Ledley  
National Biomedical Research Foundation  
Silver Spring, Maryland*

KTHECAT  
ATATETH  
ATET  
ERAT  
HERA  
MIK



KTHECAT  
ATATETH  
ATET  
-----  
MIK  
-----  
**MIKTHECATATETHERAT**

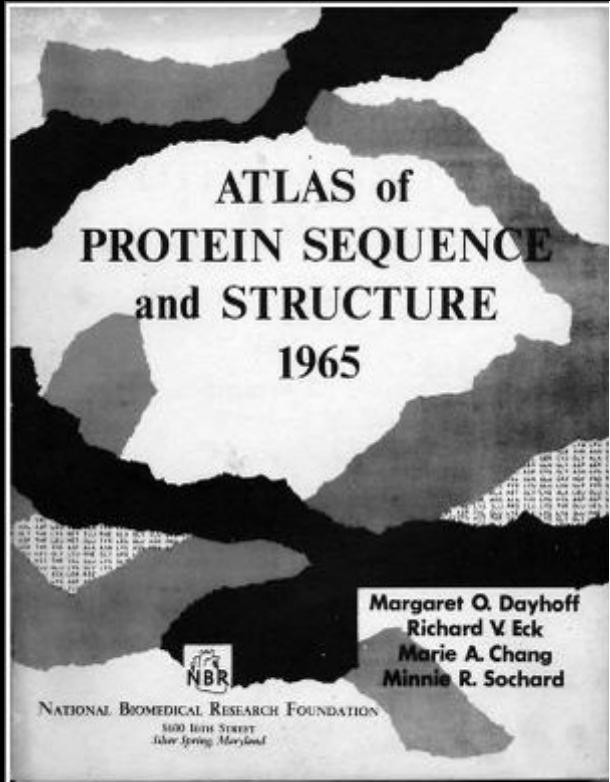
Peptide  
Peptide  
Peptide  
Peptide  
Peptide  
Peptide  
Peptide  
Protein

1965

GÉNÈSE DE LA " BIOLOGIE COMPUTATIONNELLE "

## PRINCIPALES CONTRIBUTIONS DE M.O. DAYHOFF

### 2- La première "base de données" de séquences de protéines...



BOVINE GROWTH HORMONE

5	10	15	20	25	30
1 A F P A M S L S G L F A N A V L R A Q H L H Q L A A D T F K					
31 E F E R T Y I P E G Q R Y S I Q N T Q V A F C F S E T I P A					
61 P T G K N E A Q Q K S D L E L L R I S L L I Q S W L G P L					
91 Q F L S R V F T N S L V F G T S D R V Y E K L K D L E E G I					
121 L A L M R E L E D G T P R A G Q I L K Q T Y D K F D T N M R					
151 S D D A L L K N Y G L L S C F R K D L H K T E T Y L R V M K					
181 C R R F G E A S C A R					

#### COMPOSITION

15 ALA A	11 GLN Q	27 LEU L	13 SER S
13 ARG R	13 GLU E	11 LYS K	12 THR T
6 ASN N	10 GLY G	4 MET M	1 TRP W
10 ASP D	3 HIS H	13 PHE F	6 TYR Y
4 CYS C	7 ILE I	6 PRO P	6 VAL V

MOL. WT. = 21,816

NUMBER OF RESIDUES = 191



**Maintenant,  
RETOURNONS EN 1970...**

1990

1985

1980

1975

1970

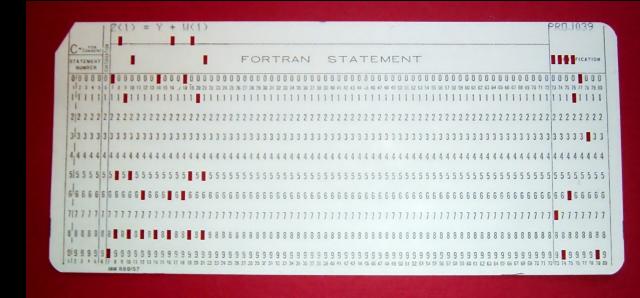
# 1960-1970

L'INFORMATIQUE : LOURDE AU SENS PROPRE ET FIGURÉ

1990

**1960-1970****L'INFORMATIQUE : LOURDE AU SENS PROPRE ET FIGURÉ**

Ordinateur typique de l'époque (IBM 7090)



Une ligne de code en 1962 (FORTRAN)



Code source d'un programme

1990  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
1970  
-

**1968**

## DÉBUTS DE LA MINIATURISATION DE L'INFORMATIQUE



**Robert Noyce**  
(co-inventeur du  
circuit imprimé)  
et  
**Gordon Moore** \*\*\*  
(chimiste)

fondent

**INTEGRATED ELECTRONICS**  
a.k.a.

**intel**®

\*\*\* Oui, c'est celui à qui nous devons la loi de Moore.

