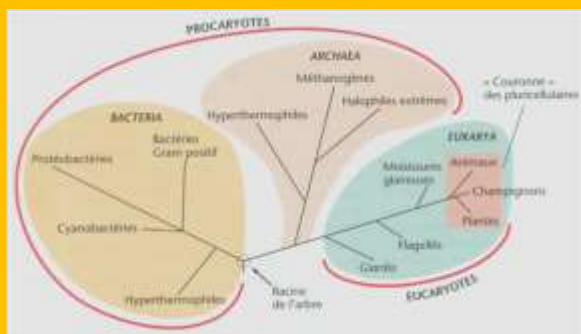
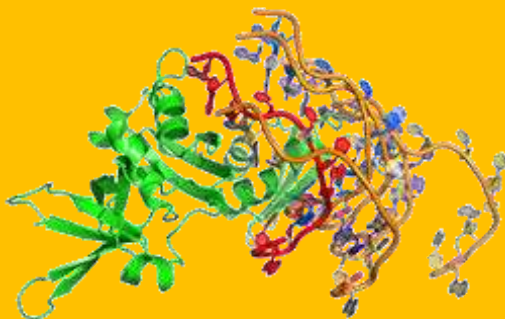


Université Frères Mentouri Constantine 1  
Institut de la Nutrition, de l'Alimentation et des Technologies Agro-alimentaires (INATAA)  
1<sup>e</sup> année Master Biotechnologie alimentaire  
2019-2020



# COURS DE BIOINFORMATIQUE



# **CE QUE NOUS ALLONS VOIR:**

- ✓ Introduction à la Bioinformatique et à ses applications ;
- ✓ Les bases de données biologiques ;
- ✓ Les alignements : comparaison de 02 séquences nucléiques et protéiques, les alignements multiples ;
- ✓ Phylogénie moléculaire et construction des arbres phylogénétiques ;
- ✓ Annotation des séquences.

# OBJECTIFS DU COURS:

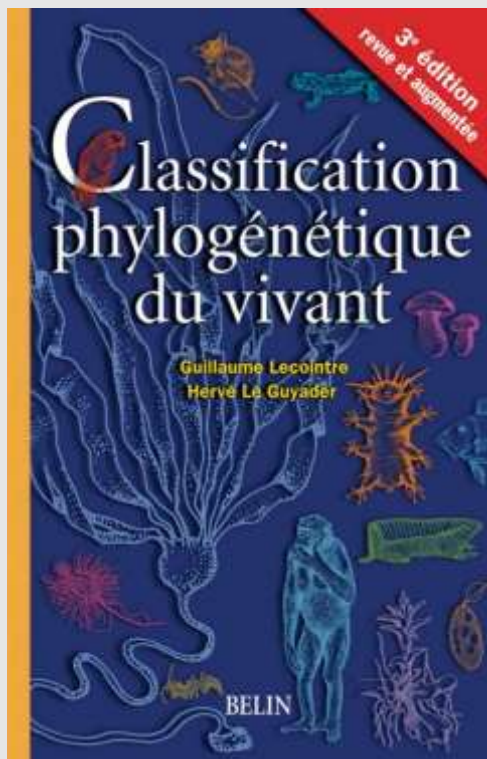
A la fin de ce cours, l'étudiant sera capable de :

- ✓ Interroger une base de données biologiques ;
- ✓ Récolter et analyser des séquences nucléiques et protéiques : aligner, annoter, réaliser un arbre phylogénétique en utilisant différents outils.

# LIVRES A CONSULTER



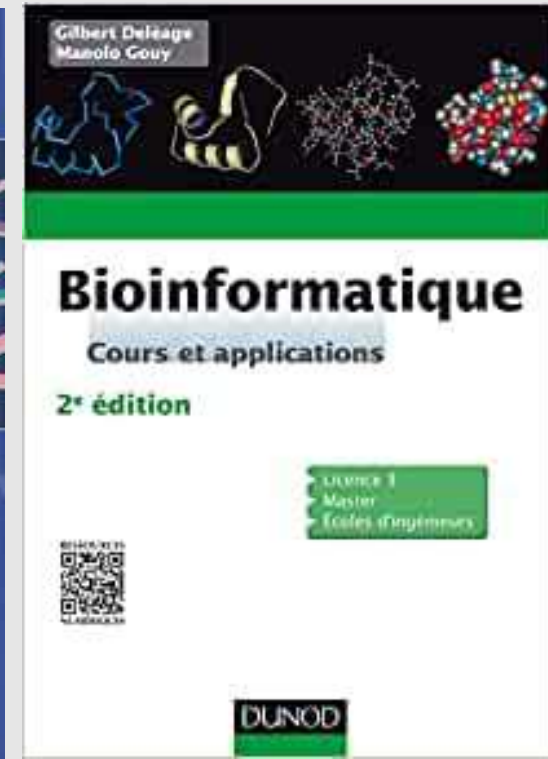
**NA/8544 ET NA/8546  
NA/5047 À NA/5049  
(CD ROM)**



**NA/8220/1 À  
NA/8226/1**

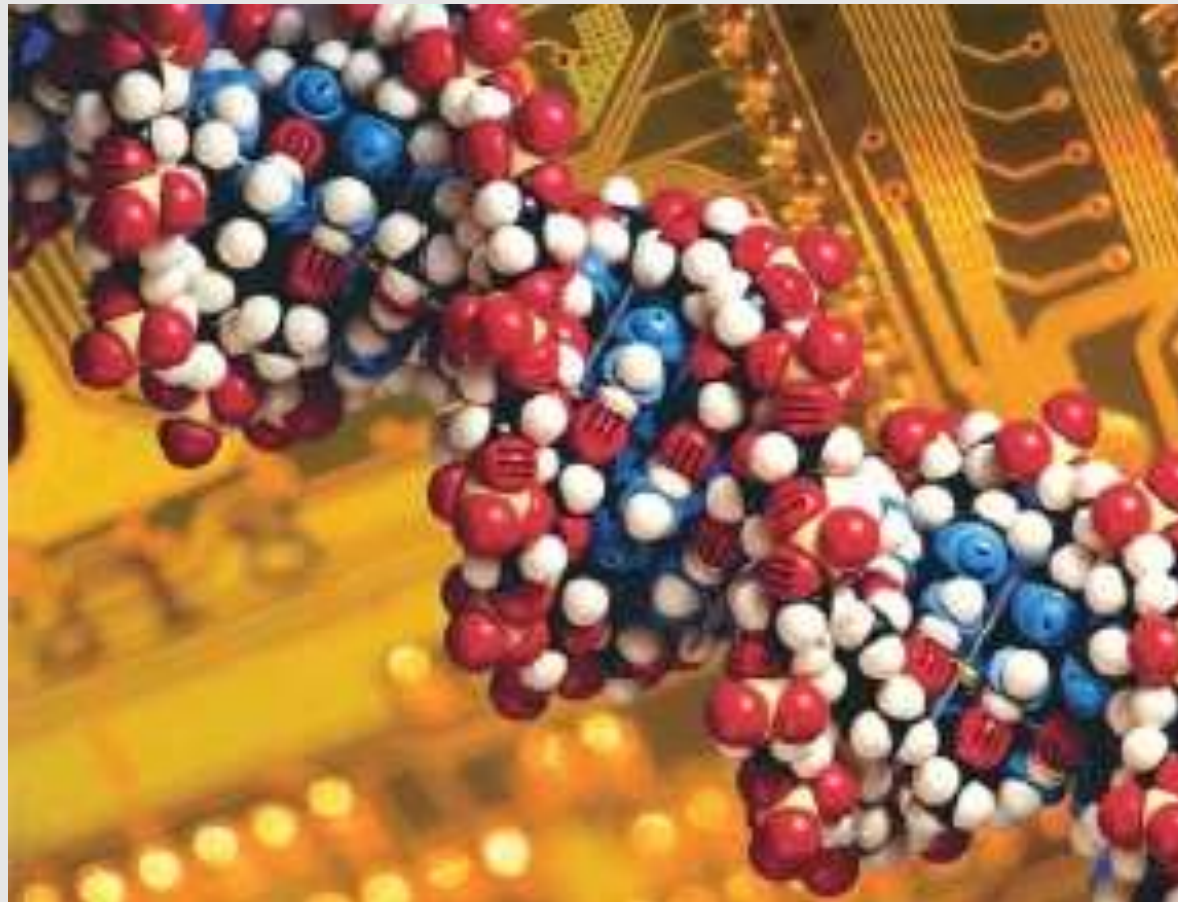


**NA/7836  
À  
NA/7839**

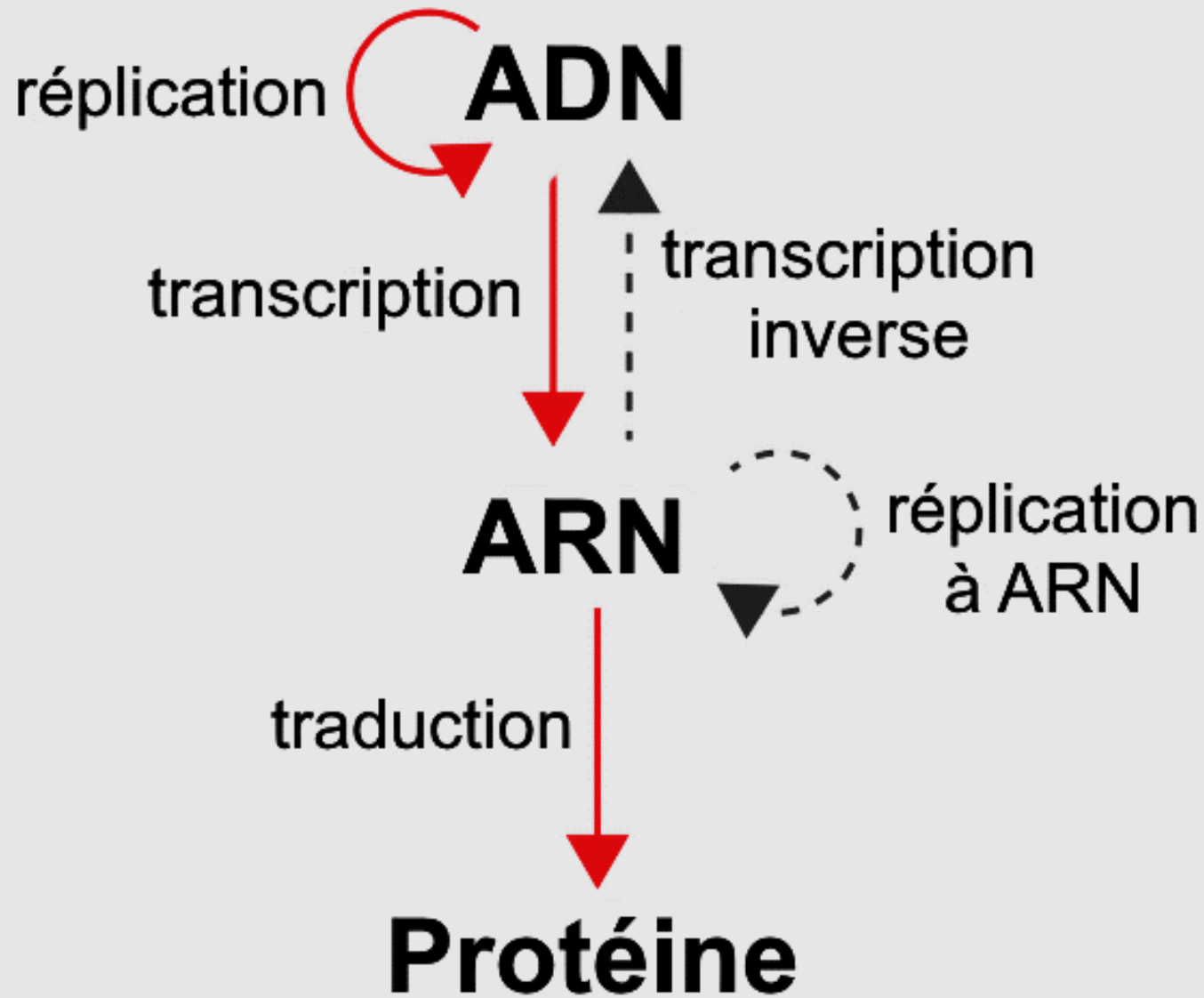




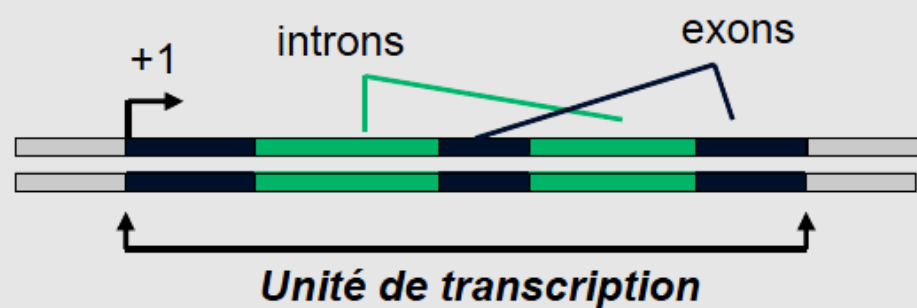
# **INTRODUCTION À LA BIOINFORMATIQUE ET À SES APPLICATIONS**



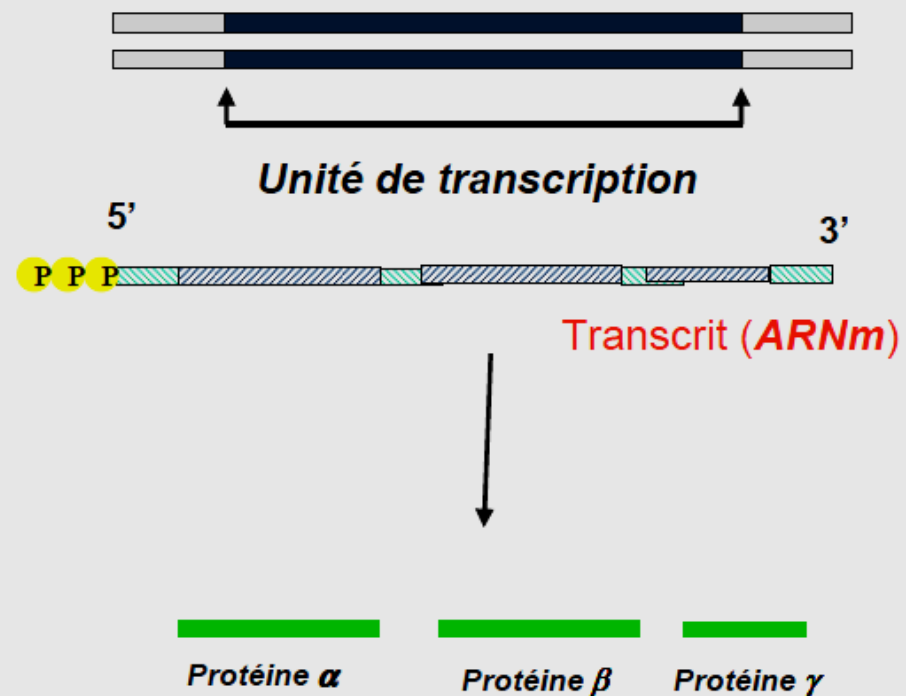
## DOGME CENTRAL DE LA BIOLOGIE MOLÉCULAIRE



## DOGME CENTRAL DE LA BIOLOGIE MOLÉCULAIRE

*Eucaryote*

ADN

*Procaryote*

## DÉFINITION DE LA BIOINFORMATIQUE

- La bioinformatique est le **traitement de l'information biologique** sous forme de données accessibles aisément et exploitables.
- Elle est également définie comme étant la (multi) discipline théorique de l'analyse "*in silico*" de l'information biologique contenue dans les **séquences** nucléiques et protéiques\*.

---

\*la bioinformatique tire sa définition de deux concepts importants : la biologie et l'information car le suffixe informatique ne veut pas forcément signifier l'utilisation des ordinateurs pour la biologie.



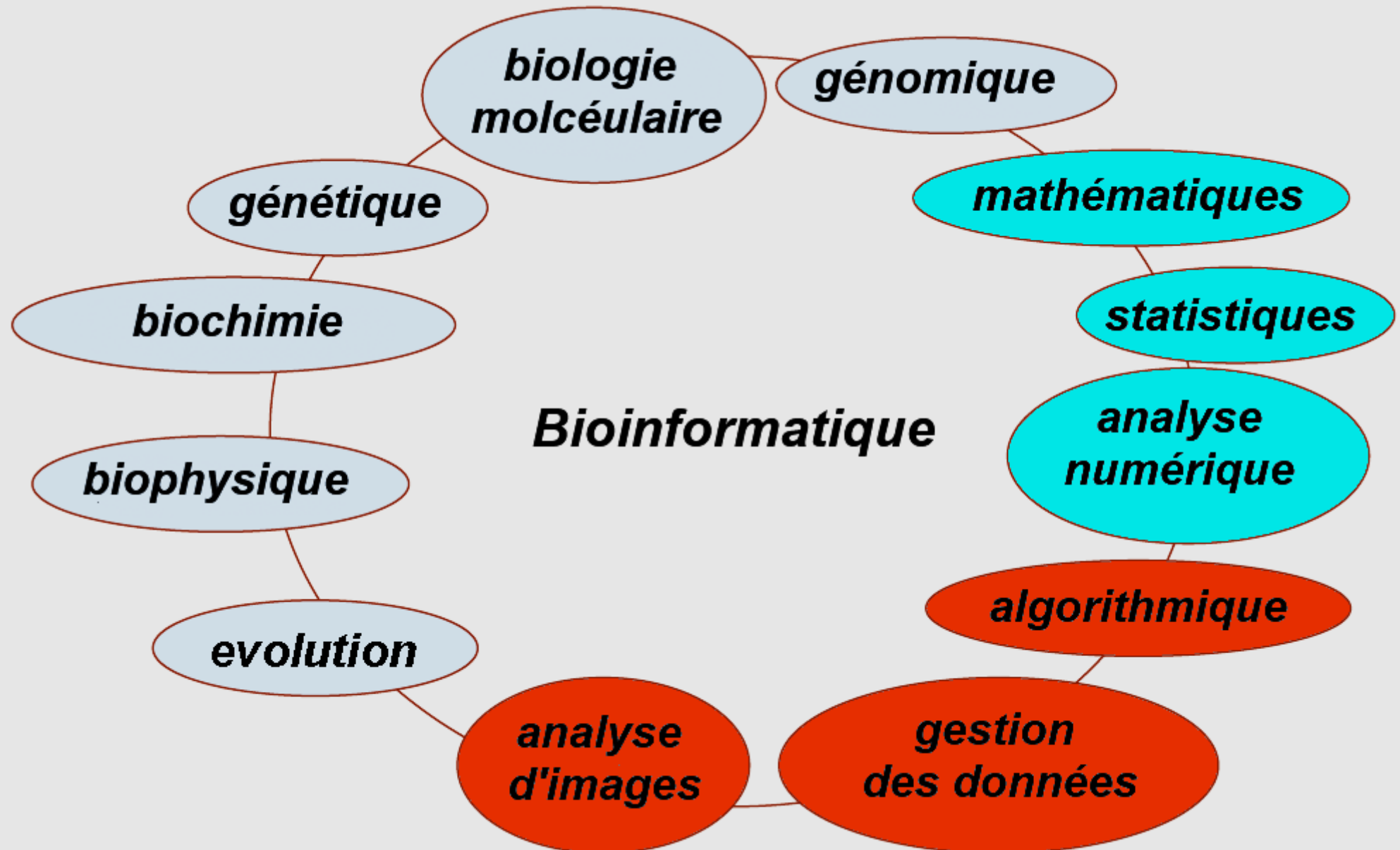
**L'INFORMATION BIOLOGIQUE**

La bioinformatique s'intéresse aux données liées au :

- ✓ génome (totalité du matériel génétique de la cellule) ;
- ✓ transcriptome (ensemble des ARNm transcrits) ;
- ✓ protéome (l'ensemble des protéines bio-synthétisées) ;
- ✓ métabolome (molécules organiques -métabolites- impliquées dans les activités métaboliques de la cellule vivante).

