

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

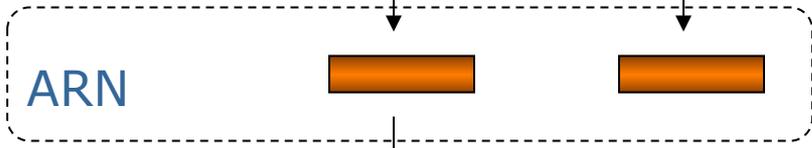
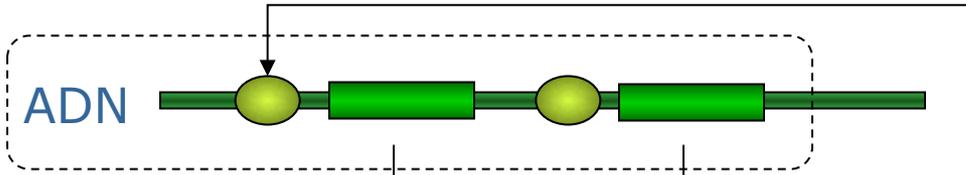
4. Comparaison de séquences

- **Comment prédire les fonctions des gènes/protéines ?**
- Évolution et similarité de séquences
- Quantifier la similarité de deux séquences
- L'alignement de séquences devient un problème d'optimisation
- Un alignement de séquences vu comme un chemin dans une grille
- Si un chemin est optimal, tous ses chemins partiels sont optimaux
- Coûts et alignement
- Un algorithme récursif
- Éviter la récursivité : une version itérative
- Cet algorithme est-il efficace ?

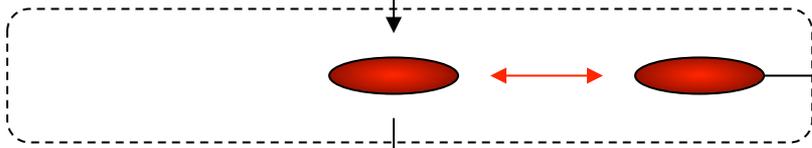
Bases de données sur les gènes et les protéines

- GeneBank, UniProt,...
- À une séquence d'un gène ou d'une protéine sont associées des **informations sur leurs fonctions**
- Sous la forme de
 - **Texte libre**
 - **Mots-clés** (termes GO)
 - Entrées de la **classification enzymatique**

Gènes

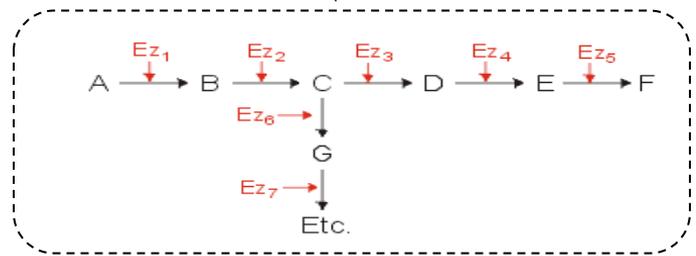


Protéines



Régulation

Enzymes



Réactions et voies
métaboliques

Bases de données sur les gènes et les protéines

- GeneBank, UniProt,...
- À une séquence d'un gène ou d'une protéine sont associées des **informations sur leurs fonctions**
- Sous la forme de
 - **Texte libre**
 - **Mots-clés** (termes GO)
 - Entrées de la **classification enzymatique**
 - ✓ tripeptide aminopeptidases : EC 3.4.11.4

Tables extraites des bases de données

Séquence	Organisme	Fonctions (<i>EC number</i>)
ACCGTTACG...	E.coli	3.4.11.4
ACTTTTGCC...	B. subtilis	2.3.4
TGGTATGCT...	H. influenzae	4.1.1.3

Prédiction des fonctions

- Pour une séquence de gène ou de protéine
 1. Rechercher dans la première colonne de la table une séquence similaire
 2. Quand une séquence similaire est trouvée, enregistrer les informations associées
 3. Poursuivre la recherche d'autres séquences similaires

Prédiction des fonctions

- Pour une séquence de gène ou de protéine
 1. Rechercher dans la première colonne de la table une séquence similaire
 2. Quand une séquence similaire est trouvée, enregistrer les informations associées
 3. Poursuivre la recherche d'autres séquences similaires

Comment évaluer la similarité
de deux séquences ?