

3. Prédiction des gènes

- Tous les gènes se terminent sur un codon stop
- Un algorithme simple de prédiction de gènes
- À la recherche des codons start et stop
- Prédiction de tous les gènes d'une séquence
- Comment améliorer la qualité des prédictions ?
- L'algorithme de Boyer-Moore
- **Index et arbre des suffixes**
- Des méthodes probabilistes à la rescousse
- Comment évaluer la qualité de prédiction des méthodes ?
- La prédiction de gènes dans les génomes eucaryotes

Pour améliorer la performance des algorithmes de recherche de motifs

- **Prétraiter le motif**
 - Algorithme de Boyer-Moore
- **Prétraiter le texte**
 - Index
 - Arbre des préfixes

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

Index de mots de longueur fixe

ACGGCTAGTTAGAA#



longueur = 14

Index de mots de longueur fixe

ACGGCTAGTTAGAA#
← longueur = 14 →

| | |
|-----|-------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

TAG

| | |
|-----|-------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

| |

TAG

| | |
|------------|--------------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

| |

GTTAG

TAG

| | |
|------------|--------------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

| |

GTTAG

TAG

| | |
|------------|--------------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

| |
GTTAG

TAG

| | |
|------------|--------------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

| |
GTTAG

TAG

| | |
|------------|--------------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

CGA

| | |
|------------|--------------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |

Index de mots de longueur fixe

ACGGCTAGTTAGAA#

CGA

| | |
|-----|-------|
| AA# | 13 |
| ACG | 1 |
| AGA | 11 |
| AGT | 7 |
| CGG | 2 |
| CTA | 5 |
| GAA | 12 |
| GCT | 4 |
| GGC | 3 |
| GTT | 8 |
| TAG | 6, 10 |
| TTA | 9 |



Arbre des suffixes

ACGGCTAGTTAGAA\$

ACGGCTAGTTAGAA\$

\$

A\$

AA\$

GAA\$

AGAA\$

TAGAA\$

TTAGAA\$

GTTAGAA\$

AGTTAGAA\$

...

Arbre des suffixes

