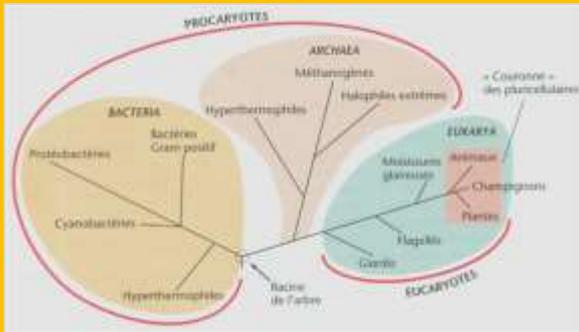
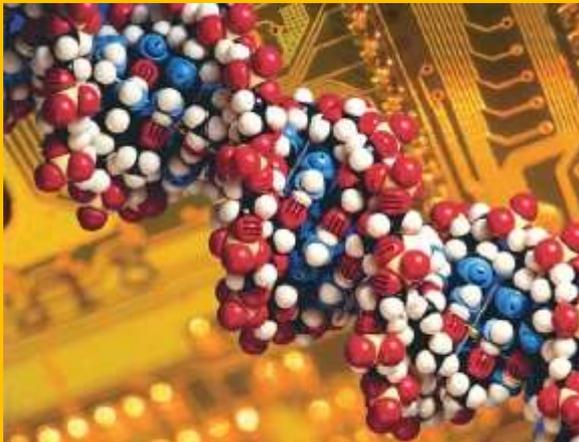


Université Frères Mentouri Constantine 1
Institut de la Nutrition, de l'Alimentation et des Technologies Agro-alimentaires (INATAA)
1^e année Master Biotechnologie alimentaire
2019-2020



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:

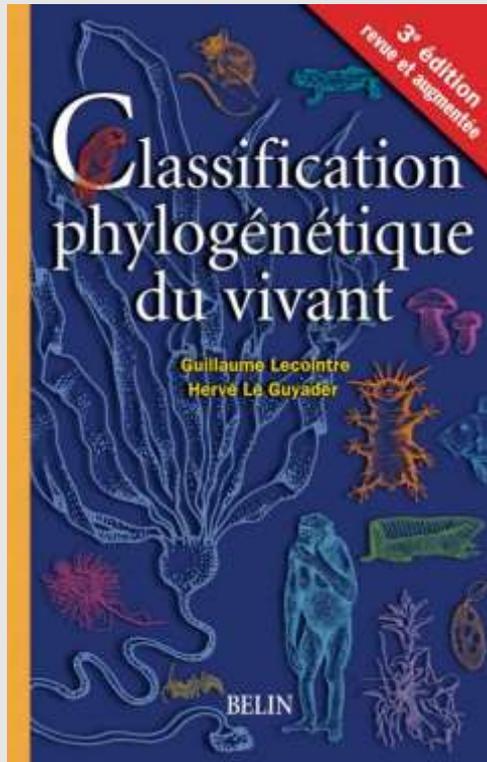
A la fin de ce cours, l'étudiant sera capable de :

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

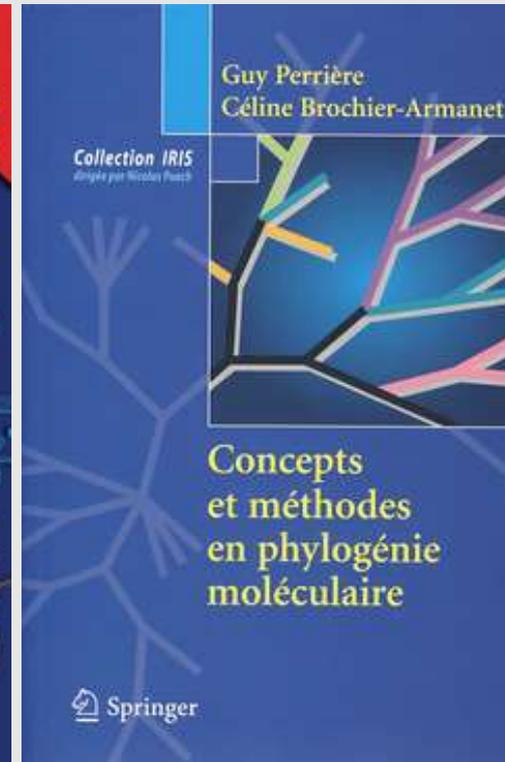
LIVRES A CONSULTER



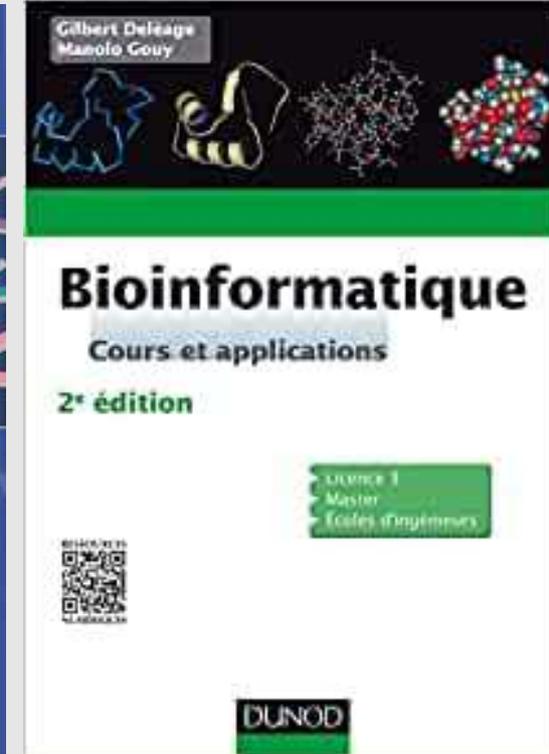
**NA/8544 ET NA/8546
NA/5047 À NA/5049
(CD ROM)**



