

Introduction à la Bioinformatique 2024



- Ezechiel Bionimian TIBIRI, Ph.D
- Biologie moléculaire – Bioinformatique



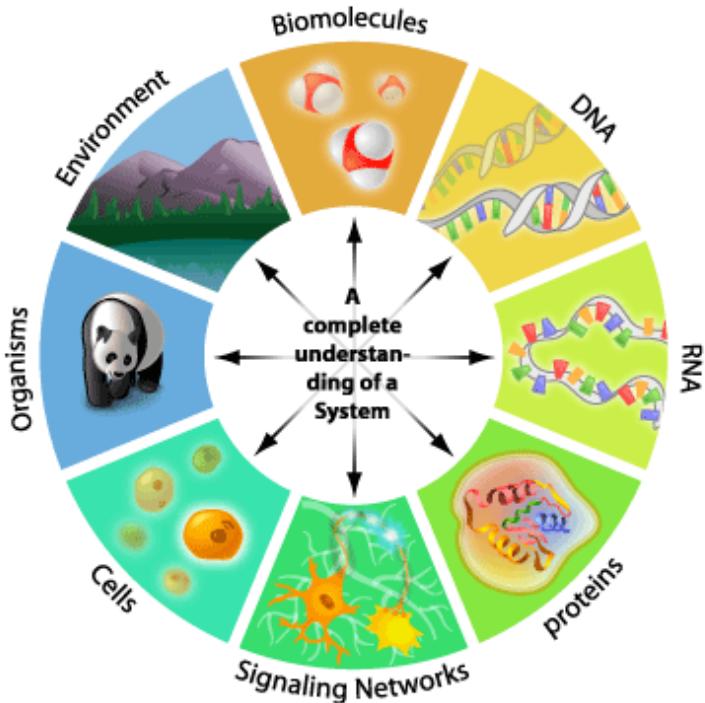
Introduction aux bases de données et aux ressources

Recherche de séquences

Objectifs du cours

- Comprendre la structure et la disposition des ressources de données du NCBI et de l'EBI
- Comprendre la différence entre les bases de données, les outils et les repositories.
- Recherche de données dans des bases de données spécifiques à l'aide de numéros d'accès, de noms de gènes, etc.
- Utiliser les ressources NCBI et EBI

Data



Introduction

- Plusieurs bases de données et ressources en ligne
- Besoin de savoir laquelle :
 - ✓ Quelles sont les bases de données et les ressources existantes
 - ✓ Quels sont les outils disponibles pour exploiter ces ressources ?
 - ✓ Quels sont les outils disponibles pour rechercher dans les ressources?

Bases de données biologiques



Bases de données biologiques

- Bases de données biologiques sont :
 - ✓ Publique ou privée
 - ✓ Protéine, nucléotide, structure, littérature, annotation...
 - ✓ Généralisée ou spécialisée
 - ✓ Centré sur la séquence (aa ou nt) ou le génome

Bases de données biologiques

Quelques noms de banques de données:

- ❖ **Séquences en acides nucléiques** (DNA et mRNA); [EMBL](#), [GenBank](#), [DDBJ](#)
- ❖ **Séquences en acides aminés** (protéines); [Swiss-Prot](#), [wwPDB](#)
- ❖ **Références bibliographiques**; [PubMed](#)
- ❖ **Informations générales sur les gènes et/ou les maladies**; [EntrezGene](#), [OMIM](#), [HMGD](#)

Bases de données primaires

- International Nucleotide Sequence Database Collaboration (INSDC)
- Données de séquences génomiques stockées dans 3 bases de données publiques
- Chacun a son propre



Bases de données secondaires

- Des bases de données spécialisées construites à partir de données de séquences primaires
- Fournissent plusieurs ressources et annotations différentes

Ressources bioinformatiques les plus populaires

- National Centre for Biotechnology Information (NCBI)



- European Bioinformatics Institute (EMBL-EBI)



Recherche dans les bases de données: NCBI

Google ncbi

Tous Vidéos Images Livres Maps Plus Outils

Environ 87 900 000 résultats (0,49 secondes)

<https://www.ncbi.nlm.nih.gov> Traduire cette page

National Center for Biotechnology Information

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information. About the NCBI | ...

PubMed
PubMed® comprises more than 32 million citations for biomedical ...

Gene
Advanced search - RefSeqGene - OMIM - Genome Workbench

BLAST
Nucleotide - Standard Protein
BLAST - Nucleotide BLAST - ...

