



**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE MINISTERE DE
L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE UNIVERSITÉ
ABOU BEKR BELKAID – TLEMCEN**

FACULTÉ :SNV

Département de Biologie

Filière :Biologie moléculaire et cellulaire (M1)

Module: La bio-informatique appliquée à la génétique:

« Les banques de données et ses applications »

Présenté par : DR SALMI T

Année 2023/2024

La bio-informatique appliquée à la génétique

1^{ere}
définition

une discipline appliquée dans tous les domaines :
physique , chimie, biologie moléculaire et la génétique

- ❖ L'association de la biologie et de l'informatique
- ❖ L'analyse et traitement des données et de l'information biologique
- ❖ L'information liée aux molécules biologiques :
 - leurs séquences
 - leurs structures
 - leurs fonctions
 - leurs interactions
 - leur intégration dans la cellule

La bio-informatique appliquée à la génétique

La génétique in silico

in silico :

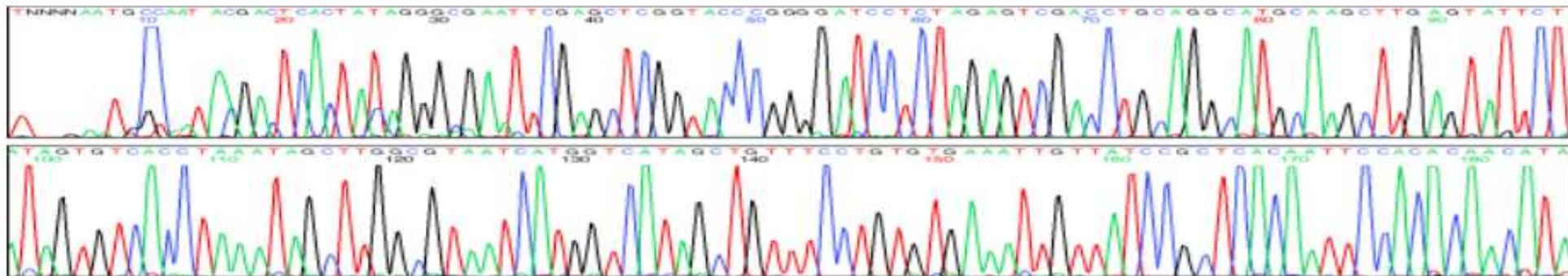
se réfère à l'outil informatique.

cela veut dire



l'utilisation des processeurs, logiciels informatiques pour gérer, traiter et analyser l'information biologique contenue essentiellement dans les séquences nucléiques et protéiques

Exemple d'enregistrement obtenu à partir d'un séquenceur automatique



La bio-informatique

Discipline récente: (dans les années quatre vingt)

Discipline hybride : issue de diverses disciplines

- la biochimie
- la génétique (la génomique structurale, la génomique fonctionnelle)
- la protéomique



Historique

Premiers développements

1965 : première compilation papier « Atlas of Protein Sequences »

Premiers supports informatiques

1971 : première version de la PDB au Brookhaven National Laboratory

1981-1982 : premières versions de EMBL et Genbank

1986 : première version de Swissprot

2002 : première version de Genbank WGS

2013 (oct) : PDB comporte 94 540 structures dont 87 516 de protéines

2023 : État actuel :

Genbank rel 197 qui contient 167 295 840 entrées et 154 192 921 011 bases.

Bio-informatique et les Sciences Des

« OMES »

ADN (Génome = génomique)

- Séquences de nucléotides
- Séquences de gènes
- Banques de données

ARN(Transcriptome = Transcriptomique)

- Séquence
- Structure

Protéines(Protéome = Protéomique)