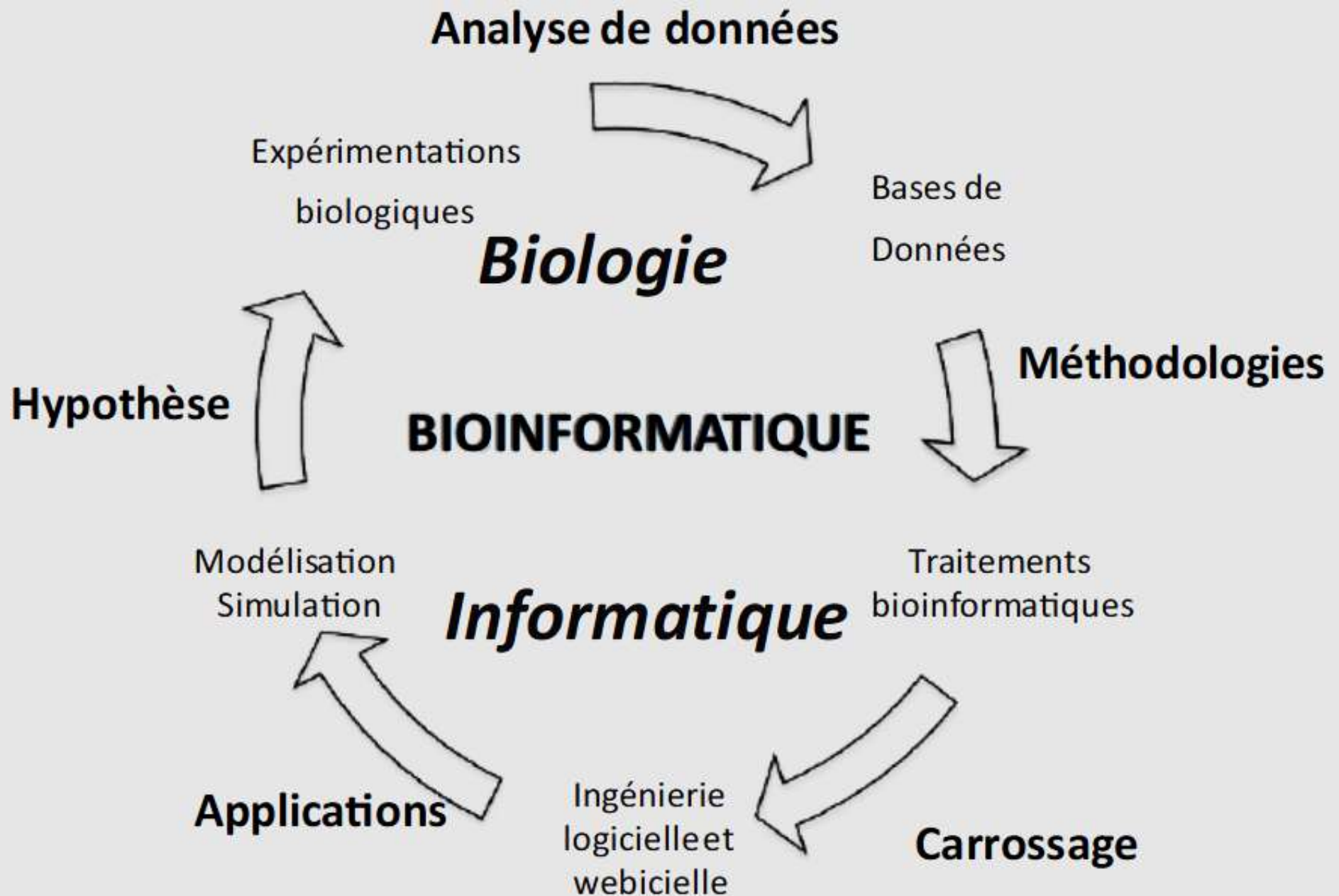
## DISCIPLINES DE LA BIOINFORMATIQUE

La bioinformatique est un champ d'étude multidisciplinaire:

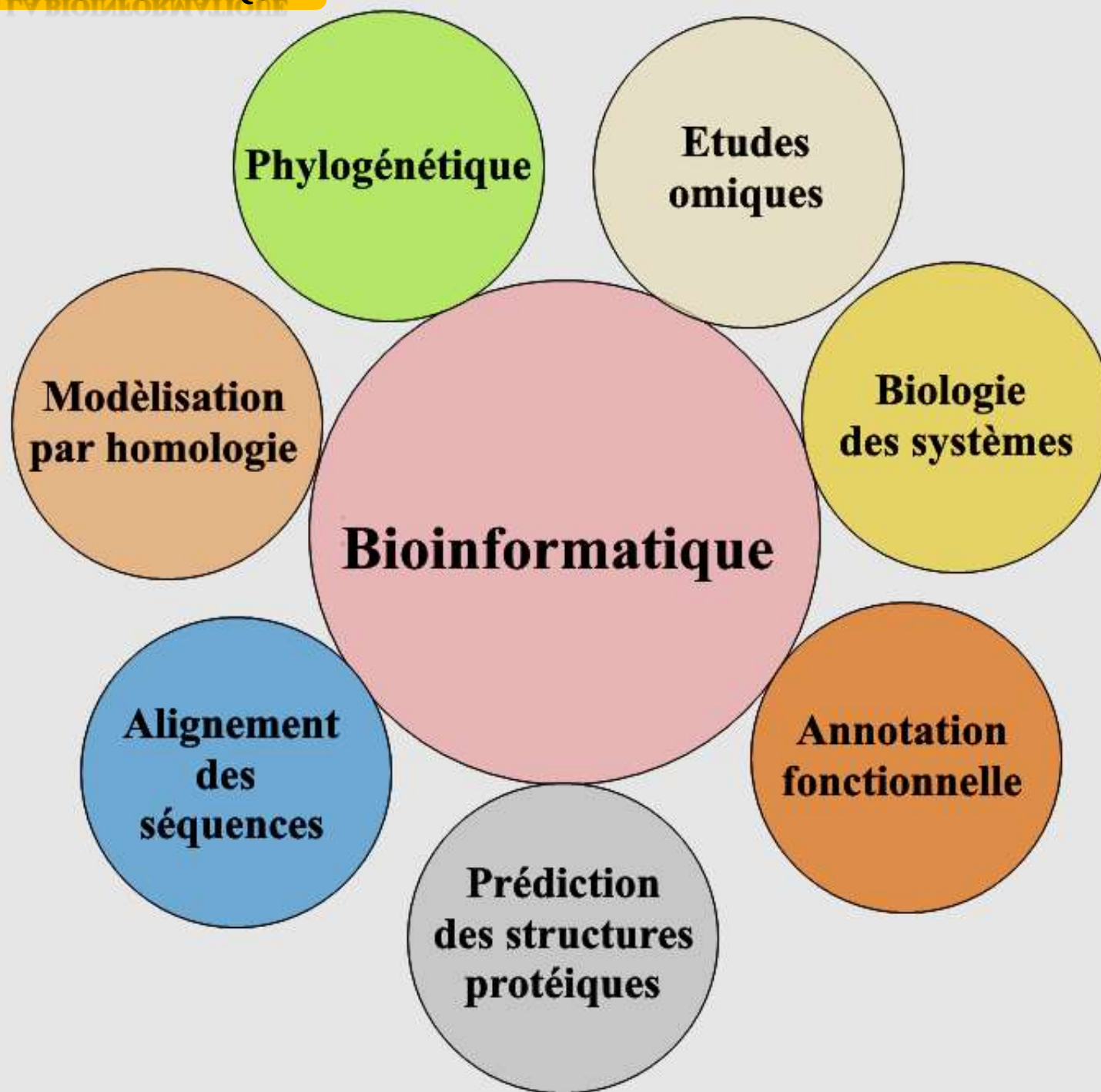


## DISCIPLINES DE LA BIOINFORMATIQUE

La bioinformatique est un champ d'étude multidisciplinaire:



## APPLICATIONS DE LA BIOINFORMATIQUE



## APPLICATIONS DE LA BIOINFORMATIQUE

La bioinformatique utilise des méthodes et des logiciels qui permettent :

- ✓ Le recueil, le stockage et la gestion des données biologiques et leur distribution à travers les réseaux ;
- ✓ Le développement des outils pour analyser les problèmes de biologie moléculaire, notamment:



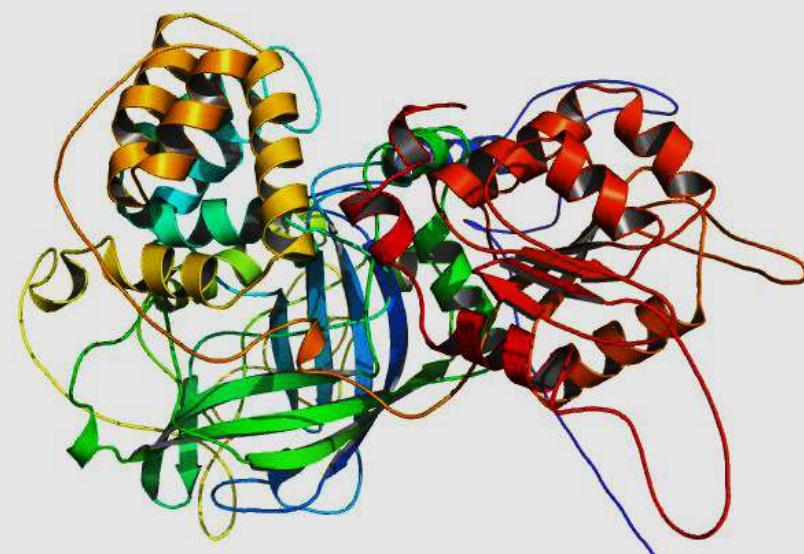
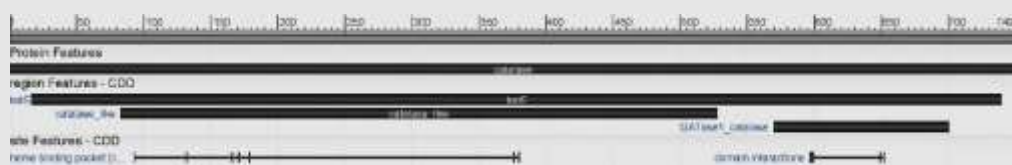
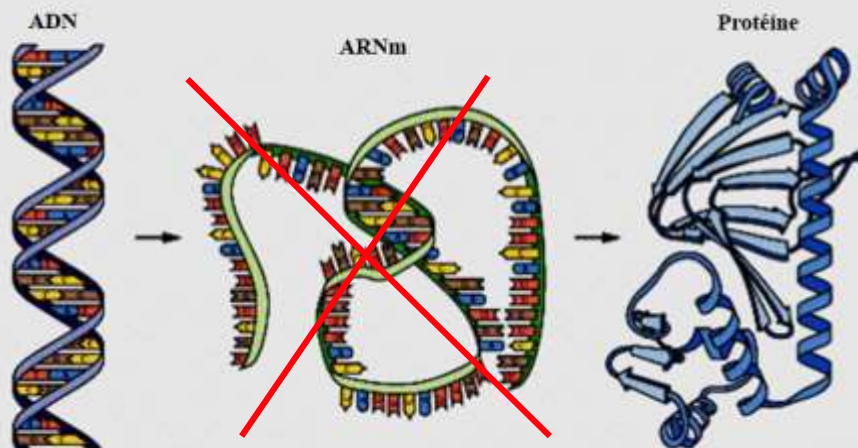
**APPLICATIONS DE LA BIOINFORMATIQUE**

- ❑ L'analyse, la comparaison et la prédiction de la structure des gènes, et des génomes ;
- ❑ La modélisation et la prédiction de la structure et de la fonction des protéines ;
- ❑ L'étude des réseaux de régulation métaboliques, génomiques, protéomiques, métagénomiques et biologiques en générale.
- ❑ Les études phylogénétiques et l'évolution moléculaires des êtres vivants.

## EXEMPLES D'APPLICATIONS

## PRÉDICTION DE STRUCTURE, FONCTION ET RÔLE

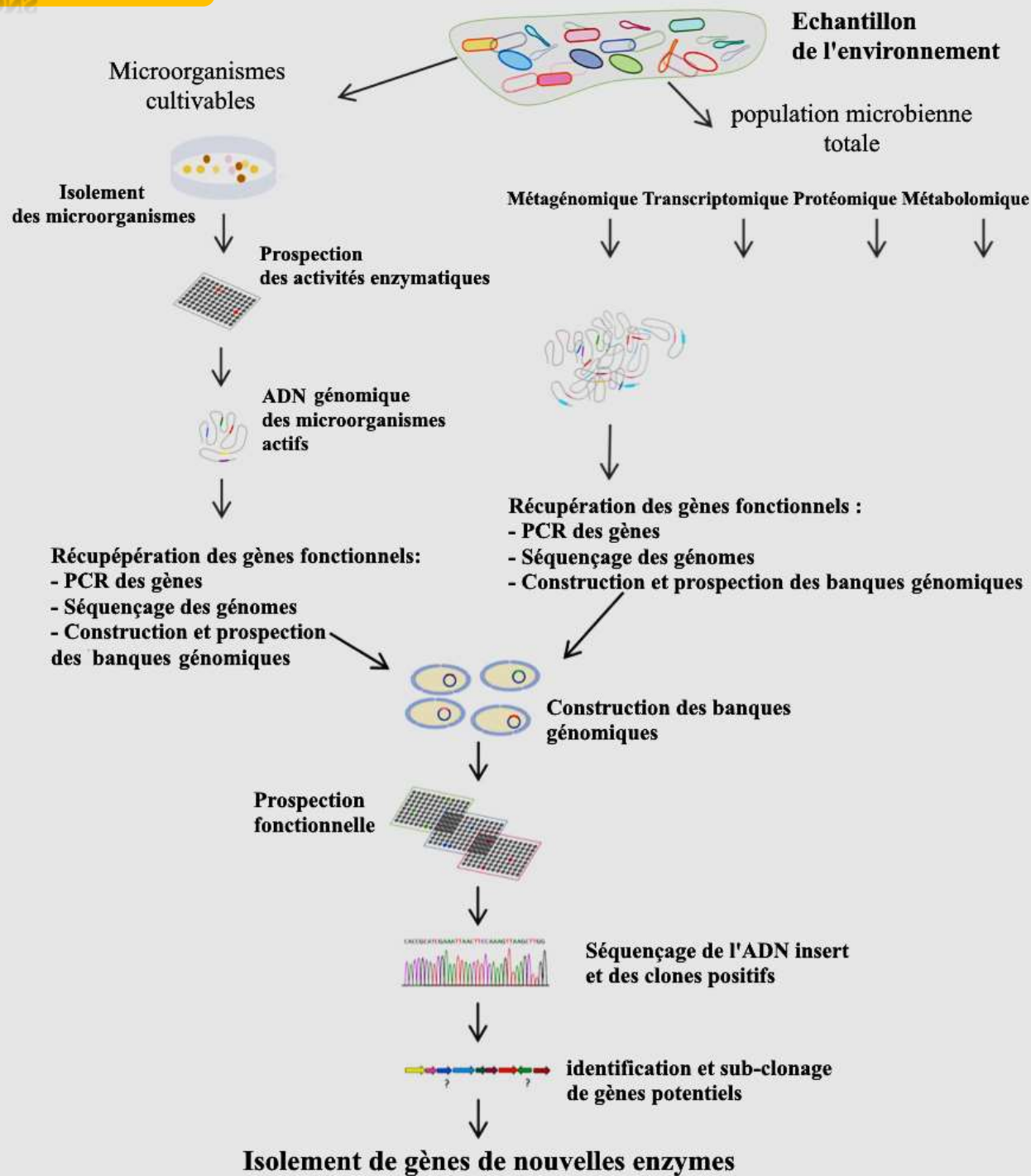
ADN  $\xrightleftharpoons{\text{Dédution}}$  Protéine



