

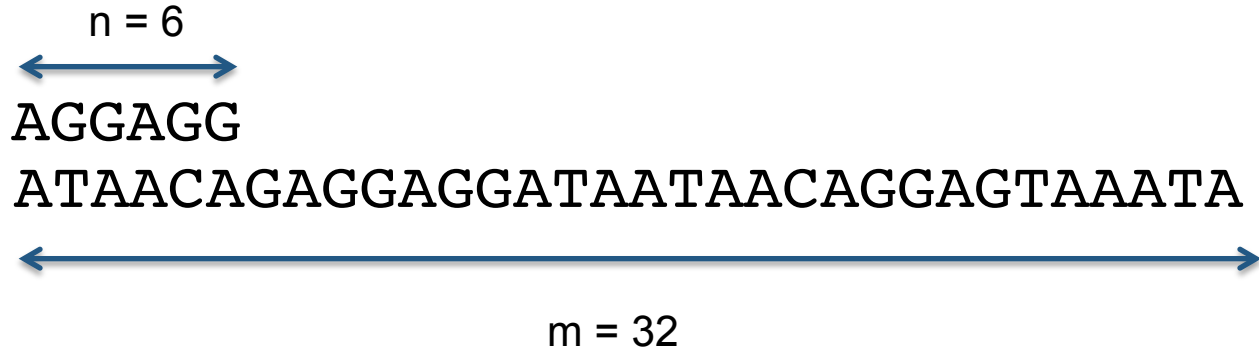
3. Prédiction des gènes

- 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

Algorithmes de recherche de motifs

- **Algorithme naïf**

- n : longueur du motif à rechercher
- m : longueur du texte
- dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$



Algorithmes de recherche de motifs

- **Algorithme naïf**
 - n : longueur du motif à rechercher
 - m : longueur du texte
 - dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**

- n : longueur du motif à rechercher
- m : longueur du texte
- dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**

- n : longueur du motif à rechercher
- m : longueur du texte
- dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**
 - n : longueur du motif à rechercher
 - m : longueur du texte
 - dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**

- n : longueur du motif à rechercher
- m : longueur du texte
- dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**

- n : longueur du motif à rechercher
- m : longueur du texte
- dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**

- n : longueur du motif à rechercher
- m : longueur du texte
- dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Algorithmes de recherche de motifs

- **Algorithme naïf**
 - n : longueur du motif à rechercher
 - m : longueur du texte
 - dans le pire des cas, nombre de comparaisons : $O(n * (m-n))$

AGGAGG

ATAACAGAGGAGGATAATAACAGGAGTAAATA

Pour améliorer la performance des algorithmes de recherche de motifs

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

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement

CGGCTG

ATAACAGGAGTAAATAACGGCTGGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- Un exemple frappant :
 - pas de A dans le motif à rechercher
 - on saute directement à la position courante + 6 (longueur du motif)

CGGCTG

ATAACAGGAGTAAATAACGGCTGGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- Un exemple frappant :
 - pas de A dans le motif à rechercher
 - on saute directement à la position courante + 6 (longueur du motif)

CGGCTG

ATAACAGGAGTAAATAACGGCTGGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- Un exemple frappant :
 - pas de A dans le motif à rechercher
 - on saute directement à la position courante + 6 (longueur du motif)

CGGCTG

ATAACAGGAGTAAATAACGGCTGGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- Un exemple frappant :
 - il existe un autre C dans le motif
 - on saute directement à ce C

CGGCTC
ATAACAGGAGTAAATAACGGCTCGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- Un exemple frappant :
 - il existe un autre G dans le motif
 - on saute directement à ce G

CGGCTC
ATAACAGGAGTAAATAACGGCTCGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- Un exemple frappant :
 - Le motif est trouvé !

CGGCTC
ATAACAGGAGTAAATA**CGGCTC**CGAGTAAATA

L'algorithme de Boyer-Moore

- Commencer les comparaisons à partir de la fin du motif
- Utiliser les informations sur les occurrences des lettres au sein du motif, obtenues lors du prétraitement
- **Plus le motif est long, plus le gain est important !**