1990

**1971****DÉBUTS DE LA MINIATURISATION DE L'INFORMATIQUE**

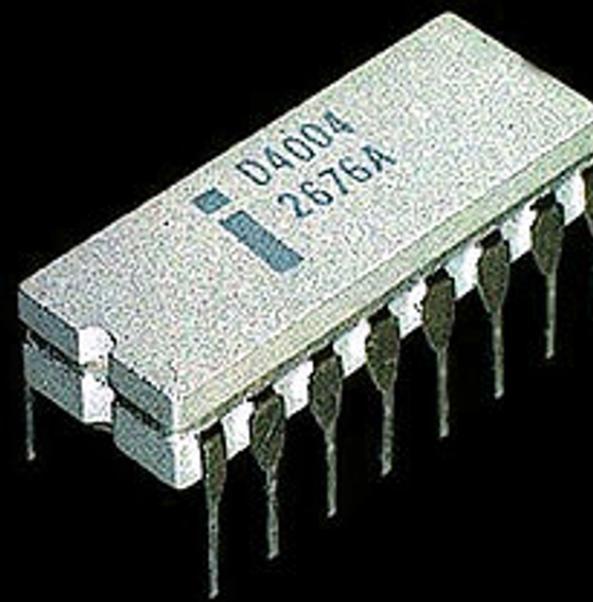
1985

**INVENTION DU PREMIER MICROPROCESSEUR**

1980

1975

1970

**intel® 4004**

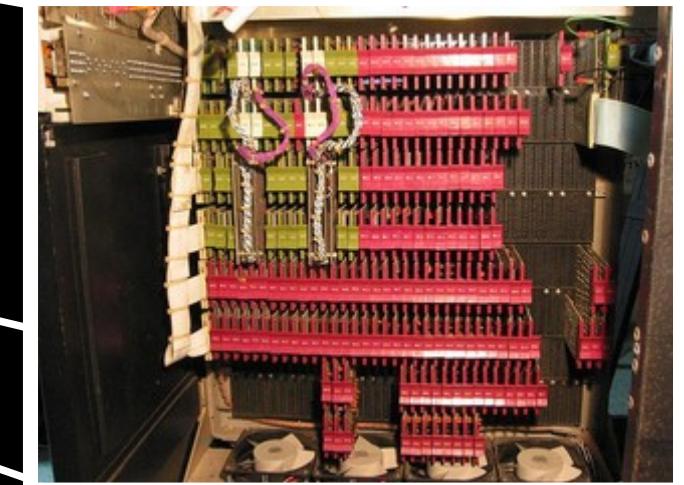


**1971**

DÉBUTS DE LA MINIATURISATION DE L'INFORMATIQUE



**PDP-8 (1968)**  
Ordinateur pré-Intel 4004



Processeur d'un PDP-8

1990

1971

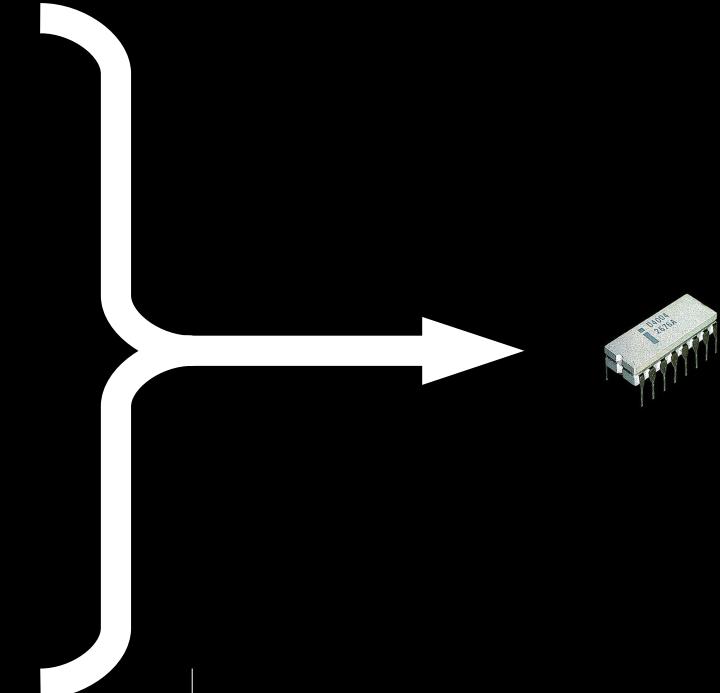
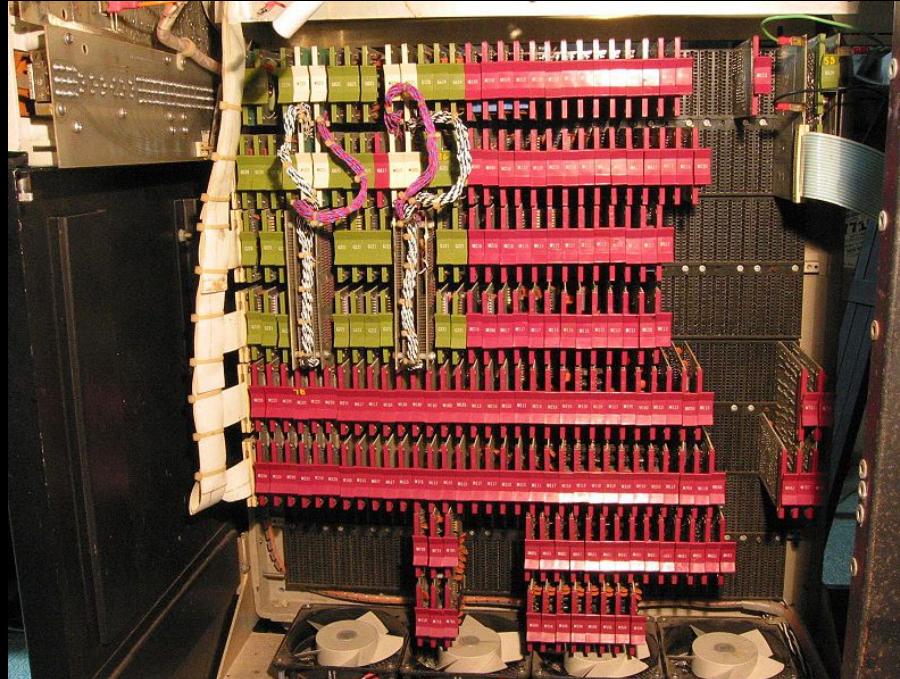
DÉBUTS DE LA MINIATURISATION DE L'INFORMATIQUE

1985

1980

1975

1970



**1975****PREMIER MICRO-ORDINATEUR VENDU AUX PARTICULIERS**

1990  
-  
-  
-  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
-  
-  
-  
1970  
-

1975

PREMIER MICRO-ORDINATEUR VENDU AUX PARTICULIERS



## ALTAIR 8800

- Processeur 2 MHz
- 256 octets de mémoire vive
- 400\$ US

...mais utilisation fastidieuse.

1990

**1975****PREMIER MICRO-ORDINATEUR VENDU AUX PARTICULIERS****...car l'ORDINATEUR EST "VIDE" !**

1985

**Il faut entrer manuellement l'interface utilisateur de deux façons:****1) en language binaire (panneau avant)**

1980

1975

1970



1990

**1975****PREMIER MICRO-ORDINATEUR VENDU AUX PARTICULIERS****...car l'ORDINATEUR EST “VIDE” !**

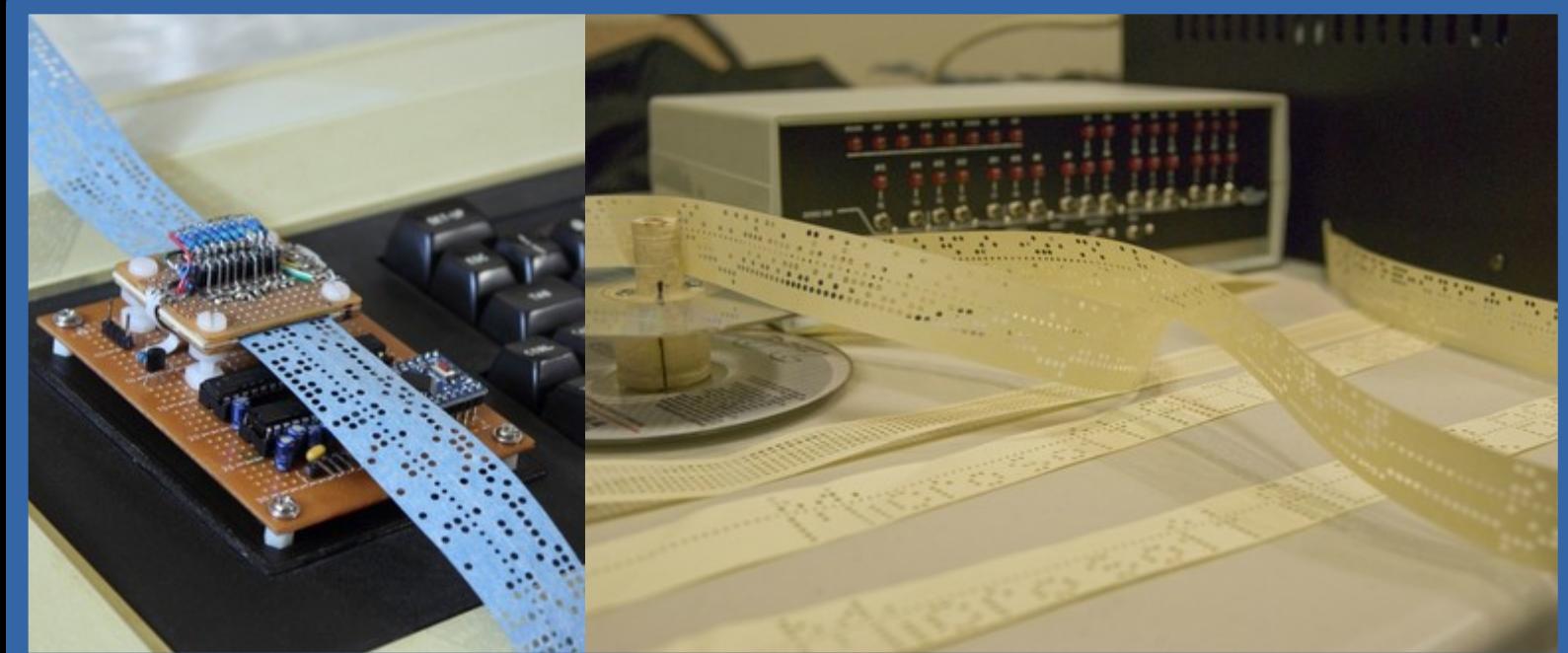
1985

**Il faut entrer manuellement l'interface utilisateur de deux façons:****2) ou en utilisant un lecteur de ruban perforé.**

1980

1975

1970



1990  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
1970

**1975**

**PREMIER MICRO-ORDINATEUR VENDU AUX PARTICULIERS**



**Seulement après ces étapes, l'ordinateur était “utilisable”...**

1990

1977

PEUT-ON FAIRE PLUS SIMPLE?

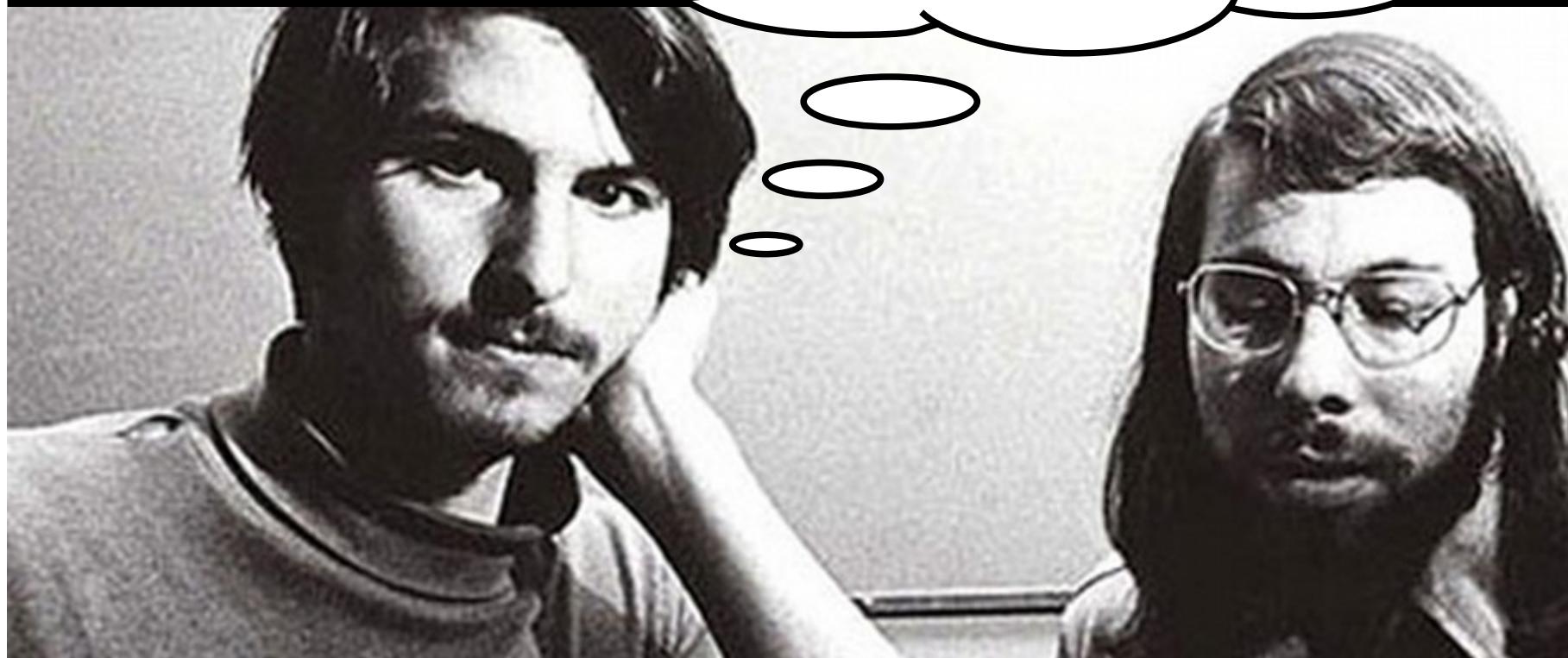
1985

Ordinateur “tout-en-un”  
déjà prêt à utiliser ?