- Séquence
- Structure

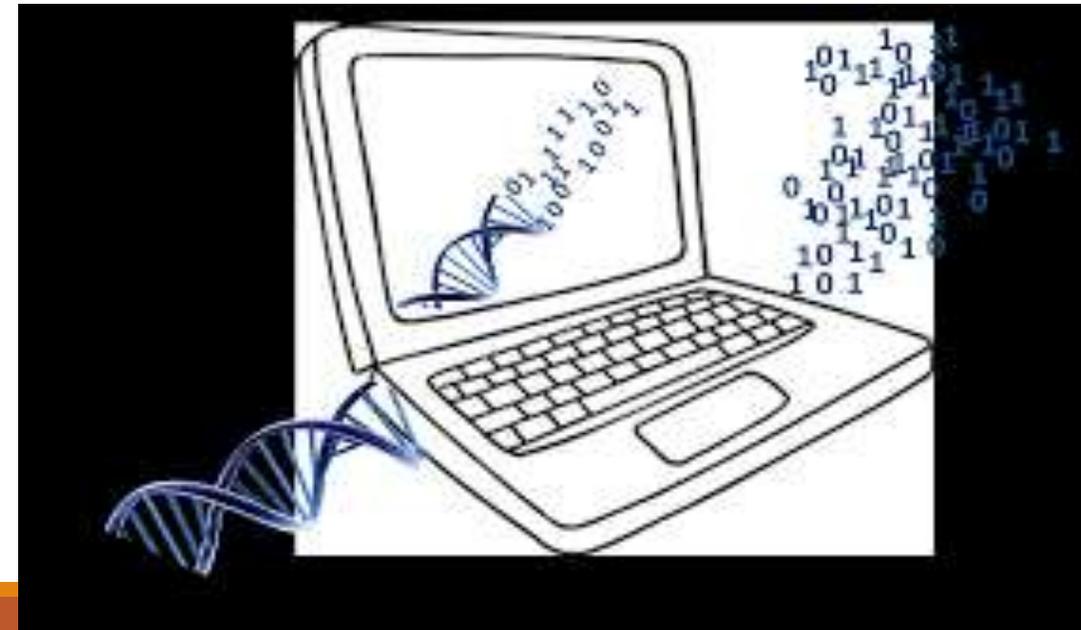
- Réseaux d'interaction

2^{ème}
définition

La bio-informatique

C'est une discipline qui permet l'analyse et l'interprétation des informations biologiques contenues soit dans le génome (séquences ADN, ARN) soit dans le protéome.

nouvellement incluse dans les
systèmes d'enseignement
biologique



Les objectifs

- ❖ **Interpréter les données biomoléculaires**
- ❖ **Analyser la structure des molécules**
- ❖ **Prédire le rôle et la fonction de cette structure**
- ❖ **Organisation des données biologiques dans des bases de données**

Les objectifs

❖ **Traitements systématiques des données** : repérer et caractériser une fonction et/ou une structure biologique importante

❖ **Elaboration de stratégies** :

- apporter des connaissances biologiques supplémentaires en combinant les données biologiques initiales et les données biologiques obtenues "in silico".

- ces connaissances permettent, à leur tour, de développer de nouveaux concepts en biologie.

- concepts qui, pour être validés, peuvent nécessiter le développement de nouvelles théories et outils en mathématiques et en informatique

Les applications de la bio-informatique

L'acquisition des données biologiques

- l'analyse de génomes entiers (exons, introns)
- la détection des SNP ou d'insertion / délétion
- les séquences nucléotidiques et les séquences polypeptidiques
- PCR in vitro (enzymes de restriction, amorces)

Le séquençage

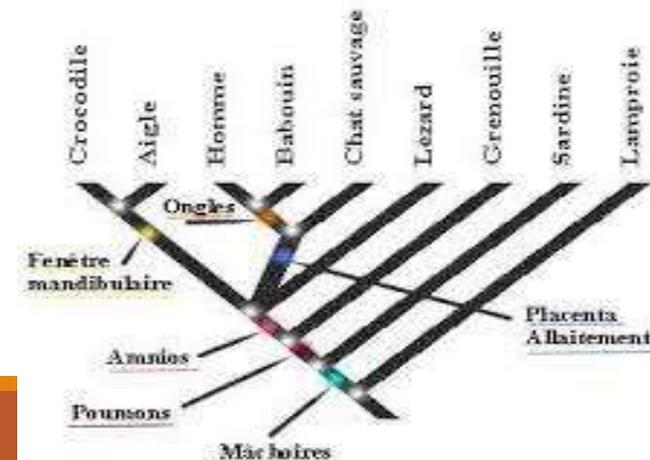
Depuis l'invention du séquençage de l'ADN par Frederick Sanger (1977) les progrès technologiques dans la bio-informatique ont été développés