Prédiction de la structure 3D d'une catalase de *Aspergillus fumigatus*

## EXEMPLES D'APPLICATIONS

## PROSPECTION DE NOUVELLES ENZYMES INDUSTRIELLES



## EXEMPLES D'APPLICATIONS

DESIGN *IN SILICO* DE PROTÉINES

## Design rationnel

Enzyme/point mutation design



Gene construction

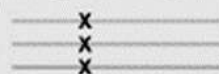


Transformation and expression of designed enzyme



## Design semi-rationnel

Construction of targeted gene library



Cloning of targeted library

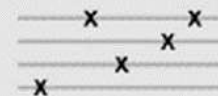


Transformation and library expression



## Evolution dirigée

Error-prone PCR for introduction of genetic diversity



Library cloning



Transformation and library expression



Screening/Selection



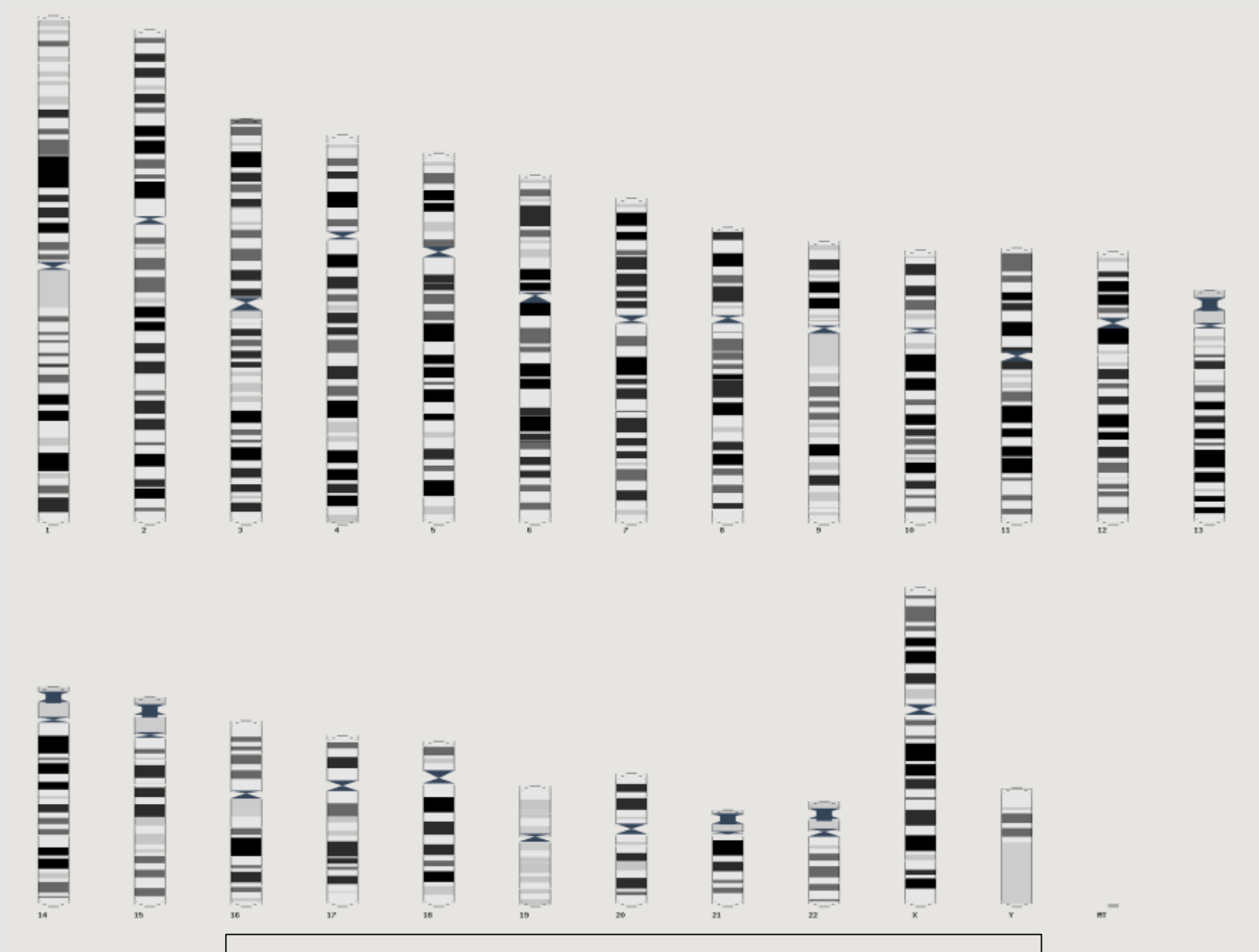
Evolved enzyme



Repeated cycles of evolution

## EXEMPLES D'APPLICATIONS

## ANNOTATION DES GÉNOMES



Vue schématique du caryotype humain



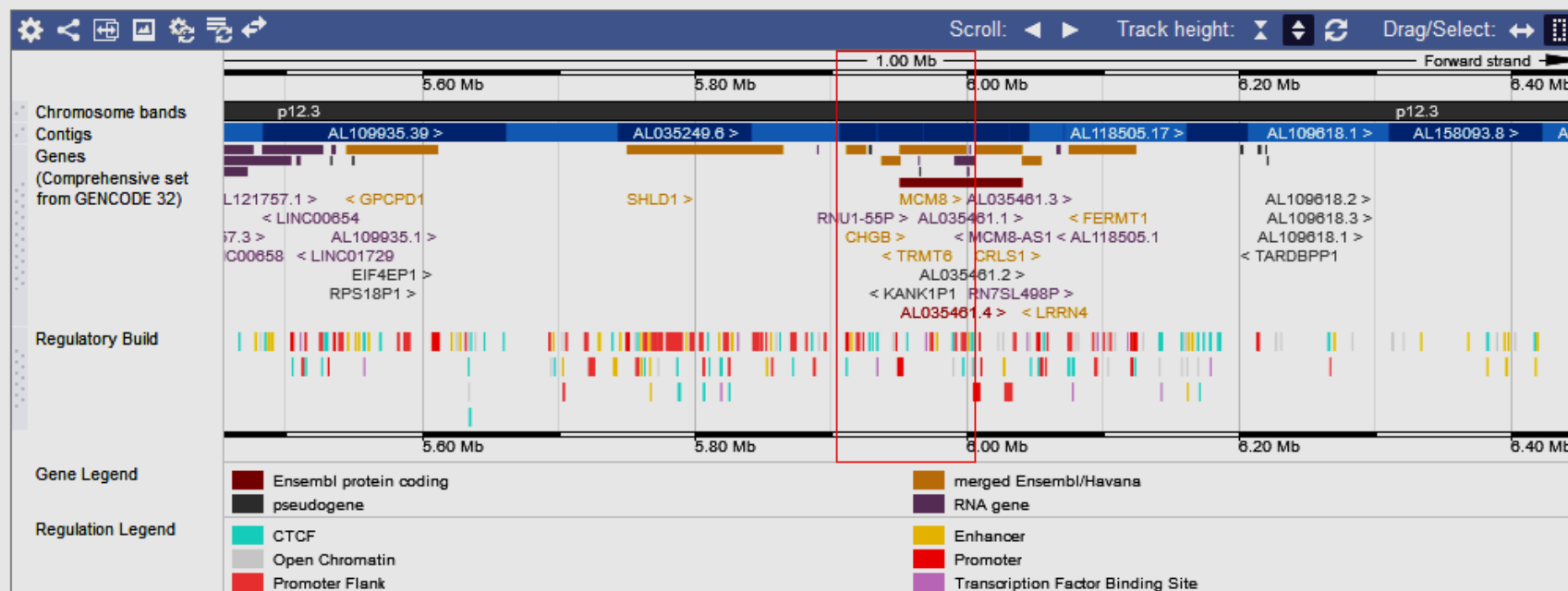
## EXEMPLES D'APPLICATIONS

## ANNOTATION DES GÉNOMES

## Chromosome 20: 5,904,970-6,004,970



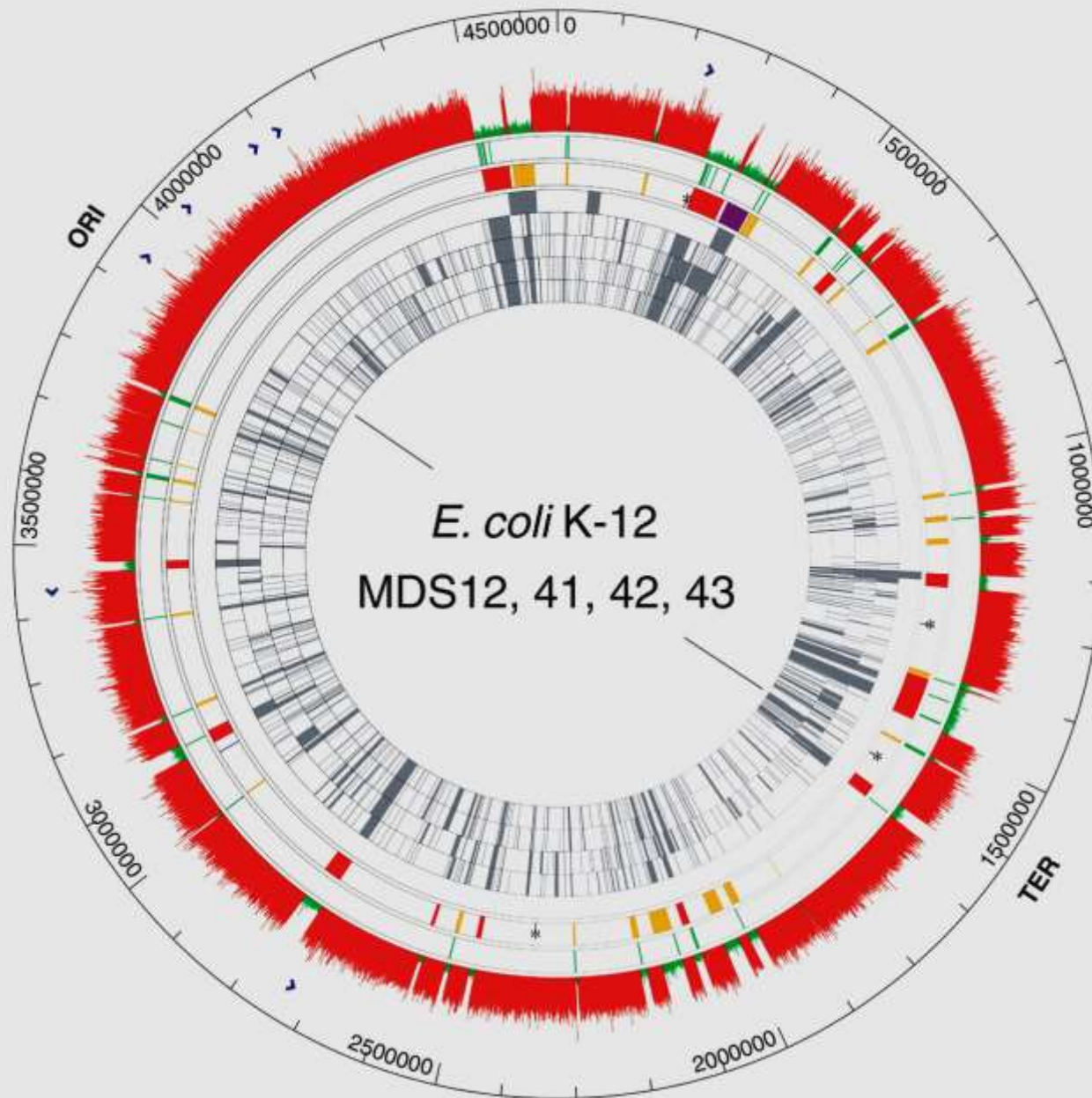
## Region in detail ?



Vue schématique du chromosome humain numéro 22

## EXEMPLES D'APPLICATIONS

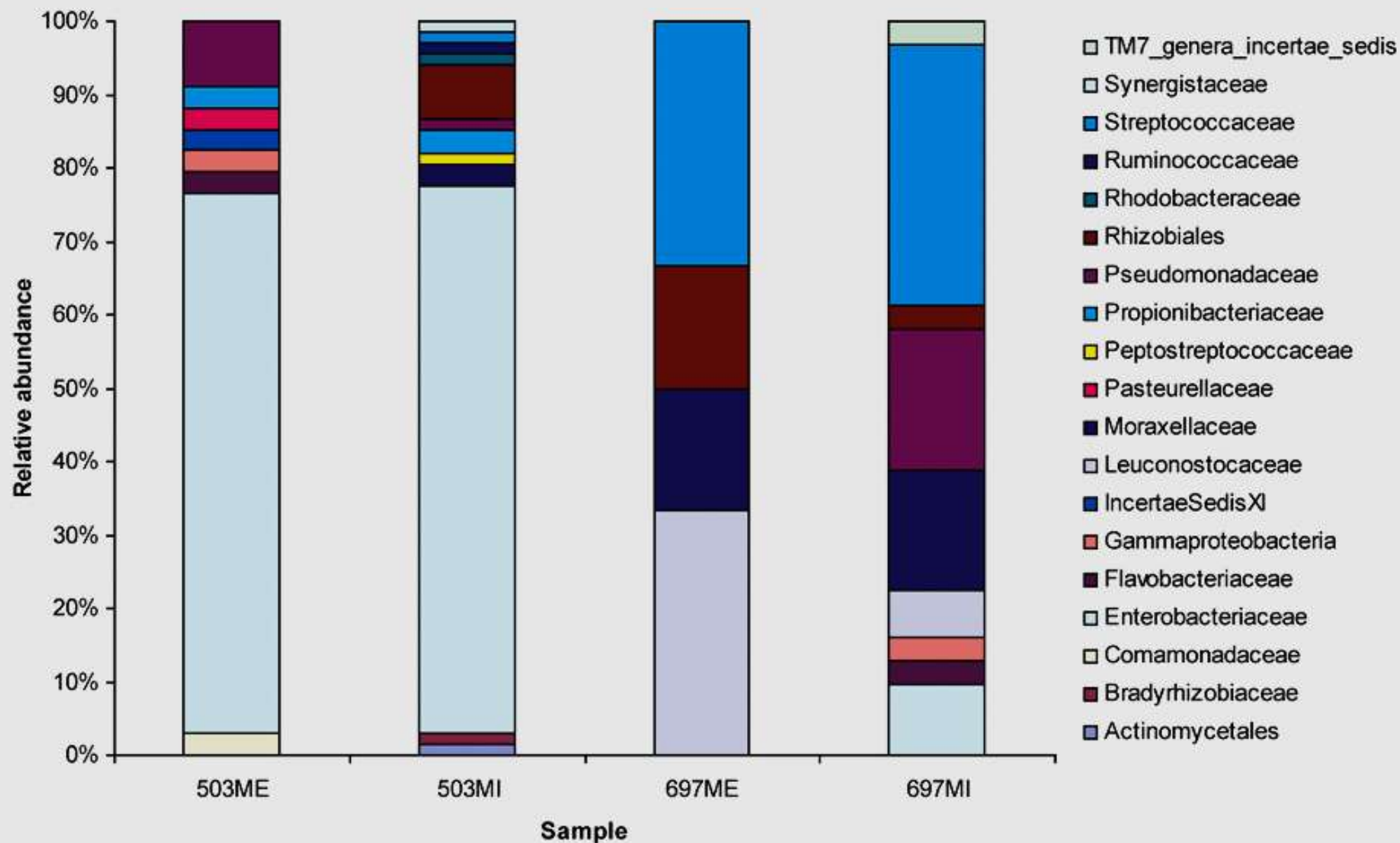
## ANNOTATION DES GÉNOMES



Vue schématique du génome d' *E. coli* K-12

## EXEMPLES D'APPLICATIONS

## ANALYSE DES SÉQUENCES NGS)

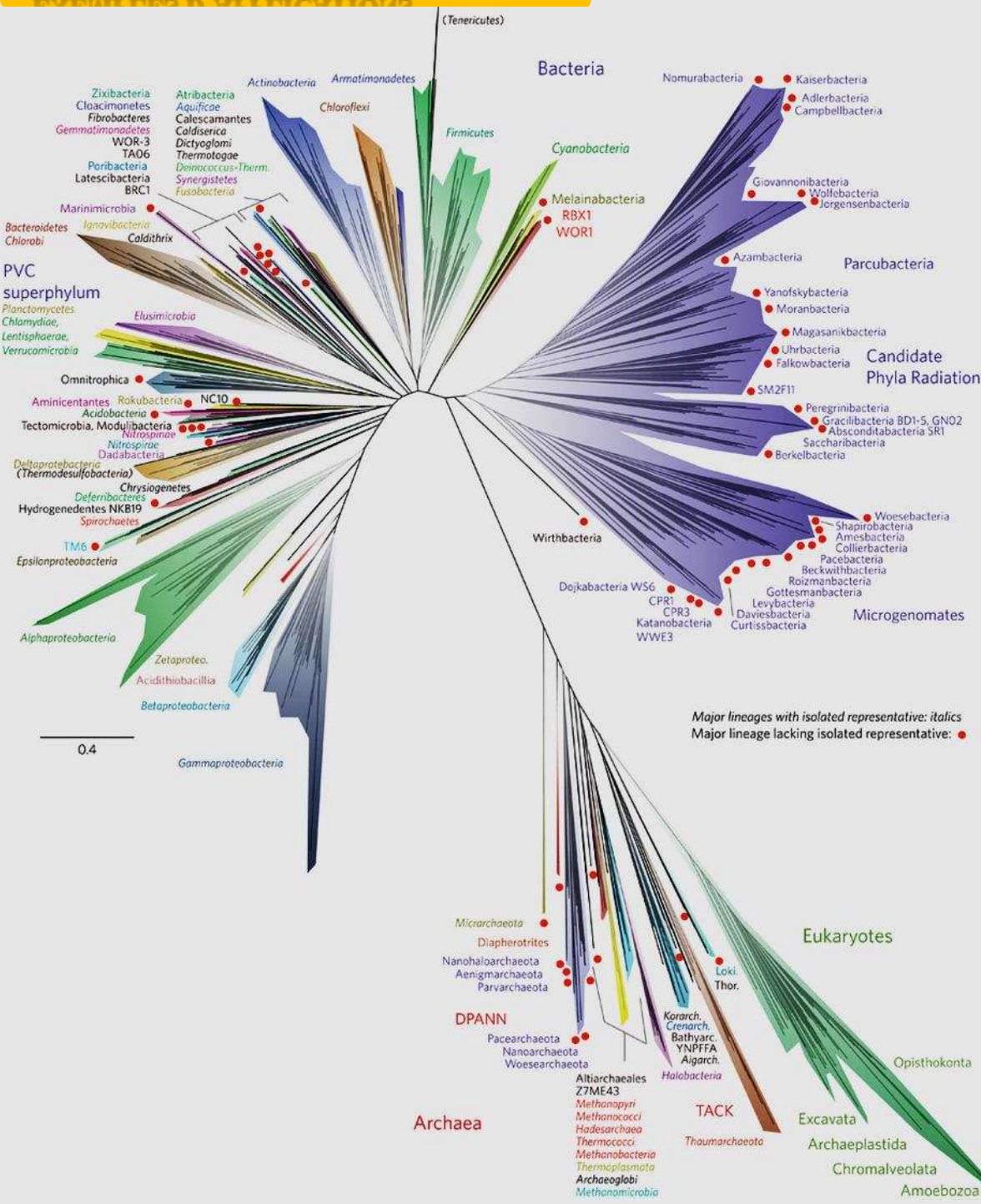


Abondance de la flore microbienne intestinale  
dans un échantillon de matières fécales infantiles



## EXEMPLES D'APPLICATIONS

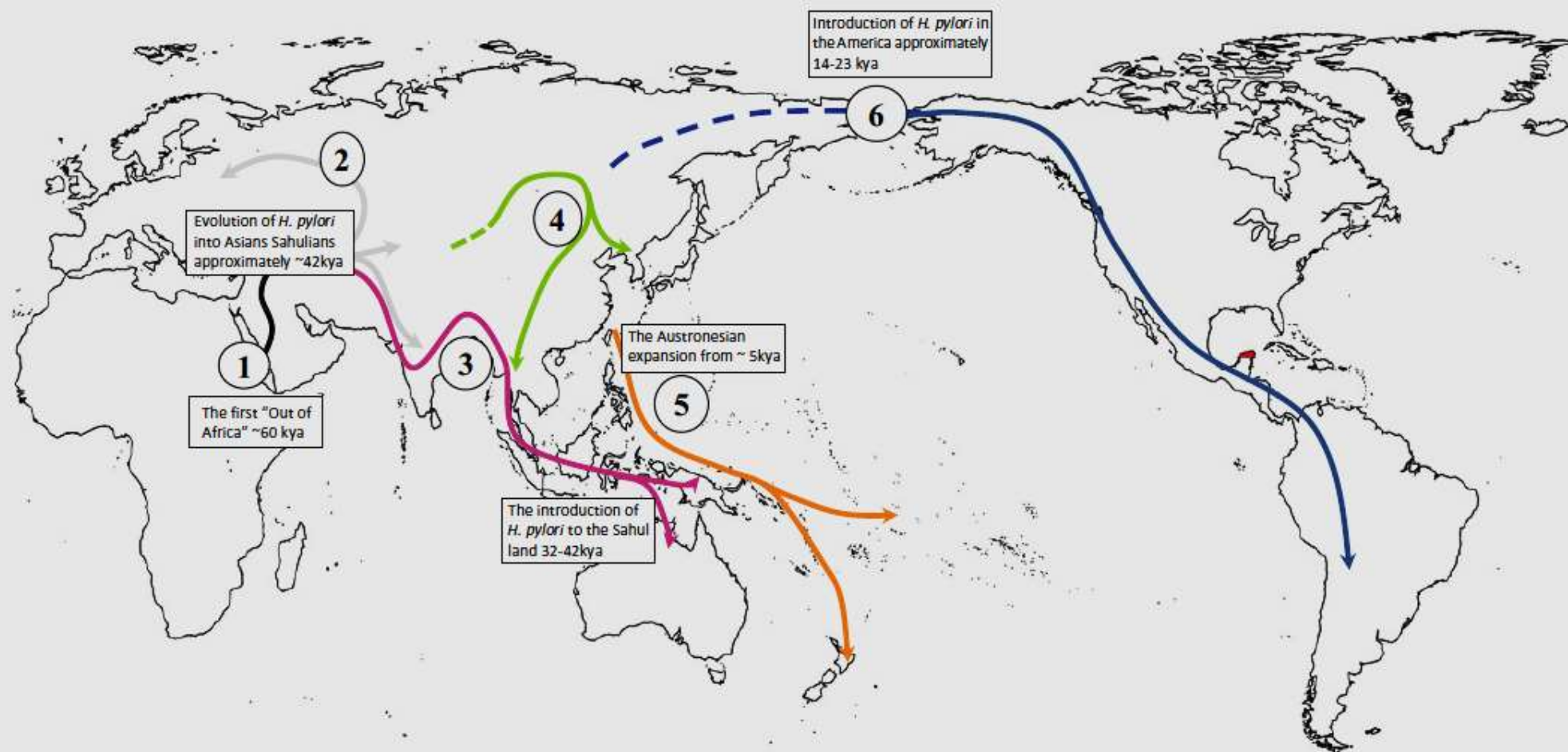
## « L'ARBRE DE LA VIE »



L'arbre phylogénétique universel (2016). L'arbre se compose de trois domaines d'organismes: les *Bacteria*, les *Archaea*, qui sont des cellules procaryotes, et des *Eukarya*. Ainsi que d'autres organismes non-cultivés ou en cours de classification.

## EXEMPLES D'APPLICATIONS

## PHYLOGÉOGRAPHIE

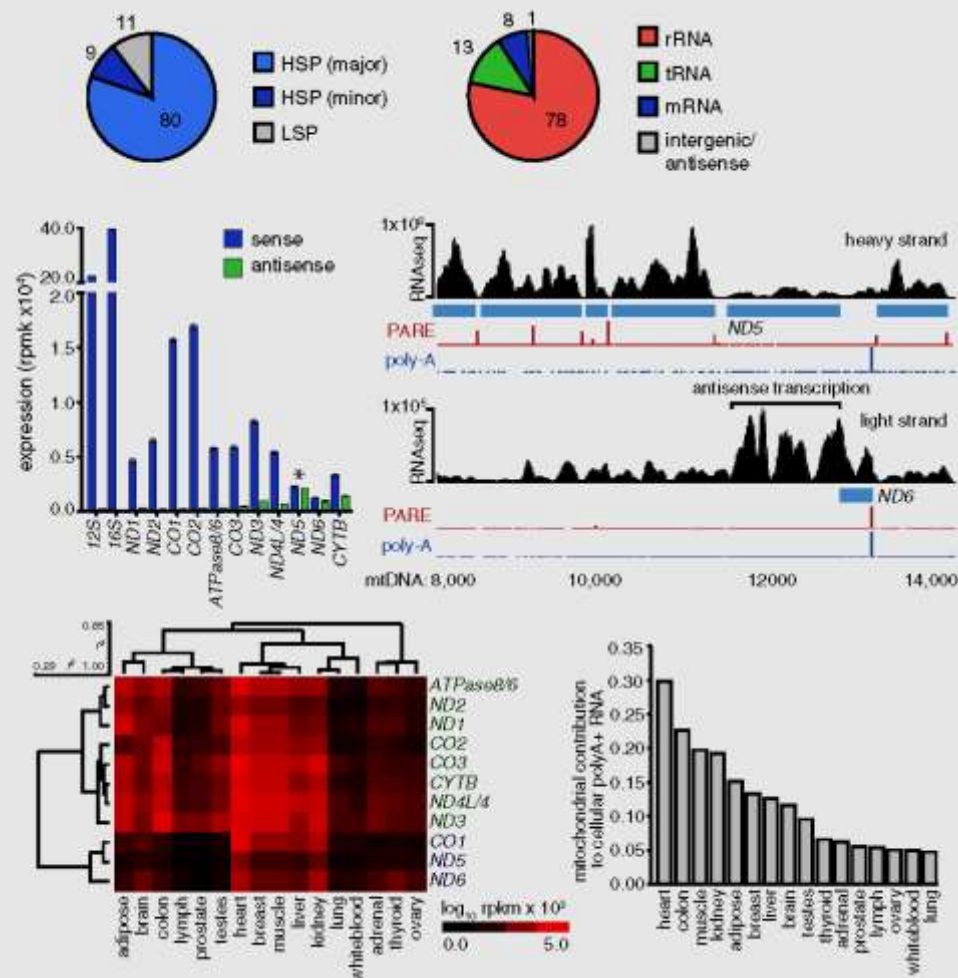
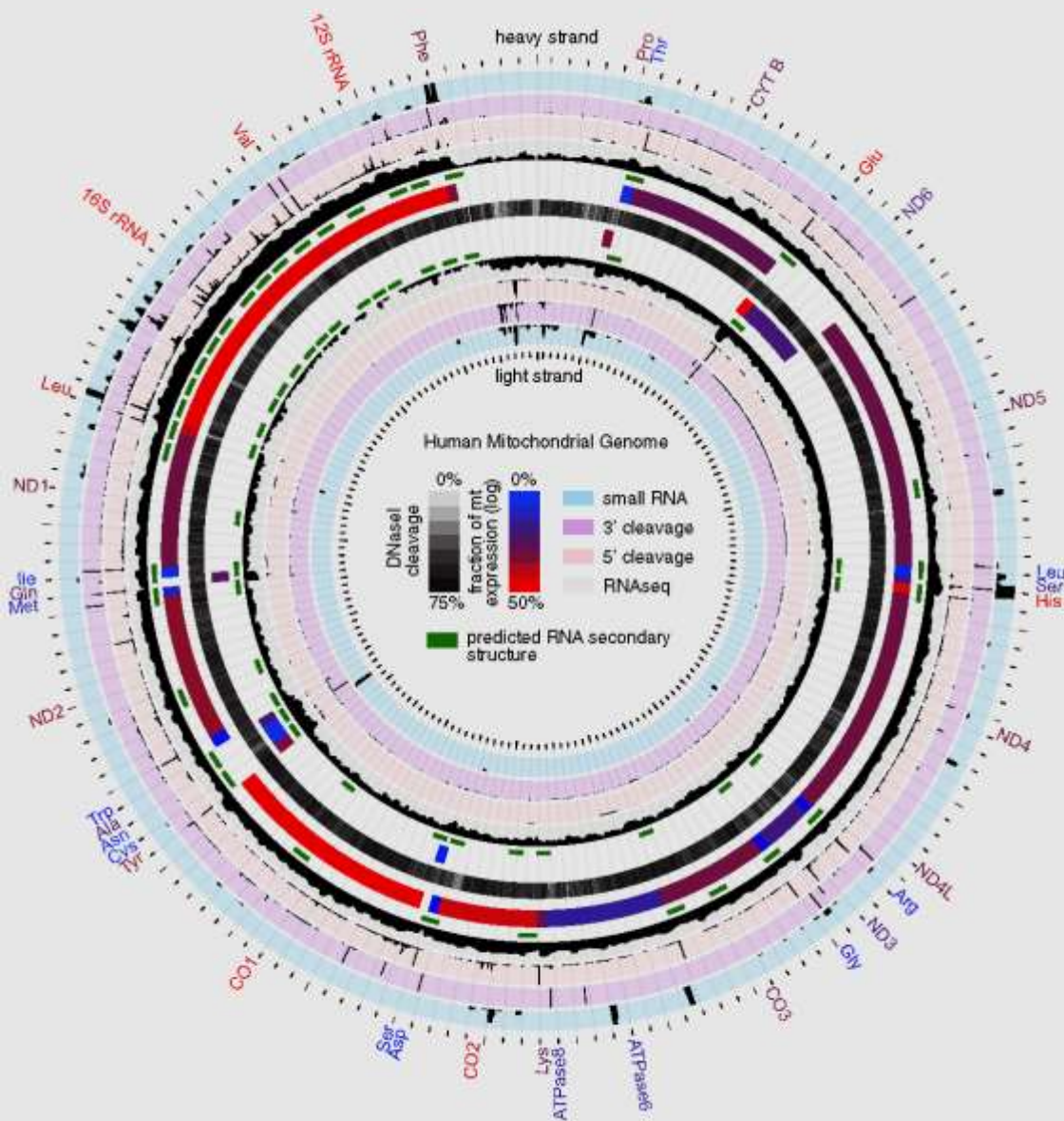


