



Recherche de motifs : Hash Table

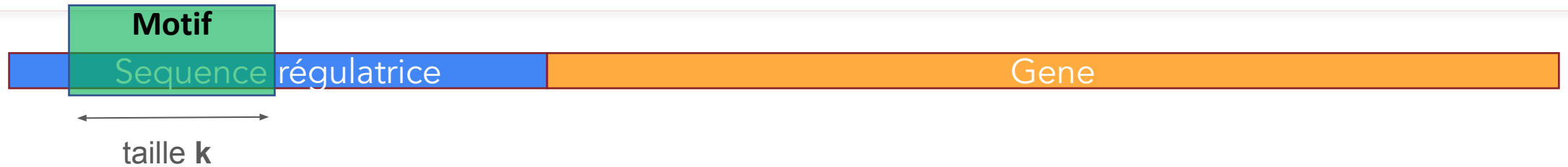
Cours 4



Plan du cours

- Algorithme : Table de Hachage
- Exemple
- Complexité

Algorithme: Table de hachage



- On cherche un motif de taille **k** dans les séquences régulatrices.
- Nous allons utiliser l'algo Table de hachage pour trouver les motifs de taille **k**, les plus fréquents.

Exemple: $k=3$, sequence = ACCATACCAGGCACC

Sequence

Table hachage

k=3

ACCATACCAGGCACC

ACCATACCAGGCACC

| | |
|-----|-----|
| ACC | = 1 |
| | |
| | |
| | |

Table hachage

Sequence

k=3

| |
|-----------------|
| ACCATACCAGGCACC |
| ACCATACCAGGCACC |
| ACCATACCAGGCACC |
| ACCATACCAGGCACC |
| ACCATACCAGGCACC |

Motif

frequence

| | |
|-----|---|
| ACC | 1 |
| CCA | 1 |
| CAT | 1 |
| ATA | 1 |

Sequence

k=3

| |
|------------------|
| ACCATAACCAGGCACC |
| ACCATAACCAGGCACC |
| ACCATAACCAGGCACC |
| ACCATAACCAGGCACC |
| ACCATAACCAGGCACC |
| ACCATAACCAGGCACC |

Table hachage

| Motif | frequence |
|-------|-----------|
| ACC | 2 |
| CCA | 1 |
| CAT | 1 |
| ATA | 1 |

Sequence

k=3

ACCATACCAGGCACC

1

ACCATACCAGGCACC

2

ACCATACCAGGCACC

3

ACCATACCAGGCACC

4

ACCATACCAGGCACC

5

ACCATACCAGGCACC

6

ACCATACCAGGCACC

Table hachage

Motif

frequence

ACC

2

CCA

2

CAT

1

ATA

1

Sequence

k=3

1

2

3

4

5

6

7

ACCATACCAGGCACC

ACCATACCAGGCACC

ACCATACCAGGCACC

ACCATACCAGGCACC

ACCATACCAGGCACC

ACCATACCAGGCACC

ACCATACCAGGCACC

ACCATACCAGGCACC

Table hachage

Motif

frequence

ACC

2

CCA

2

CAT

1

ATA

1

CAG

1

k=3

ACCATACCAGGCACC

1

ACCATACCAGGCACC

2

ACCATACCAGGCACC

3

ACCATACCAGGCACC

4

ACCATACCAGGCACC

5

ACCATACCAGGCACC

6

ACCATACCAGGCACC

7

ACCATACCAGGCACC

8

ACCATACCAGGCACC

9

ACCATACCAGGCACC

10

ACCATACCAGGCACC

11

ACCATACCAGGCACC

12

ACCATACCAGGCACC

Table hachage

| | |
|-----|---|
| ACC | 3 |
| CCA | 2 |
| CAT | 1 |
| ATA | 1 |
| CAG | 1 |
| AGG | 1 |
| GGC | 1 |
| GCA | 1 |
| CAC | 1 |

Le motif de taille 3 le plus fréquent es ACC

Hash Table : complexité

$$(n - k + 1) * (k) = O(n*k - k^2 + k)$$

A retenir

- L'algorithme table de Hachage peut trouver des motifs **invariables** de taille **k** dans les séquences régulatrices.
- Il trouve les f motifs les plus fréquents, $f \geq 1$
- Complexité $\sim O(n \cdot k)$