Modélisation moléculaire

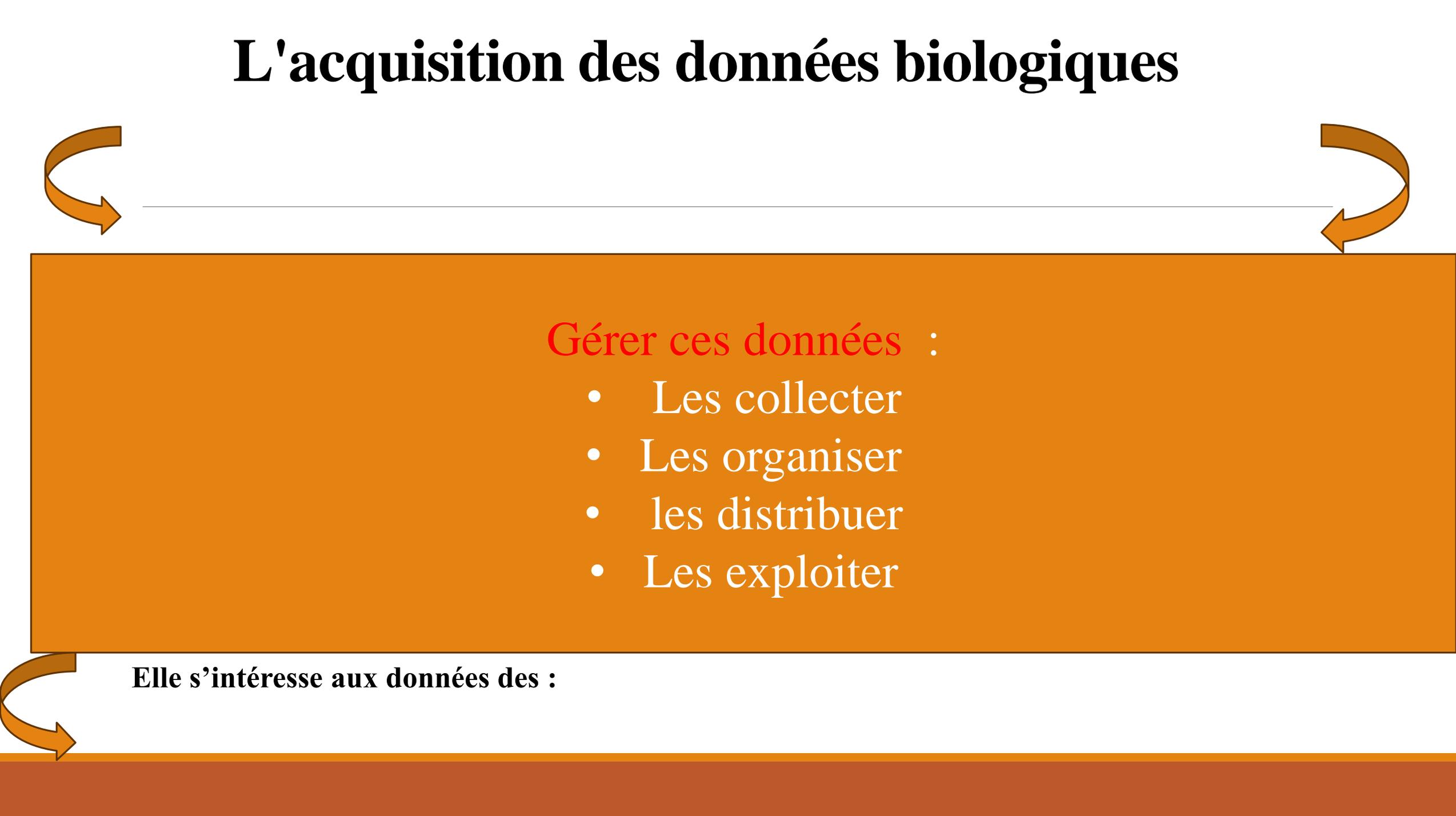
L'objectif est d'obtenir une reconstruction tridimensionnelle présentant la meilleure adéquation avec les résultats expérimentaux. Ces données sont issues principalement d'analyses cristallographiques

Construction d'arbres phylogénétiques

une représentation schématique permettant de mettre en avant une parenté entre [espèces](#) ou groupes d'espèces.