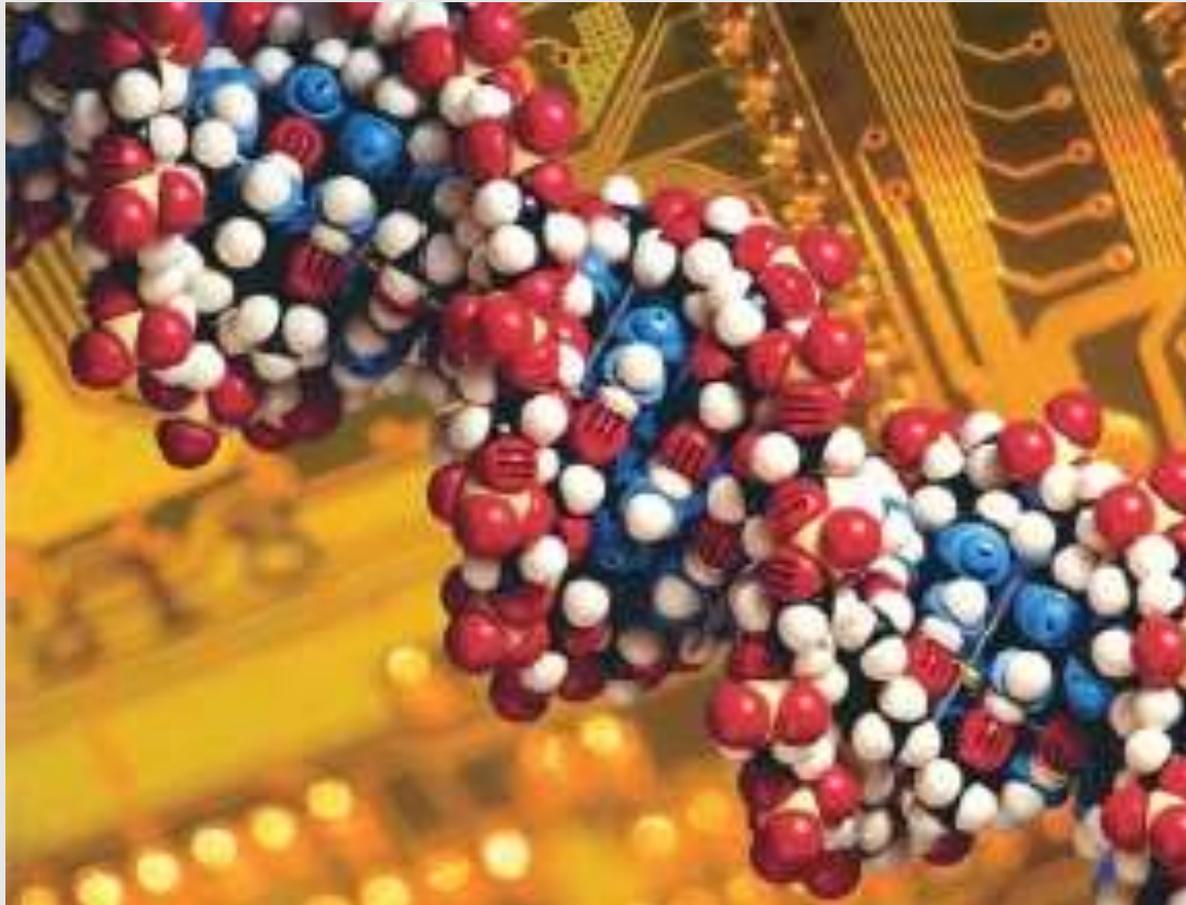
**NA/8220/1 À
NA/8226/1**



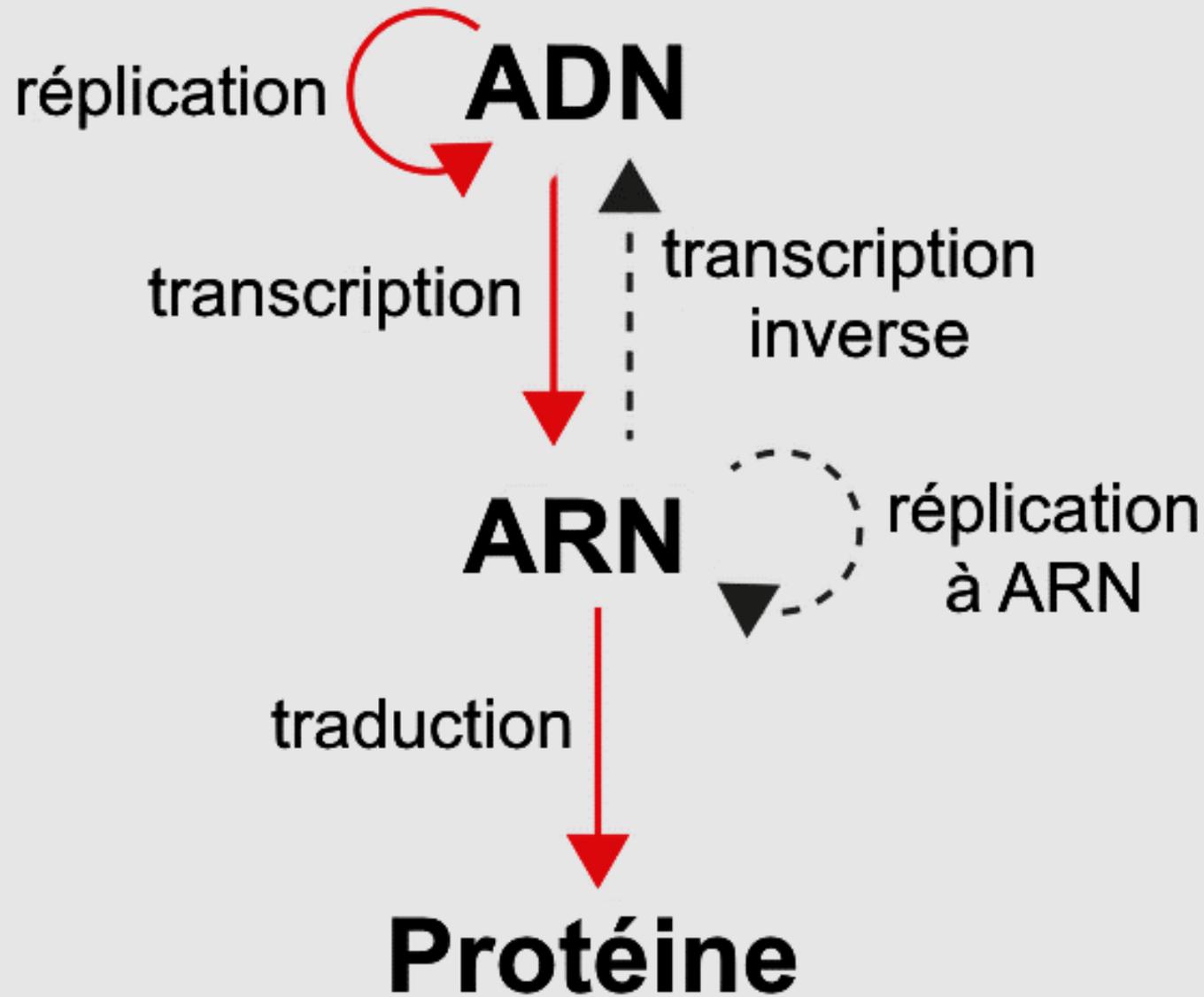
**NA/7836
À
NA/7839**



INTRODUCTION À LA BIOINFORMATIQUE ET À SES APPLICATIONS

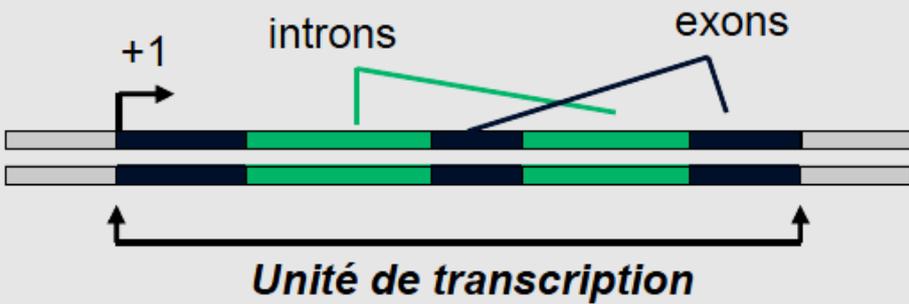


DOGME CENTRAL DE LA BIOLOGIE MOLÉCULAIRE



DOGME CENTRAL DE LA BIOLOGIE MOLÉCULAIRE

Eucaryote



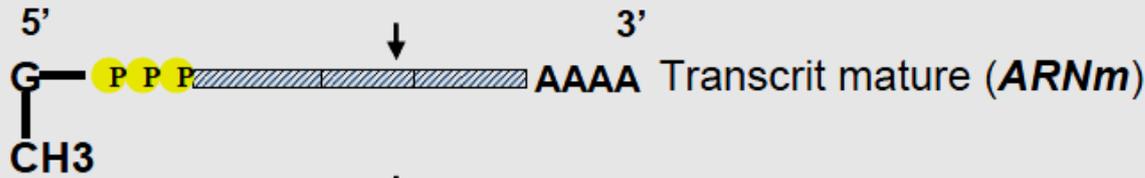
ADN

Procaryote



Transcrit primaire (ARNm)

Transcrit (ARNm)



Protéine



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.

