

2. Gènes et protéines

- La séquence est-elle un bon modèle de l'ADN ?
- Les gènes, de Mendel à la biologie moléculaire
- Le code génétique
- Un algorithme de traduction
- Implémenter le code génétique
- Algorithmes + structures de données = programmes
- Les compromis de la conception d'algorithmes
- Les technologies de séquençage de l'ADN
- Le séquençage de génomes complets
- **Comment trouver les gènes ?**

Annotation de génomes

- **Prédiction des**
 - **positions** de gènes
 - **fonctions** des protéines codées par ces gènes
- **Un génome bactérien**
 - typiquement (*E. coli*) 4,5 Mb
 - ≈ 4 500 gènes
- **Un génome humain**
 - 3,5 Gb
 - ≈ 20 000 gènes « seulement »

De quelles connaissances disposons-nous ?

- Un **gène** (région codante) commence par un **codon start** : ATG

Mais

situé au sein d'une région codante, ATG code la méthionine

- Un gène se termine toujours par **un des 3 codons stop** : TAA, TAG, TGA
- Un **codon stop** ne peut pas apparaître au sein d'une région codante

Mais

des triplets start et stop peuvent apparaître en dehors des régions codantes, les dites « régions inter-géniques »

Et ce n'est pas tout !

- Une région codante est une **succession de triplets**, alors appelés **codons**
 - Les codons **start** et **stop** doivent être espacés d'un nombre de bases multiple de 3
- Il y a **trois manières de grouper 3 par 3 les bases** d'une séquence, en commençant soit :
 1. par la base i
 2. par la base $i+1$
 3. par la base $i+2$

Et ce n'est pas tout !

- Une région codante est une **succession de triplets**, alors appelés **codons**
 - Les codons **start** et **stop** doivent être espacés d'un nombre de bases multiple de 3
- Il y a **trois manières de grouper 3 par 3 les bases** d'une séquence, en commençant soit :

1. par la base i

CCTAGCTAATTGCTATTAATTGTGTCATGACGTCTAG

2. par la base $i+1$

CCTAGCTAATTGCTATTAATTGTGTCATGACGTCTAG

3. par la base $i+2$

CCTAGCTAATTGCTATTAATTGTGTCATGACGTCTAG

Et ce n'est pas tout !

- Une région codante est une **succession de triplets**, alors appelés **codons**
 - Les codons **start** et **stop** doivent être espacés d'un nombre de bases multiple de 3
- Il y a **trois manières de grouper 3 par 3 les bases** d'une séquence, en commençant soit :

1. par la base i

CCTAGCT**TA**ATTGCTA**TTA**ATTGTGTCAT**TGA**CGTCTAG

2. par la base $i+1$

CCTAGCTAATTGCTAT**TA**ATTGTGTCATGACGTC**TAG**

3. par la base $i+2$

CC**TAG**CTAATTGCTATTAATTGTGTC**ATG**ACGTCTAG

- **Trois phases de lecture**

Et ce n'est pas tout !

- Une région codante est une **succession de triplets**, alors appelés **codons**
 - Les codons **start** et **stop** doivent être espacés d'un nombre de bases multiple de 3
- Il y a **trois manières de grouper 3 par 3 les bases** d'une séquence, en commençant soit :

1. par la base i

CCTAGCT**TA**ATTGCTA**TTA**ATTGTGTCAT**GAC**GTCTAG

2. par la base $i+1$

CCTAGCTAATTGCTAT**TA**ATTGTGTCATGACGT**CTAG**

3. par la base $i+2$

C**CTAG**CTAATTGCTATTAATTGTGTC**ATG**ACGTCTAG

- **Trois phases de lecture**
- Les **codons start** et **stop** d'un gène doivent être dans la même phase
- Les **deux brins d'ADN portent des gènes**