L'acquisition des données biologiques



Gérer ces données :

- Les collecter
- Les organiser
- les distribuer
- Les exploiter

Elle s'intéresse aux données des :

L'acquisition des données biologiques

Différents types de données :

- génomique
- transcriptome
- protéome
- métabolome (molécules organiques telles que lipides, glucides, faisant partie des activités métaboliques de la cellule)

Gérer l'information

Les banques de données

Ensemble de données relatif à un domaine défini des connaissances, généralement organisé et structuré



en base de données

- 1) Ensemble de données structuré généralement en champs
- 2) un fichier permettant le stockage des informations ainsi que l'accès à ces informations devenues structurées
- 3) un tableau dans lequel on intègre des informations de manière logique et structurée

La base de données

Ce sont des bibliothèques électroniques et informatisées qui contiennent des informations :

- sur les sciences de la vie
- collectées grâce à des expériences scientifiques
- à la littérature publiée
- aux technologies expérimentales à haut débit
- et aux analyses informatiques.

La différence entre base et banque ?



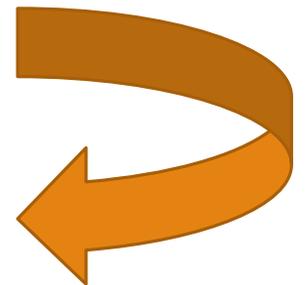
- ❖ **Une banque de données** est une base de données (tableau structuré) qui contient des informations biologiques hétérogènes (virus, bactéries, champignons, végétaux, animaux)
- ❖ **Une base de données** est plus spécialisée (base spécifique à E. coli, à Bacillus, etc.)

La différence entre base et banque ?

Les banques de données ou **bases de données** biologiques sont des bibliothèques contenant des données de séquences biologiques collectées grâce à des expériences scientifiques, largement diffusées par le réseau internet.