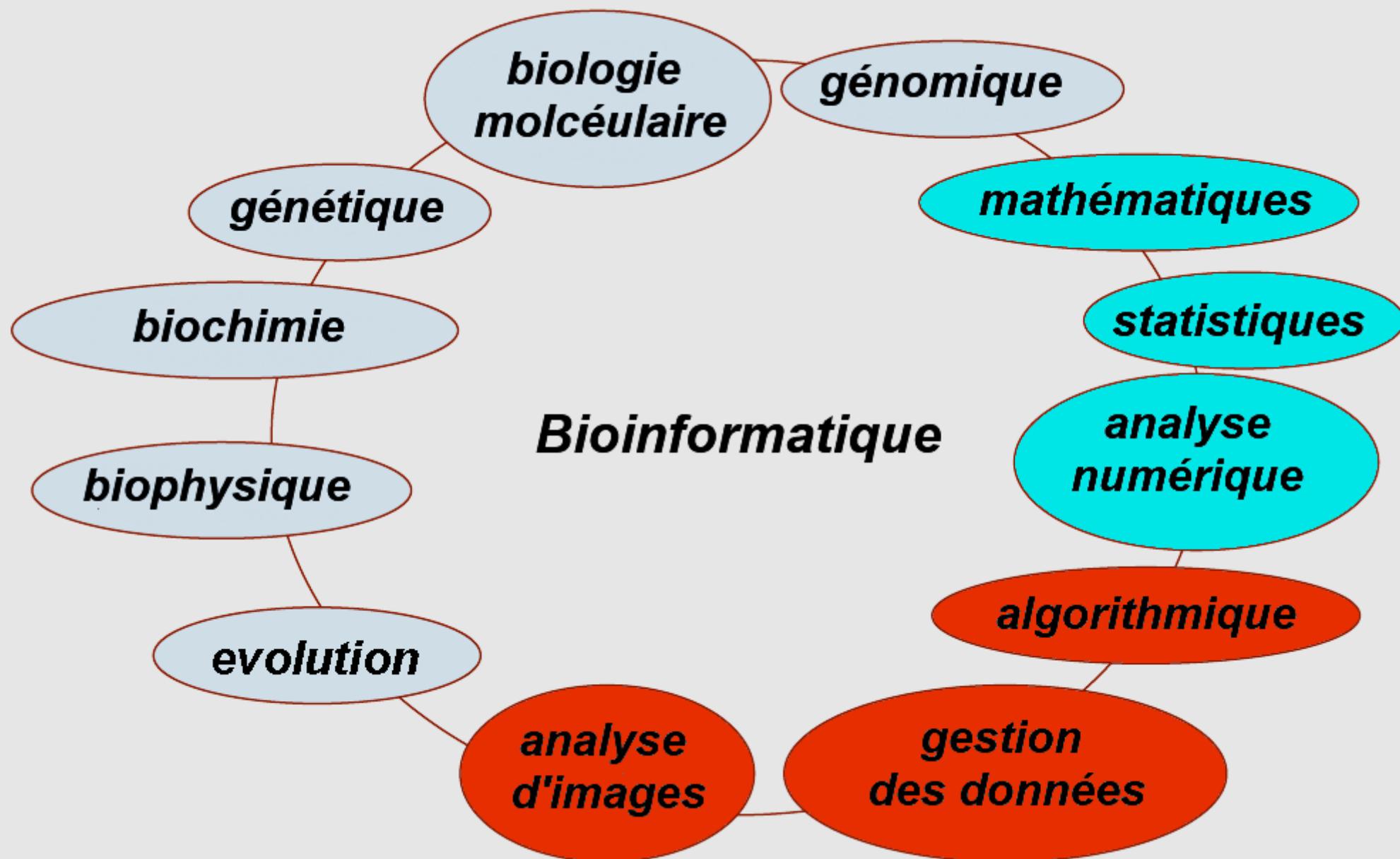
L'INFORMATION BIOLOGIQUE

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).

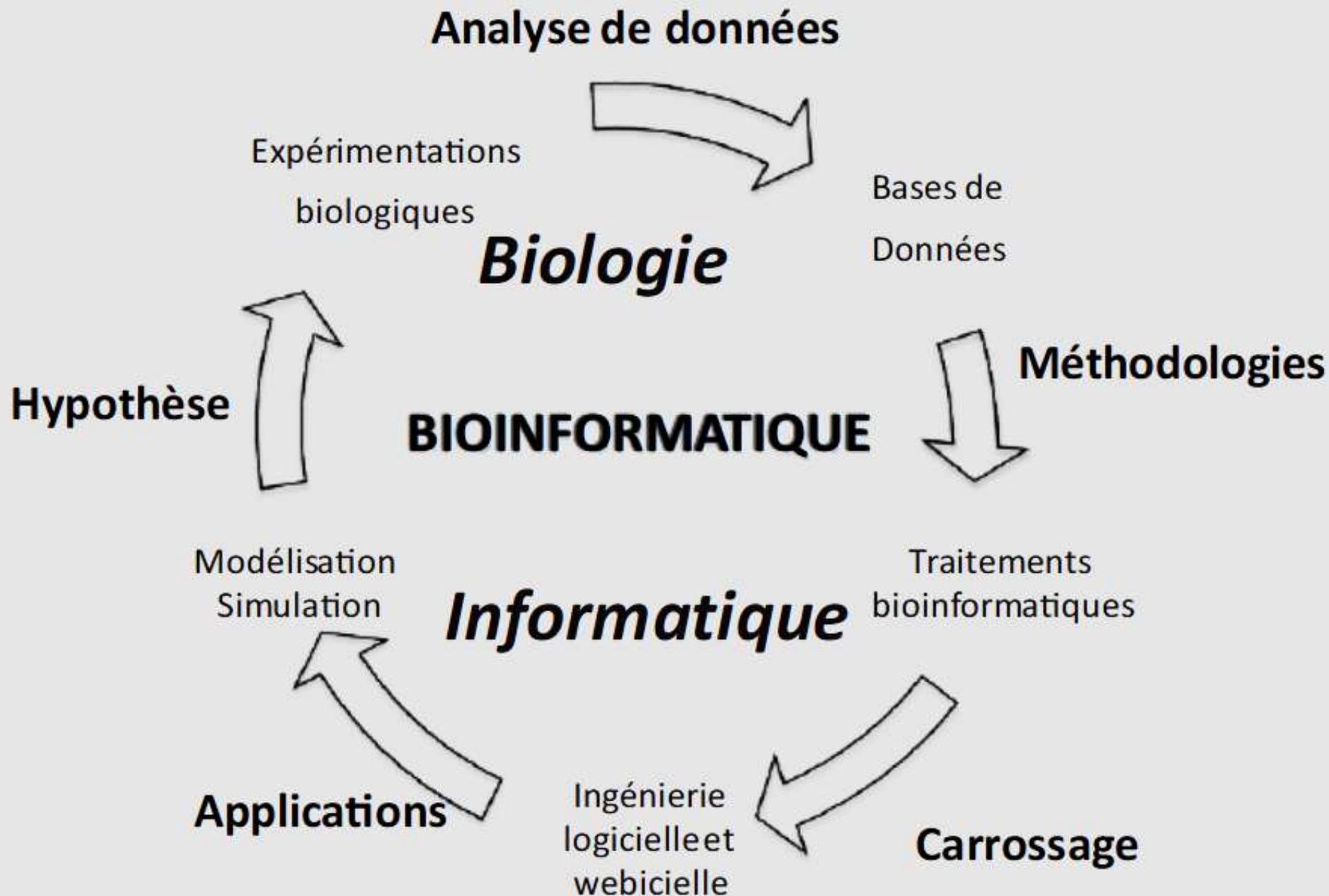
DISCIPLINES DE LA BIOINFORMATIQUE

La bioinformatique est un champ d'étude multidisciplinaire:

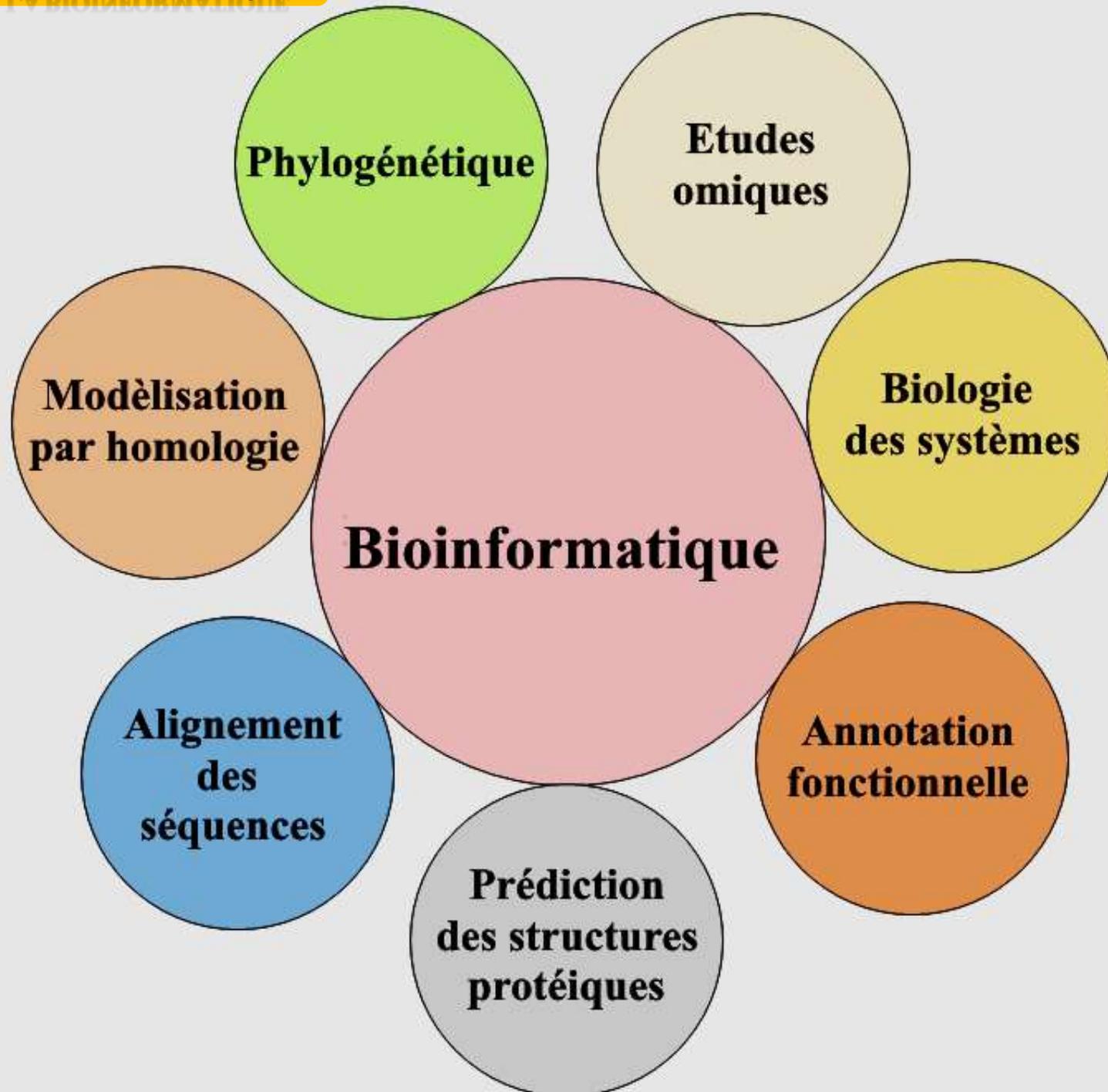


DISCIPLINES DE LA BIOINFORMATIQUE

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:

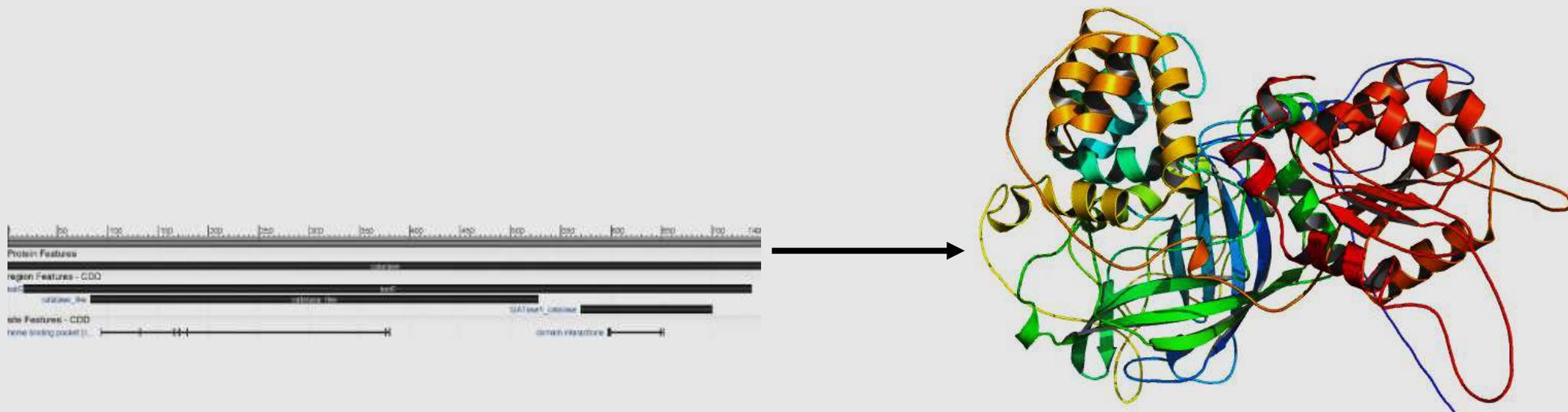
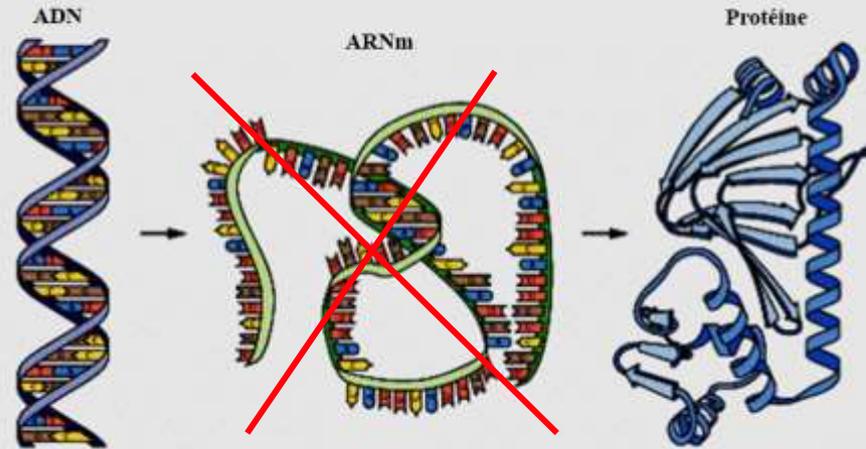
APPLICATIONS DE LA BIOINFORMATIQUE