1980

1975

1970



Steve Jobs

Steve Wozniak

1990  
-  
-  
-  
1985  
-  
-  
-  
1980  
-  
-  
-  
1975  
-  
-  
-  
1970

# 1977

## PEUT-ON FAIRE PLUS SIMPLE?



QG  
d'Apple Computer  
à l'époque

(a.k.a. le garage de  
la famille Jobs)



Apple Computer®





1977

PEUT-ON FAIRE PLUS SIMPLE?



# Apple I

“ preuve de concept ”

Matériel requis:



1990  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
1970

**1978-1985**

**LE MARCHÉ DES MICRO-ORDINATEURS EXPLOSE**



Premier micro-ordinateur  
d'Apple vendu au grand public

Modèles concurrents



**L'informatique est  
désormais accessible  
à tous...**

**1970****PREMIERS ALGORITHMES D'ALIGNEMENT DE SÉQUENCES**

1990  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
1970



1970

## PREMIERS ALGORITHMES D'ALIGNEMENT DE SÉQUENCES

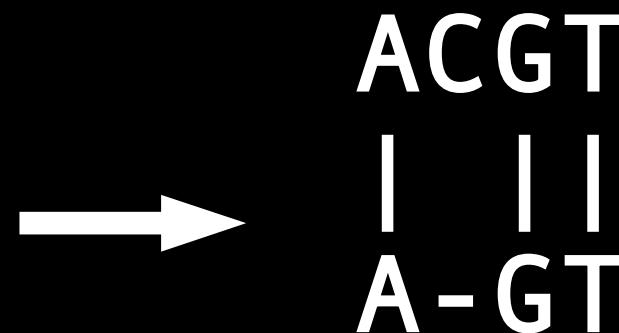
Algorithme dynamique  
NEEDLEMAN-WUNSCH

Pourquoi ?

Faire ressortir de manière non-biaisée les ressemblances  
entre deux séquences biologiques (Homologie !)

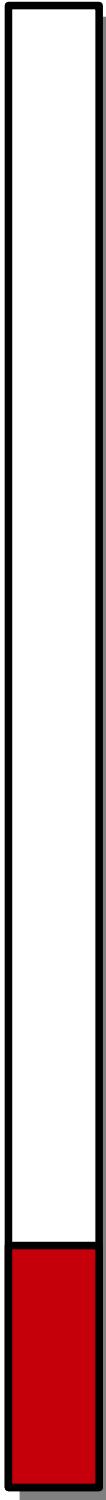
	a	c	g	t
0	-2	-4	-6	-8
a	-2	3 ← 1	-1	-3
g	-4	-1	0	4
t	-6	-3	-2	2

(match, mismatch, gap) = (3, -3, -2)



**1972****DÉBUTS DE LA BIOLOGIE MOLÉCULAIRE**

1990  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
1970



1990

1972

## DÉBUTS DE LA BIOLOGIE MOLÉCULAIRE



PAUL BERG

Prix Nobel de chimie 1980

*“for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA”*

1985

1980

1975

1970

*Proc. Nat. Acad. Sci. USA*  
Vol. 69, No. 10, pp. 2904–2909, October 1972

**Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli***

(molecular hybrids/DNA joining/viral transformation/genetic transfer)

DAVID A. JACKSON\*, ROBERT H. SYMONS†, AND PAUL BERG

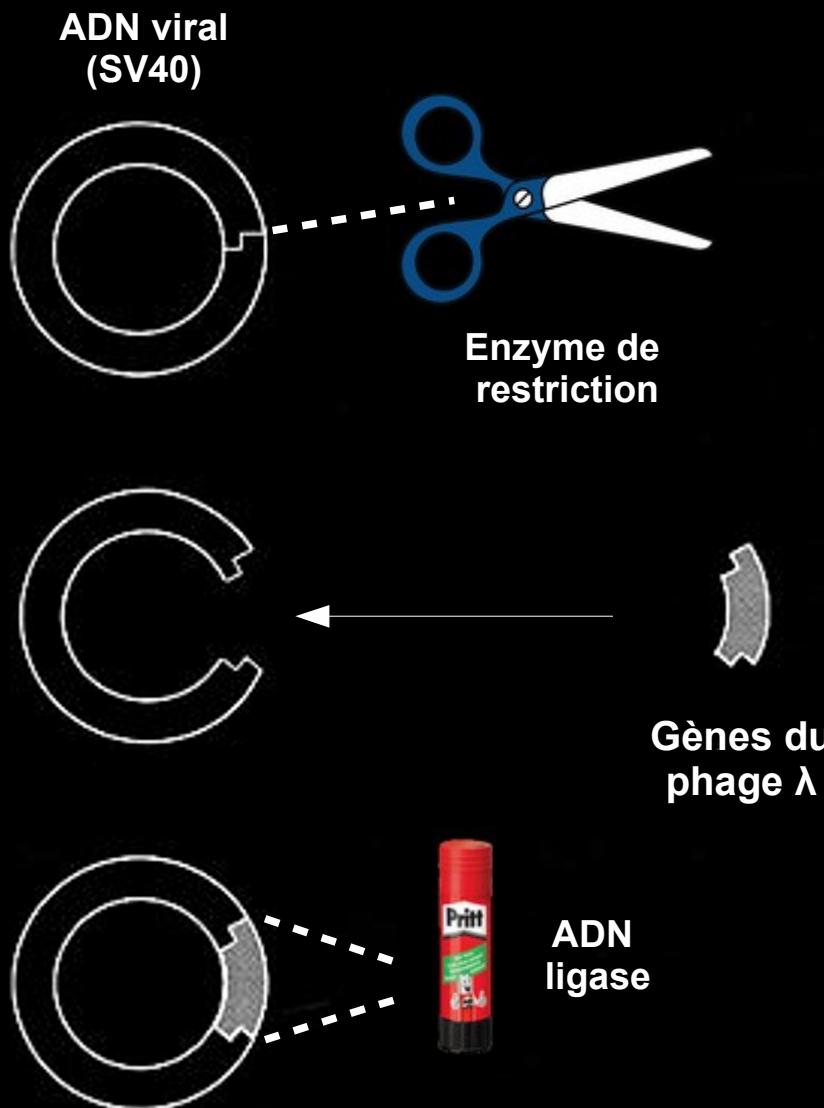
Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305

*Contributed by Paul Berg, July 31, 1972*



**1972**

## DÉBUTS DE LA BIOLOGIE MOLÉCULAIRE



**Il s'agissait de la première molécule artificielle**

**D'ADN RECOMBINANT DE L'HISTOIRE.**

**La suite ?**

1990

**1975****DÉBUTS DE LA BIOLOGIE MOLÉCULAIRE**

1985

**1975: CONFÉRENCE  
D'ASILOMAR SUR  
L'ADN RECOMBINANT**

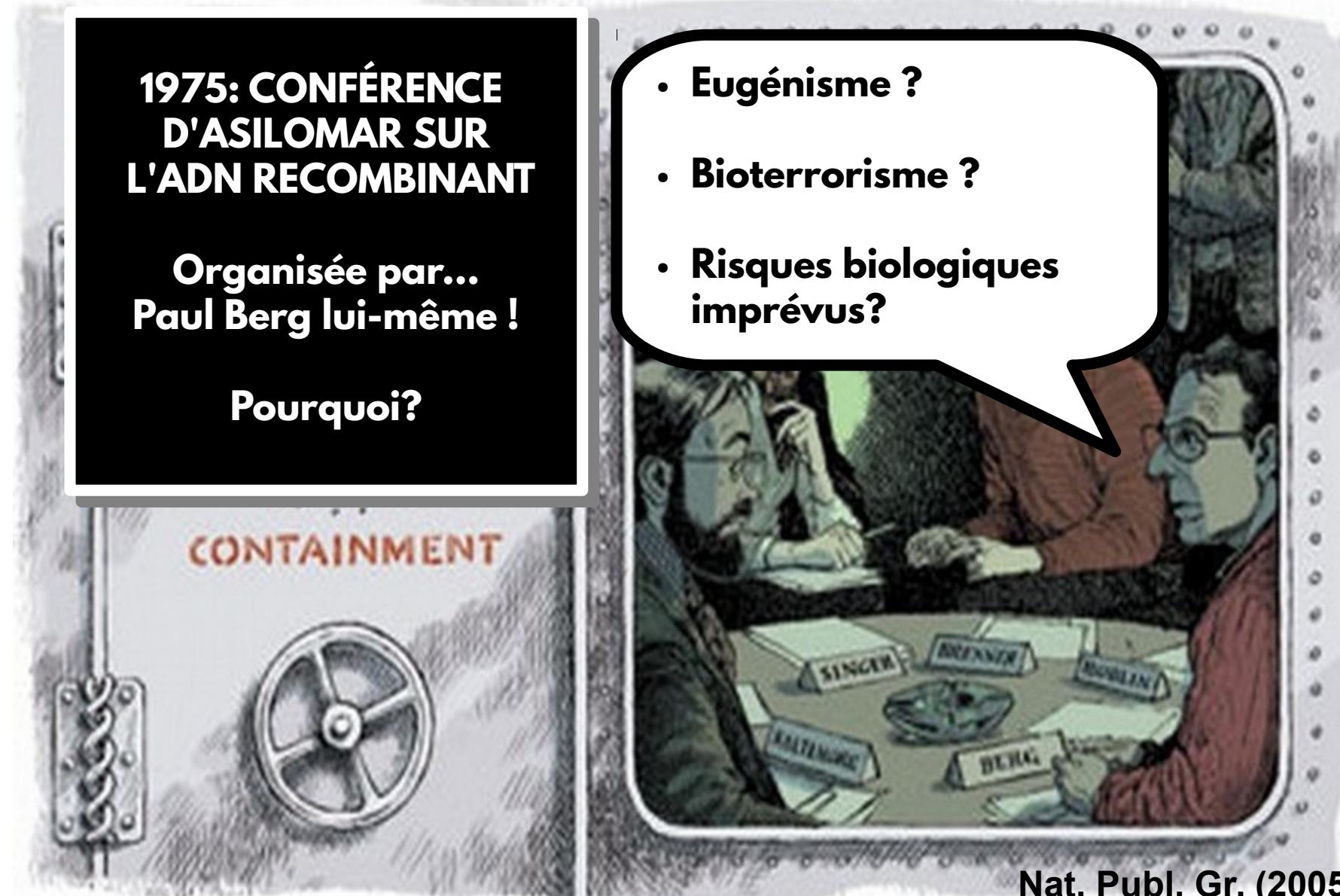
Organisée par...  
Paul Berg lui-même !