Vagues de migration humaines estimées à partir de l'analyse de marqueurs moléculaires



## EXEMPLES D'APPLICATIONS

## TRANSCRIPTOMIQUE



Carte du transcriptome du génome mitochondrial humain



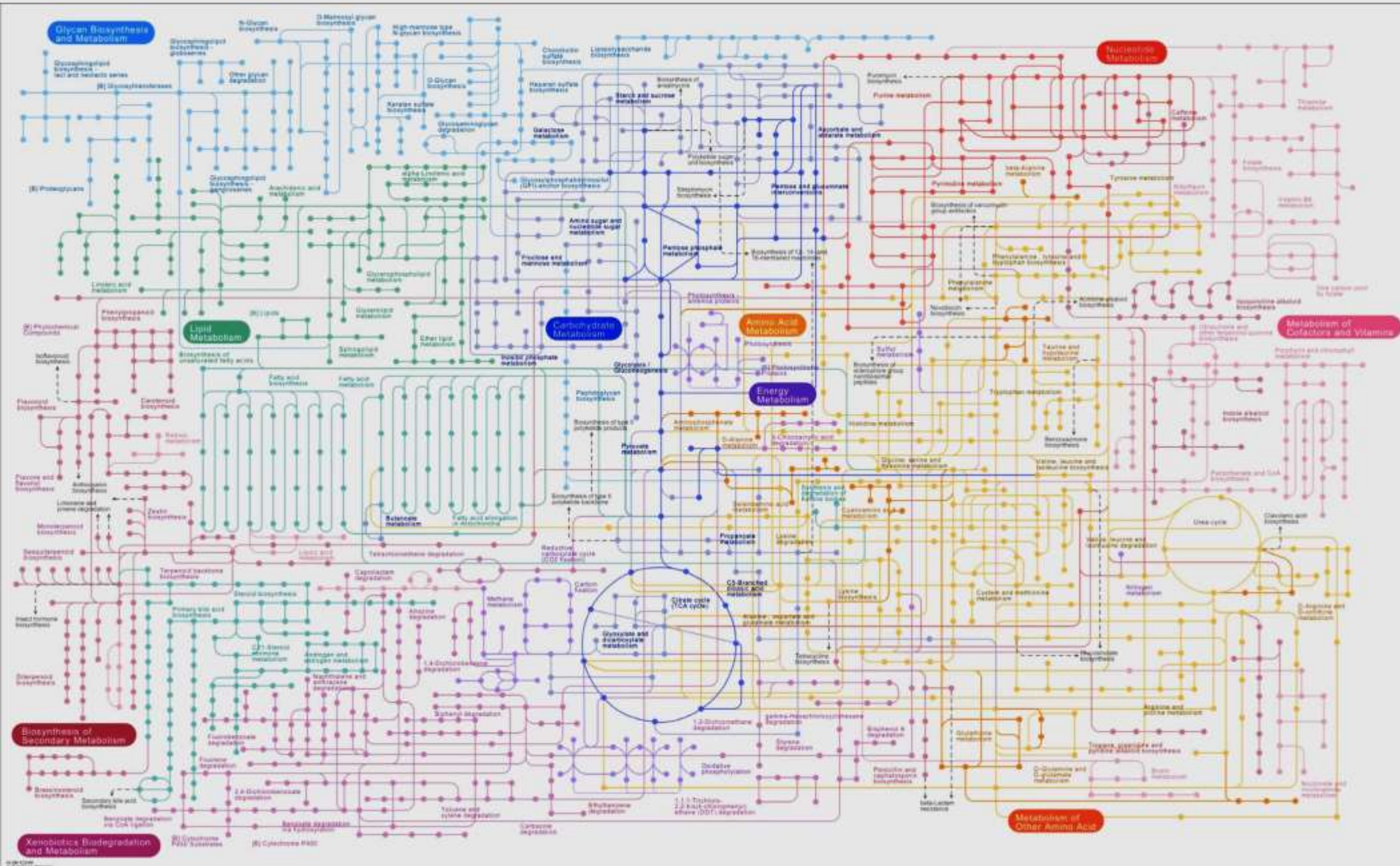


Schéma général du réseau métabolique humain

Exemples de secteurs d'activités :

- biotechnologies diverses : animaux, agriculture, environnement, industrie, etc.
- médecine / recherche de médicaments / pharmacologie ;
- thérapie génique ;
- recherche académique ;
- médecine légale.



# CHAPITRE I

# LES BASES DE DONNÉES BIOLOGIQUES

**SIB** Swiss Institute of Bioinformatics

**ExPASy Proteomics Server**

Databases Tools Services Mirrors About Contact

You are here: ExPASy CH > Databases > Around UniProtKB

**swissprot** **UniProt** **Swiss-Prot** **Protein knowledgebase** **TrEMBL** **Computer-annotated supplement to Swiss-Prot**

The UniProt Knowledgebase consists of:

- **UniProtKB/Swiss-Prot**, a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domain structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases. [More details / References / Linking to UniProtKB/Swiss-Prot / User manual / Recent changes / Disclaimer](#)
- **UniProtKB/TrEMBL**, a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups **at SIB** and **at EBI**

UniProt Knowledgebase Release 2010\_05 consists of:  
 UniProtKB/Swiss-Prot Release 2010\_05 of 20-Apr-10: 516603 entries ([More statistics](#))  
 UniProtKB/TrEMBL Release 2010\_05 of 20-Apr-10: 10706472 entries ([More statistics](#))

Access to the UniProt Knowledgebase

• UniProtKB with sites

**DDBJ** DNA Data Bank of Japan

Accession: DNA, Protein, AIDs, Taxonomy, Site Search

Accession numbers

☒ DDJ ☐ UniProt ☐ PDB ☐ DAD ☐ PRF ☐ Patent [History](#)

**HOME** Submission How to Use Search/Analysis FTP/Web API Report/Statistics Contact Us [RSS](#) [Japanese](#)

• About DDJ  
 • How to Use  
 • Q and A

**Sequence Submission**

• [SAKURA](#)  
 • [Mass Submission](#)  
 • [Data Updates](#)  
 • [DDJ Read Archive](#)  
 • [DDJ Trace Archive](#)

**Search**

• [Identity](#)  
 • [ARSA](#)  
 • [TVSearch](#)  
 • [BLAST](#)  
 • [PS-BLAST](#)  
 • [FASTA](#)  
 • [SSEARCH](#)

**Phylogenetics**

• [ClustalW](#)

**DDJ - DNA Data Bank of Japan**

DDJ (DNA Data Bank of Japan) is one of the three databanks that constitute DDJ/EMBL/GenBank International Nucleotide Sequence Database, which was established through cooperative work with EBI in Europe and NCBI in the USA.

**Hot Topics** [More](#)

- Apr. 15, 2010 [The Chinese Academy of Science professors visited DDJ](#)
- Apr. 12, 2010 [Release of the raw and assembled sequence data set from ngs](#)
- Apr. 12, 2010 [DAD \(DDJ amino acid database\) Rel. 51.0 Released](#)

**Maintenance** [More](#)

- Apr. 21, 2010 [Suspension of some DDJ activities in Japanese holidays \(4/29-5/1-5\)](#)
- Mar. 16, 2010 [\(Apr. 23\) ARSA database search \(DDJ, DAD\) temporary unavailable](#)
- Feb. 03, 2010 [\(Important\) Termination of a part of DDJ services](#)

**Sequence Data Submission**

- [Submit my sequences](#)  
Orientation for the data submission
- [Update my entries](#)

**FTP/Web API**

- [FTP \(ftp.ddbj.nig.ac.jp\)](#)  
Download data files
- [Web API](#)

**EMBL-EBI** [All Databases](#)  [Go](#) [Reset](#) [Advanced Search](#) [What's new](#)

Databases Tools EBI Groups Training Industry About Us Help

• **EMBL Bank Home**

- Access
- Documentation
- News
- Submission
- Publications
- People
- Contact

**EMBL Fetch**

Fetch an EMBL record by ID

[Go](#)

**News**

9th January 2010: NSDC and Genome Reference Consortium discussed in Bioinform... [more](#)

**Collaborations**

- [EMBL - International Nucleotide Sequence Database Collaboration](#)
- [EMBL - The Nucleotide Sequence Database is produced in collaboration with](#)

**EMBL Nucleotide Sequence Database**

The EMBL Nucleotide Sequence Database (also known as EMBL-Bank) constitutes Europe's primary nucleotide sequence resource. Main sources for DNA and RNA sequences are [direct submissions](#) from individual researchers, genome sequencing projects and patent applications.

The database is produced in an international [collaboration](#) with GenBank (USA) and the DNA Database of Japan (DDJ). Each of the three groups collects a portion of the total sequence data reported worldwide, and all new and updated database entries are exchanged between the groups on a daily basis. The [current database release](#) (Release 103, March 2010), with according [Release notes](#) and [user manual](#) are available from the EBI servers. A sample database entry is shown [here](#).

A publication in [Nucleic Acids Research](#) 2009 37: D18-D25 provides further information and details.

The EMBL nucleotide sequence database forms part of the [European Nucleotide Archive](#), an EBI project led by [Guy Cauchemez](#) as part of the [The Protein and Nucleotide Database Group \(PANDA\)](#) under [Ewan Birney](#).

Link	Explanation
<a href="#">Access</a>	Database queries, Contextual genomes, webserver, FTP archives (EMBL release, alignments etc), <a href="#">EMBL sequence version archive (SVA)</a> , <a href="#">Browse by geography</a>
<a href="#">Submission</a>	Primary sequence submissions, third party annotation, updates
<a href="#">Documentation</a>	Release notes, user manual, information for Submitters, FAQ, Release information, <a href="#">Formatting Changes</a> , <a href="#">EMBL database statistics</a> , <a href="#">Feature lists</a> , <a href="#">EMBL documentation</a> , <a href="#">Sequence entry</a> , <a href="#">Submission</a> , <a href="#">Submitter's Code</a> , <a href="#">Examples of annotations</a> , <a href="#">EMBL Features &amp; Qualifiers</a> , <a href="#">ORF line standards</a> , <a href="#">Database Policies</a>
<a href="#">Publications</a>	Group publications
<a href="#">People</a>	Group members
<a href="#">Contact</a>	How to contact the EMBL Nucleotide Sequence Database
<a href="#">News</a>	List of recent changes on this site

## NOTION DE BASE DE DONNÉES

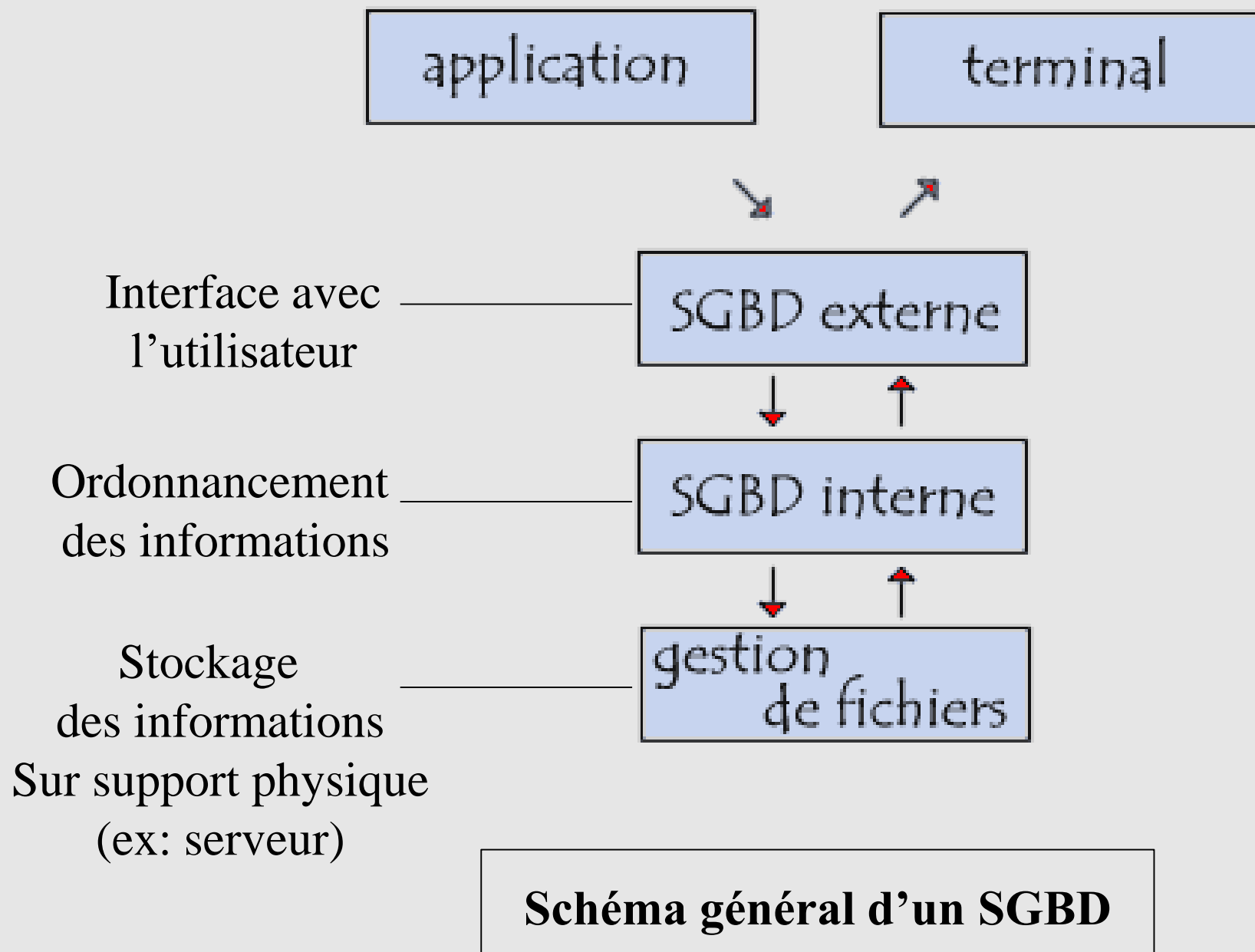
**Base de données (*database* ou **BD**)** : Collection de données interreliées, stockées ensemble pour servir une ou plusieurs applications, en parallèle, de façon optimale, accessibles en ligne et à distance ou pas.

Le terme **banque de données (*atabank*)** peut décrire un ensemble de fichiers texte sans relation entre eux ou de données hétérogènes (on parle de fichier « plat »). L'usage du terme est obsolète.

**Système de Gestion de Bases de Données (SGBD)** : Ensemble des programmes assurant la structuration, le stockage, la maintenance, la mise à jour et la recherche des données d'une base et l'interrogation de la base.



## NOTION DE BASE DE DONNÉES



## EXEMPLES DE BASES ET DE DONNÉES



Est un logiciel de gestion de **bases de données** utilisant un langage informatique d'exploitation de bases de données rationnelles appelé SQL (*Structured Query language*):

## EXEMPLES DE BASES ET DE DONNÉES

File Home Create External Data Database Tools Design Arrange Format

Views Themes Controls Header / Footer Tools

All Access Objects Search...

Queries

Forms

Reports

ContactsExtended

ContactCard

ContactDetails

ContactDS

ContactList

Contacts

DefaultImage

GettingStarted

Main

NameCard

NameList

ReportCenter

ReportCenterDe...

VideoPlayer

Comments

ContactAddress...

ContactDetails

ContactList

ContactPhoneB...

