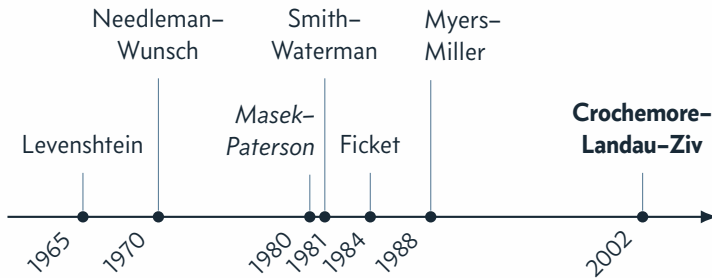


CROCHEMORE-LANDAU-ZIV
ALIGNEMENT GLOBAL ET LOCAL
EN TEMPS SOUS-QUADRATIQUE

Mattéo Delabre

IFT-6291 — Bio-informatique génomique
Université de Montréal
15 novembre 2021

QUELQUES ALGORITHMES D'ALIGNEMENT



MASEK-PATERSON / MÉTHODE DES QUATRE RUSSES

		A	A	C	G	A	C	G	A
C									
T					0	-1	-1		
A					-1		+2		
C					+2	-1	+2		
G									
A									
G									
A									

- ▶ Découpage de la matrice d'alignement en t -blocs chevauchants
- ▶ **Pré-calcul** du bord « sortie » pour chaque bord « entrée » possible
- ▶ Codage des bords en **valeurs relatives** plutôt qu'absolues
- ▶ $O(n^2 / \log n)$

MASEK-PATERSON / MÉTHODE DES QUATRE RUSSES

- ▶ Limites de l'approche
 - Ne permet pas de faire d'alignement local
 - Pré-calcul pour toutes les combinaisons de bords et de mots
 - Sensible aux valeurs des pondérations

- ▶ Idée d'amélioration
 - **Calculer chaque bloc à partir de blocs précédents**
 - Exploiter la compressibilité des séquences comparées

COMPRESSION SANS PERTE DE SÉQUENCES

- ▶ Tirer parti de la prédictibilité des séquences pour les représenter avec moins de *bits*
- ▶ Des algorithmes et logiciels qu'on utilise tous les jours
- ▶ Famille **LZ** : utilise des **références à des parties antérieures** de la séquence pour économiser de la place

The logo for 7zip, featuring a large white number '7' inside a black square, followed by the word 'ZIP' in a bold, black, sans-serif font inside a white rectangular box.The logo for bzip2, featuring the text 'bzip2' in a blue, cursive, lowercase font.The logo for gzip, featuring the text 'gzip' in a blue, cursive, lowercase font.

ALGORITHME LZ78

Séquence initiale :

↓ A C C T A G T C T A C G

Dictionnaire :



Séquence compressée :

ALGORITHME LZ78

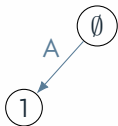
Séquence initiale :

| A ↓ C C T A G T C T A C G

Séquence compressée :

∅A

Dictionnaire :



ALGORITHME LZ78

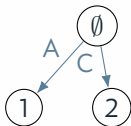
Séquence initiale :

| A | C ↓ C T A G T C T A C G

Séquence compressée :

∅A ∅C

Dictionnaire :



ALGORITHME LZ78

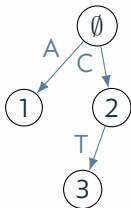
Séquence initiale :

| A | C | C T ↓ A G T C T A C G

Séquence compressée :

∅A ∅C 2T

Dictionnaire :



ALGORITHME LZ78

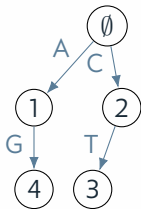
Séquence initiale :

| A | C | C T | A G ↓ T C T A C G

Séquence compressée :

∅A ∅C 2T ∅T

Dictionnaire :



ALGORITHME LZ78

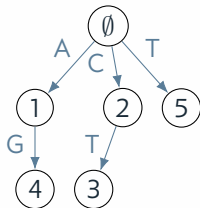
Séquence initiale :

| A | C | C T | A G | T ↓ C T A C G

Séquence compressée :

∅A ∅C 2T ∅T 3A

Dictionnaire :



ALGORITHME LZ78

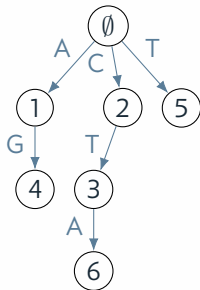
Séquence initiale :

| A | C | C T | A G | T | C T A ↓ C G

Séquence compressée :

∅A ∅C 2T ∅T 3A 2G

Dictionnaire :



ALGORITHME LZ78

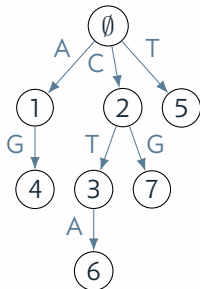
Séquence initiale :

| A | C | C T | A G | T | C T A | C G ↓

Séquence compressée :

∅A ∅C 2T ∅T 3A 2G

Dictionnaire :



