

# DNA PROJECT

Gabriel Misrachi & Raphaël Montaud

# Plus longue sous-séquence

- Intérêt biologique: mutations
- Programmation naïve: on test tout.  $O(m \cdot 2^n)$
- Programmation dynamique:  $O(m \cdot n)$   
Si :  $a_i = b_j$  alors :  $c[i,j] = 1 + c[i-1,j-1]$   
Sinon :  $c[i,j] = \max (c[i-1,j], c[i,j-1])$
- Programmation non exacte:  
Optimized Keyword Trees

# Exemple:

- B A T O N
- A L G O

# Zoom: récupérer la solution

- Remonter le tableau des opérations

		B	A	T	O	N
	0	0	0	0	0	0
		←	←	←	←	←
A	0	0	1	1	1	1
	↑	↖	↖	←	←	←
L	0	0	1	1	1	1
	↑	↖	↑	↖	↖	↖
G	0	0	1	1	1	1
	↑	↖	↑	↖	↖	↖
O	0	0	1	1	2	2
	↑	↖	↑	↖	↖	←

# Zoom: récupérer la solution

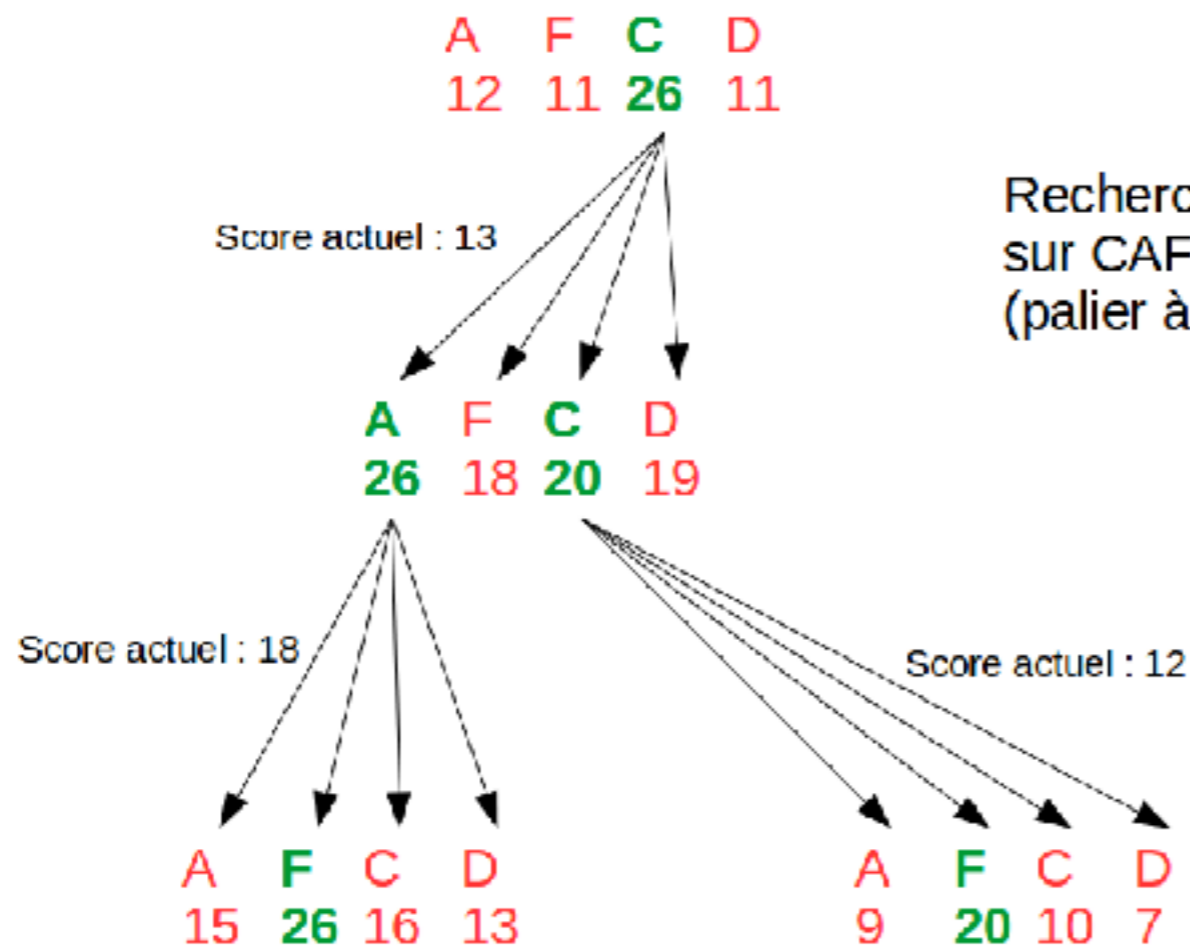
- Remonter le tableau des opérations

		B	A	T	O	N
		0	0	0	0	0
			←	←	←	←
A	0	0	1	1	1	1
		↑	↖	←	←	←
L	0	0	1	1	1	1
		↑	↑	↖	↖	↖
G	0	0	1	1	1	1
		↑	↑	↖	↖	↖
O	0	0	1	1	2	2
		↑	↑	↖	↖	←

# Alignement Optimal

- Equivalence avec le problème de plus longue sous-séquence
- Récurrence légèrement modifiée:  
 $c[i,j]$  est le maximum de :
  - $c[i-1,j] + \text{score}(a_j, -)$
  - $c[i, j-1] + \text{score}(-, b_j)$
  - $c[i-1, j-1] + \text{score}(a_i, b_j)$

# Programmation non exacte



Recherche de graines  
sur CAF avec  $th = 0,76$   
(palier à 20)

	A	F	C	D
A	5	-3	-1	-2
F		8	-2	-5
C			13	-4
D				8

# Résultats

	th = 0,5	th = 0,7	th = 0,9
k = 3	15%	13%	2%
k = 4	3,60%	0,80%	0,30%
k = 5	1%	0,10%	0,02%

Rapport des temps de calcul entre les deux méthodes