Label

Main Contact Details

Contact Details

Save & New Save & Close

First Name Anna

Last Name Bedecs

Job Title Owner

Company Company A

E-mail

Web Page

Business Phone (123)555-0100

Fax (123)555-0101

Home Phone

Mobile Phone

Address 123 1st Street

City Seattle

State/Province WA

ZIP/Postal Code 99999

Country/Region USA

Record: 1 of 29 No Filter Search

Microsoft Office



## EXEMPLES DE BASES ET DE DONNÉES

Issue Tracking Application - Microsoft Access

Data Insert Page Layout External Data Advanced Tools

Table Table Templates SharePoint Lists Forms Simple Form Split Form Multiple Items More Forms Reports Simple Report Labels Blank Report Query Wizard Advanced New Field Lookup Column Add Existing Fields

All Access Objects

Tables

- Contacts
- Customers
- Issues
- MSysCompactError
- Settings
- Table1

Queries

- Contacts Extended
- Issues Extended
- Open Issues

Forms

- Contacts
- Issue PopUp
- Issues by Status Chart
- Issues Datasheet
- Open Issues
- Open Issues Split View
- Template Setup

Reports

- Closed Issues
- Copy Of Issue Details
- Copy Of Open Issues
- Open Issues by Assignment

Issues by Status Chart

Customer Name	Product Name	Customer Contact	Contact E-mail	Account Manager	AM Manager
Datum	MacroPod, Softy	Alberts, Amy E.	Amy@datum.com	Polly, Laura Steele	Laura@Fabrikam.com
Adventure Works	Gillow	Hanson, Mark	Mark@AdventureW.com	Sankovic, Barbara	Barbara@Fabrikam.com
Alpine Ski House	MacroPod	Barber, David	David@alpine.com	Kharatishvili, Tengiz	Tengiz@Fabrikam.com
Baldwin Museum of Science	Gillow, Softy	Holt, Holly	Holly@baldwinmofs.com	Wycoff, Pieter	Pieter@Fabrikam.com
Blue Yonder Airlines	MacroPod	Price, Julian	Julian@blueyonder.com	Wycoff, Pieter	Pieter@Fabrikam.com
City Power & Light	Gillow	Brunner, Daniel	Daniel@cpl.com	Polly, Laura Steele	Laura@Fabrikam.com
Coho Vineyard	MacroPod	Karnik, Sachin	Sachin@cohov.com	Trukawka, Adam	adam@Fabrikam.com
Coho Winery	MicroPod	Saddow, Peter	Peter@cohow.com	Speckmann, Melanie	Melanie@Fabrikam.com
Contoso, Ltd	Jetso	Javier, Francisco	Francisco@contoso.com	Rettig, Bjorn	Bjorn@Fabrikam.com
Contoso Pharmaceuticals	Gillow	Dell, Eduard	Eduard@contosoph.com	Stehmann, Victor	Victor@Fabrikam.com
Consolidated Messenger	MicroPod	Lidman, Anna	Anna@consolidatedmsgr.com	Rettig, Bjorn	Bjorn@Fabrikam.com
Fourth Coffee	Gillow, Jetso	Gruber, Eric	Eric@4thcoffee.com	Wycoff, Pieter	Pieter@Fabrikam.com
Graphic Design Institute	Jetso	Oliveira, Manuel	Manuel@gdi.com	Rovik, Dag	dag@Fabrikam.com
Humongous Insurance	MacroPod	Aalling, Lene	Lene@humongousins.com	Paiha, Dominik	Dominik@Fabrikam.com
Litware, Inc.	MicroPod, Jetso	Haddock, Rich	Rich@litware.com	Rettig, Bjorn	Bjorn@Fabrikam.com
Lucerne Publishing	MicroPod, Jetso	Hrebicek, Ondrej	Ondrej@lucernep.com	Wycoff, Pieter	Pieter@Fabrikam.com
Margie's Travel	Softy	Bermejo, Antonio	Antonio@Margiestravel.com	Speckmann, Melanie	Melanie@Fabrikam.com
Northwind Traders	MicroPod	Barreto, Paula	Paula@northwindt.com	Purcell, Sean	Sean@Fabrikam.com
Proseware, Inc.	Gillow	Bott, Jörg	Jorg@proseware.com	Rovik, Dag	dag@Fabrikam.com
School of Fine Art	MacroPod, Jetso	Juhl, Claus	Claus@soff.com	Wycoff, Pieter	Pieter@Fabrikam.com
Southridge Video	Gillow	Axen, Thomas	Thomas@southridge.com	Polly, Laura Steele	Laura@Fabrikam.com
Tailspin Toys	MacroPod	Hill, Christopher	Chris@Tailspintoys.com	Trukawka, Adam	adam@Fabrikam.com
Trey Research	Gillow	Czernek, Pawel	Pawel@treyresearch.com	Speckmann, Melanie	Melanie@Fabrikam.com
The Phone Company	MacroPod	Lee, Oliver	Oliver@thephonecomp.com	Stehmann, Victor	Victor@Fabrikam.com
Wide World Importers	Softy	Fakhouri, Fadi	Fadi@Wwimporters.com	Trukawka, Adam	adam@Fabrikam.com
Wingtip Toys	Softy	McKay, Yvonne	Yvonne@wingtips.com	Sankovic, Barbara	Barbara@Fabrikam.com
Woodgrove Bank	MicroPod	Campbell, David	David@woodgrove.com	Sankovic, Barbara	Barbara@Fabrikam.com

Record: 1 of 27 No Filter Search

Datasheet View

## EXEMPLES DE BASES ET DE DONNÉES

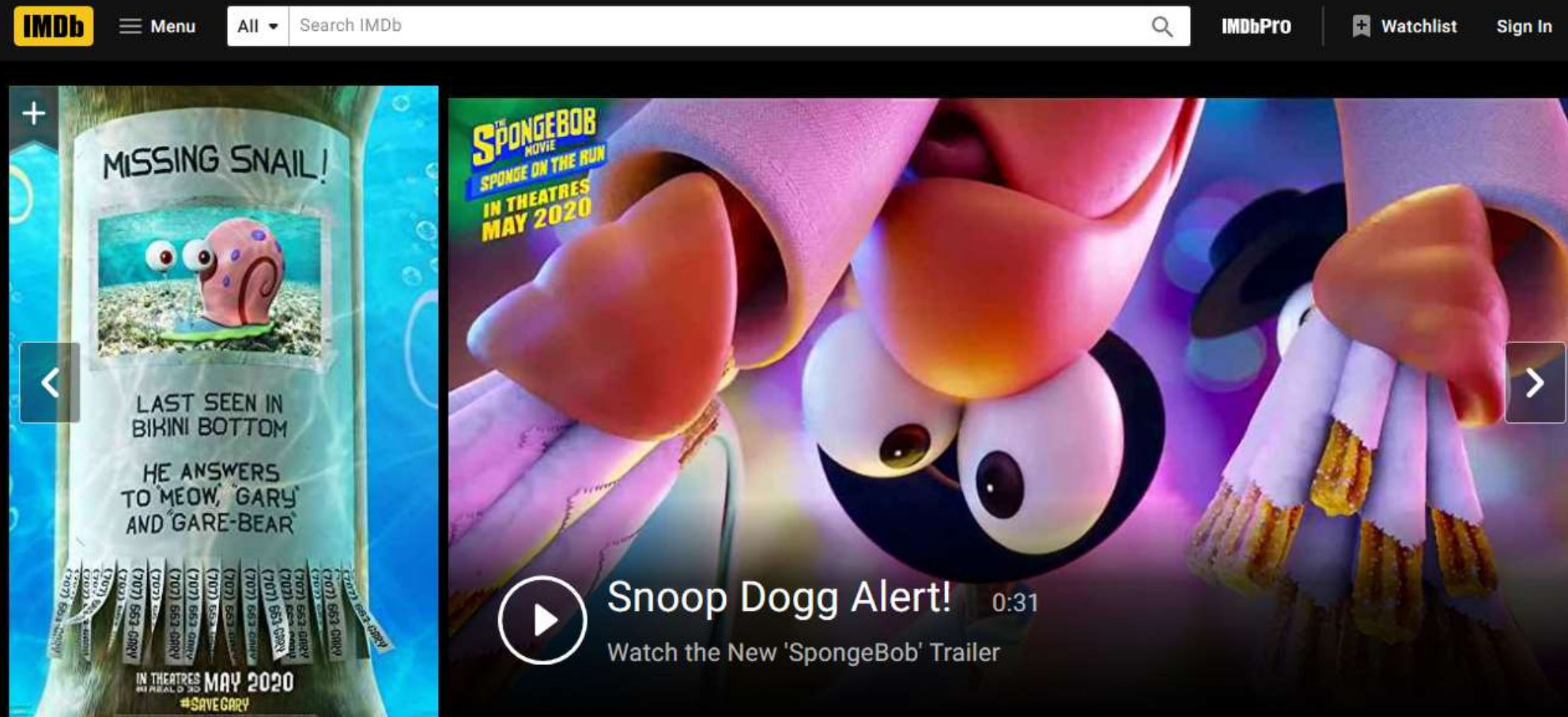


Effectuer une recherche sur Google ou saisir une URL

Google est la plus grande **base** de pages **indexées** sur internet. C'est un puissant instrument de recherche automatisé (moteur de recherche) qui permet de recenser et de classer dans des bases de données accessibles gratuitement une grande partie du contenu web.



## EXEMPLES DE BASES DE DONNÉES



<https://www.imdb.com/> est une base de données filmographiques.

## EXEMPLES DE BASES DE DONNÉES



## Cherchez des ouvrages dans les bibliothèques près de chez vous

2 milliard d'ouvrages disponibles ici via une bibliothèque

Tout Livres DVD CD Articles

Tout chercher

[Recherche avancée](#) | [Trouver une bibliothèque](#)

WorldCat vous connecte aux collections et aux services de plus de 10 000 bibliothèques à travers le monde [Détails >>](#)

**Identifiez-vous** pour écrire des critiques et créer des listes et des bibliographies sur les ouvrages de bibliothèques

Nom d'utilisateur:

Mot de passe:

☐ Se souvenir de moi sur cet ordinateur.

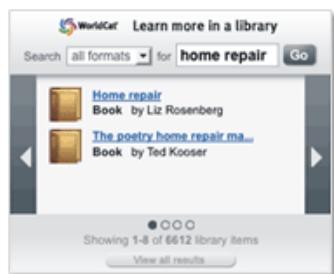
[Mot de passe oublié?](#) | [Créez un compte gratuit](#)

**Avis de confidentialité** : OCLC traite vos données personnelles dans les États-Unis d'Amérique pour fournir ce service. Pour plus d'informations sur l'utilisation, la divulgation et la protection de vos données personnelles, veuillez consulter la [Politique de confidentialité](#) d'OCLC.

[Abonnez-vous aux nouvelles par courriel](#)

## Utilisez les outils WorldCat

Accédez rapidement aux ressources de bibliothèques à partir de votre navigateur ou de votre page Web personnalisée. Ajoutez les bibliothèques du monde votre téléphone portable et vos applications Facebook.



[Widgets et services Web](#)

[API de recherche WorldCat](#)

[Applications et extensions pour navigateurs](#)

[Créez un lien vers WorldCat](#)

[Ajoutez un logo WorldCat](#)

## Créez une bibliographie

Obtenez des références bibliographiques dans les cinq styles les plus utilisés et exportez ces références bibliographiques en divers formats, dont EndNote, Reference Manager et RefWorks.



Tutoriel vidéo

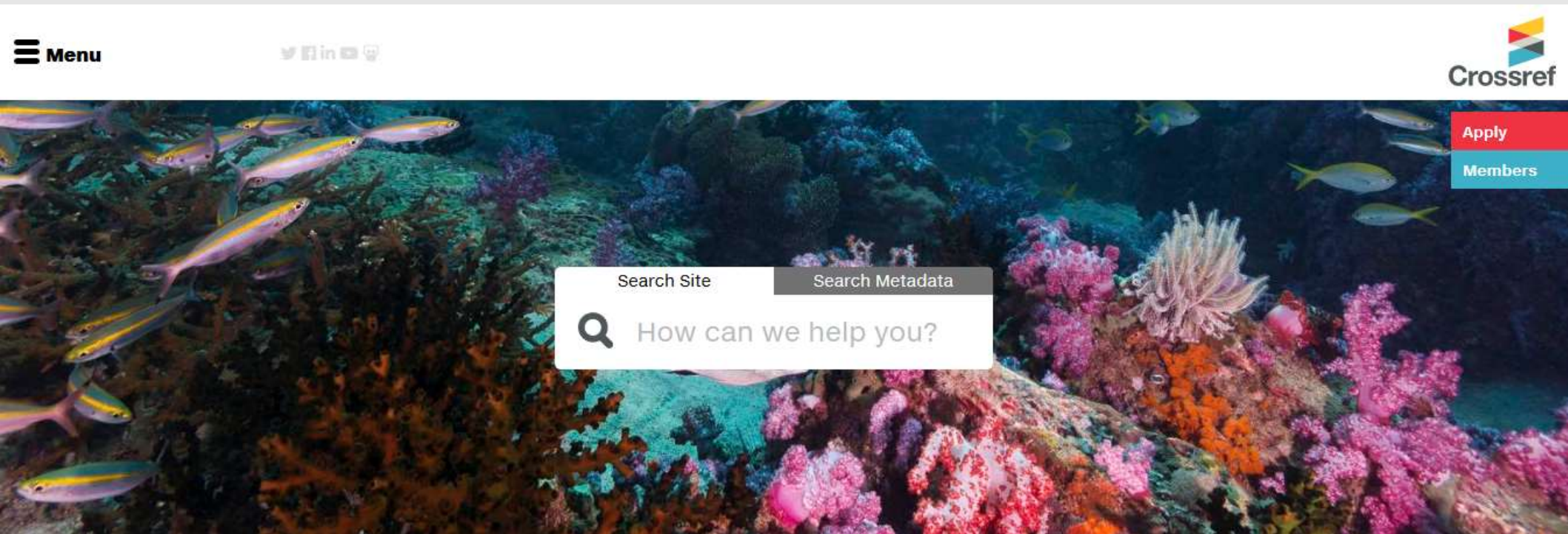
[Comment utiliser les références bibliographiques de WorldCat.org \(en anglais\) \(YouTube\)](#)

Durée: 1:47

<https://www.worldcat.org/> est une base de données bibliographiques.



## EXEMPLES DE BASES DE DONNÉES



<https://www.crossref.org/> est une base de **DOI** (identifiant numérique d'objet).

## CATÉGORIES DE BASES DE DONNÉES BIOLOGIQUES

**BD biologiques:** grandes bibliothèques de données de biologie et des sciences de la vie résultant des expériences entreprises dans les laboratoires du monde entier aussi bien que de l'analyse des simulations et des documents édités.

On distingue deux types de bases :

1. **Bases de données généralistes** : correspondent à une collecte de données la plus exhaustive et la plus large possible ;
2. **Bases de données spécialisées** : correspondent à des données plus homogènes établies autour d'une thématique particulière.

## CATÉGORIES DE BASES DE DONNÉES BIOLOGIQUES

# 1. Les bases de données généralistes

Ces bases contiennent des données hétérogènes

- Bases de séquences nucléiques (**ENA**, **GenBank**, **DDBJ**) ;
- Bases de séquences protéiques (**Uniprot**, **PIR**, **SwissProt**) ;
- Bases génomiques et de localisation (**Genome**) ;
- Bases de structures 3D de macromolécules (**PDB**) ;
- Bases de données bibliographiques (**Medline**)

**Avantage** : tout est consultable en une fois

**Inconvénients** : difficiles à maintenir, difficiles à interroger, problèmes de redondance



- Origine des données :
  - séquençage d'ADN.
- Les données stockées : séquences + annotations
  - gènes, fragments de gènes, séquences non-exprimées, etc.
  - génomes entiers ;
  - traduction en séquences protéiques des séquences ADN.

## BASES DE DONNÉES GÉNÉRALISTES

## BD DE SÉQUENCES NUCLÉIQUES

Depuis 1987, 3 principales bases sont **interconnectées** dans le cadre d'un consortium appelé :

**International Nucleotide Sequence Database Collaboration:**

<http://www.insdc.org/>



\*En Décembre 2019, les bases du consortium archivaient plus de 2 milliards 400 millions de séquences nucléiques.

**Genbank:** base de données américaine, diffusée par le  
NCBI (*National Center for Biotechnology Information*, Los Alamos, USA)  
<http://www.ncbi.nlm.nih.gov/genbank/>

NCBI Resources ▾ How To ▾ [Sign in to NCBI](#)

GenBank

GenBank ▾ Submit ▾ Genomes ▾ WGS ▾ Metagenomes ▾ TPA ▾ TSA ▾ INSDC ▾ Other ▾

## GenBank Overview

### What is GenBank?

GenBank<sup>®</sup> is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research, 2013 Jan;41\(D1\):D36-42](#)). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

### Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see [BLAST info](#) for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using [NCBI e-utilities](#).
- The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server: <ftp://ftp.ncbi.nlm.nih.gov/ncbi-asn1> and <ftp://ftp.ncbi.nlm.nih.gov/genbank>.

### GenBank Resources

[GenBank Home](#)

[Submission Types](#)

[Submission Tools](#)

[Search GenBank](#)

[Update GenBank Records](#)

## ENA

*(European Nucleotide Archive)*base européenne diffusée par **EMBL-EBI** (Cambridge, UK)*(European Molecular Biology Library-European Bioinformatics Institute)*<http://www.ebi.ac.uk/ena>

The screenshot shows the ENA website interface. At the top, there's a navigation bar with 'EMBL-EBI' and links for 'Services', 'Research', 'Training', and 'About us'. Below this is a teal header with the ENA logo and a search bar. The search bar has a 'Search' button and links for 'Advanced' and 'Sequence'. Below the header is a dark green navigation bar with links: 'Home', 'Search & Browse', 'Submit & Update', 'Software', 'About ENA', and 'Support'. A green banner below the navigation bar announces the new ENA Browser. The main content area has a heading 'European Nucleotide Archive' and a paragraph describing the archive. Below this is a 'Text Search' section with a search bar, examples, a 'search' button, and a link to 'Advanced search'. There is also a 'Sequence Search' section with a larger text area for pasting sequences, a 'Search' button, and a link to 'Advanced search'. On the right side, there's a 'Popular' section with a list of links and a 'Latest ENA news' section with a news item dated 17 Dec 2019.

EMBL-EBI

Services Research Training About us

ENA  
European Nucleotide Archive

Search

Examples: BN000065, histone

Advanced  
Sequence

Home Search & Browse Submit & Update Software About ENA Support

The new ENA Browser is now live, with improved features for searching & downloading data!  
Please visit <https://www.ebi.ac.uk/ena/browser/>.

## European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

### Text Search

Examples: BN000065, histone

search

[Advanced search](#)

### Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

[Advanced search](#)

### Popular

- Submit and update
- Sequence submissions
- Genome assembly submissions
- Submitting environmental sequences
- Citing ENA data
- Rest URLs for data retrieval
- Rest URLs to search ENA

### Latest ENA news

**17 Dec 2019:** [Retirement of periodic ENA assembled/annotated sequence release in March 2020](#)

The European Nucleotide Archive (ENA) captures,...

**DDBJ (*DNA Data Bank of Japan*):** Base de données japonaise, diffusée par le NIG (*National Institute of Genetics*, Japon)

<https://www.ddbj.nig.ac.jp>

The screenshot shows the DDBJ website interface. At the top, there is a navigation bar with the DDBJ logo and 'Services' menu on the left, and links for 'Login & Submit', 'Policies and Disclaimers', 'Contact', and 'Japanese' on the right. Below the navigation bar, the main header reads 'Bioinformation and DDBJ Center'. To the right of this header is a dropdown menu set to 'BI-DDBJ Web Sites' and a 'Google Custom Search' box with a magnifying glass icon. A yellow banner below the header contains the text: '(Feb 13 - 21) Suspension of BI-DDBJ services due to NIG Supercomputer maintenance'. The main content area features a grid of eight service categories, each with an icon and a title: 'Search & Analysis' (monitor with graph), 'Submissions' (cylinder with arrow), 'Downloads' (download arrow), 'SuperComputer' (server rack), 'Statistics' (bar chart), 'Activities' (pencil), 'Training' (graduation cap), and 'About Us' (three people silhouettes).

DDBJ Services

Login & Submit Policies and Disclaimers Contact Japanese

Bioinformation and DDBJ Center

BI-DDBJ Web Sites Google Custom Search

(Feb 13 - 21) Suspension of BI-DDBJ services due to NIG Supercomputer maintenance

Bioinformation and DDBJ Center provides sharing and analysis services for data from life science researches and advances science.

Search & Analysis Submissions Downloads SuperComputer

Statistics Activities Training About Us



- Origine des données
  - Traduction de séquences d'ADN
  - Séquençage de protéines
  - Protéines dont la structure 3D est connue
- Les données stockées : séquences + annotations
  - Protéines entières
  - Fragments de protéines

## BASES DE DONNÉES GÉNÉRALISTES

## BD DE SÉQUENCES PROTÉIQUES

**UniProt** (*Universal Protein Resource*)\*: Consortium regroupant les données de plusieurs bases de données protéiques (séquences protéiques annotées): SwissProt (base suisse)-TrEMBL (*Traduced EMBL*) et PIR (*Protein Information Resource*) ainsi que d'autres bases protéiques

<http://www.uniprot.org>

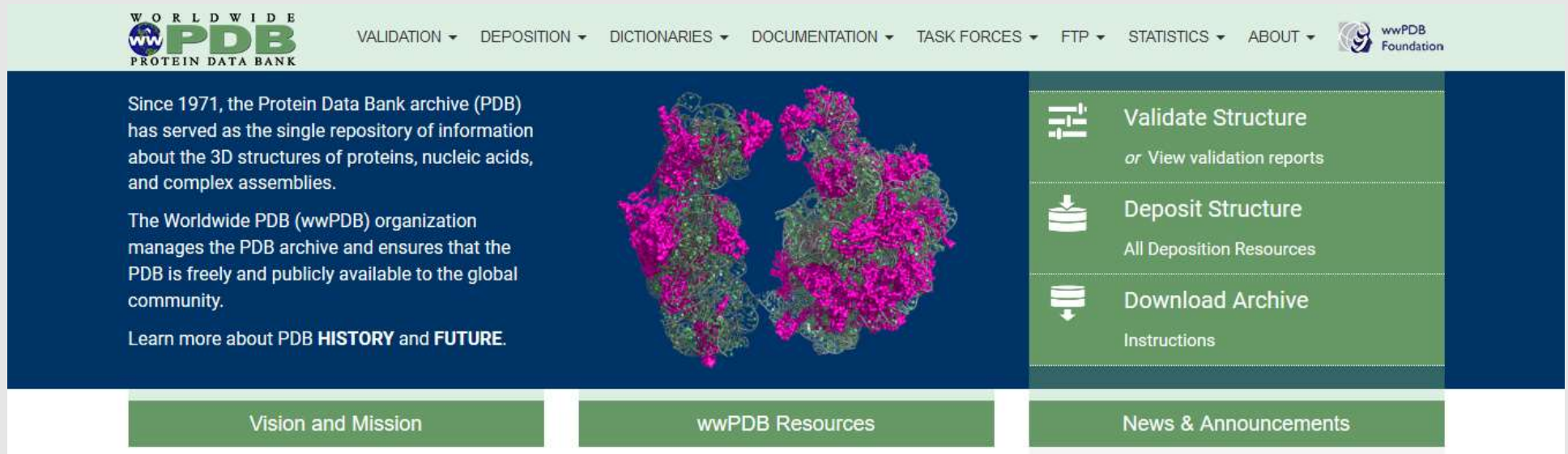
The screenshot shows the UniProt website interface. At the top, there's a navigation bar with the UniProt logo and a search bar containing 'UniProtKB'. Below the search bar are links for 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. A 'Help' and 'Contact' link is also present. The main content area starts with the mission statement: 'The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.' Below this, there are four colored boxes representing different databases: UniProtKB (561,568 records, manually annotated), UniRef (Sequence clusters), UniParc (Sequence archive), and Proteomes (Proteome sets). A 'Supporting data' section follows, listing literature citations, taxonomy, subcellular locations, cross-ref. databases, diseases, and keywords. On the right side, there's a 'News' section with links to 'Forthcoming changes', 'UniProt release 2019\_11', and 'UniProt release 2019\_10'. Social media icons for Blog, Twitter, Facebook, and RSS are also visible.

