

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

Recherche d'un triplet donné

TAG

ATTGCTTACTAGAAATCGTACGGGTACGTAAATCGTATTCCGAT

Recherche d'un triplet donné

TAG

ATTGCTTACTAGAAATCGTACGGGTACGTAAATCGTATTCCGAT

Recherche d'un triplet donné

TAG

ATTGCTTACTAGAAATCGTACGGGTACGTAAATCGTATTCCGAT

Recherche d'un triplet donné

TAG

ATTGCTTAC**TAG**AATCGTACGGGTACGTAAATCGTATTCCGAT

Recherche d'un triplet donné

TAG

ATTGCTTAC**TAG**AATCGTACGGGTACGTAAATCGTATTCCGAT

Nombre de comparaisons dans le cas le plus défavorable pour la recherche d'un codon stop

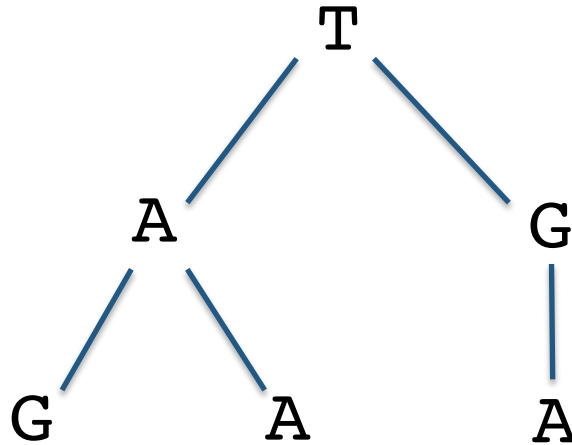
= longueur de la séquence

Pour les 3 codons stop

= 3 * longueur de la séquence

Codons stop

TGA
ou
TAG
ou
TAA



La fonction NextStopCodon

```
Function NextStopCodon (index: integer) returns integer  
  if index + 2 ≤ lengthSequence  
  then  
    repeat  
      if Sequence [index] = "T"  
        then  
          if Sequence [index+1] = "A"  
            then if sequence [index+2] = "G" or sequence [index+2] = "A"  
              then return index  
              else index ← index + 3  
            else if Sequence [index+1] = "G" and sequence [index+2] = "A"  
              then return index  
              else index ← index + 3  
          else index ← index + 3  
        until index ≥ lengthSequence - 2  
      return 0  
    else return 0  
  end NextStopCodon
```