ALGORITHME LZ78

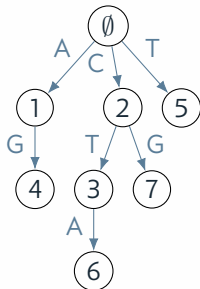
Séquence initiale :

| A | C | C T | A G | T | C T A | C G |

Séquence compressée :

∅A ∅C 2T ∅T 3A 2G

Dictionnaire :



- ▶ Nombre de facteurs moyen : $O(hn / \log n)$
- ▶ $0 < h \leq 1$ mesure la « compressibilité » de la séquence

IDÉE PRINCIPALE DE CROCHEMORE-LANDAU-ZIV

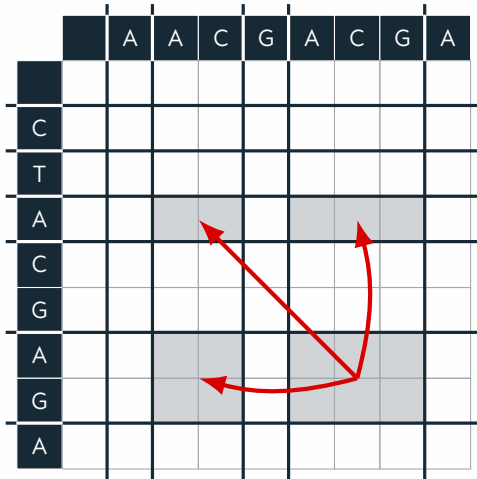
	A	A	C	G	A	C	G	A
C								
T								
A								
C								
G								
A								
G								
A								

IDÉE PRINCIPALE DE CROCHEMORE-LANDAU-ZIV

		A	A	C	G	A	C	G	A
C									
T									
A									
C									
G									
A									
G									
A									

- ▶ Découper en **blocs** suivant la **factorisation** des séquences par LZ78

IDÉE PRINCIPALE DE CROCHEMORE-LANDAU-ZIV



- ▶ Découper en **blocs** suivant la **factorisation** des séquences par LZ78
- ▶ Chaque bloc a déjà été vu en partie à trois autres endroits de la table
- ▶ Réutiliser du travail déjà fourni pour les trois **préfixes**

MATRICES DE CHEMINEMENT

Préfixe gauche

		A	C
	iii	iv	v
A	ii		4
G	i	2	3

Préfixe diagonal

		A	C
	ii	iii	iv
A	i	2	3

Préfixe haut

		A	C	G
	ii	iii	iv	v
A	i	2	3	4

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i						
ii						
iii						
iv						
v						
vi						

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2			
ii	-1	-1	-2			
iii	-2	0	0			
iv		-2	-2			
v			-2			
vi						

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2			
ii	-1	-1	-2		-3	
iii	-2	0	0		-1	-3
iv		-2	-2		-2	-2
v			-2		-1	-1
vi					-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2		-3	
iii	-2	0	0		-1	-3
iv		-2	-2		-2	-2
v			-2		-1	-1
vi					-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	-2	-3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0


Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0		-1	-3
iv		-2	-2		-2	-2
v			-2		-1	-1
vi					-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4



CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2		-2	-2
v			-2		-1	-1
vi					-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2		-1	-1
vi					-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi					-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL RÉCURSIF DE LA MATRICE DE CHEMINEMENT

Préfixe gauche

	1	2	3	4	5
i	0	-1	-2		
ii	-1	-1	-2	-2	
iii	-2	0	0	0	-2
iv		-2	-2	-1	-1
v			-2	-1	0

Préfixe diagonal

	1	2	3	4
i	0	-1	-2	
ii	-1	1	0	-2
iii		-1	-1	-1
iv			-1	0

Préfixe haut

	1	2	3	4	5
i	0	-1	-2	-3	
ii	-1	1	0	-1	-3
iii		-1	-1	-2	-2
iv			-1	-1	-1
v				-1	0

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	iii	iv	v	vi
A	ii			5
G	i	2	3	4

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1			

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	-2		

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	-1		

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1		




CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	-2		



CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1		

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	-3	


CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	-2	



CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	


CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	-2	




CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	-3	



CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	-4

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	-1

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	2

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	0



CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	-1




CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	-1



CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			
G	-1	1	1	2

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			-3
G	-1	1	1	2

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			0
G	-1	1	1	2

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			-2
G	-1	1	1	2



CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			-2
G	-1	1	1	2

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			0
G	-1	1	1	2

CALCUL DU BORD DE SORTIE

Matrice de cheminement

	1	2	3	4	5	6
i	0	-1	-2	-3		
ii	-1	-1	-2	-1	-3	
iii	-2	0	0	1	-1	-3
iv		-2	-2	0	-2	-2
v			-2	0	-1	-1
vi				-2	-1	0

Bloc

		A	C	G
	1	0	-1	1
A	0			0
G	-1	1	1	2

VUE D'ENSEMBLE DE L'ALGORITHME

- 1 **Factorisation** des séquences avec LZ78 $O(n)$
- 2 Pour chaque bloc de périmètre t :
 - 2.1 Récupération des **blocs préfixes** $O(1)$
 - 2.2 Calcul de la colonne de **cheminement** qui lui est propre $O(t)$
 - 2.3 **Maximum** des colonnes pour obtenir le bord de sortie $O(t^2)$?

