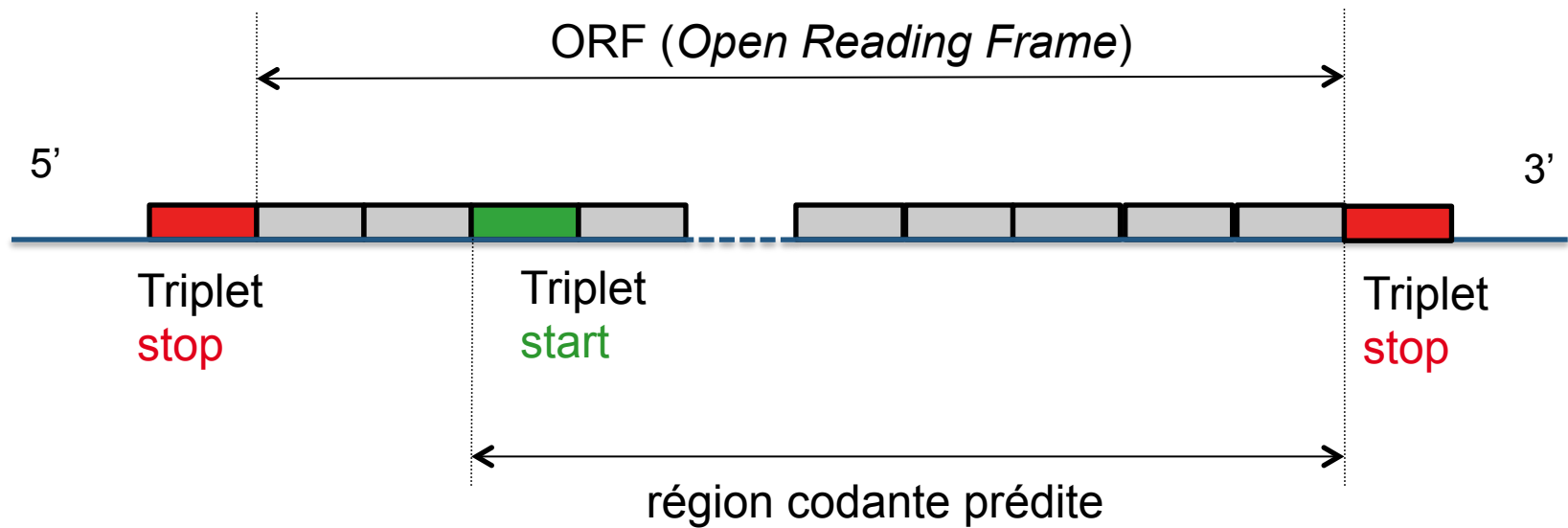


3. Prédiction des gènes

- 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



```

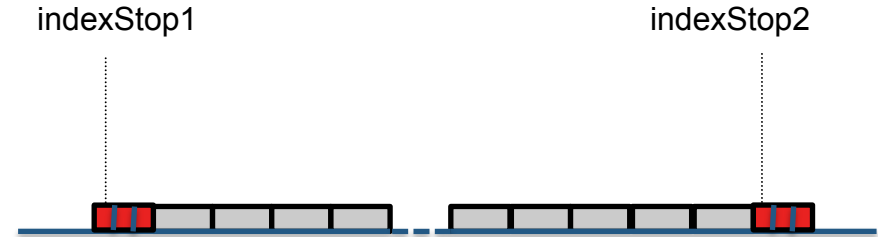
indexStop1, indexStop2, indexStart, indexDNA, iGene: integer
Gene: array [1:*, 1:2] of integer
indexDNA ← 1
iGene ← 1
repeat
  indexStop1 ← NextStopCodon (indexDNA)
  if indexStop1 > 0 then
    indexStop2 ← NextStopCodon (indexStop1+3)
    Length = IndexStop2 - IndexStop1 + 1
    if Length ≥ LengthMin then
      IndexStart ← NextStartCodon (indexStop1+3)
      if indexStart > 0 and indexStart < IndexStop2 then
        Gene [iGene, 1] ← indexStart
        Gene [iGene, 2] ← indexStop2
        iGene ← iGene +1
      endif
    endif
  endif
  indexDNA ← max (indexStop2, indexStop1, indexDNA+3)
until (indexStop1 = 0) or (indexDNA > LengthSequence)

```

```

indexStop1, indexStop2, indexStart, indexDNA, iGene: integer
Gene: array [1:*, 1:2] of integer
indexDNA ← 1
iGene ← 1
repeat
  indexStop1 ← NextStopCodon (indexDNA)
  if indexStop1 > 0 then
    indexStop2 ← NextStopCodon (indexStop1+3)
    Length = IndexStop2 - IndexStop1 + 1
    if Length ≥ LengthMin then
      IndexStart ← NextStartCodon (indexStop1+3)
      if indexStart > 0 and indexStart < IndexStop2 then
        Gene [iGene, 1] ← indexStart
        Gene [iGene, 2] ← indexStop2
        iGene ← iGene + 1
      endif
    endif
  endif
  indexDNA ← max (indexStop2, indexStop1, indexDNA+3)
until (indexStop1 = 0) or (indexDNA > LengthSequence)

```



```

indexStop1, indexStop2, indexStart, indexDNA, iGene: integer
Gene: array [1:*, 1:2] of integer
indexDNA ← 1
iGene ← 1
repeat
  indexStop1 ← NextStopCodon (indexDNA)
  if indexStop1 > 0 then
    indexStop2 ← NextStopCodon (indexStop1+3)
    Length = IndexStop2 - IndexStop1 + 1
    if Length ≥ LengthMin then
      IndexStart ← NextStartCodon (indexStop1+3)
      if indexStart > 0 and indexStart < IndexStop2 then
        Gene [iGene, 1] ← indexStart
        Gene [iGene, 2] ← indexStop2
        iGene ← iGene + 1
      endif
    endif
  endif
  indexDNA ← max (indexStop2, indexStop1, indexDNA+3)
until (indexStop1 = 0) or (indexDNA > LengthSequence)

```