\*En Décembre 2019, Uniprot archivait plus de 180 millions séquences protéiques annotées automatiquement ou manuellement, non revues ou revues, respectivement.

## PDB (*Protein Data Bank*) \*:

Structure 3D de protéines, acides nucléiques et autres molécules

<http://www.wwpdb.org/>



The screenshot shows the PDB website homepage. At the top, there is a navigation bar with links: VALIDATION, DEPOSITION, DICTIONARIES, DOCUMENTATION, TASK FORCES, FTP, STATISTICS, and ABOUT. The PDB logo is on the left, and the wwPDB Foundation logo is on the right. The main content area has a dark blue background. On the left, there is text about the PDB's history and mission. In the center, there is a 3D molecular structure visualization. On the right, there are three green buttons: 'Validate Structure' (with a sub-link 'or View validation reports'), 'Deposit Structure' (with a sub-link 'All Deposition Resources'), and 'Download Archive' (with a sub-link 'Instructions'). At the bottom, there are three green buttons: 'Vision and Mission', 'wwPDB Resources', and 'News & Announcements'.

WORLDWIDE PDB PROTEIN DATA BANK

VALIDATION ▾ DEPOSITION ▾ DICTIONARIES ▾ DOCUMENTATION ▾ TASK FORCES ▾ FTP ▾ STATISTICS ▾ ABOUT ▾ wwPDB Foundation

Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Learn more about PDB **HISTORY** and **FUTURE**.

Validate Structure  
or View validation reports

Deposit Structure  
All Deposition Resources

Download Archive  
Instructions

Vision and Mission

wwPDB Resources


News & Announcements

\*Le 28 Janvier 2020, PDB archivait près de 167 mille structures 3D, obtenues essentiellement par expérimentation (cristallographie à rayon X, spectroscopie RMN, Cryo-microscopie électronique, etc.).

## Genome

Base de génomes annotés appartenant à plus de 307 000 organismes (eucaryotes, procaryotes, virus, plasmides, organelles) (données de janvier 2020)

<https://www.ncbi.nlm.nih.gov/genome>

 NCBI
 Resources ☒ How To ☒
Sign in to NCBI


Genome

Genome

Search

Limits Advanced

Help



### Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

#### Using Genome

[Help](#)

[Browse by Organism](#) UPDATED

[Download / FTP](#)

[Download FAQ](#)

[Submit a genome](#)

#### Custom resources

[Human Genome](#)

[Microbes](#)

[Organelles](#)

[Viruses](#)

[Prokaryotic reference genomes](#)

#### Other Resources

[Assembly](#)

[BioProject](#)

[BioSample](#)

[Genome Data Viewer](#) NEW

#### Genome Tools

[BLAST the Human Genome](#)

[Microbial Nucleotide BLAST](#)

#### Genome Annotation and Analysis

[Eukaryotic Genome Annotation](#)

[Prokaryotic Genome Annotation](#)

[PASC \(Pairwise Sequence Comparison\)](#)

#### External Resources

[GOLD - Genomes Online Database](#)

[Bacteria Genomes at Sanger](#)

[Ensembl](#)



## 2. Les bases de données spécialisées

Ces bases contiennent des données homogènes, sont établis autour

❑ D'une thématique :

- bases spécialisées dans certaines voies métaboliques, de structures particulières, d'expression de gènes, etc.



❑ D'un organisme :

- génome d'*Arabidopsis thaliana*, génome de *Bacillus subtilis*, etc.
- **Avantages** : facilité de mettre à jour les données, de vérifier leur intégrité,...
- **Inconvénients** : ne ciblent pas toujours ce que l'on veut; toutes les bases possibles n'existent pas.

# IMG (Integrated Microbial Genomes and Microbiomes) :

## Base de génomes microbiens annotés


<https://img.jgi.doe.gov/>


INTEGRATED MICROBIAL GENOMES & MICROBIOMES

[JGI HOME](#)
[CONTACT US](#)
[LOGIN / SIGN-ON](#)

[Home](#)
[IMG/M](#)
[Find Genomes](#)
[Find Genes](#)
[Find Functions](#)
[Compare Genomes](#)
[OMICS](#)
[My IMG](#)
[Help](#)

 NERSC, JGI and IMG will be down during a planned outage, starting from Feb 21, 2020 to Feb 25, 2020

### Integrated Microbial Genomes and Microbiomes




The **mission** of the Integrated Microbial Genomes & Microbiomes(IMG/M) system is to support the annotation, analysis and distribution of microbial genome and microbiome datasets sequenced at DOE's Joint Genome Institute (JGI).

IMG/M is also open to scientists worldwide for the annotation, analysis, and distribution of their own genome and microbiome datasets, as long as they agree with the IMG/M data release policy and follow the metadata requirements for integrating data into IMG/M (see IMG/M submission site).

If you use IMG web resources or data to assist in research publications or proposals, please cite Chen et al., 2018 (PMID: [30289528](#)).

[Review Data Usage Policy](#)
[Submit Your Data](#)

 IMG only accepts submission of assembled sequence data. Users who have un-assembled data can use [KBase Assembly Service](#) before submitting to IMG.

### Announcements

Thank you for making IMG submission so popular! Due to the high volume of submissions and a current system limitation, we have a backlog of 3000+ submissions. It is not necessary to submit additional tickets regarding these problems - we are aware of the issue and are working on it. We will not be able to answer e-mails regarding the status of specific submissions at this time.

### IMG Content

Genes


57 907 337 756

Bases

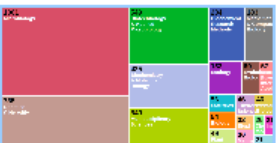
17 166 428 477 809

Scaffolds



53 155 729 405

 [IMG Statistics](#)


### IMG & GOLD Citations



### IMG Data Marts


 **IMG/ABC** - biosynthetic gene clusters and novel secondary metabolites in IMG
  **IMG/VR (Viral)** - viral fragments derived from metagenomic samples.

### Learn How to




- Find unique genes in genome
- Find common genes
- Find abundant protein families
- Use "My workspace"
- Submit Your Data

### Find and Analyze



- Genomes
- Metagenomes
- Proteins
- Pathways

### Save and Manage



- Systemwide Preferences
- My Object Sets
- Submit your Data
- My Jobs
- My Groups

6

## KEGG PATHWAY :

Base de données relatives aux voies métaboliques

<https://www.genome.jp/kegg/pathway.html>



### KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions and relations

Menu PATHWAY BRITE MODULE KO GENES LIGAND NETWORK DISEASE DRUG DBGET

Select prefix

map

Organism

Enter keywords

Go

Help

[ [New pathway maps](#) | [Update history](#) ]

### Pathway Maps

**KEGG PATHWAY** is a collection of manually drawn [pathway maps](#) representing our knowledge on the molecular interaction, reaction and relation networks for:

#### 1. Metabolism

[Global/overview](#) [Carbohydrate](#) [Energy](#) [Lipid](#) [Nucleotide](#) [Amino acid](#) [Other amino](#) [Glycan](#)  
[Cofactor/vitamin](#) [Terpenoid/PK](#) [Other secondary metabolite](#) [Xenobiotics](#) [Chemical structure](#)

#### 2. Genetic Information Processing

#### 3. Environmental Information Processing

#### 4. Cellular Processes

#### 5. Organismal Systems

#### 6. Human Diseases

#### 7. Drug Development

KEGG PATHWAY is the reference database for pathway mapping in [KEGG Mapper](#).

## FlyBase :

Base de séquences nucléotidiques de la drosophile

<https://flybase.org/>

FB2019\_06, released Dec 19, 2019

**FlyBase**  
A Database of *Drosophila* Genes & Genomes

Home Tools Downloads Links Community Species About Help Archives J2G Jump to Gene Go

BLAST GBrowse JBrowse Resources RNA-Seq Vocabularies ImageBrowse Batch Download

Using JBrowse on FlyBase.

Please pay your YEARLY website access fee! (annual renewal is automatic)  
**THANK YOU TO THOSE WHO HAVE ALREADY HELPED**

The NHGRI is reducing the funding of FlyBase by 20-25% (normalized to 30-35%). These cuts will make it difficult to deliver high quality, essential curation and tools. We are calling on you to help with a **FlyBase website access fee PER person / PER year**.

Questions? See this [FAQ](#) or [email us](#). Optional general tax-deductible contributions [here](#).

ALL countries	\$150.00	Commercial	\$750.00	Institutional Rate (covering several labs)	<a href="#">CONTACT US</a>
---------------	----------	------------	----------	--	----------------------------

PLEASE COMPLETE THIS **FORM** to sign up to pay fee. (Qualtrics external link)

QuickSearch

Human Disease Protein Domains Gene Groups Pathways GO Data Class

Search FlyBase Homologs GAL4 etc Expression Phenotype References

Everything Search

Click [here](#) to submit multiple IDs/symbols.

Note: Wild cards (\*) can be added to your search term

Resources for:

PUBLIC, TEACHERS, STUDENTS BIOMEDICAL RESEARCH POSITIONS AVAILABLE

FAST-TRACK YOUR PAPER FLYBASE NEWS FLY BOARD COMMUNITY NEWS MEETINGS COURSES FLYBOOK

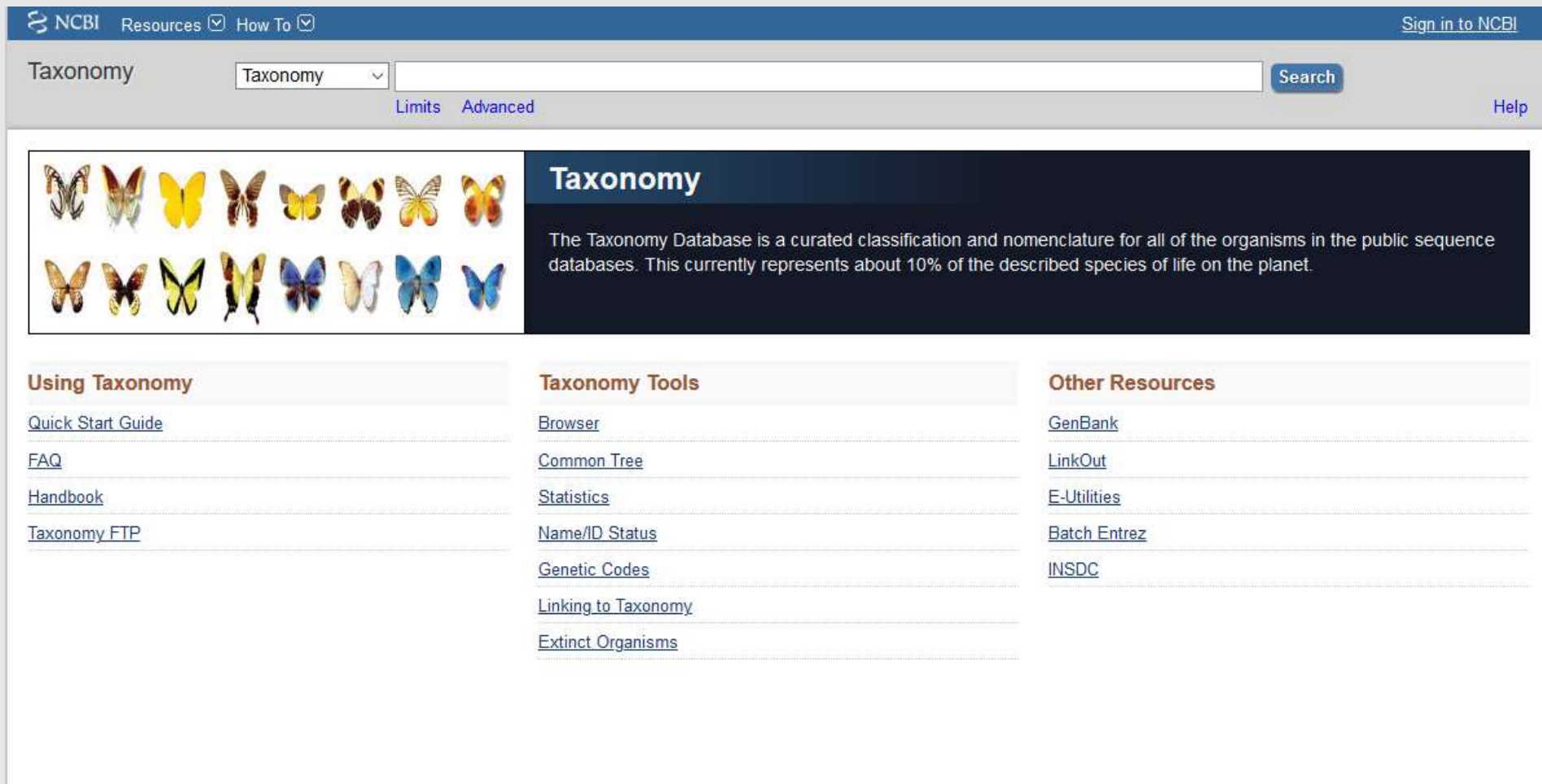
Commentary [see all Commentaries](#)



## Taxonomy

Base de données taxonomiques de plus de 653 000 organismes génétiquement identifiées  
(données de Janvier 2020)

<https://www.ncbi.nlm.nih.gov/taxonomy>



The screenshot shows the NCBI Taxonomy database homepage. At the top, there is a navigation bar with links to NCBI, Resources, and How To. A search bar is prominently displayed with a dropdown menu set to 'Taxonomy' and a 'Search' button. Below the search bar, there are links for 'Limits' and 'Advanced'. The main content area features a header with the word 'Taxonomy' and a description: 'The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.' To the left of this text is a grid of 16 colorful butterfly images. Below the main content, there are three columns of links. The first column, 'Using Taxonomy', includes links to 'Quick Start Guide', 'FAQ', 'Handbook', and 'Taxonomy FTP'. The second column, 'Taxonomy Tools', includes links to 'Browser', 'Common Tree', 'Statistics', 'Name/ID Status', 'Genetic Codes', 'Linking to Taxonomy', and 'Extinct Organisms'. The third column, 'Other Resources', includes links to 'GenBank', 'LinkOut', 'E-Utilities', 'Batch Entrez', and 'INSDC'.

NCBI Resources How To Sign in to NCBI

Taxonomy Taxonomy Search Limits Advanced Help

### Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

#### Using Taxonomy

- [Quick Start Guide](#)
- [FAQ](#)
- [Handbook](#)
- [Taxonomy FTP](#)

#### Taxonomy Tools

- [Browser](#)
- [Common Tree](#)
- [Statistics](#)
- [Name/ID Status](#)
- [Genetic Codes](#)
- [Linking to Taxonomy](#)
- [Extinct Organisms](#)

#### Other Resources

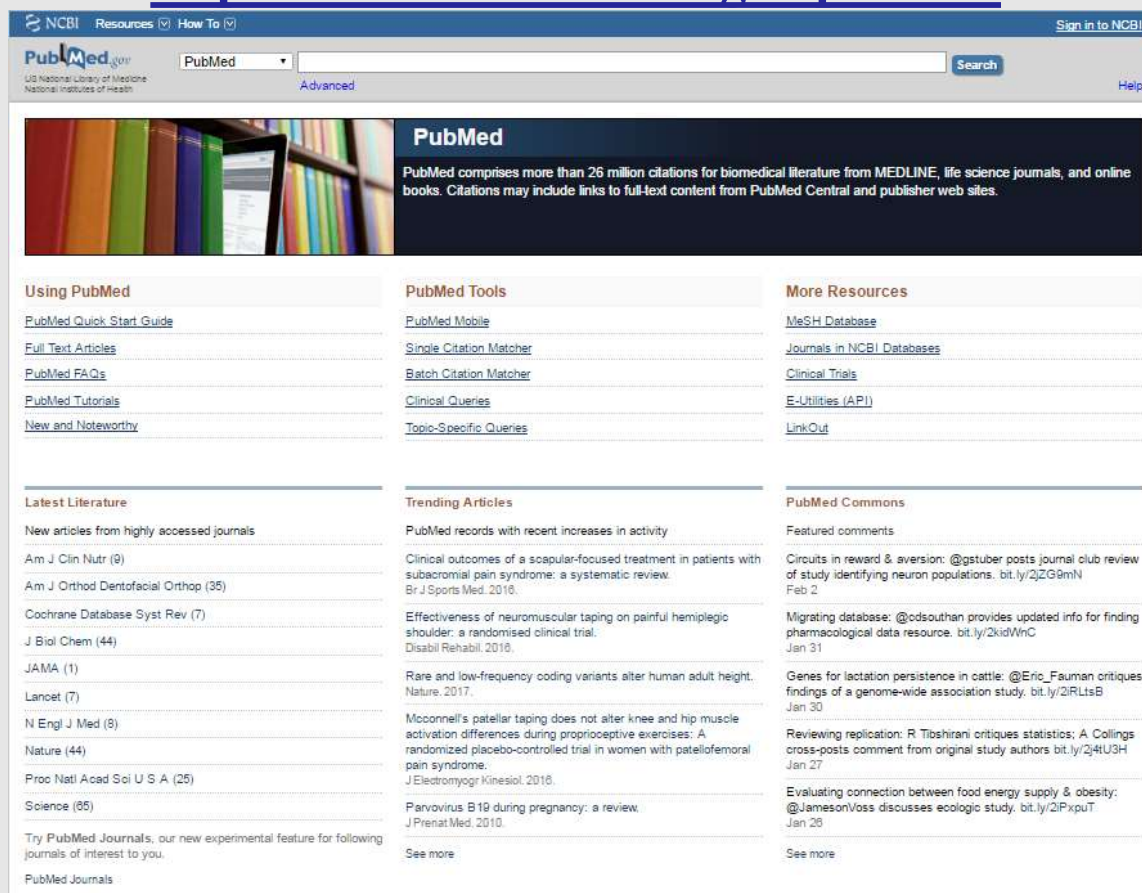
- [GenBank](#)
- [LinkOut](#)
- [E-Utilities](#)
- [Batch Entrez](#)
- [INSDC](#)

## BASES DE DONNÉES SPÉCIALISÉES

## BD BIBLIOGRAPHIQUES

- ✓ **Medline** (*Medical Literature Analysis and Retrieval System Online*) : la plus grande base de données bibliographiques de littérature relative aux sciences biologiques et médicales (biologie, biochimie, médecine clinique, pharmacologie, psychiatrie, toxicologie, etc.), gérée par la bibliothèque nationale de médecine des Etats Unis d'Amérique (NLM).