Proteins
Protein - Protein Clusters - Identical Protein Groups - ...

Nucleotide
Nucleotide. The Nucleotide database is a collection of ...

All Resources
A database of human genes and genetic disorders. NCBI ...

Autres résultats sur nih.gov »

National Center for Biotechnology Information

Entreprise

Le National Center for Biotechnology Information, en français « Centre américain pour les informations biotechnologiques », est un institut national américain pour l'information biologique moléculaire. [Wikipédia](#)

Fondateur : Claude Denson Pepper

Création : 4 novembre 1988

Organisation mère : United States National Library of Medicine



Recherche dans les bases de données NCBI

NCBI Resources How To

All Databases 

Sign in to NCBI

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

UNITE

A new NIH initiative to end structural racism and achieve racial equity in the biomedical research enterprise.

Ending Structural Racism NIH

LEARN MORE

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit

Deposit data or manuscripts into NCBI databases

Download

Transfer NCBI data to your computer

Learn

Find help documents, attend a class or watch a tutorial

Popular Resources

PubMed

Bookshelf

PubMed Central

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

Develop

Use NCBI APIs and code libraries to build applications

Analyze

Identify an NCBI tool for your data analysis task

Research

Explore NCBI research and collaborative projects

NCBI News & Blog

BLAST+ 2.12.0 now available with more efficient multithreaded searches

09 Jul 2021

BLAST+ 2.12.0 programs feature better multithreaded searches and support a

https://github.com/Ezechiel-Tibiri/bioinformatique_2023

Codeathon from the Couch — NCBI

Menu déroulant des différentes BD de NCBI

Recherche dans les bases de données: NCBI

The screenshot shows the NCBI homepage with a search bar at the top. Below the search bar, there is a banner for COVID-19 Public health information and another for Ending Structural Racism. A sidebar on the left contains links for various databases, with 'Gene' highlighted in blue. The main content area features a mission statement about advancing science and health through genomic information. At the bottom, there are sections for 'Submit', 'Download', and 'Learn'.

NCBI Resources ▾ How To ▾ Sign in to NCBI

All Databases ▾

- Assembly
- Biocollections
- BioProject
- BioSample
- BioSystems
- Books
- ClinVar
- Conserved Domains
- dbGaP
- dbVar
- Gene**
- Genome
- GEO DataSets
- GEO Profiles
- GTR
- HomoloGene
- Identical Protein Groups
- MedGen
- MeSH
- NCBI Web Site

COVID-19
Public health information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

Ending structural racism and achieve racial equity in the biomedical research enterprise.

NCBI

National Center for Biotechnology Information advances science and health by providing access to genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein

Bases de données de NCBI

- NCBI comprend plus de 30 bases de données
- la littérature : [PubMed Central \(PMC\)](#), [Bookshelf](#) et [PubReader](#)
- La santé: [ClinVar](#), [dbGaP](#), [dbMHC](#), the [Genetic Testing Registry](#), [HIV-1/Human Protein Interaction Database](#) et [MedGen](#)
- Les génomes: [BioProject](#), [Assembly](#), [Genome](#), [BioSample](#), [dbSNP](#), [dbVar](#), [Nucleotide](#), [Probe](#) et [RefSeq](#).
- Les gènes: [Gene](#), [Gene Expression Omnibus \(GEO\)](#), [HomoloGene](#), [PopSet](#), [Refseq](#) et [UniGene](#).
- Les protéines: [Protein](#), the [Conserved Domain Database \(CDD\)](#), [COBALT](#), [Conserved Domain Architecture Retrieval Tool \(CDART\)](#), the [Molecular Modeling Database \(MMDB\)](#), [Refseq](#) et [Protein Clusters](#).
- Les produits chimiques: [Biosystems](#) et [PubChem](#)

EMBL - EBI

- Maintenir la gamme la plus complète au monde de bases de données moléculaires librement accessibles et actualisées
- Proposer des formations en ligne et en direct pour l'utilisation de leurs ressources.
- <https://www.ebi.ac.uk/training>

EMBL - EBI

The EMBL-EBI website has been redesigned. Please [send us feedback](#) about this page.