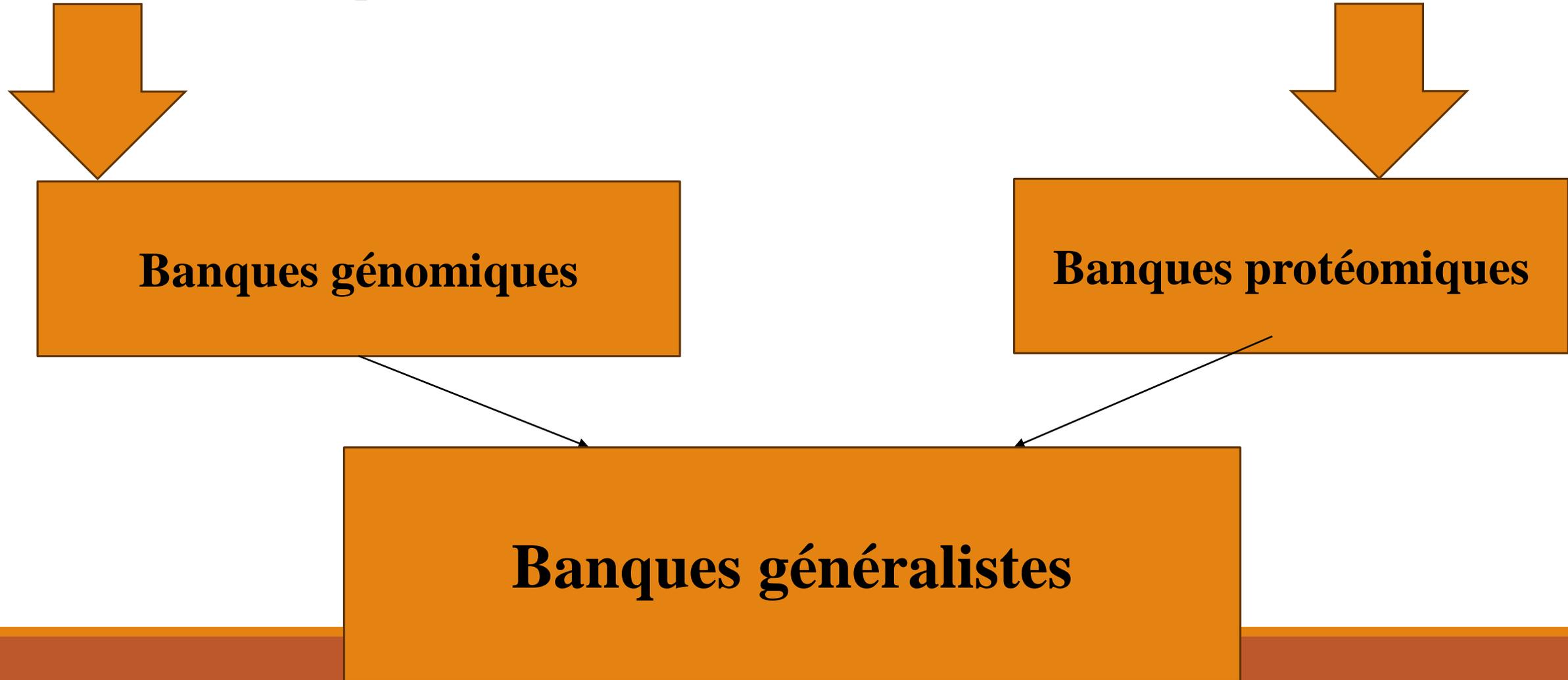


Elles sont généralement reliées entre elles par des liens.