<http://www.ncbi.nlm.nih.gov/pubmed>

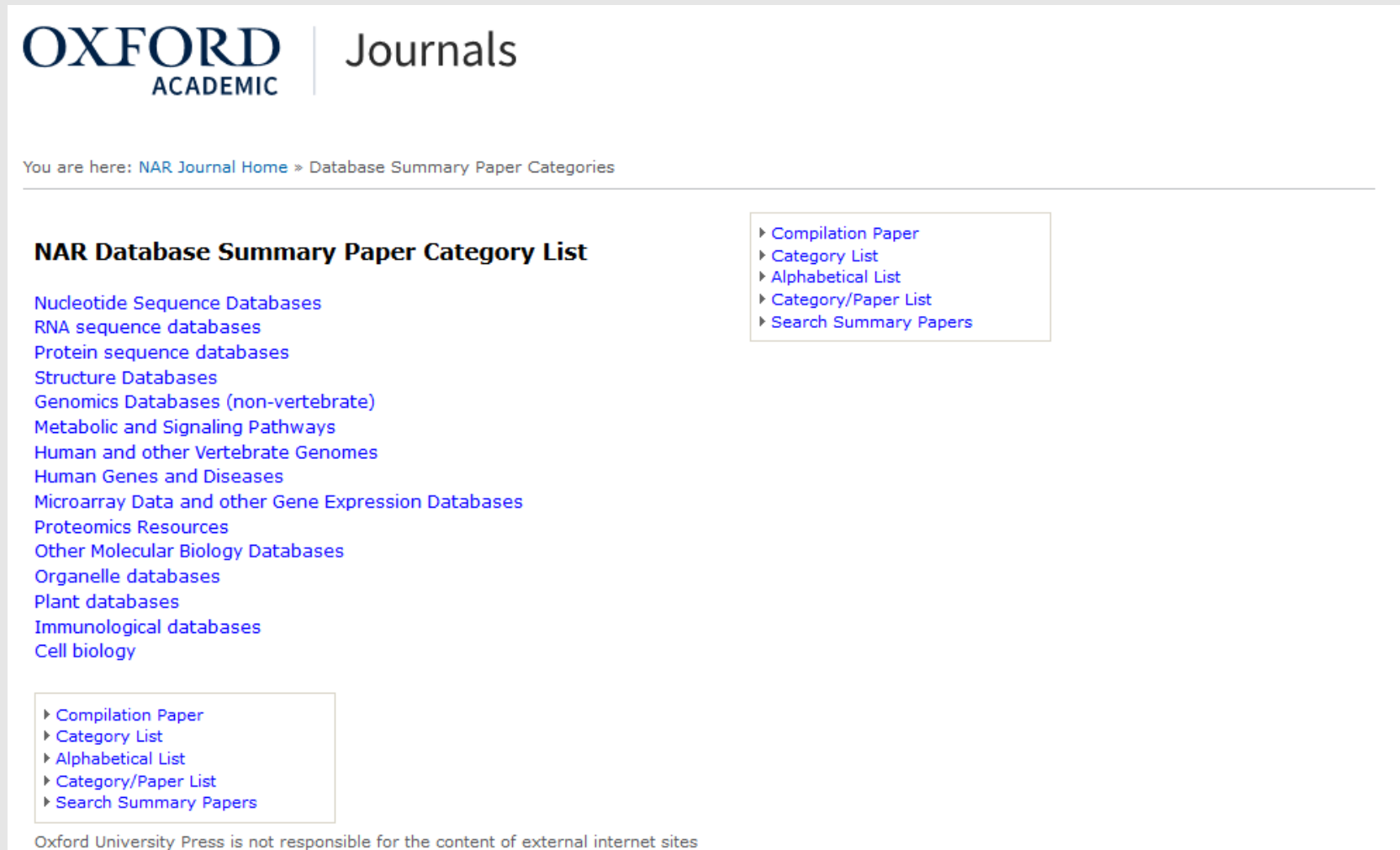


Fin 2016, cette base de données contenait plus de 23,5 millions de citations, publiées depuis 1948 dans environ 5100 revues biomédicales en 60 langues différentes

## AUTRES EXEMPLES DE BASES DE DONNÉES BIOLOGIQUES

La revue *Nucleic Acids Research* recense en janvier 2019 plus de 1600 bases de données biologiques spécialisées. La liste est consultable sur ce lien :

<http://www.oxfordjournals.org/nar/database/c>



The screenshot shows the Oxford Academic Journals website. The header includes the Oxford Academic logo and the word "Journals". Below the header, a breadcrumb trail reads "You are here: NAR Journal Home » Database Summary Paper Categories". The main content area is titled "NAR Database Summary Paper Category List". It contains a list of database categories, each with a blue link: "Nucleotide Sequence Databases", "RNA sequence databases", "Protein sequence databases", "Structure Databases", "Genomics Databases (non-vertebrate)", "Metabolic and Signaling Pathways", "Human and other Vertebrate Genomes", "Human Genes and Diseases", "Microarray Data and other Gene Expression Databases", "Proteomics Resources", "Other Molecular Biology Databases", "Organelle databases", "Plant databases", "Immunological databases", and "Cell biology". To the right of this list is a box containing a vertical list of links: "Compilation Paper", "Category List", "Alphabetical List", "Category/Paper List", and "Search Summary Papers". At the bottom left, there is another box with the same vertical list of links. At the bottom of the page, a footer states "Oxford University Press is not responsible for the content of external internet sites".

OXFORD  
ACADEMIC | Journals

You are here: [NAR Journal Home](#) » Database Summary Paper Categories

### NAR Database Summary Paper Category List

- [Nucleotide Sequence Databases](#)
- [RNA sequence databases](#)
- [Protein sequence databases](#)
- [Structure Databases](#)
- [Genomics Databases \(non-vertebrate\)](#)
- [Metabolic and Signaling Pathways](#)
- [Human and other Vertebrate Genomes](#)
- [Human Genes and Diseases](#)
- [Microarray Data and other Gene Expression Databases](#)
- [Proteomics Resources](#)
- [Other Molecular Biology Databases](#)
- [Organelle databases](#)
- [Plant databases](#)
- [Immunological databases](#)
- [Cell biology](#)

- [Compilation Paper](#)
- [Category List](#)
- [Alphabetical List](#)
- [Category/Paper List](#)
- [Search Summary Papers](#)

Oxford University Press is not responsible for the content of external internet sites

## AUTRES EXEMPLES DE BASES DE DONNÉES BIOLOGIQUES

La revue *Nucleic Acids Research* recense en janvier 2019 plus de 1600 bases de données biologiques spécialisées. La liste est consultable sur ce lien :

<http://www.oxfordjournals.org/nar/database/c>

The screenshot shows a web browser window with the URL [www.oxfordjournals.org/nar/database/c](http://www.oxfordjournals.org/nar/database/c). The page is titled "OXFORD ACADEMIC Journals". Below the header, a breadcrumb trail reads "You are here: NAR Journal Home » Database Summary Paper Categories". The main content area is titled "NAR Database Summary Paper Category List". It lists various biological databases and resources, including:

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression Databases
- Proteomics Resources
- Other Molecular Biology Databases
- Organelle databases
- Plant databases
- Immunological databases
- Cell biology

On the right side of the list, there is a box containing the following links:

- Compilation Paper
- Category List
- Alphabetical List
- Category/Paper List
- Search Summary Papers

At the bottom of the page, there is a footer with the following links:

- About Us
- Careers
- Gateways
- Oxford Academic Books & Online
- Subjects
- Arts & Humanities
- Resources
- Agents
- Customer Services
- Contact Us
- Oxford University Press is a department of the



## LES BASES DE DONNÉES BIOLOGIQUES

## LES SYSTÈMES D'INTERROGATION

Chaque base de séquences possède son propre système d'interrogation, chaque système utilise sa propre syntaxe de requête (langage).

Beaucoup de ces système permettent d'interroger plusieurs bases de données généralistes et spécialisées:

- ✓ **SRS** (*Sequence Retrieval System*) : permet d'interroger des centaines de bases de données biologiques, généralistes et spécialisées, il est notamment utiliser pour Uniprot ;
- ✓ **Entrez** : appartenant au NCBI, permet l'interrogation des bases de séquences Medline et PubMed, GenBank, et les autres bases affiliées au portail américain ;
- ✓ **ACNUC** : système d'interrogation des bases GenBank, ENA/EMBL ou UniProtKB/SwissProt, etc.
- ✓ **EB-Eye** : système d'interrogation des bases EBI (*European Bioinformatics Institute*).

## LES BASES DE DONNÉES BIOLOGIQUES

## LES SYSTÈMES D'INTERROGATION

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide OR Mus musculus[All Fields] AND insulin[All Fields] AND (animals[filter] AND biomol\_mrna[PROP]) Search

Create alert Advanced Help

Species clear Summary 20 per page Sort by Default order

Animals (2,835) Customize ...

Molecule types clear

mRNA (2,835) Customize ...

Source databases

INSDC (GenBank) (1,137) RefSeq (1,698) Customize ...

Sequence Type

Nucleotide (2,320) EST (515)

Sequence length

Custom range...

Release date

Custom range...

Revision date

Custom range...

Clear all

Show additional filters

Items: 1 to 20 of 2835

Filters activated: Animals, mRNA. Clear all

1. [Mus musculus insulin degrading enzyme \(Ide\), mRNA](#)  
5,087 bp linear mRNA  
Accession: NM\_031156.3 GI: 459352740  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

2. [Mus musculus insulin receptor substrate 3 \(Irs3\), mRNA](#)  
2,337 bp linear mRNA  
Accession: NM\_010571.3 GI: 146134339  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

3. [Mus musculus insulin receptor \(Insr\), transcript variant 2, mRNA](#)  
9,391 bp linear mRNA  
Accession: NM\_001330056.1 GI: 1052292401  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

4. [Mus musculus insulin receptor \(Insr\), transcript variant 1, mRNA](#)  
9,355 bp linear mRNA  
Accession: NM\_010568.3 GI: 1052292400  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

5. [Mus musculus insulin II \(Ins2\), transcript variant 2, mRNA](#)  
485 bp linear mRNA  
Accession: NM\_008387.5 GI: 915410360  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

6. [Mus musculus insulin II \(Ins2\), transcript variant 3, mRNA](#)  
587 bp linear mRNA  
Accession: NM\_001185084.2 GI: 915410508  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Send to: Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree](#)

Mus musculus (2822)  
Rattus norvegicus (8)  
Danio rerio (2)  
Homo sapiens (1)  
Gallus gallus (1)  
All other taxa (1)  
More...

Find related data

Database: Select

Find items

Search details

```
((("Mus musculus"[Organism] OR  
("Mus musculus"[Organism] OR Mus  
musculus[All Fields])) AND  
insulin[All Fields]) AND  
(animals[filter] AND  
biomol_mrna[PROP]) AND
```

Search See more...

Recent activity

Turn Off Clear

Q ((("Mus musculus"[Organism] OR Mus musculus[All Fields]) AND insulin[All Fields]) AND biomol\_mrna[PROP]) (2835) Nucleotide

Q Mus musculus insulin AND (animals[filter] AND biomol\_mrna[PROP]) (2835) Nucleotide

Q mouse insulin AND (animals[filter] AND biomol\_mrna[PROP]) (4253) Nucleotide

Q mouse insulin AND (biomol\_mrna[PROP]) (4258) Nucleotide

Q mouse insulin (110409) Nucleotide

See more...

Exemple d'une requête sous le langage ENTREZ de Genbank

## Définition d'un format

Les séquences sont stockées en général sous forme de fichiers texte, accessibles par des systèmes d'interrogations (SRS pour UniProt, ACNUC pour EMBL, Entrez pour Genbank, etc.).

Le format correspond à **l'ensemble des règles de présentation** auxquelles sont soumises la ou les séquences dans un fichier donné. Ainsi, le format permet :

- une mise en forme automatisée;
- le stockage homogène de l'information;
- le traitement informatique ultérieur de l'information. Pour lire et traiter les séquences, les logiciels d'analyse autorisent un ou plusieurs formats des données.

## LES BASES DE DONNÉES BIOLOGIQUES

## SYNTAXE D'UNE ENTRÉE

Une entrée : fiche signalétique d'une séquence donnée

Contient trois parties :

- 1- Description générale de la séquence
- 2- Features : Description des objets biologiques présents sur la séquence
- 3- La séquence

**Description générale de la  
séquence**

**« Features »**  
**Description des objets  
biologiques présents sur la  
séquence**

### La séquence

```
ctccggcagc ccgaggtcat cctgctagac tcagacctgg atgaacccat agacttgcgc    60
tcggtcaaga gccgcagcga ggccggggag ccgccagct cctccaggt gaagcccag    120
acaccggcgt cggcggcggt ggcggtggcg gcggcagcgg caccaccac gacggcggag    180
```

- **Chaque ligne commence par un mot-clé**
  - **Deux lettres pour EMBL**
  - **Maximum 12 lettres pour Genbank et DDBJ**
- **Fin d'une entrée : //**



# Saccharomyces cerevisiae strain JZ1C invertase (SUC2) gene, complete cds

GenBank: JQ836661.1

[FASTA](#) [Graphics](#)

LOCUS JQ836661 1599 bp DNA linear PLN 26-DEC-2012  
DEFINITION Saccharomyces cerevisiae strain JZ1C invertase (SUC2) gene, complete cds.  
ACCESSION JQ836661  
VERSION JQ836661.1 GI:393395465  
KEYWORDS -  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM [Saccharomyces cerevisiae](#)  
Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;  
Saccharomycetes; Saccharomycetales; Saccharomycetaceae;  
Saccharomyces.  
REFERENCE 1 (bases 1 to 1599)  
AUTHORS Wang, S.-A. and Li, F.-L.  
TITLE Invertase SUC2 is the Key Hydrolase for Inulin Degradation in  
Saccharomyces cerevisiae  
JOURNAL Appl. Environ. Microbiol. 79 (1), 403-406 (2013)  
PUBMED [23104410](#)  
REFERENCE 2 (bases 1 to 1599)  
AUTHORS Wang, S.-A. and Li, F.-L.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAR-2012) Key Laboratory of Biofuels, Qingdao  
Institute of Bioenergy and Bioprocess Technology, 189 Songling  
Road, Qingdao, Shandong 266101, China

FEATURES  
source Location/Qualifiers  
1..1599  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="JZ1C"  
/db\_xref="taxon:4932"  
c1..1599  
/gene="SUC2"  
c1..1599  
/gene="SUC2"  
/product="invertase"  
1..1599  
/gene="SUC2"  
/codon\_start=1  
/product="invertase"  
/protein\_id="AFN08663.1"  
/db\_xref="GI:393395465"  
/translation="MLQAFLLAGFAAKIASMTNETSDFLVHFTYNEGWMNDPN  
GLWYDEKDAKWHLYPQYHNDTVMGTPLPWGHATSDDLTHNDEPIAIAPKRNDSGAF  
SGSMVVDYNNITSGFPNDIIPRQKCVAIWNTYTPKSEHQYISYELDGGTTFYQKNP  
VLAARSTQVHDPRVFWYEPSSQWMTAAKSGQYKIEIYSSDOLSKWLESAFAMEGFL  
GYQECFGLIEVPTKQOPSKSYWVFISIMPAPAGGSSFNQYFVGSFNGTHFEAFDNG  
SRVYDFGRDYALQTFYNTDPTYGSALGIWASRWETSAFVFTNPWSSMLVSNFSL  
NTEYQANPETELINLKAEPILNISNAGPMSRATNITLTRANSYVDLSNSTGTLEFE  
LVYAVNTTQTISKSVFPLSLMFEGLEDPEKYLNGFEASASFFLDKGNSEKVFYKE  
NPFYTNQMSVNNQPFKSENLSYKVGLLDQNIILELYPNQGVVSTNTYPMITGNAL  
GSVMITGVNLEFYIDKPVVERVK"

ORIGIN  
1 atgttttttc aagtttttct ttctcttttg gctggttttg cagccaaaat atctgcatca  
61 atgcacaaag aaactagcgc tagacttttg gtccacttca ccccacaaa ggcctggatg  
121 aatgacccaa atgggttttg gtacgtgtaa aaagatgcca atggcctctc gtactttcaa  
181 tacaacccaa atgacacngt atgggttagc ccattgtttt ggggcctatc tacttccgat  
241 gatttgcctc atggggaaga tgaacccatt gctatngctc ccaagcgtaa cgaattcagg  
301 gctttctctc gctccatggt ggttgattac aacacacaga gtgggttttt caatgtactl  
361 attgatccaa gacaaagatg cgttgcgatt tggacttata acactcctga aagtgaagag  
421 caatcacctt gctattctct tgaatgggtt tacactttta ctgaatccca aaagaacctt  
481 gttttagctg ccaactccac tcaattcaga gatccaaagg tgttctgata tgaacctlnt  
541 caaaaatgga ttatgacggc tgcacaaatc caagatccca aaattgaatt ttactctctt  
601 gatgacttga agtactggaa gctagaatct gcaattgcta atgaaggttt cttaggctac  
661 caatatgaat gtccgggttt gattgaagtc ccaactgagc aagatccttc caaatctcat  
721 tgggtcctgt ttatttctat caatccaggt gcaactgctg ggggttcttt caacccat  
781 ttgtttgget ccttccatgg tactctattt gaagcgtttg acaatcaatc tagagtggta  
841 gattttggta aggaactata tgccttgcaa actttcttca acacagaacc aacgtacggt  
901 tcagcattag gtattgctgt ggccttcaaa tgggagtaca gtgcttttgt cccaactaac  
961 ccatggagat catccatgtc ttgtgtccgc aagttttctt tgaadactga atatccagct  
1021 aatccagaga ctgaattgat caatttgaaa gccgaaccaa tattgaacct tagtaatgct  
1081 ggtccctggt ctggttttgc tactaacaca acttcaacta aggcacattc ttcaatgtc  
1141 gatttgagca actgcactgg tactctagag tttagagttg tttagcgtgt taacaccaca  
1201 caaacccat ccaactcgtt ctttcccgag ttctcacttt ggttcaaggg tttagaagat  
1261 cutgagaagt atttaagaat yggttttgaa gccagtgttt ctctcttttt tttagccgt  
1321 ggttaactta aggtcaagtt tgtcaagggc aacccatatt tcaacaaag aatgtctgtc  
1381 acaacacac catccaagtc tgaagacgac ctcaattact ataaagtgtc cggcctactg  
1441 gatcaaaaca tcttggaatt gtacttcaac gatggagatg tggtttctac aatatctac  
1501 ttcatgacca ccggtaacgc tctaggatct gtgaacatga ccaatggtgt cgataattg  
1561 ttctacattg caaagtacca agtaagggaa gtaaatag

Description générale de la séquence

« Features »

Description des objets biologiques présents sur la séquence

La séquence

## Description générale de la séquence (Genbank)

LOCUS JQ836661 1599 bp DNA linear PLN 26-DEC-2012  
 DEFINITION *Saccharomyces cerevisiae* strain JZ1C invertase (SUC2) gene,  
 complete cds.  
 ACCESSION JQ836661  
 VERSION JQ836661.1 GI:393395465  
 KEYWORDS .  
 SOURCE *Saccharomyces cerevisiae* (baker's yeast)  
 ORGANISM [Saccharomyces cerevisiae](#)  
 Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;  
 Saccharomycetes; Saccharomycetales; Saccharomycetaceae;  
 Saccharomyces.  
 REFERENCE 1 (bases 1 to 1599)  
 AUTHORS Wang,S.A. and Li,F.L.  
 TITLE Invertase SUC2 Is the Key Hydrolase for Inulin Degradation in  
*Saccharomyces cerevisiae*  
 JOURNAL Appl. Environ. Microbiol. 79 (1), 403-406 (2013)  
 PUBMED [23104410](#)  
 REFERENCE 2 (bases 1 to 1599)  
 AUTHORS Wang,S.-A. and Li,F.-L.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-MAR-2012) Key Laboratory of Biofuels, Qingdao  
 Institute of Bioenergy and Bioprocess Technology, 189 Songling  
 Road, Qingdao, Shandong 266101, China

## « Features »

### Description des objets biologiques présents sur la séquence

FEATURES	Location/Qualifiers
source	1..1599 /organism="Saccharomyces cerevisiae" /mol_type="genomic DNA" /strain="JZ1C" /db_xref="taxon: <a href="#">4932</a> "
<a href="#">gene</a>	<1..>1599 /gene="SUC2"
<a href="#">mRNA</a>	<1..>1599 /gene="SUC2" /product="invertase"
<a href="#">CDS</a>	1..1599 /gene="SUC2" /codon_start=1 /product="invertase" /protein_id=" <a href="#">AFN08663.1</a> " /db_xref="GI:393395466" /translation="MLLQAFLLFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPN GLWYDEKDAKWHLYFQYNPNDTVWGTPLEFWGHATSDDLTHWEDEPIAIAPKRND SGAF SGSMVVDYNNNTSGFFNDTIDPRQRCVAIWYNTPESEEQYISYSLDGGYTFTEYQKNP VLAANSTQFRDPKVFWYEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFL GYQYECPLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQ SRVVDGKDYALQTFNTDPTYGSALGIAWASNWEYSASFVPTNPWRSSMSLVKRFSL NTEYQANPETELINLKAEPILNISNAGPWSRFATNTTLTKANSYNVDLSNSTGTLEFE LVYAVNTTQTISKSVFPDLSLWFKGLEDP E EYLRMGFEASASSFFLDRGNSKVKEVKE NPYFTNRMSVNNQPFKSENDLSYYKVYGLLDQNILELYFNDGDVVSTNTYFMTTGNAL GSVNMTTGVDNLFYIDKFQVREVK"



## La séquence (format Genbank)

ORIGIN

```

1  atgctttttgc aagcttttcct tttcctttttg gctgggttttg cagccaaaat atctgcatca
61  atgacaaacg aaactagcga tagacctttg gtccacttca caccacaaca gggctggatg
121  aatgacccaa atgggtttgtg gtacgatgaa aaagatgcca aatggcatct gtactttcaa
181  tacaacccaa atgacaccgt atgggggtacg ccattgtttt ggggccatgc tacttccgat
241  gatttgactc attgggaaga tgaaccatt gctatcgctc ccaagcgtaa cgattcaggt
301  gctttctctg gctccatggg ggttgattac aacaacacga gtgggttttt caatgatact
361  attgatccaa gacaaagatg cgttgcgatt tggacttata acactcctga aagtgaagag
421  caatacatta gctattctct tgatgggtgg taccctttta ctgaatacca aaagaaccct
481  gtttttagctg ccaactccac tcaattcaga gatccaaagg tgttctggta tgaaccttct
541  caaaaatgga ttatgacggc tgccaaatca caagactaca aaattgaaat ttactcctct
601  gatgacttga agtcctggaa gctagaatct gcatttgcta atgaagggtt cttaggctac
661  caatatgaat gtccagggtt gattgaagtc ccaactgagc aagatccttc caaatcctat
721  tgggtcatgt ttatttctat caatccagggt gcacctgctg gcggttcctt caaccaatat
781  tttgttggat cettcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggta
841  gatttttggt aggactacta tgccttgcaa actttcttca acacagaccc aacgtacggg
901  tcagcattag gtattgcctg ggcttcaaac tgggagtaca gtgcctttgt cccaactaac
961  ccatggagat catccatgtc tttgggtccgc aagttttctt tgaacactga atatcaagct
1021  aatccagaga ctgaattgat caatttgaaa gccgaaccaa tattgaacat tagtaatgct
1081  ggtccctggg ctcgttttgc tactaacaca actctaacta aggccaattc ttacaatgtc
1141  gatttgagca actcgactgg taccctagag tttgagttgg tttacgctgt taacaccaca
1201  caaaccatat ccaaatccgt ctttcccgac ttatcacttt ggttcaaggg tttagaagat
1261  cctgaagaat atttaagaat gggttttgaa gccagtgtt cttccttctt tttggaccgt
1321  ggtaactcta aggtcaagtt tgtcaaggag aacccatatt tcacaaacag aatgtctgtc
1381  aacaaccaac cattcaagtc tgagaacgac ctaagttact ataaagtgtc cggcctactg
1441  gatcaaaaca tcttggaatt gtacttcaac gatggagatg tggtttctac aaatacctac
1501  ttcatgacca ccggtaacgc tctaggatct gtgaacatga ccactgggtg cgataatttg
1561  ttctacattg acaagttcca agtaagggaa gtaaaatag