Pourquoi?

1980

1975

1970



**1976**

LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

1990  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
1970

**1976**

# LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

# Méthode de **MAXAM-GILBERT**

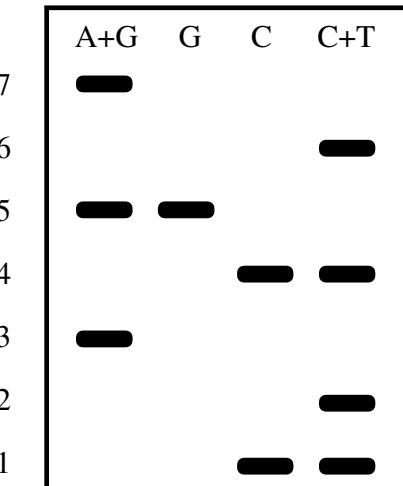
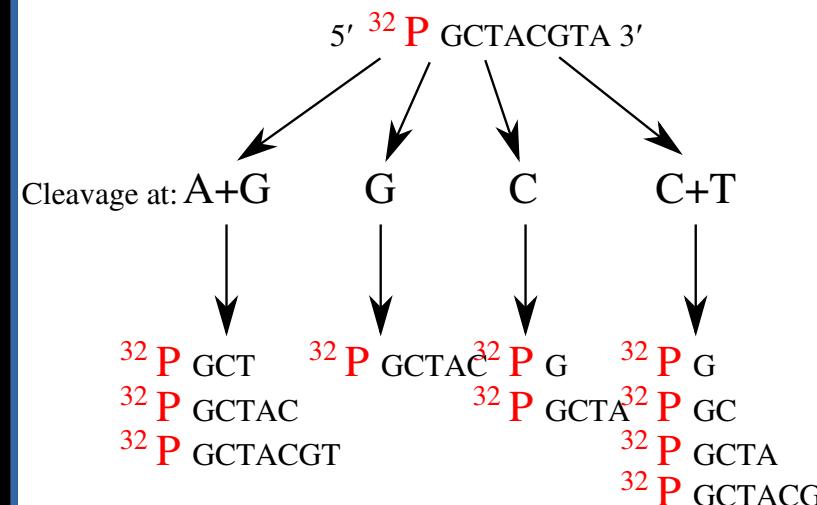
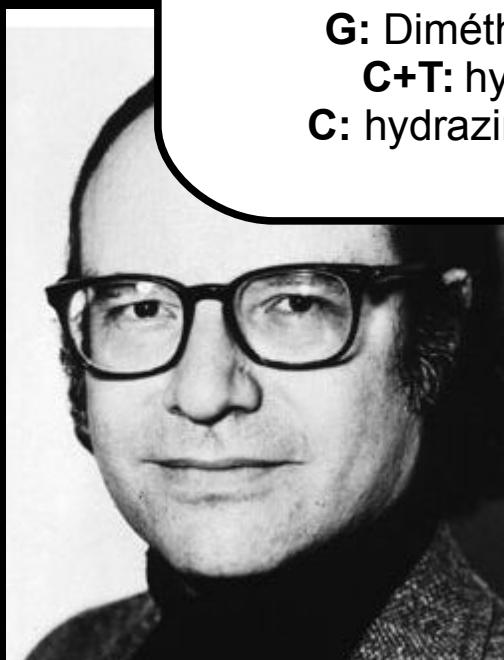
Principe: dégradation chimique de fragments d'ADN marqués au  $^{32}\text{P}$ . au niveau de certaines bases.

## A+G: Acide formique

## G: Diméthyl sulfate

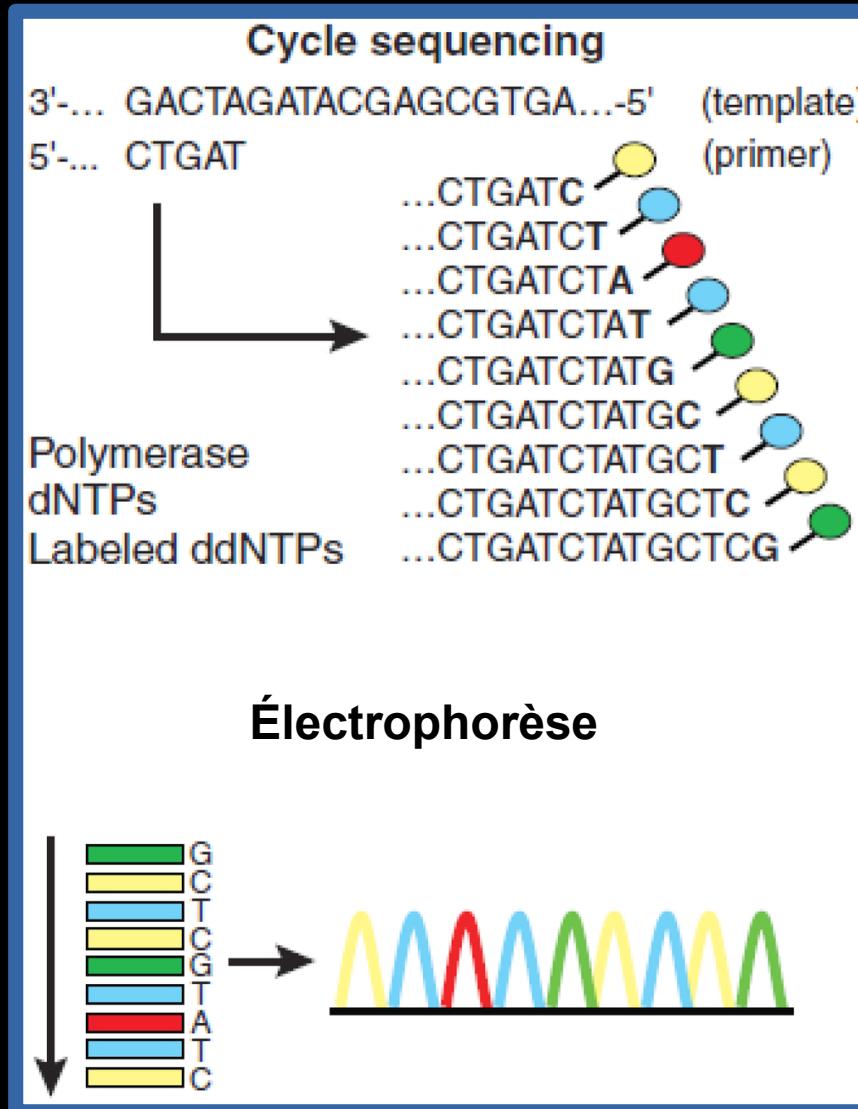
## C+T: hydrazine

C: hydrazine + NaCl



**1977**

# LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

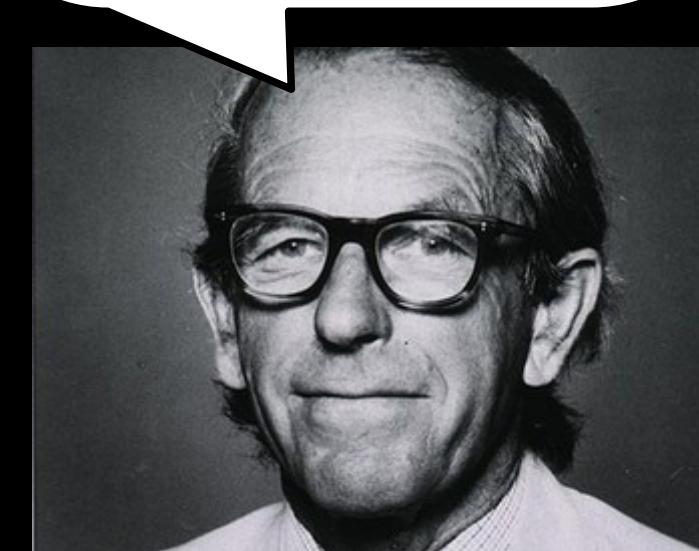


# Méthode de Frederick SANGER

## **“Séquençage par synthèse”**

## Réactifs requis

- une ADN polymérase;
  - une amorce;
  - des didéoxynucléotides.