- ❑ 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

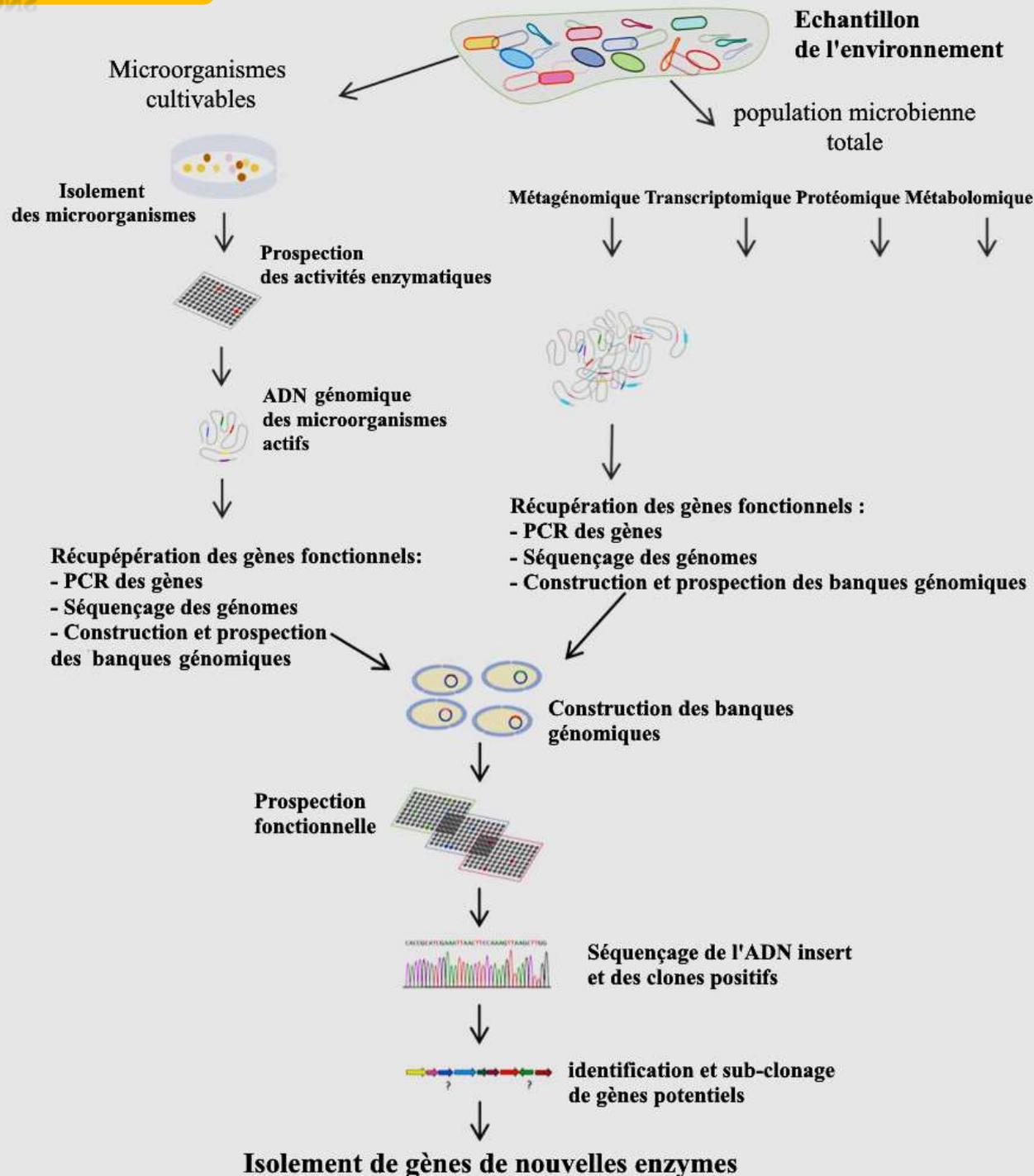
ADN $\xrightarrow{\text{Dédution}}$ Protéine



