

BIOINFORMATIQUE : GÉNOMES ET ALGORITHMES

Analyse informatique de l'information génétique

François
Rechenmann



GÉNOMES ET ALGORITHMES

1. ADN et séquences génomiques
- 2. Gènes et protéines**
3. Prédiction des gènes
4. Comparaison de séquences
5. Arbres phylogénétiques

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 ?

L'ADN est une molécule

- Organisée en **chromosomes, plasmides, segments,...**
- Compactée **au sein de la cellule**
 - Dans le noyau des cellules eucaryotes
 - Directement au sein du cytoplasme des cellules procaryotes
- Elle peut subir des **modifications chimiques**
 - Exemple: méthylation
- Une simple séquence de caractères peut-elle
constituer un bon modèle de cette molécule ?

Qu'est-ce qu'un bon modèle ?

- **Un modèle exact et complet ?**
 - Aucun modèle ne peut être une représentation exacte de son objet
 - Seul l'objet est un modèle exact et complet de lui-même
 - Rappelez-vous l'exemple mythique de la carte à l'échelle 1:1
 - ✓ Infaisable
 - ✓ Inutilisable
- Un modèle est une **description simplifiée et orientée** de la réalité
- Orientée par l'ensemble des questions auxquelles il doit contribuer à répondre

Qu'est-ce qu'un bon modèle ?

- **Un modèle exact et complet ?**
 - Aucun modèle ne peut être une représentation exacte de son objet
 - Seul l'objet est un modèle exact et complet de lui-même
 - Rappelez-vous l'exemple mythique de la carte à l'échelle 1:1
 - ✓ Infaisable
 - ✓ Inutilisable
- Un modèle est une **description simplifiée et orientée** de la réalité
- Orientée par l'ensemble des questions auxquelles il doit contribuer à répondre

Un bon modèle est
un modèle utile

La séquence est-elle un bon modèle de l'ADN ?

- C'est un **texte linéaire** qui ignore la structure tridimensionnelle de la molécule
 - Des régions éloignées dans le texte peuvent se retrouver proches dans la molécule compactée
- Elle ne retient pertinente que la **succession des nucléotides**

La séquence est-elle un bon modèle de l'ADN ?

- C'est un **texte linéaire** qui ignore la structure tridimensionnelle de la molécule
 - Des régions éloignées dans le texte peuvent se retrouver proches dans la molécule compactée
- Elle ne retient pertinente que la **succession des nucléotides**

Oui, mais ce modèle se révèle utile,
et même très utile.

La séquence est-elle un bon modèle de l'ADN ?

- C'est un texte linéaire qui ignore la structure tridimensionnelle de la molécule
 - Des régions éloignées dans le texte peuvent se retrouver proches dans la molécule compactée
- Elle ne retient pertinente que la succession des nucléotides

Oui, mais ce modèle se révèle utile,
et même très utile.

- Un exemple vu : la **prédiction de l'origine de réplication**
- Un nouvel exemple à venir : la **prédiction des gènes**

Le vocabulaire reflète les niveaux de modélisation

- Le **biologiste** : « molécule d'ADN », « nucléotides »
- Le **bioinformaticien** : « séquence génomique », « bases »
- L'**algorithmicien** : « chaîne de caractères », « lettres »