1990  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1970  
-

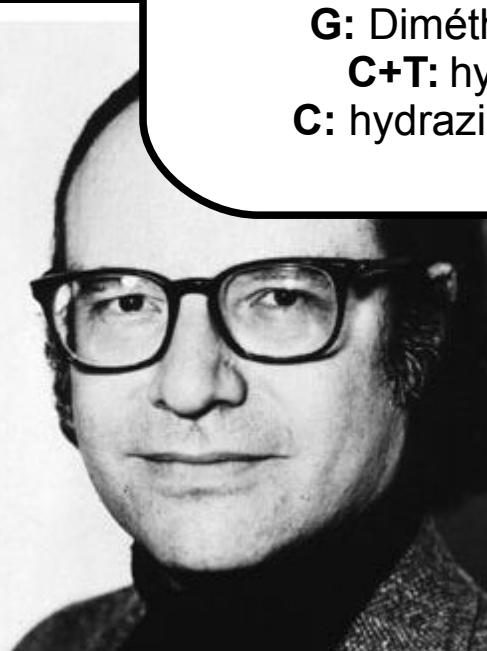
1977

## LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

### Méthode de **MAXAM-GILBERT**

Principe: dégradation chimique de fragments d'ADN marqués au  $^{32}\text{P}$ . au niveau de certaines bases.

**A+G:** Acide formique  
**G:** Diméthyl sulfate  
**C+T:** hydrazine  
**C:** hydrazine + NaCl

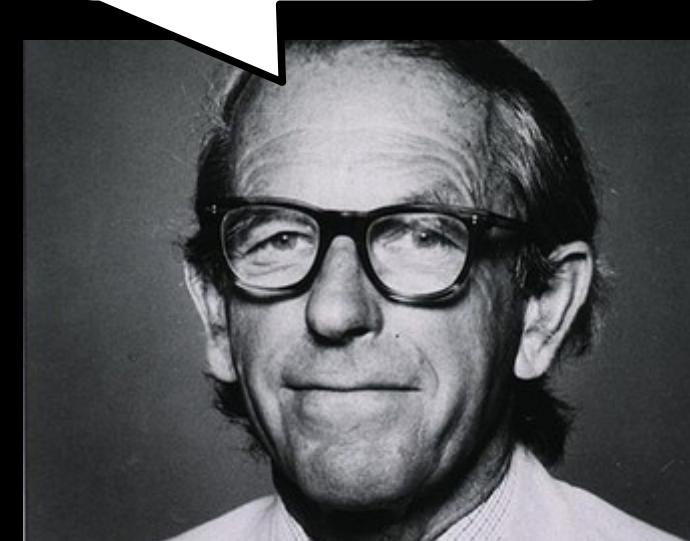


### Méthode de **Frederick SANGER**

“Séquençage par synthèse”

Réactifs requis:

- une ADN polymérase;
- une amorce;
- des didéoxynucléotides.





## 1977 LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

Principe: dégradation chimique de fragments d'ADN marqués au  $^{32}P$ .  
au niveau de certaines bases.  
A+G: Acide formique  
G: Diméthyl sulfate  
C+T: hydrazine  
C: hydrazine + NaCl

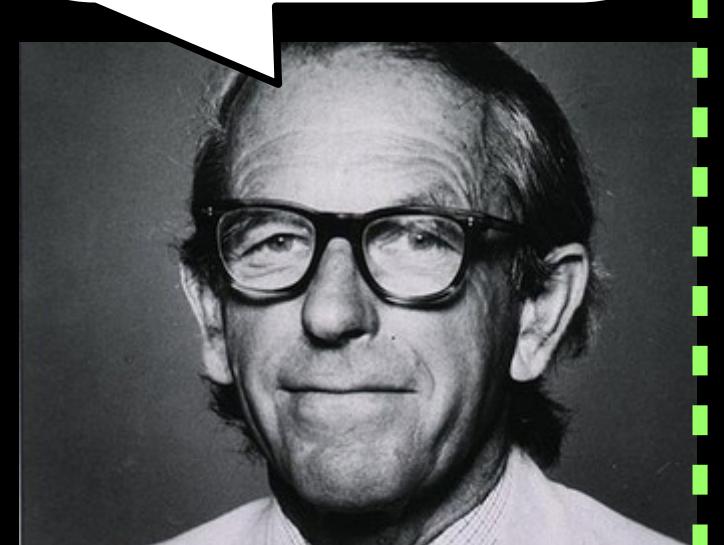
Méthode de  
V.-GILBERT

## Méthode de Frederick SANGER

“Séquençage par synthèse”

Réactifs requis:

- une ADN polymérase;
- une amorce;
- des didéoxynucléotides.





**1977**

LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

5' ... ATAGGTCCAT ... 3'  
3' ... TATCCAGGTA ... 3'

L'ADN est  
désormais  
**LISIBLE.**

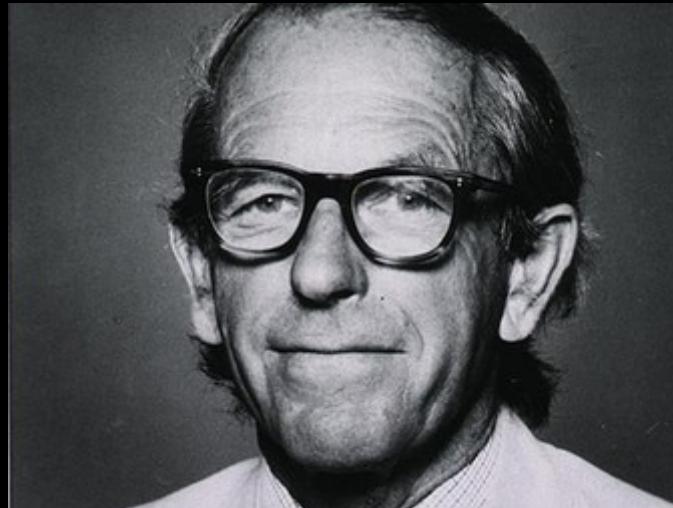


1990  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
1970  
-

**1977**

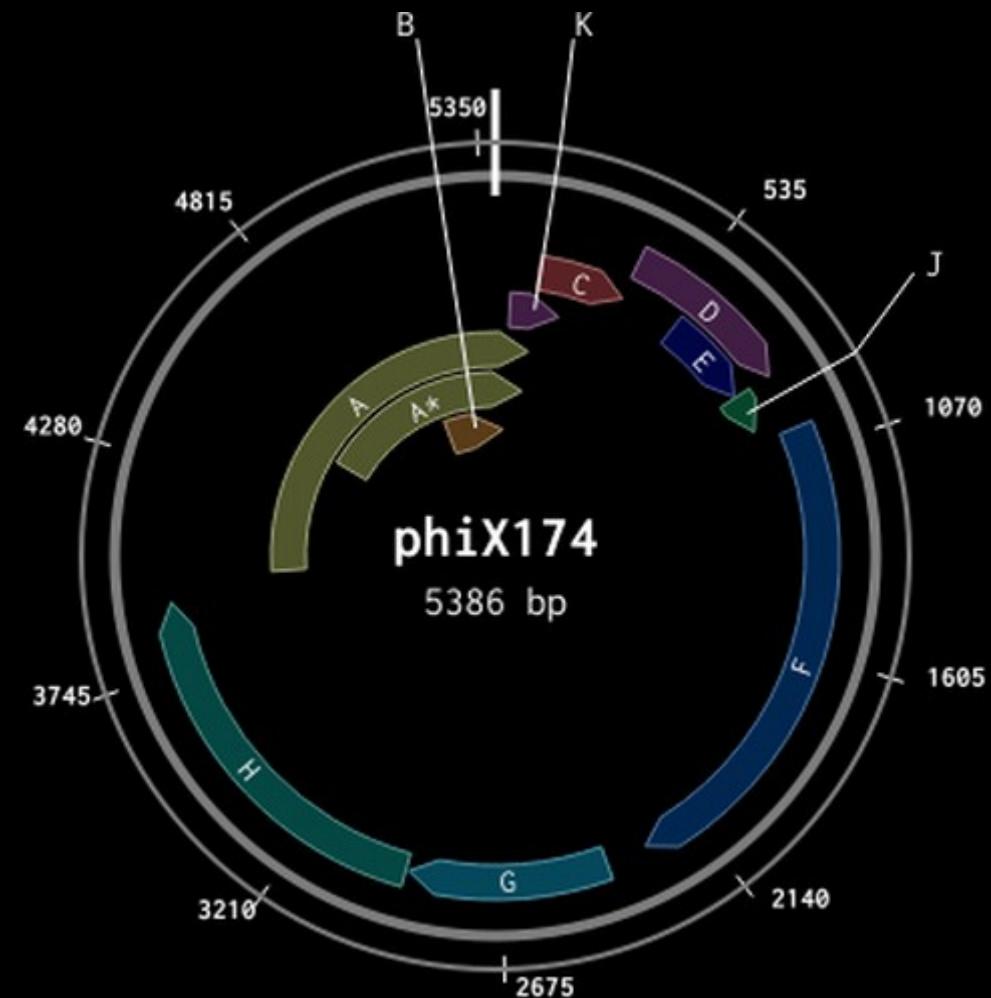
**LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ**