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

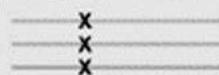


Transformation and expression of designed enzyme



Design semi-rationnel

Construction of targeted gene library



Cloning of targeted library

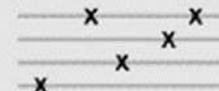


Transformation and library expression



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

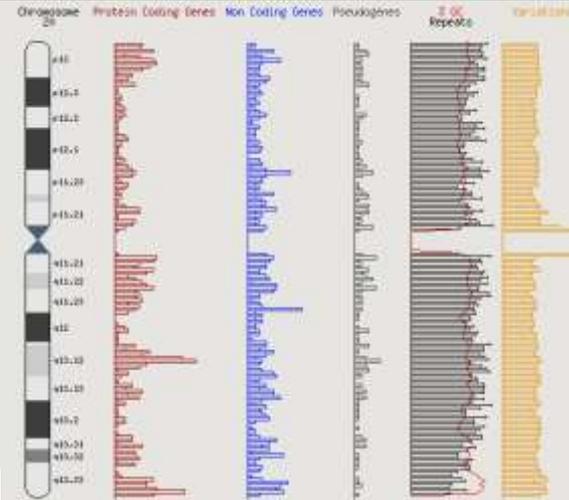
ANNOTATION DES GÉNOMES



Vue schématique du caryotype humain

EXEMPLES D'APPLICATIONS

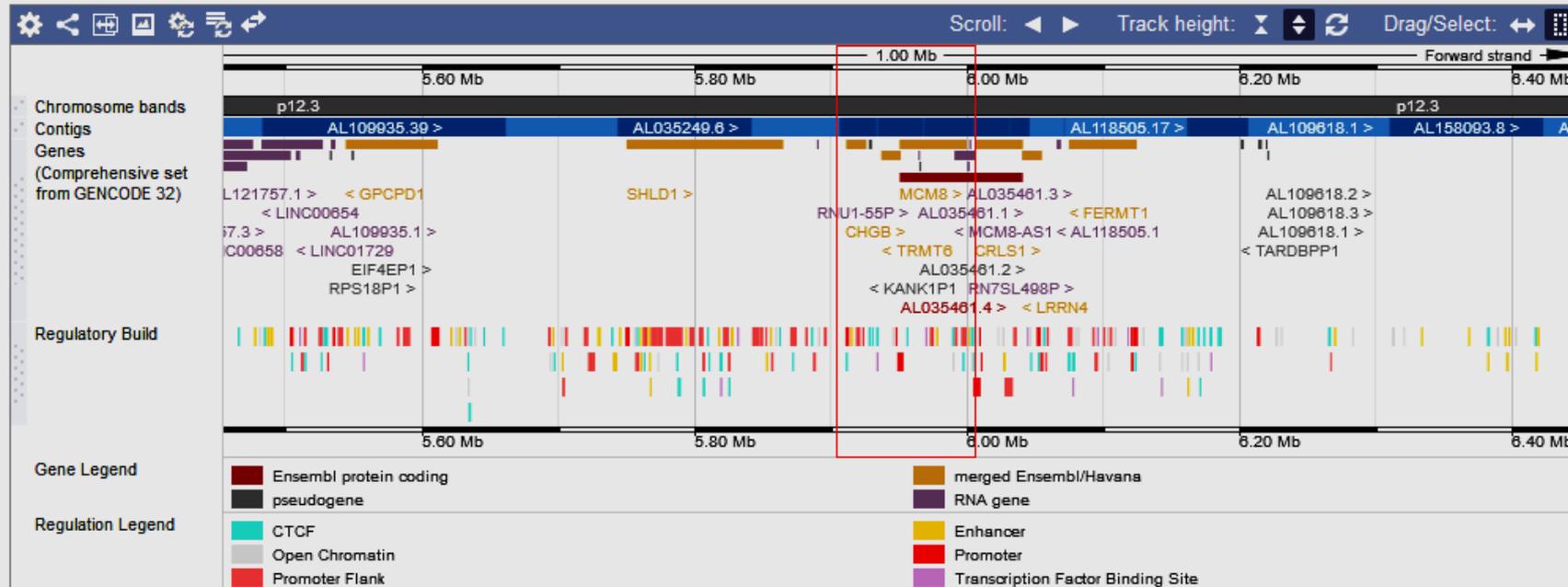
ANNOTATION DES GÉNOMES



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



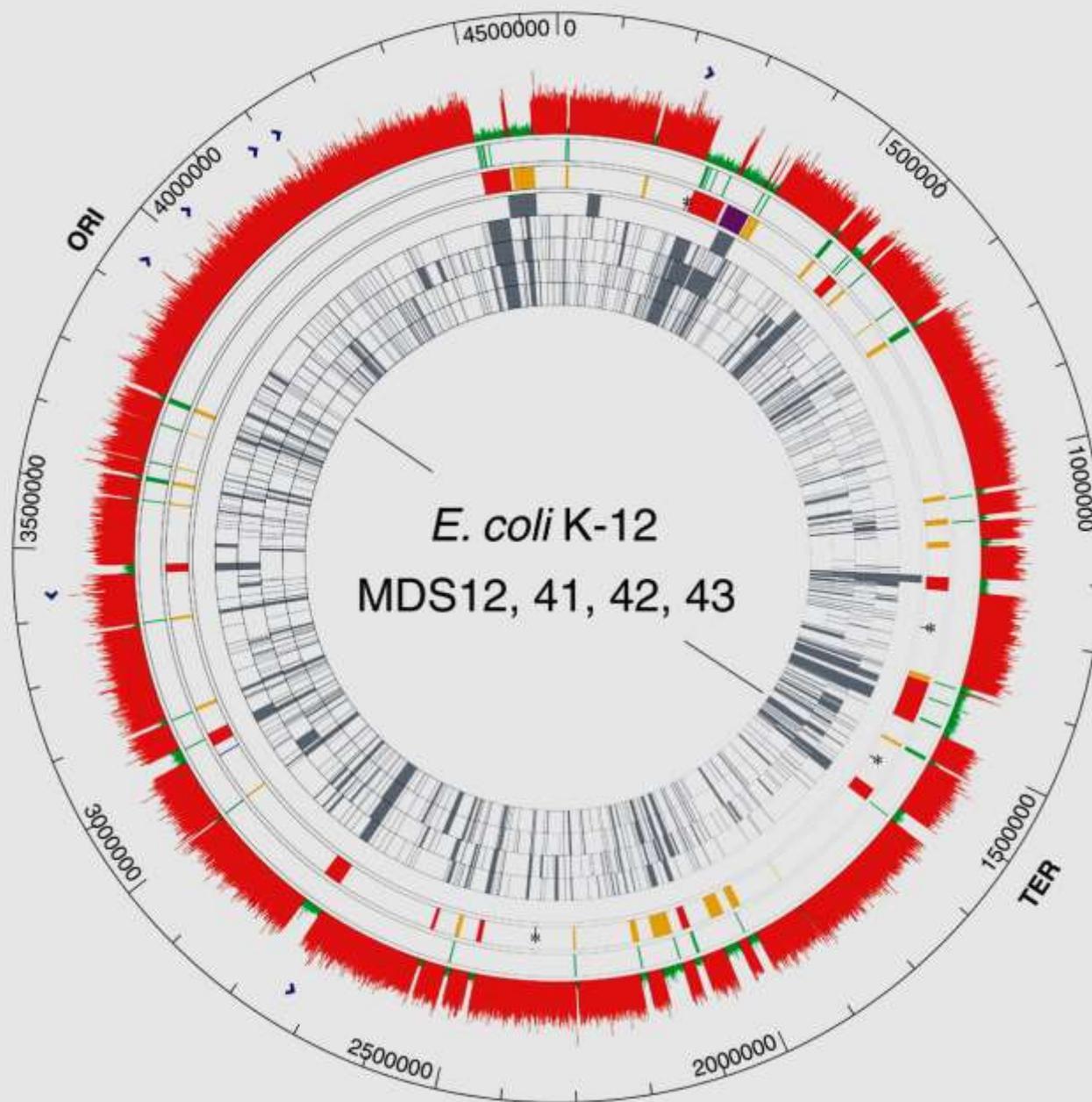
Region in detail ?