CALCUL EFFICACE DES MAXIMA MARGINAUX

	iv	v	vi	vii	viii
	iii				7
	ii				6
	i	2	3	4	5

	1	2	3	4	5	6	7	8
i								
ii								
iii			A		B			
iv								
v			C		D			
vi								
vii								
viii								

CALCUL EFFICACE DES MAXIMA MARGINAUX

	iv	v	vi	vii	viii
iii					7
ii					6
i	2	3	4	5	

$$B = B_1 + B_2 \quad C = C_1 + C_2$$

	1	2	3	4	5	6	7	8
i								
ii								
iii			A		B			
iv								
v			C		D			
vi								
vii								
viii								

CALCUL EFFICACE DES MAXIMA MARGINAUX

	iv	v	vi	vii	viii
iii					7
ii					6
i	2	3	4	5	

Diagram illustrating the calculation of marginal maxima. Red arrows labeled A and B point from cell (iii, 2) to (ii, 3) and (ii, 5) respectively. Green arrows labeled C and D point from cell (v, 3) to (ii, 3) and (ii, 5) respectively.

$$B = B_1 + B_2 \quad C = C_1 + C_2$$

$$B_1 + C_2 \leq A$$

$$C_1 + B_2 \leq D$$

	1	2	3	4	5	6	7	8
i								
ii								
iii			A		B			
iv								
v			C		D			
vi								
vii								
viii								

CALCUL EFFICACE DES MAXIMA MARGINAUX

	iv	v	vi	vii	viii
iii					7
ii					6
i	2	3	4	5	

Diagram illustrating the calculation of marginal maxima. Red arrows labeled A and B point from row iii to column 3 and from row ii to column 5, respectively. Green arrows labeled C and D point from row v to column 3 and from row vi to column 5, respectively.

$$B = B_1 + B_2 \quad C = C_1 + C_2$$

$$B_1 + C_2 \leq A$$

$$C_1 + B_2 \leq D$$

	1	2	3	4	5	6	7	8
i								
ii								
iii			A		B			
iv								
v			C		D			
vi								
vii								
viii								

Condition de Monge

$$B + C \leq A + D$$

CALCUL EFFICACE DES MAXIMA MARGINAUX

- ▶ Matrices de cheminement respectent la condition de Monge
 - $B + C \leq A + D$
- ▶ Implique qu'elles sont totalement monotones
 - $A \leq C \Rightarrow B \leq D$
- ▶ **SMAWK** : Maximum de toutes les colonnes en $O(t)$!

COMPLEXITÉ DE L'ALGORITHME

- ▶ Bloc de périmètre t calculé en temps $O(t)$
 - 1 Récupération des **blocs préfixes** $O(1)$
 - 2 Calcul de la colonne de **cheminement** qui lui est propre $O(t)$
 - 3 **Maximum** des colonnes via SMAWK $O(t)$
- ▶ Nombre de facteurs LZ78 : $O(hn / \log n)$
- ▶ Taille totale des périmètres de blocs : $O(hn^2 / \log n)$
- ▶ Complexité en temps et espace : $O(hn^2 / \log n)$

EXTENSION : RETROUVER UN ALIGNEMENT

- ▶ Des **pointeurs supplémentaires** doivent être maintenus
 - Pour chaque case du *bord de sortie*
 - ▶ Retenir la *case d'entrée* optimale
 - ▶ Retenir quel bloc a vu apparaître en premier cette case
 - Pour chaque case de la *colonne de cheminement*, retenir la direction d'origine optimale
- ▶ Complexité de l'algorithme inchangée
- ▶ Temps pour retrouver un alignement : $O(n)$

EXTENSION : ALIGNEMENT LOCAL

- ▶ Un chemin d'alignement local peut
 - ▶ 1 Être entièrement contenu dans un bloc
 - ▶ 2 Traverser plusieurs blocs
 - ▶ Commence à l'intérieur d'un bloc
 - ▶ Traverse zéro ou plusieurs blocs de part en part
 - ▶ Termine à l'intérieur d'un bloc
- ▶ Des **valeurs supplémentaires** doivent être retenues
- ▶ Complexité de l'algorithme inchangée

CONCLUSION

- ▶ **Factorisation** des séquences avec LZ78
- ▶ Partitionnement de la matrice d'alignement selon les facteurs
- ▶ Chaque bloc calculé en $O(t)$ en **réutilisant les blocs précédents**
- ▶ Complexité globale en temps & espace $O(hn^2 / \log n)$

AVANTAGES

- ▶ Exploite la compressibilité des séquences
- ▶ Peut travailler directement sur des séquences compressées
- ▶ Pas de pré-calcul à faire
- ▶ Permet l'alignement local

LIMITES

- ▶ Implémentation assez subtile
- ▶ *Gaps* linéaires uniquement
- ▶ Utilise beaucoup de mémoire
- ▶ Pas cache-efficace