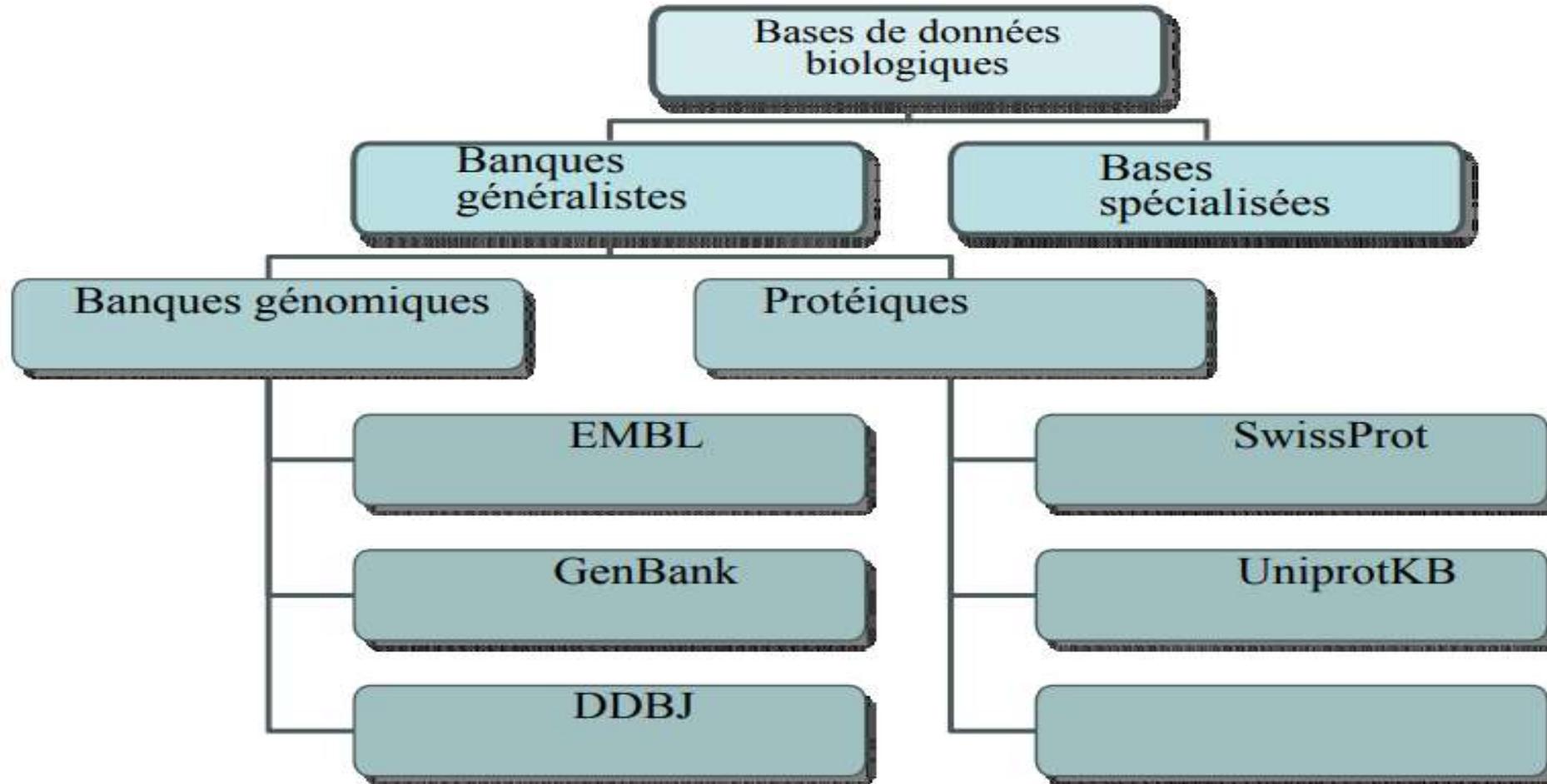


La bio-informatique est définie aussi comme l'utilisation **de base de données** pour analyser:

- les gènes
- les protéines



Exemples



Les banques de données généralistes, spécifiques :

Généralistes

Spécialistes

contiennent des données
hétérogènes

contiennent des données
homogènes

Collecte la plus
exhaustive possible

Collecte établie autour d'une
thématique particulière

Exemples

→ Banques de séquences nucléiques généralistes			
Nom	Lien	Date de création	Description
EMBL	http://www.ebi.ac.uk/embl/	1980	Banque européenne (European Molecular Biology Laboratory) diffusée par l'EBI (European Bioinformatics Institute, Cambridge)
GenBank	http://www.ncbi.nlm.nih.gov/	1982	Banque américaine diffusée par NCBI (National Center for Biotechnology Information, Los Alamos)
DDBJ	http://www.ddbj.nig.ac.jp/	1986	DNA Data Bank of Japan diffusée par le NIG (National Institute of Genetics)
→ Banques de séquences protéiques généralistes			
UniProt	https://www.uniprot.org/	1986	Séquences annotées & séquences codantes traduite de l'EMBL

Exemples

→ Banques de données spécialisées		
Ensembl	https://www.ensembl.org/index.html	Banque intégrative génomique
Prosite	http://prosite.expasy.org/	Recense les motifs protéiques ayant une signification biologique
Reactome	https://reactome.org/PathwayBrowser/	Banque intégrative métabolique
Kegg Pathway	http://www.genome.jp/kegg/pathway.html	Interactions moléculaires et réactions
PFAM	http://xfam.org/	Domaines protéiques
Interpro	http://www.ebi.ac.uk/interpro/	Regroupe plusieurs banques existantes

Exemples:

Base de données	Description	Site web (ou adresse internet)
Genbank	Banques de séquences nucléotidiques publiquement disponibles et leur traduction en protéines.	www.ncbi.nlm.nih.gov/enter (Acides nucléiques)
EMBL (European Molecular Biology Library).		www.srs.ebi.ac.jp (Acides nucléiques)
NCBI (national center for biotechnology Information)	Portail américain de bioinformatique: BLAST, banques de séquences nucléotidiques et génomiques, PubMed.	www.ncbi.nlm.nih.gov (Portail)
Pubmed-Medline	Articles, publications en biologie et en médecine	www.ncbi.nlm.nih.gov (Bibliographie)
Google Scholar	Publication scientifiques avec google	http://scholar.google.com (Bibliographie)
Connectsciences	Articles, thèses	http://connectsciences.inist.fr (Bibliographie)

Les structures tridimensionnelles des protéines et aussi, malgré leur nombre plus restreint, des acides nucléiques (en particulier les ARN de transfert).	Protein Data Bank
Les données obtenues en protéomique (gels d'électrophorèse bidimensionnel).	SWISS-2DPAGE
Le changement d'un nucléotide dans un gène quelconque ("Single Nucleotide Polymorphism").	SNP
La taxonomie (classification) des organismes.	Taxonomy
Les réseaux d'interactions qu'établissent les molécules biologiques.	BioCarta
Les voies métaboliques .	KEGG
L'ontologie : l'organisation hiérarchique de la connaissance sur un ensemble d'objets par leur regroupement en sous-catégories suivant leurs caractéristiques essentielles.	GO
Les données bibliographiques (diffusion des résultats de la recherche par les articles).	PubMed

**MERCI
DE VOTRE
ATTENTION**