**La même année...**

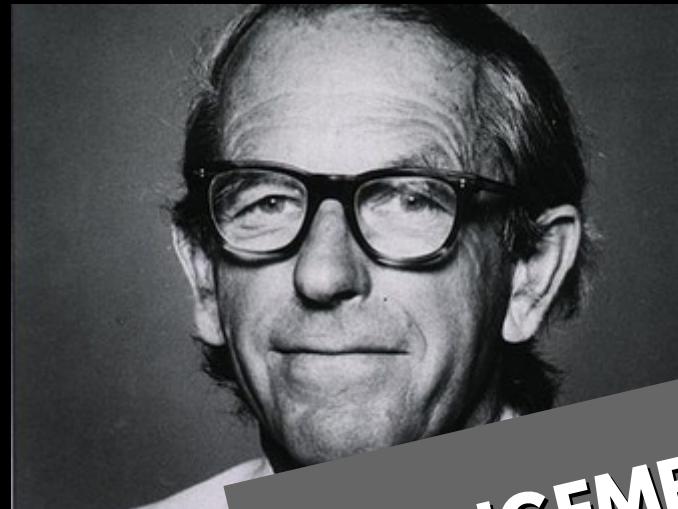


**SANGER *et al.***

**Premier génome  
complet à être  
séquencé**



1990

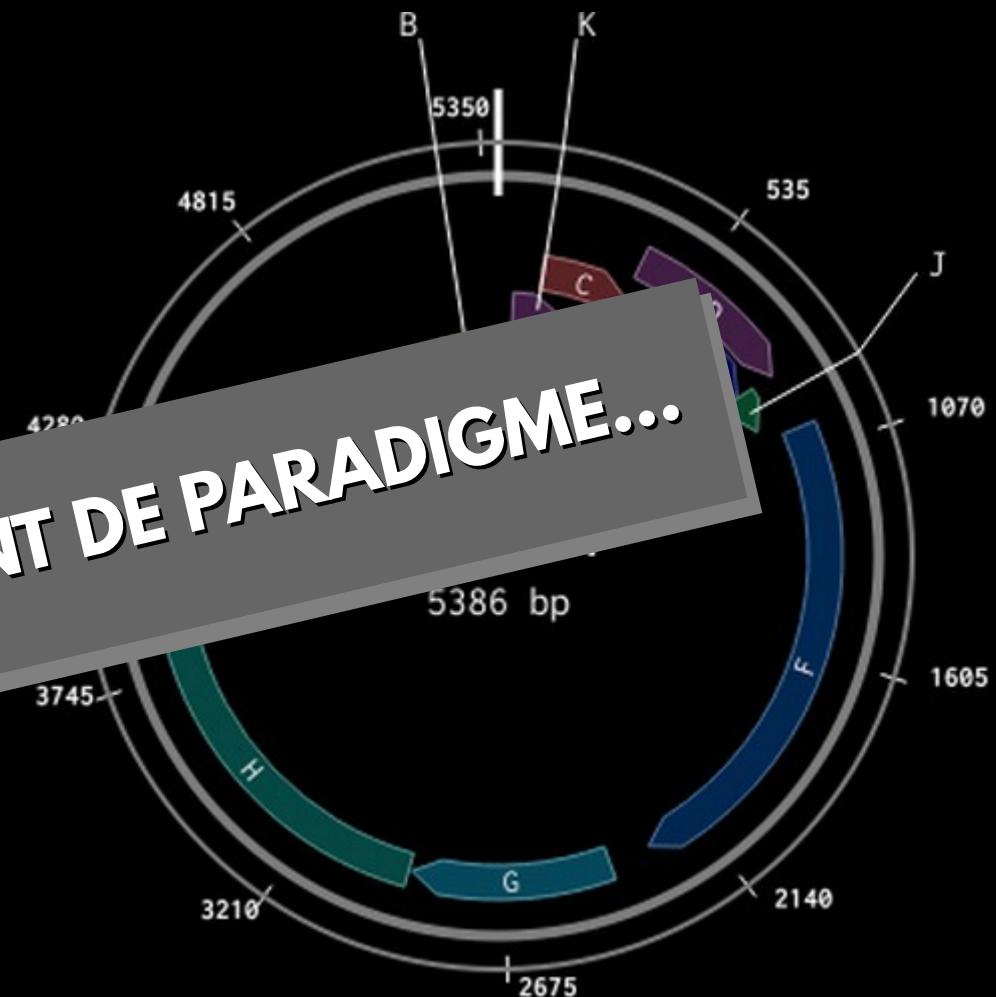
**1977****LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ****La même année...****SACCHAROMYCES CERVENIS  
CHANGEMENT DE PARADIGME...****Premier genome  
complet à être  
séquencé**