Vue schématique du chromosome humain numéro 22

EXEMPLES D'APPLICATIONS

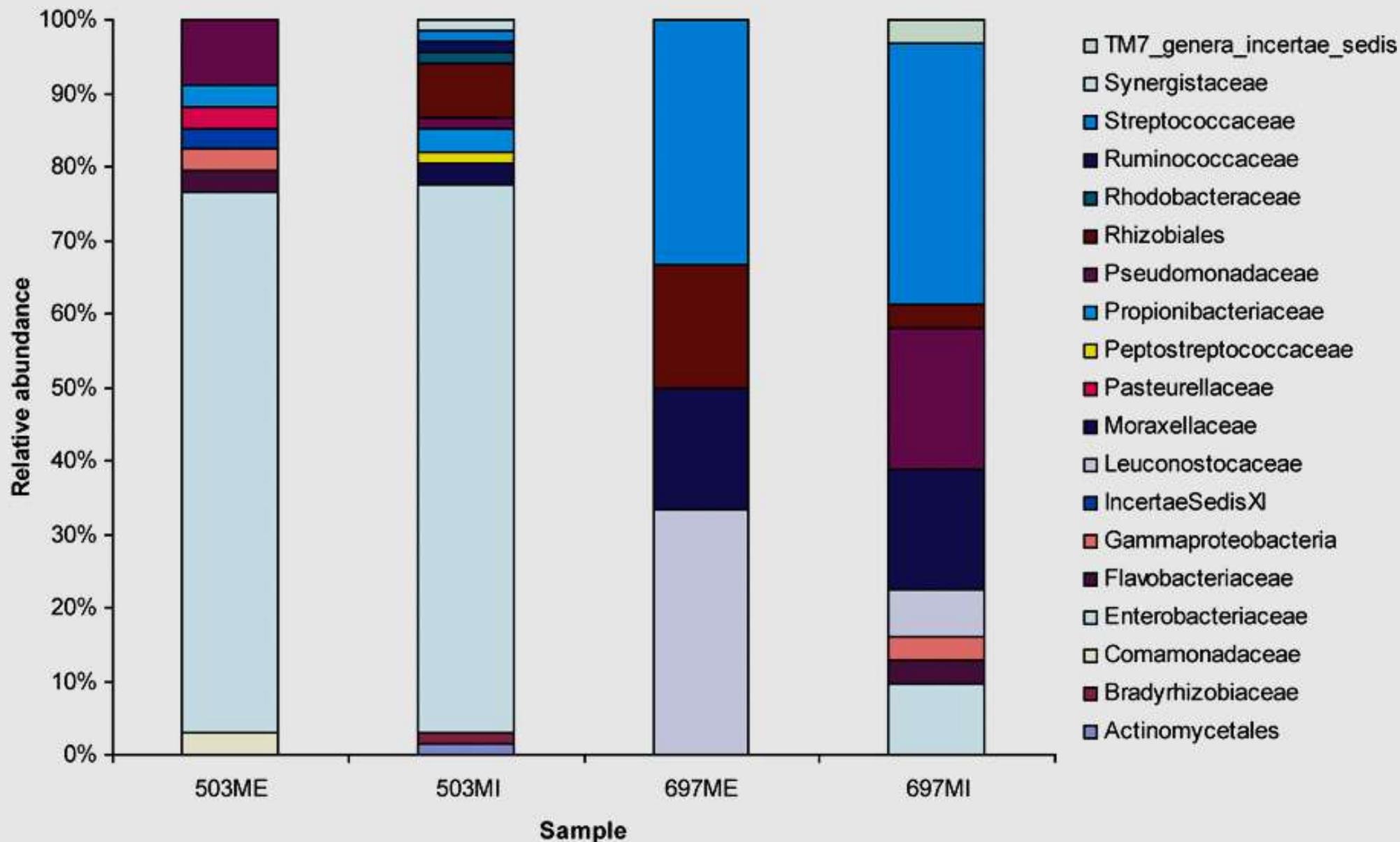
ANNOTATION DES GÉNOMES



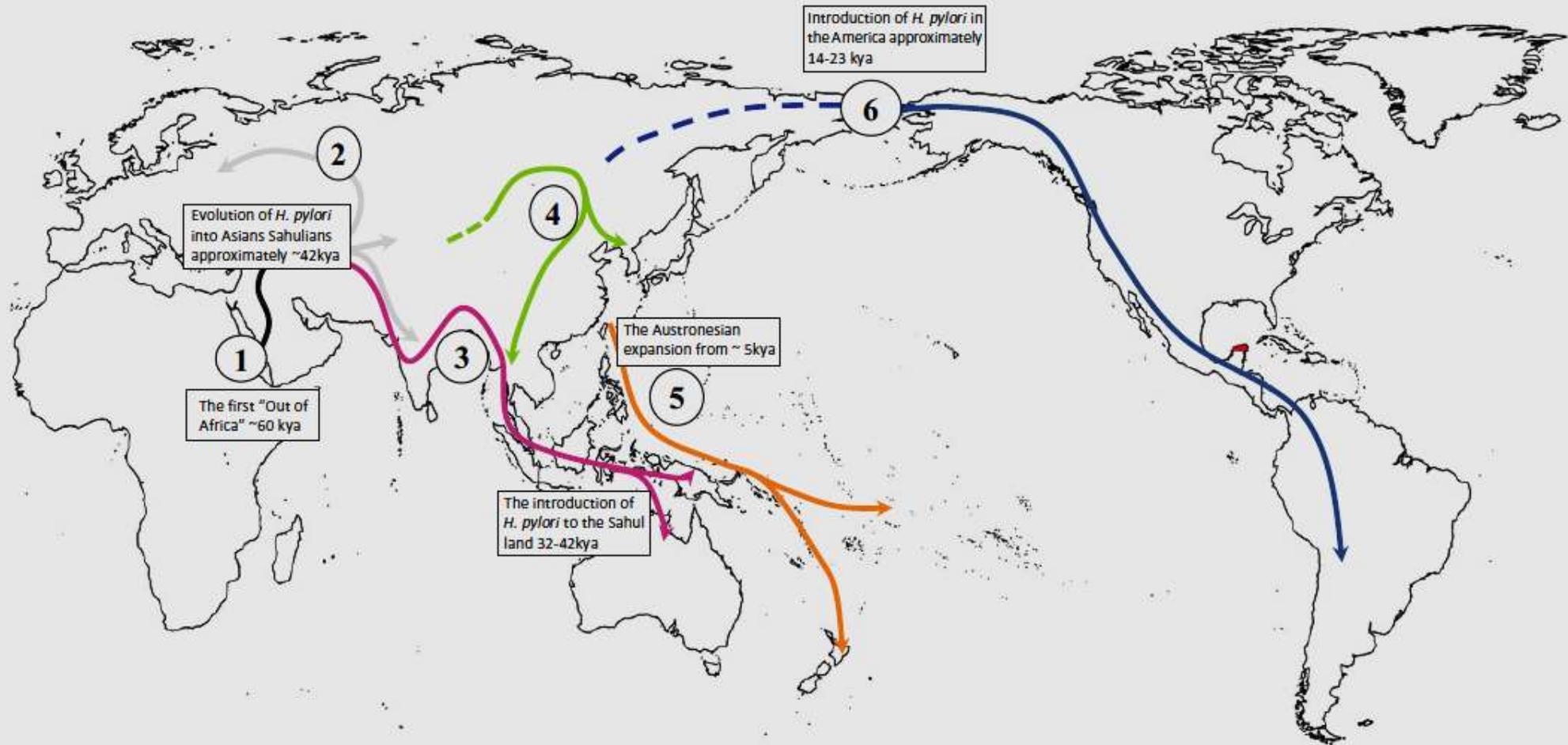
Vue schématique du génome d' *E. coli* K-12

EXEMPLES D'APPLICATIONS

ANALYSE DES SÉQUENCES NGS)



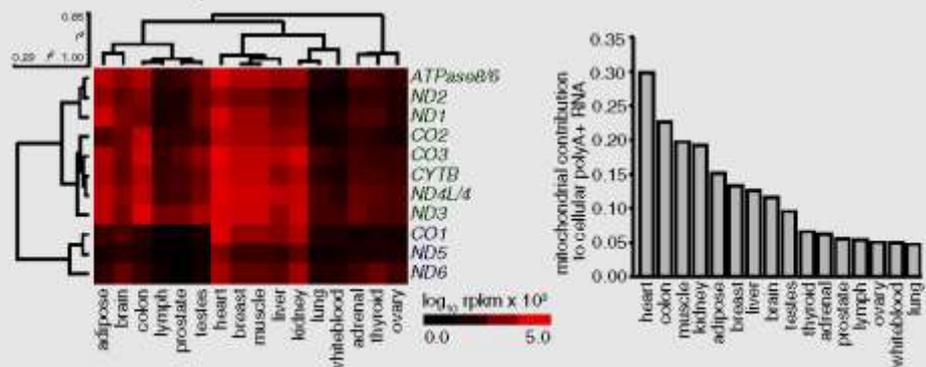
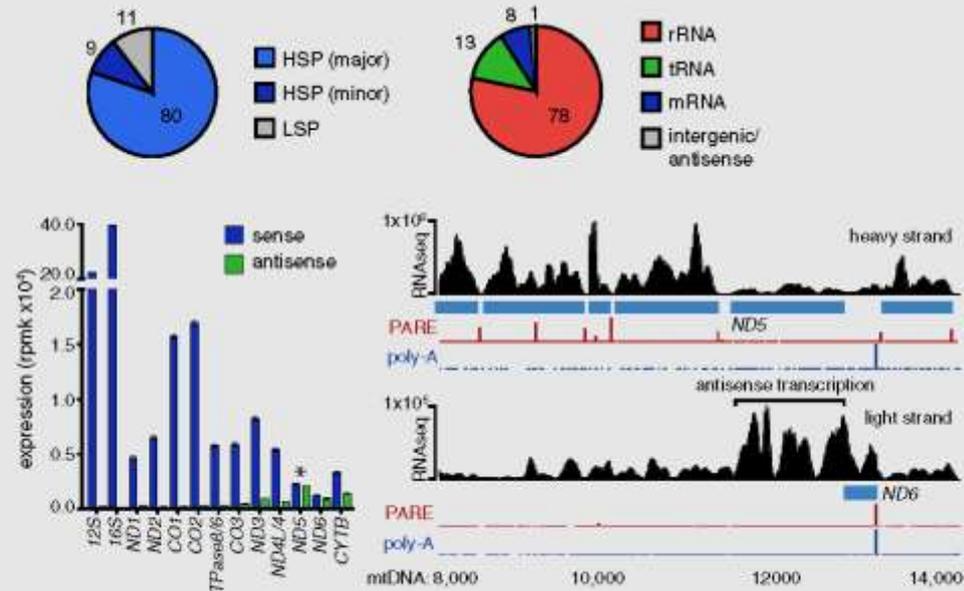
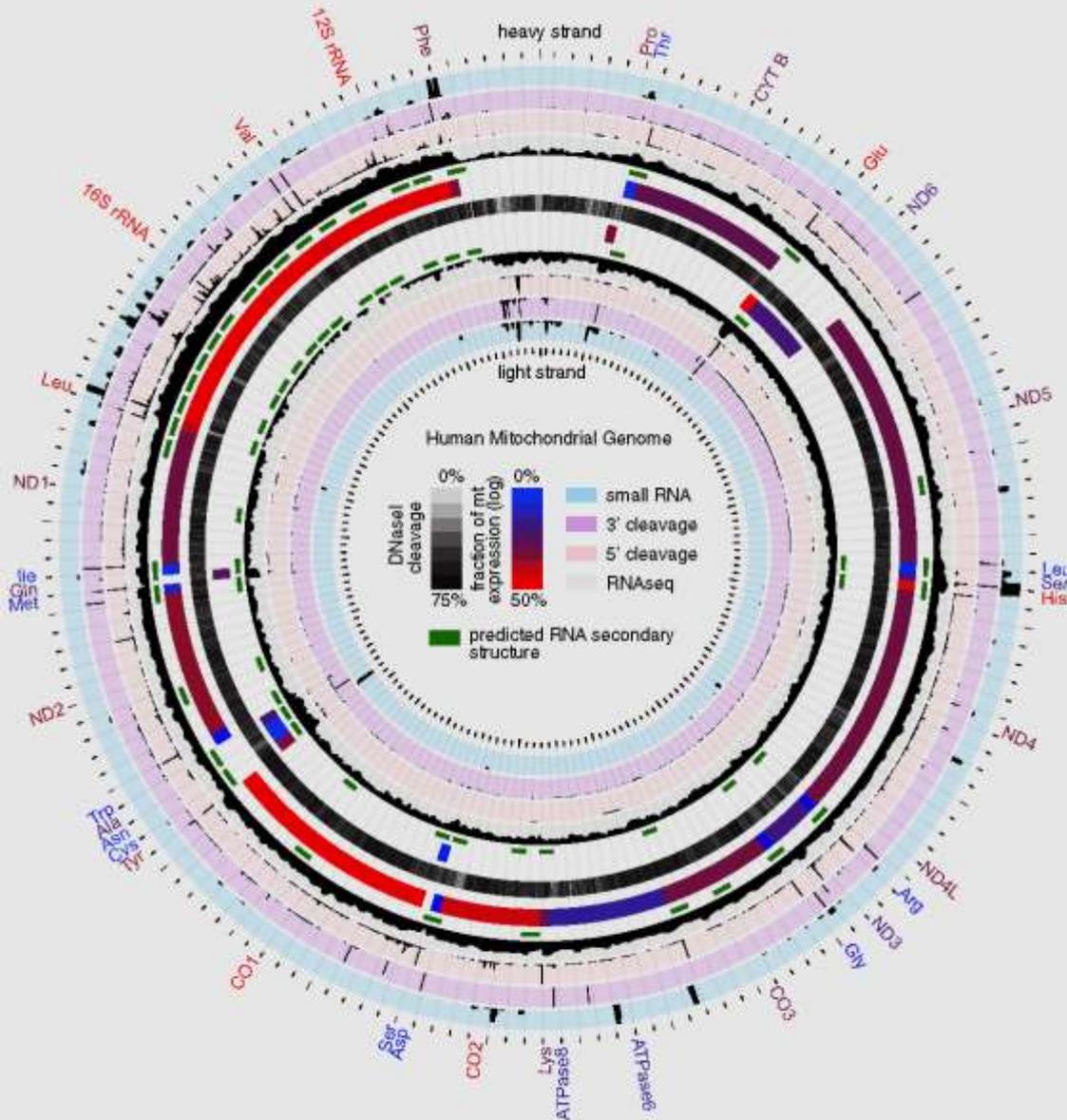
Abondance de la flore microbienne intestinale dans un échantillon de matières fécales infantiles