EMBL's European Bioinformatics Institute

EMBL-EBI

Unleashing the potential of big data in biology

Find a gene, protein or chemical

Example searches: [blast keratin bfl1](#) | [About EBI Search](#)

[Find data resources](#) [Submit data](#) [Explore our research](#) [Train with us](#)

Latest news



Organisations should embrace open science faster – interview with Prof. Dame Janet Thornton

17 May 2022



Europe PMC: Harnessing the power of text mining to accelerate life sciences research

12 May 2022



2.4 billion sequences now available in the latest MGnify protein database release

11 May 2022



[Predicted complexes from ModelArchive now on PDBe-KB pages](#)

6 May 2022

Services

[Overview](#)[A to Z](#)[Data submission](#)[Research infrastructure development programme](#)[Support](#)

The European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our web services to access our resources programmatically.
— You can read more about our services in the journal *Nucleic Acids Research*

Tools & Data Resources

Tools

Clustal Omega



Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

[Web API](#)[Multiple sequence alignment](#)

InterProScan



InterProScan searches sequences against InterPro's predictive protein signatures.

[Web API](#)[Protein feature detection](#)[Sequence motif recognition](#)

BLAST [protein]



Fast local similarity search tool for protein sequence databases.

[Web API](#)[Sequence similarity search](#)

BLAST [nucleotide]



Fast local similarity search tool for nucleotide sequence databases.

Data resources

Ensembl



Genome browser, API and database, providing access to reference genome annotation

[Web API](#)[EMBL-EBI Terms of use](#)

UniProt



A comprehensive resource for protein sequence and functional annotation.

[Web API](#)[CC-BY](#)

PDBe



The European resource for the collection, organisation and dissemination of 3D structural data (from PDB and EMDB) on biological macromolecules and their complexes.

[Web API](#)[CC0](#)

Europe PMC



A database to search the worldwide life sciences literature

[Web API](#)[EMBL-EBI Terms of use](#)

Browse by type

XXX	DNA & RNA	Gene Expression
Structures	Systems	Proteins
Ontologies	Literature	Chemical biology
		Cross domain

Programmatic access

EMBL-EBI web services allow you to query our large biological data resources programmatically, so that you can develop data analysis pipelines or integrate public data with your own applications. The Web Services technology we use are built on open standards to ensure client and server software from various sources will work well together.

[Browse EMBL-EBI web services](#)

Principles of service provision

Open

Our data and tools are freely available, without restriction. The only exception is access to the highly sensitive human genetic information, for which access depends

https://github.com/Ezechiel-Tibiri/bioinformatique_2023

Bases de données spécialisées

- Il existe un grand nombre de bases de données spécialisées
 - ❖ La plupart des séquences sont également dans la banque GenBank/EMBL
 - ❖ Peut contenir des génomes entiers
 - ❖ Peut contenir des ressources spécialisées
 - ❖ Contient des outils spécifiques pour l'exploitation des données

Bases de données spécialisées

- Plasmodium <https://plasmodb.org/plasmo/app>
- Les collections spécialisées de Sanger
<https://www.sanger.ac.uk>
- Base de données sur les hépatites
https://hcv.lanl.gov/content/sequence/HCV/news/old_news.html
- Base de données de recherche sur la grippe influenza
<https://www.fludb.org/brc/home.spg?decorator=influenza>

Design d'amorce utilisant Primer Blast

The screenshot shows the Primer-BLAST search interface. At the top, there is a header with the NIH logo, the text "An official website of the United States government [Here's how you know](#)", and a user email "tibionez@gmail.com". Below the header, the title "Primer-BLAST" is displayed, followed by the subtitle "A tool for finding specific primers". A sub-instruction "Finding primers specific to your PCR template (using Primer3 and BLAST)." is also present.

The main form is divided into several sections:

- Primers for target on one template**: This section includes fields for "PCR Template" (with options to enter accession, gi, or FASTA sequence, or upload a file), "Range" (for forward and reverse primers), and "Primer Parameters" (such as PCR product size, number of primers to return, and primer melting temperatures).
- Exon/intron selection**: This section requires a refseq mRNA sequence as PCR template input and includes settings for exon junction span, exon junction match, and intron inclusion.
- Primer Pair Specificity Checking Parameters**: This section includes a "Specificity check" field with a dropdown menu showing "1000" and "1000L" with a question mark icon.

At the bottom right of the interface, the URL https://github.com/Ezechiel-Tibiri/bioinformatique_2023 is displayed.

Take home

- Une grande quantité de données existent
- Les bases de données primaires stockent les données brutes des séquences
- Les bases de données secondaires fournissent des informations sur l'annotation des données de séquence.
- Il est important de savoir comment et où les données sont stockées
- NCBI et EBI sont les deux ressources les plus populaires pour obtenir des données biologiques.