1985

1980

1975

1970

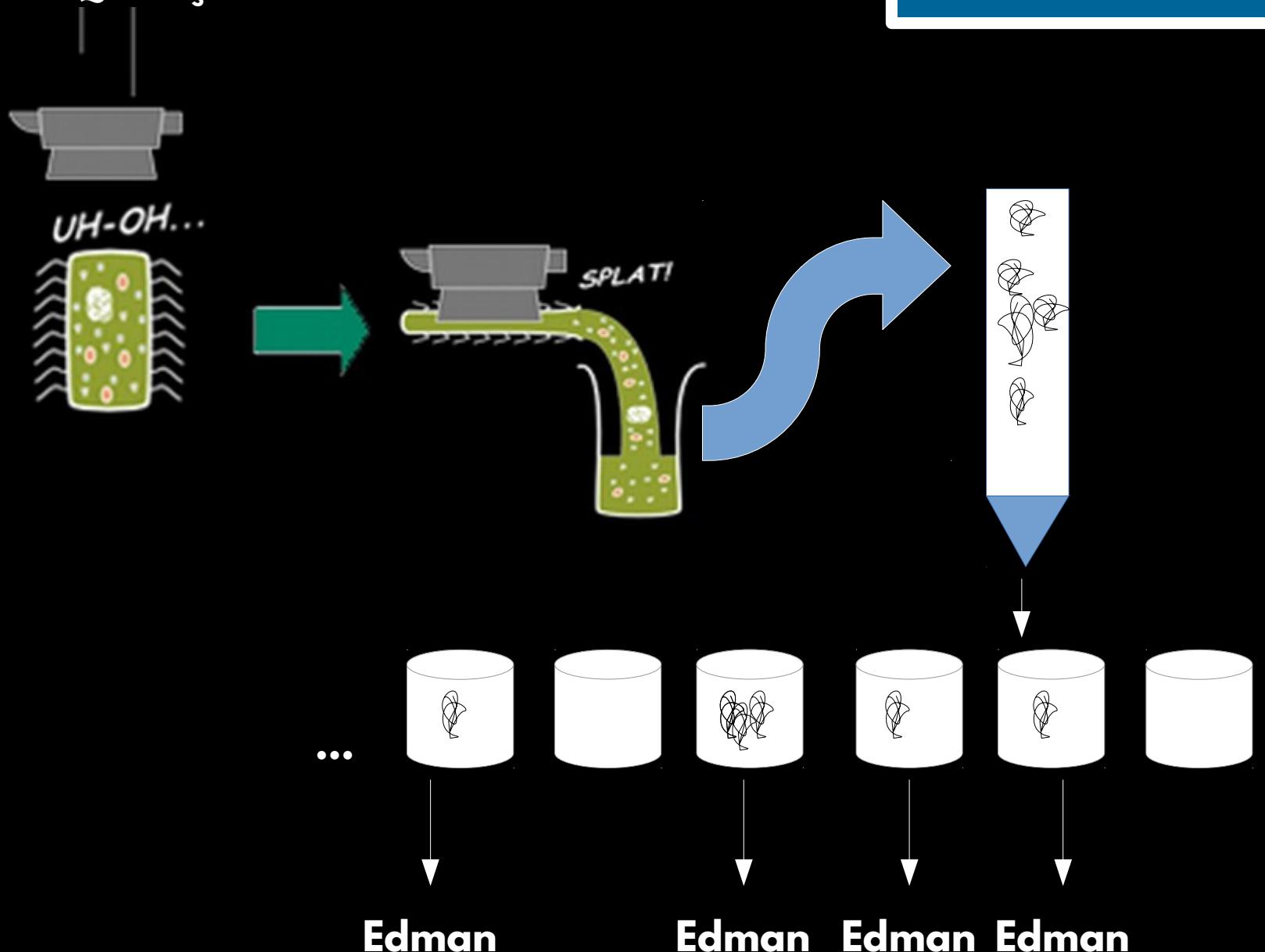




1977

LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

PROTÉINES

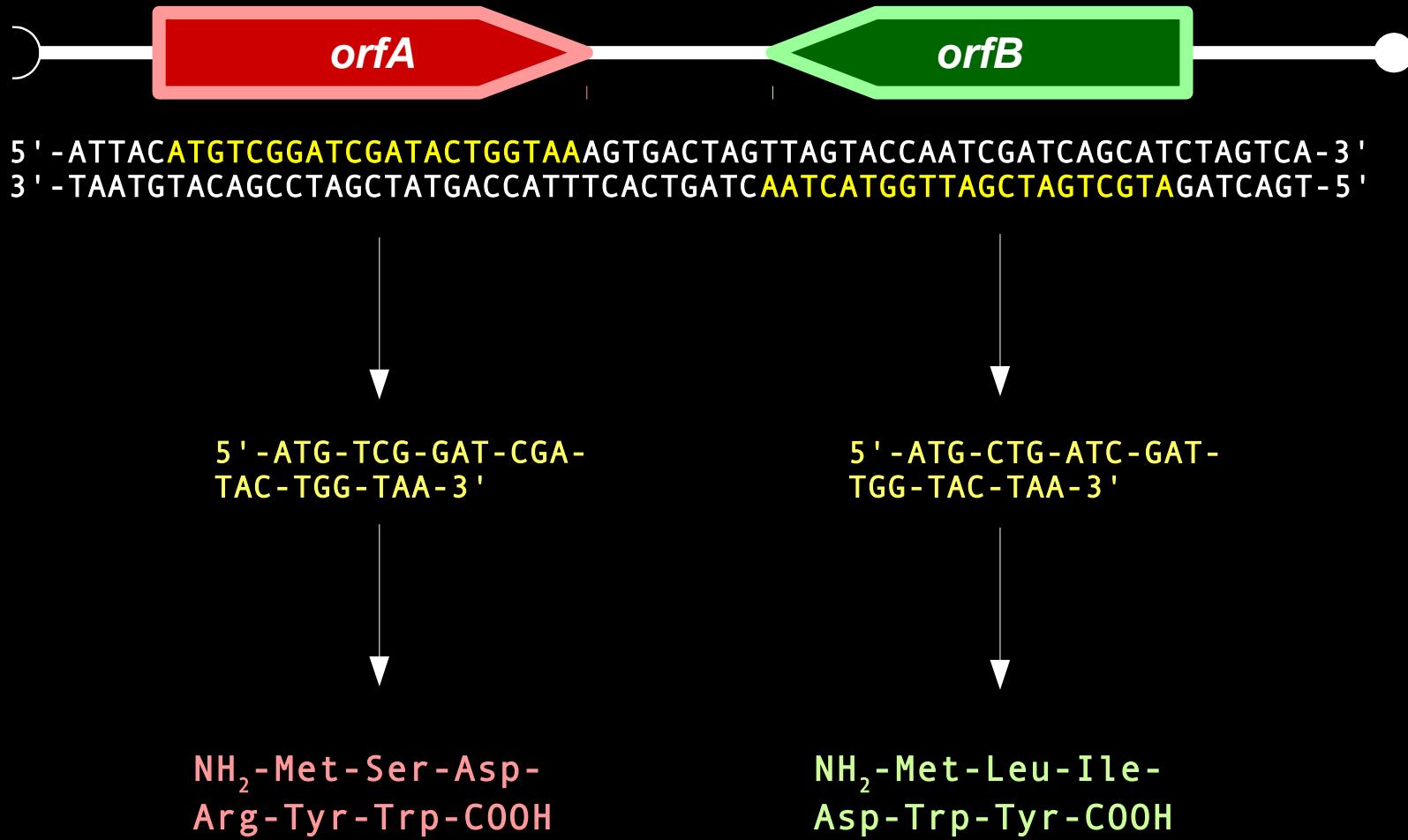


1990  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
1970  
-

1977

LE SÉQUENÇAGE D'ADN DEVIENT RÉALITÉ

ADN



1990  
-  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1985  
-  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1980  
-  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1975  
-  
-  
-  
-  
-  
-  
-  
-  
-  
-  
1970  
-

1983

## NOUVELLE MÉTHODE D'AMPLIFICATION D'ADN



**Amplification d'ADN  
fastidieuse à ses débuts !**

**ADN POLYMERASE  
UTILISÉE AVANT 1983:  
Fragment de Klenow**



**... thermosensible.**

**Il faut rajouter de  
l'enzyme à CHAQUE  
cycle.**



1983

## NOUVELLE MÉTHODE D'AMPLIFICATION D'ADN

Kary Mullis:

Polymerase

Chain

Reaction

...avec enzyme  
thermostable  
(Taq).



1990

1985

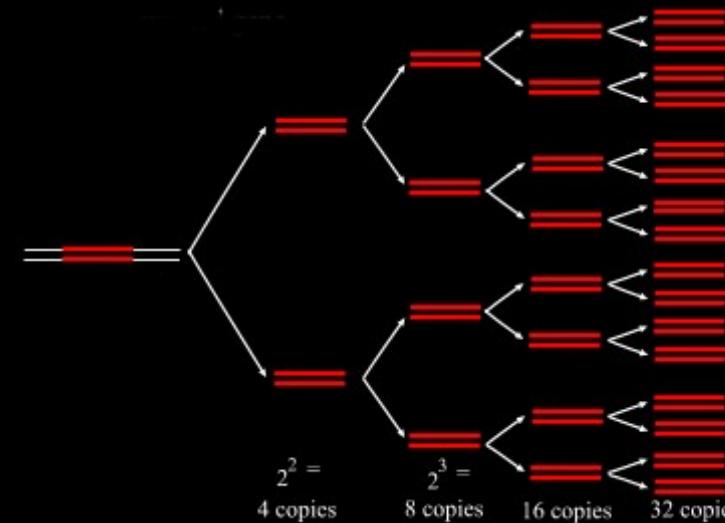
1980

1975

1970

**1983****NOUVELLE MÉTHODE D'AMPLIFICATION D'ADN****Kary Mullis:****P**olymerase**C**hain**R**eaction

...avec enzyme  
thermostable  
(*Taq*).

**Amplification exponentielle d'ADN...****...en une étape (de 25-40 cycles).****Thermocycleur à système Peltier (1986)**

**1981**

# PREMIÈRE BASE DE DONNÉES DE SÉQUENCES D'ADN

**1981**

# PREMIÈRE BASE DE DONNÉES DE SÉQUENCES D'ADN

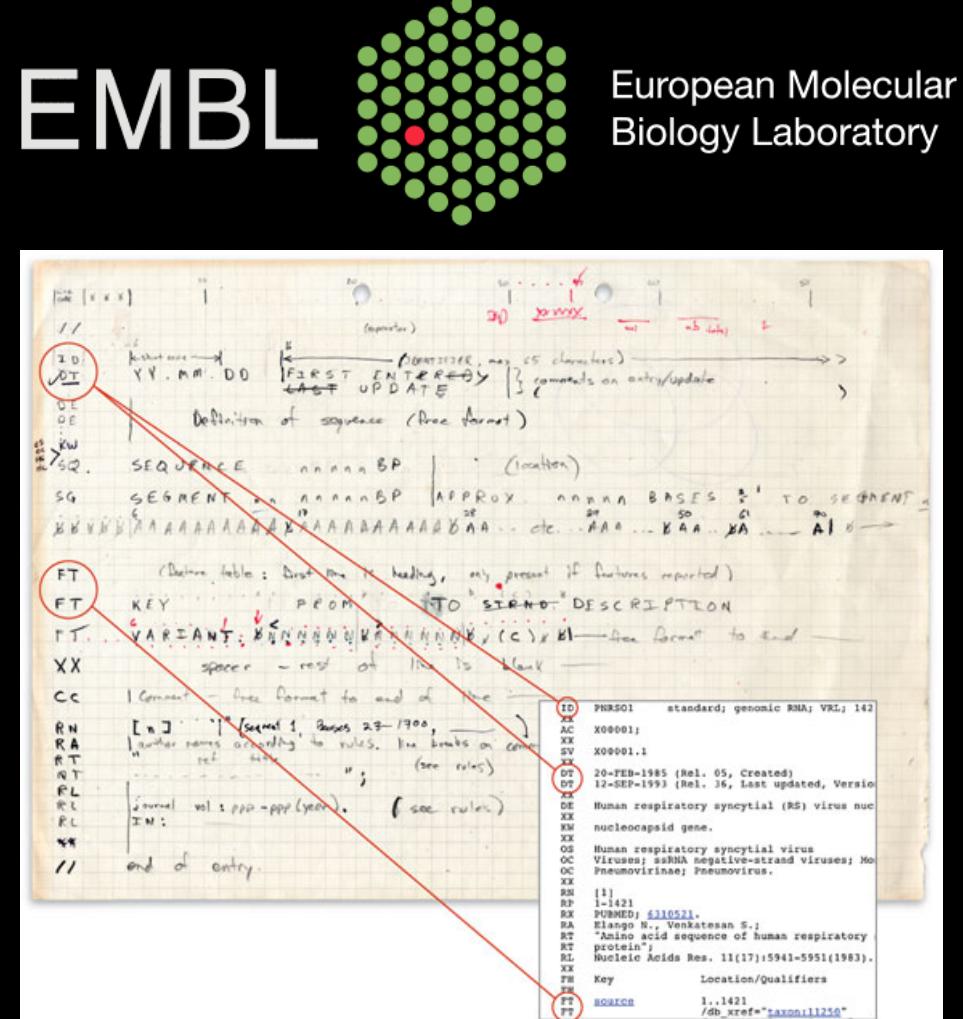
# *EMBL Data Library (Europe)*

# **MISSION:**

## **Faciliter le transfert de données de séquences.**

**(le stockage n'est pas encore un problème...)**

# NCBI GenBank (USA): 7 ans après.



# Prototype du format “EMBL” par Greg Hamm



**1981**

## NOUVEL ALGORITHME D'ALIGNEMENT

Problème avec NEEDLEMAN-WUNSCH

```
>seq1  
GATTGGTTCGTGATTACATTGA  
TGA
```

```
>seq2  
GATTACA
```

SEQ1: GATTGGTTCGTGATTACATTGA  
SEQ2: GATT-----A---C-----A

Excellent pour comparer de manière globale

Pas conçu pour la recherche de motifs locaux

1990

**1983****NOUVEL ALGORITHME D'ALIGNEMENT****Solution: algorithme de SMITH-WATERMAN**

>seq1  
GATTGGTCGTGATTACATTGA  
TGA

>seq2  
GATTACA

SEQ1 : GATTGGTCGTGATTACATTGA  
SEQ2 : -----GATTACA-----

**Principe: recherche du meilleur alignement “local”.**



**1985**

**PREMIERS “MOTEURS DE RECHERCHE” DE SÉQUENCES**



William R. Pearson

DNA and PROTEIN

Google ?