```



## Sequence: JQ836661.1 : Saccharomyces cerevisiae strain JZ1C invertase (SUC2) gene, complete cds.

```

ID: JQ836661.1 SV: 1; linear; genomic DNA; STD; RM: 1599 bp.
XX
AC: JQ836661;
XX
DT: 06-JUL-2012 (Rel. 113, Created)
DT: 28-DEC-2012 (Rel. 115, Last updated, Version 2)
XX
SR: Saccharomyces cerevisiae strain JZ1C invertase (SUC2) gene, complete cds.
XX
KW:
XX
OR: Saccharomyces cerevisiae (baker's yeast)
OC: Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetaceae;
OC: Saccharomycetaceae; Saccharomycetaceae; Saccharomycetaceae.
XX
RN: [1]
RP: 1-1599
RX: DOI: 10.1101/ADM.02456-12.
RX: PMID: 23104410.
RA: Wang S.-A., Li F.-L.
RT: "Invertase SUC2 Is the Key Hydrolase for Inulin Degradation in
RT: Saccharomyces cerevisiae".
RL: Appl. Environ. Microbiol. 79(11):403-406(2013).
XX
RH: [2]
RP: 1-1599
RA: Wang S.-A., Li F.-L.
RT:
RL: Submitted (27-MAR-2012) to the INSDC.
RL: Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess
RL: Technology, 189 Dongling Road, Qingdao, Shandong 266101, China
XX
SR: EuropePMC; PMC3536088; 23104410.
XX
FH: Key Location/Qualifiers
FH

```

```

FT: source 1..1599
FT: /organism="Saccharomyces cerevisiae"
FT: /strain="JZ1C"
FT: /mol_type="genomic DNA"
FT: /db_xref="taxon:4932"
FT: gene 1..1599
FT: /gene="SUC2"
FT: mRNA 1..1599
FT: /gene="SUC2"
FT: /product="invertase"
FT: cds 1..1599
FT: /codon_start=1
FT: /gene="SUC2"
FT: /product="invertase"
FT: /db_xref="GeneID:10909"
FT: /db_xref="InterPro:IPR001362"
FT: /db_xref="InterPro:IPR000995"
FT: /db_xref="InterPro:IPR01040"
FT: /db_xref="InterPro:IPR01314"
FT: /db_xref="InterPro:IPR01319"
FT: /db_xref="InterPro:IPR01403"
FT: /db_xref="InterPro:IPR02096"
FT: /db_xref="UniProtKB/TrEMBL:JZ1C"
FT: /protein_id="AF086661.1"
FT: /translation="MLIQAFLLLAGFAKTSAMTSTSGRPLVHTPHKGMKNDPWS
FT: LNYKATAGNHLVPGKNDHTVWSTPLWMAATSDLTQGRDEPIATAKKSGHGAAGS
FT: SMVYVYMTSGPNDTTPGKQVAMTYNTPESEYVTSYALDNGVTPTEYQNPVLA
FT: ANRTQPGKPVWTEPSGKQMTAAKSGDYKCEVYASCEKSMKLSAFANSGFLVQY
FT: KCPGLTVPTSGSKSYNMFISINQAPAGSGNGVFGSGPMTHTFAFGQKRVN
FT: PMSGYALGFTWSDPTGALITAMKQSGKSTAPPTWQKSGMULSAPFSLTETVA
FT: NPEKELINLAKAPTLATSNAGPMSRTATSTTLTQNGTQVCLSNSTGLTEFLVYVNT
FT: TQTTSKSVFVSLGNTWGLKQPEKYLKSGFASASGFLIRGKSKVKPVKNTPTNRW
FT: SVNKGPPKGNICSTPVYGLDQNTLILYNDGQVYVNTYMTGALGGVWNTGV
FT: DNLPTIDKQVREK"
XX

```

```

SQ: Sequence 1599 bp: 159 34 71 116 61 476 71 9 other
atgctcttgc agcttctcct tctctcttgc gctggttttc gctggttttc atctgcacac
atgacaaagc aactatagca tagactcttc gctcacttcc taccacacac ggcctgagc
aatgacacaa atggttttgc gtagatgaa aactatgcaa atggtcactt gctcttcaa
taccacacaa atgacacagt atggttttgc cactgctttt ggcctcagtc tacttctgct
gatttctgct atggttgaag tgaacacatt gctatctgtc ccaagatgta agattcagct
gctttctctg gctcactggt gcttgcattc aacacacaga gctggttttt caatgatact
atgacacaaa gacaaagatg cgttgcattc tggacttata aactatctga aactgacagc
caatcactta gctatctctc tggctggttc taccatttca ctgactacaa aactgacact
gctttcagtc cactcactcc tcaactcaga gctcactcag tcttctgta tgaactctct
caaaatgaga tctatgagtc tgcacatcca caagatctca aactgacatt tctactctct
gatgacttga agctcagcag gctatgactc gacttctgta atgacggttt cttagcttgc
caatcagatt gctcaggttt gattgagtc ccaactgagc agactcttc caactcctat
tggctcagtc tacttctct caatcaggtt gcaactgctg ggccttcttc caactcctat
tcttctggtt cttcactatg tactcatttt gtagcttctc aactcacttc tagactgcta
gatttctgta agactacta tgccttgcac acttcttcca aacacagacc aactgacggt
tactcttgc gcttctgctc gcttctgctc tggactgata tggctctctc ccaactaac
cactgagatt cactcacttc tcttctcttc tgaactcttc tgaactcttc atactaagct
aactcagaga ctgacttctc caatttgaac gcaacaccaa tattgacact tagtaactgt
gctcctctct cctgcttctc tactaacaca actcacttca aggcacttc tcaactgctc
gatttgcaga actgacttgc taccatagtc tttgacttgc tttactgctc taacacacaa
caaacctatc caaacctcgt ctttctctgc tttactcttc gtttcaagtc tttgacagt
cctgacagat atttcaagat ggttcttgc gcaactgctt cttctctctt tttgacagt
gcttacttca agtctcagct tgaacagagc aactcacttc taccacacag aactgcttgc
taccacacaa cactcacttc tgaactcaga gcttacttca caactgata ggcctcactc
gatcacaaca tcttgcagat gtaactcaca gatgagatg tggctcttgc aactcactc
tctatgacaa cgggttaagc tctatgactc gtagacatg cactgcttgc gataacttgc
tctcacttgc aactgcttgc agtgcagaga gtaactatg

```

## Description générale de la séquence

## « Features » Description des objets biologiques présents sur la séquence

## La séquence

## Description générale de la séquence (EMBL)

```

ID      JQ836661; SV 1; linear; genomic DNA; STD; FUN; 1599 BP.
XX
AC      JQ836661;
XX
DT      08-JUL-2012 (Rel. 113, Created)
DT      28-DEC-2012 (Rel. 115, Last updated, Version 2)
XX
DE      Saccharomyces cerevisiae strain JZ1C invertase (SUC2) gene, complete cds.
XX
KW      .
XX
OS      Saccharomyces cerevisiae (baker's yeast)
OC      Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
XX
RN      [1]
RP      1-1599
RX      DOI; 10.1128/AEM.02658-12.
RX      PUBMED; 23104410.
RA      Wang S.A., Li F.L.;
RT      "Invertase SUC2 Is the Key Hydrolase for Inulin Degradation in
RT      Saccharomyces cerevisiae";
RL      Appl. Environ. Microbiol. 79(1):403-406(2013).
XX
RN      [2]
RP      1-1599
RA      Wang S.-A., Li F.-L.;
RT      ;
RL      Submitted (27-MAR-2012) to the INSDC.
RL      Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess
RL      Technology, 189 Songling Road, Qingdao, Shandong 266101, China
XX
DR      EuropePMC; PMC3536088; 23104410.
XX

```

## « Features »

### Description des objets biologiques présents sur la séquence

```

----
FH      Key                Location/Qualifiers
FH
FT      source              1..1599
FT                               /organism="Saccharomyces cerevisiae"
FT                               /strain="JZ1C"
FT                               /mol_type="genomic DNA"
FT                               /db_xref="taxon:4932"
FT      gene                <1..>1599
FT                               /gene="SUC2"
FT      mRNA                <1..>1599
FT                               /gene="SUC2"
FT                               /product="invertase"
FT      CDS                 1..1599
FT                               /codon_start=1
FT                               /gene="SUC2"
FT                               /product="invertase"
FT                               /db_xref="GOA:I6VAV9"
FT                               /db_xref="InterPro:IPR001362"
FT                               /db_xref="InterPro:IPR008985"
FT                               /db_xref="InterPro:IPR011040"
FT                               /db_xref="InterPro:IPR013148"
FT                               /db_xref="InterPro:IPR013189"
FT                               /db_xref="InterPro:IPR018053"
FT                               /db_xref="InterPro:IPR023296"
FT                               /db_xref="UniProtKB/TrEMBL:I6VAV9"
FT                               /protein_id="AFN08663.1"
FT                               /translation="MLLQAFLELLAGFAAKISASMTNETSDRPLVHFTPNGKGMNDPNG
FT                               LWYDEKDAKWHLYFQYNPNDTVWGTPLFWGHATSDDLTHWEDEPIAIAPKRNDSGAFSG
FT                               SMVVDYNNNTSGFFNDTIDPRQRCVAIWYNTPESEEQYISYSLDGGYTFTEYQKNPVL
FT                               ANSTQFRDPKVFVWYEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFLGYQY
FT                               ECPGLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQSRVVD
FT                               FGKDYALQTFNTDPTYGSALGIAWASNWEYSAFVPTNPWRSSMSLVRKFSLNTEYQA
FT                               NPETELINLKAEPILNISNAGPWSRFATNTTLTKANSYNVDLSNSTGTLEFELVYAVNT
FT                               TQTISKSVFPDLSLWFKGLEDPEEYLRMGFEASASSFFLDRGNSKVKEFVKNPYFTNRM
FT                               SVNQPFKSENDLSYYKVYGLLDQNILELYFNDGDVVSNTYFMTTGNALGSVNMTTGV
FT                               DNLFYIDKFQVREVK"
XX

```

## La séquence (format EMBL)

```

SQ      Sequence 1599 BP; 459 A; 348 C; 316 G; 476 T; 0 other;
atgctttttgc aagcttttct tttccttttg gctggttttg cagccaaaat atctgcatca      60
atgacaaacg aaactagcga tagacctttg gtccacttca caccacaaca gggctggatg      120
aatgacccaa atgggttgtg gtacgatgaa aaagatgcca aatggcatct gtactttcaa      180
tacaacccaa atgacaccgt atgggggtacg ccattgtttt ggggccatgc tacttccgat      240
gatttgactc attgggaaga tgaacccatt gctatcgctc ccaagcgtaa cgattcaggt      300
gctttctctg gctccatggg ggttgattac aacaacacga gtgggttttt caatgatact      360
attgatccaa gacaaagatg cgttgcgatt tggacttata acactcctga aagtgaagag      420
caatacatta gctattctct tgatgggtggg tacactttta ctgaatacca aaagaaccct      480
gttttagctg ccaactccac tcaattcaga gatccaaagg tgttctggta tgaaccttct      540
caaaaatgga ttatgacggc tgccaaatca caagactaca aaattgaaat ttactcctct      600
gatgacttga agtcctggaa gctagaatct gcatttgcta atgaagggtt cttagggtac      660
caatatgaat gtccagggtt gattgaagtc ccaactgagc aagatccttc caaatcctat      720
tgggtcatgt ttatttctat caatccaggt gcacctgctg gcggttcctt caaccaatat      780
tttggttgat ccttcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggta      840
gattttggta aggactacta tgccttgcaa actttcttca acacagaccc aacgtacggg      900
tcagcattag gtattgcctg ggcttcaaac tgggagtaca gtgcctttgt cccaactaac      960
ccatggagat catccatgtc tttggtcctc aagttttctt tgaacactga atatcaagct     1020
aatccagaga ctgaattgat caatttgaaa gccgaaccaa tattgaacat tagtaatgct     1080
ggtcctctggg ctcgtttttg tactaacaca actctaacta aggccaattc ttacaatgtc     1140
gatttgagca actcgactgg taccctagag tttgagttgg ttacgctgt taacaccaca     1200
caaaccatat ccaaatccgt ctttcccgcac ttatcacttt ggttcaaggg tttagaagat     1260
cctgaagaat atttaagaat gggttttgaa gccagtgett cttecttctt tttggaccgt     1320
ggtaactcta aggtcaagtt tgtcaaggag aaccatatt tcacaaacag aatgtctgtc     1380
aacaaccaac cattcaagtc tgagaacgac ctaagttact ataaagtgtg cggcctactg     1440
gatcaaaaca tcttggaatt gtacttcaac gatggagatg tggtttctac aaatacctac     1500
ttcatgacca ccggtaacgc tctaggatct gtgaacatga ccactgggtg cgataatttg     1560
ttctacattg acaagttcca agtaagggaa gtaaaatag                                1599

```



## LES BASES DE DONNÉES BIOLOGIQUES

## SYNTAXE D'UNE ENTRÉE

- ID : nom de l'entrée , ...
  - Unique (propre à une entrée)
  - Non permanent (peut changer au cours des versions)
- AC : numéro d'accession
  - Unique, plusieurs ou pour une même entrée (fusion d'entrées)
  - Permanent (ne disparaît jamais de la base)
- SV : version de la séquence (Acc.version)
- DT : date d'incorporation dans la base et de dernière mise à jour
- DE : description du contenu de l'entrée

ID I6VAV9 YEASX Unreviewed; 532 AA.  
 AC I6VAV9;  
 DT 03-OCT-2012, integrated into UniProtKB/TrEMBL.  
 DT 03-OCT-2012, sequence version 1.  
 DT 05-DEC-2018, entry version 20.  
 DE SubName: Full=Invertase {ECO:0000313|EMBL:AFN08663.1};  
 GN Name=SUC2 {ECO:0000313|EMBL:AFN08663.1};  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;  
 OC Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932 {ECO:0000313|EMBL:AFN08663.1};  
 RN [1] {ECO:0000313|EMBL:AFN08663.1}  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JZ1C {ECO:0000313|EMBL:AFN08663.1};  
 RA Tserendorj M., Badgar B., Tserendorj N., Thillaiampalam S.,  
 RA AbouLaila M., Banzragch B., Byambaa P., Yokoyama N., Igarashi I.;  
 RT "A field study on the prevalence of equine piroplasmiasis in Mongolian  
 RT horses.";  
 RL Submitted (MAR-2012) to the EMBL/GenBank/DDBJ databases.  
 RN [2] {ECO:0000313|EMBL:AFN08663.1}  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JZ1C {ECO:0000313|EMBL:AFN08663.1};  
 DR EMBL; JQ836661; AFN08663.1; -; Genomic\_DNA.  
 DR eggNOG; KOG0228; Eukaryota.  
 DR eggNOG; COG1621; LUCA.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl compounds; IEA:InterPro.  
 DR GO; GO:0005975; P:carbohydrate metabolic process; IEA:InterPro.  
 DR Gene3D; 2.115.10.20; -; 1.  
 DR InterPro; IPR013320; ConA-like\_dom\_sf.  
 DR InterPro; IPR001362; Glyco\_hydro\_32.  
 DR InterPro; IPR018053; Glyco\_hydro\_32\_AS.  
 DR InterPro; IPR013189; Glyco\_hydro\_32\_C.  
 DR InterPro; IPR013148; Glyco\_hydro\_32\_N.  
 DR InterPro; IPR023296; Glyco\_hydro\_beta-prop\_sf.  
 DR Pfam; PF08244; Glyco\_hydro\_32C; 1.  
 DR Pfam; PF00251; Glyco\_hydro\_32N; 1.  
 DR SMART; SM00640; Glyco\_32; 1.  
 DR SUPFAM; SSF49899; SSF49899; 1.  
 DR SUPFAM; SSF75005; SSF75005; 1.  
 DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F32; 1.  
 PE 3: Inferred from homology;  
 KW Glycosidase {ECO:0000256|RuleBase:RU362110};  
 KW Hydrolase {ECO:0000256|RuleBase:RU362110};  
 KW Signal {ECO:0000256|SAM:SignalP}.  
 FT SIGNAL 1 19 {ECO:0000256|SAM:SignalP}.  
 FT CHAIN 20 532 {ECO:0000256|SAM:SignalP}.  
 FT /FTID=PRO\_5003706312.  
 FT DOMAIN 32 336 Glyco\_hydro\_32N.  
 FT {ECO:0000259|Pfam:PF00251}.  
 FT DOMAIN 372 507 Glyco\_hydro\_32C.  
 FT {ECO:0000259|Pfam:PF08244}.  
 SQ SEQUENCE 532 AA; 60661 MW; 96D9F8D5518BA977 CRC64;  
 MLLQAFLL AGFAAKISAS MTNETSDRPL VHFTPNKGWM NDPNGLWYDE KDAKWHLYFQ  
 YNPNDTVWGT PLFWGHATSD DLTHWEDEPI AIAPKRND SG AFSGSMVVDY NNTSGFFNDT  
 IDPRQRCVAI WYNTPESEE QYISYSLDGG YTFTEYQKNP VLAANSTQFR DPKVFWYEPS  
 QKWIMTAASK QDYKIEIYSS DDLKSWKLES AFANEGFLGY QYECPLIEV PTEQDPSKSY  
 WVMFISINPG APAGGSFNQY FVGSFNGTHF EAFDNQSRVV DFGKDYALQ TFFNTDPTYG  
 SALGIAWASN WEYSAFVPTN PWRSSMSLVR KFSLNTEYQA NPETELINLK AEPILNISNA  
 GPWSRFATNT TLTKANSYNV DLSNSTGTLE FELVYAVNTT QTISKSVFPD LSLWFKGLED  
 PEEYLRMGFE ASASSFFLDR GNSKVKFVKE NPYFTNRMSV NNQPFKSEND LSYYKVYGLL  
 DQNIILELYFN DGDVYSTNTY FMTTGNALGS VNMTTGVDNL FYIDKFQVRE VR

## Quelques formats de données biologiques

- ✓ Format des bases, exemples :
  - Séquences ADN/ARN : EMBL, GenBank et DDBJ
  - Séquences protéiques : Uniprot, SwissProt et TrEMBL, PIR...
- ✓ Formats de certains logiciels : PHYLIP (*PHYLogeny Inference Package*), FOSN (*Files Of Sequence Names*), RSF (*Rich Sequence Format files*), RSF (*Rich Sequence Format files*), MSF (*Multiple Sequence Format*), Fitch, DNA strider, AnTheProt, Olsen, etc.
- ✓ Formats lus par la plupart des outils en bioinformatique
  - FASTA
  - Séquence brute (*plain/raw sequence*)

## Le format FASTA

- ✓ Une ligne de commentaires précédé de « > »
- ✓ La séquence brute (pas d'espace, ni de nombre)

```
>Human Polycomb 2 homolog (hPc2) mRNA, partial cds  
ctccggcagcccgagggtcatcctgctagactcagacctggatgaacccat  
agacttgcgctcgggtcaagagccgcagcgaggccggggagccgcccagct  
ccctccagggtgaagcccgagacaccgggcgtcggcgggcgggtggcggtggcg  
gcggcagcgggcaccacccacgacggcgggagagaagcctccagccgaggccca  
ggacgaacctgcagagtcgctgagcgagttcaagcccttctttgggaata  
taattatcacccgacgtcacccgcgaactgcctcacccgttactttcaaggag  
tacgtgacgggtg
```



Exemple d'une entrée sur Genbank

## LES BASES DE DONNÉES BIOLOGIQUES

## ACTIVITÉ 01

Explorez une base de données biologique figurant sur la liste 2019 de la revue *Nucleic Acids Research* :

- Allez sur l'adresse <http://www.oxfordjournals.org/nar/database/c> ;
- Choisissez 01 base de données biologique ;
- Explorez la base et complétez le tableau ci-dessous ;
- Faites une recherche sur la base et copiez le contenu d'une entrée dans le tableau ;
- Travail à faire en groupe (2 à 4 étudiants) ;
- Envoyez le fichier word par mail avant le 25/02/2020 sur l'adresse [gomrima@umc.edu.dz](mailto:gomrima@umc.edu.dz) .

Nom de la base de données	
Adresse URL	
Brève présentation de la banque	
Nature des données	
Origine des données	
Exemple d'une entrée	
Format de l'entrée	