**1985****PREMIERS “MOTEURS DE RECHERCHE” DE SÉQUENCES**

William R. Pearson

**Algorithme FASTA**

1990

1985

## PREMIERS "MOTEURS DE RECHERCHE" DE SÉQUENCES



William R. Pearson

>query  
GATTACA

Mots-clés (k=3)

## Algorithme FASTA

1) Recherche par  
mots-clés...>Singe  
GATTCTGTGATTACATTGATGA>Rat  
GATTGGTTCTGTGATTGATGA>Pieuvre  
GATTGGTTCTGTGACATTGAGA>Levure  
GATTGGTTCTTACATTGTGA

1980

1975

1970

Requête

Base de données



1985

## PREMIERS "MOTEURS DE RECHERCHE" DE SÉQUENCES



William R. Pearson

>query  
GATTACA

Mots-clés (k=3)

GAT  
ATT  
TTA  
TAC  
ACA

Requête

Algorithme FASTA

1) Recherche par mots-clés...

#1

>Singe  
GATTCGTGATTACATTGATGA

#3

>Rat  
GATTGGTTCGTGATTGATGA

#4

>Pieuvre  
GATTGGTTCGTGACATTGAGA

#2

>Levure  
GATTGGTTCTACATTGTGA

Base de données

1990

1985

1980

1975

1970

**1985****PREMIERS "MOTEURS DE RECHERCHE" DE SÉQUENCES**

William R. Pearson

>query  
GATTACA

Mots-clés (k=3)

GAT  
ATT  
TTA  
TAC  
ACA**Requête****Algorithme FASTA****2) Puis, alignement  
Smith-Waterman**

#1

>Singe  
GATTCGTGATTACATTGATGAGATT  
|||  
GATTACAGATTACATTGAT  
|||||||  
GATTACA

1990

1985

1980

1975

1970

**1985****PREMIERS "MOTEURS DE RECHERCHE" DE SÉQUENCES**

William R. Pearson

>query  
GATTACA

Mots-clés (k=3)

GAT  
ATT  
TTA  
TAC  
ACA**Requête****Algorithme FASTA****2) Puis, alignement  
Smith-Waterman**

#1

>Singe  
GATTCGTGATTACATTGATGAGATT  
|||  
GATTACAGATTACATTGAT  
|||||||  
GATTACA**Alignement optimal**



## 1988-1990 ARRIVÉE EN SCÈNE DU NCBI



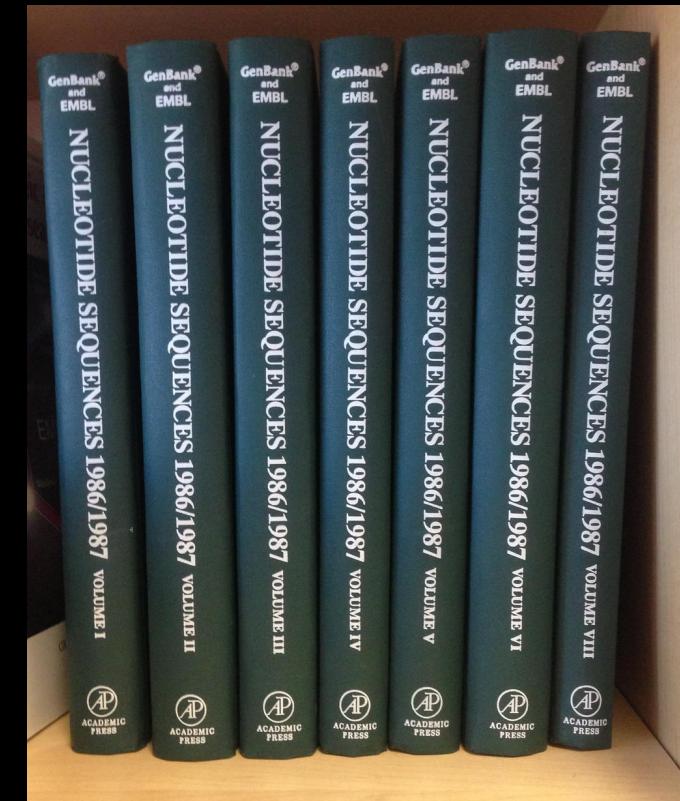
National Center for  
Biotechnology  
Information  
(1988-présent)

1990  
-  
1985  
-  
1980  
-  
1975  
-  
1970

## 1988-1990 ARRIVÉE EN SCÈNE DU NCBI



National Center for  
Biotechnology  
Information  
(1988-présent)



GenBank (format papier, 1988)

**1990**

**1985**

**1980**

**1975**

1970

# **1988-1990**

## **ARRIVÉE EN SCÈNE DU NCBI**



# National Center for Biotechnology Information (1988-présent)

J. Mol. Biol. (1990) 215, 403–410

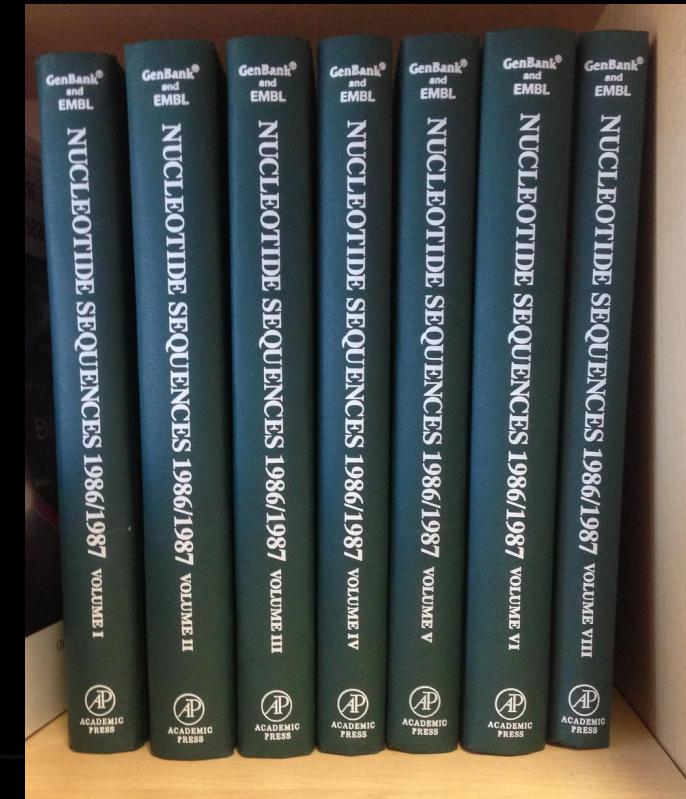
## Basic Local Alignment Search Tool

Stephen F. Altschul<sup>1</sup>, Warren Gish<sup>1</sup>, Webb Miller<sup>1</sup>  
Eugene W. Myers<sup>3</sup> and David J. Lipman<sup>1</sup>

<sup>1</sup>National Center for Biotechnology Information  
National Library of Medicine, National Institutes of Health  
Bethesda, MD 20894, U.S.A.

*<sup>2</sup>Department of Computer Science  
The Pennsylvania State University, University Park, PA 16802, U.S.A.*

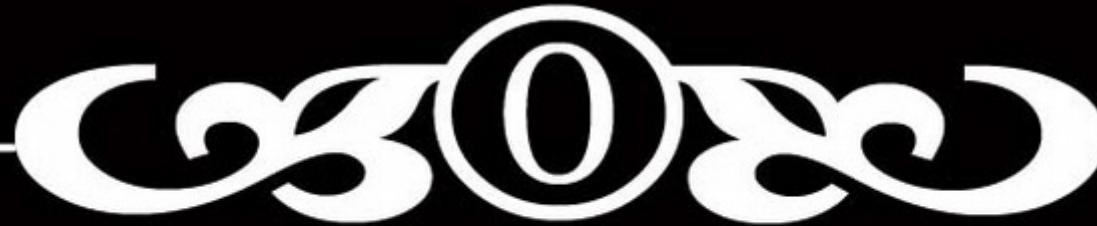
<sup>3</sup>Department of Computer Science



## GenBank (format papier, 1988)



**DE RETOUR EN 2015 ...**



**La session prochaine...**

## **PARTIE III (1990-présent)**

- **L'INFORMATIQUE LIBRE (Projet GNU/Linux, logiciels libres);**
- **PREMIERS GÉNOMES PRO- ET EUKARYOTES;**
- **INTERNET;**
- **SÉQUENÇAGE À HAUT-DÉBIT.**