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

EXEMPLES D'APPLICATIONS

TRANSCRIPTOMIQUE



Carte du transcriptome du génome mitochondrial humain

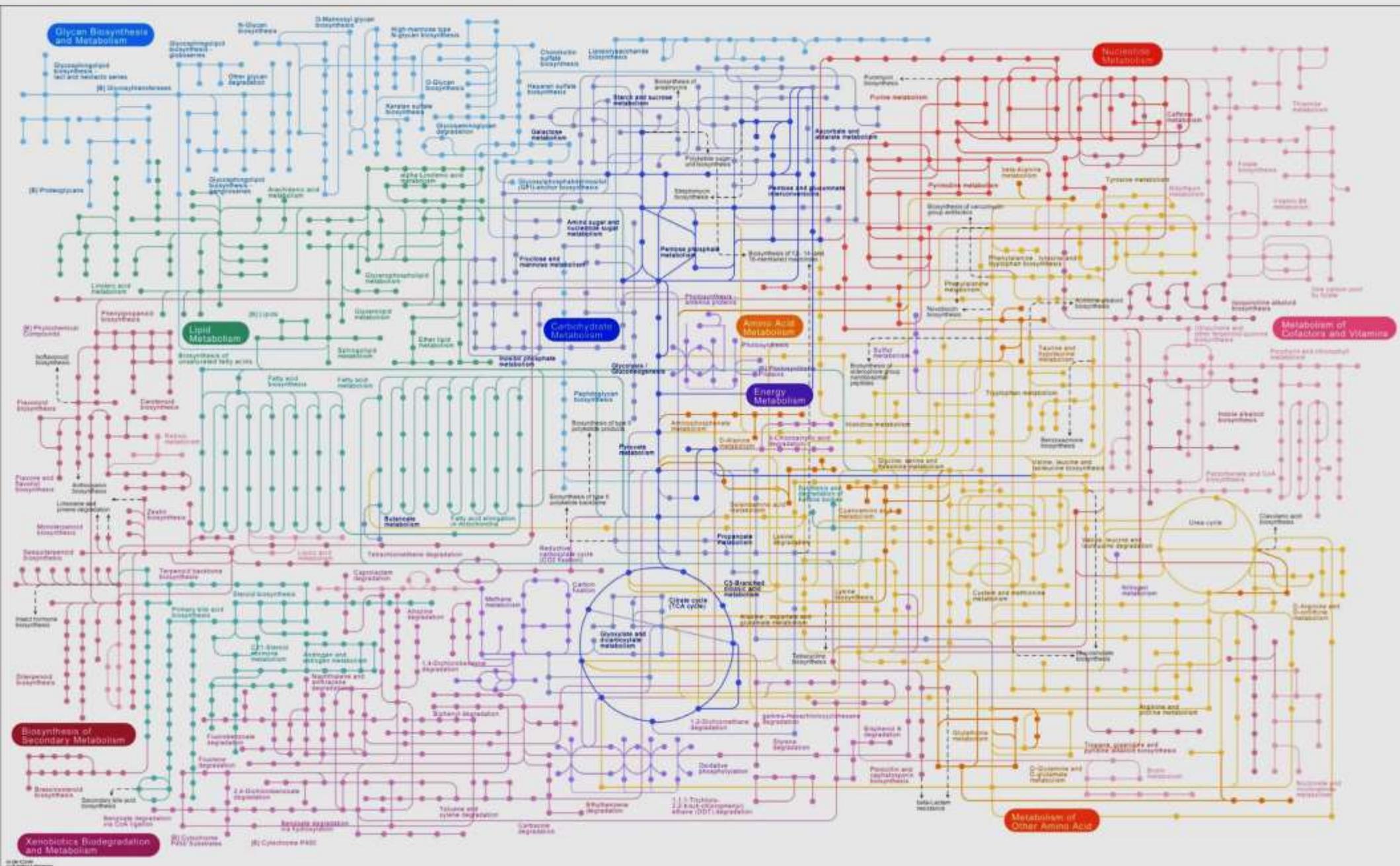


Schéma général du réseau métabolique humain

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

ExPASy Proteomics Server

Search: UniProtKB [Go] [Clear]

Databases Tools Services Mirrors About Contact

You are here: ExPASy CH > Databases > Around UniProtKB

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

[UniProt web site](#)

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 accompanying [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 Cochrane](#) as part of the [The Protein and Nucleotide Database Group \(PANDA\)](#) under [Ewan Birney](#).

Link	Explanation
Access	Database queries, Contacted genomes, webserver, FTP archives (EMBL release, alignments etc), EMBL resource version archive (SVA), Browse by geography
Submission	Primary sequence submissions, third party annotation, updates.
Documentation	Release notes , user manual , Information for Submitters , FAQ , Release information , Footprints Changes , EMBL database statistics , Feature lists , EML documentation , Search entry , Accession Number Prefix Codes , Examples of annotation , EMBL Features & Qualifiers , ORF line standards , Database Policies
Publications	Group publications
People	Group members
Contact	How to contact the EMBL Nucleotide Sequence Database
News	List of recent changes on this site

DDBJ
DNA Data Bank of Japan

Accession: DNA, Protein, AIDs, Taxonomy, Site Search
 Accession numbers [Go]
 DDBJ UniProt PDB DAD PRF Patent [History](#)

HOME Submission How to Use Search/Analysis FTP/WebAPI Report/Statistics Contact Us RSS Japanese

► About DDBJ
 ► How to Use
 ► Q and A

Sequence Submission

- [BAKURA](#)
- [Mass Submission](#)
- [Data Uploads](#)
- [DDBJ Read Archive](#)
- [DDBJ Trace Archive](#)

Search

- [Identity](#)
- [ARSA](#)
- [TxSearch](#)
- [BLAST](#)
- [PR-BLAST](#)
- [FASTA](#)
- [SSEARCH](#)

Phylogenetics

- [ClustalW](#)

Hot Topics

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

Maintenance

- Apr. 21, 2010 [Suspension of some DDBJ activities in Japanese holidays\(A/29.5/1-5\)](#)
- Mar. 16, 2010 [\(Apr. 23\)ARSA database search \(DDBJ, DAD\) temporary unavailable](#)
- Feb. 03, 2010 [\(Important\) Termination of a part of DDBJ 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](#)

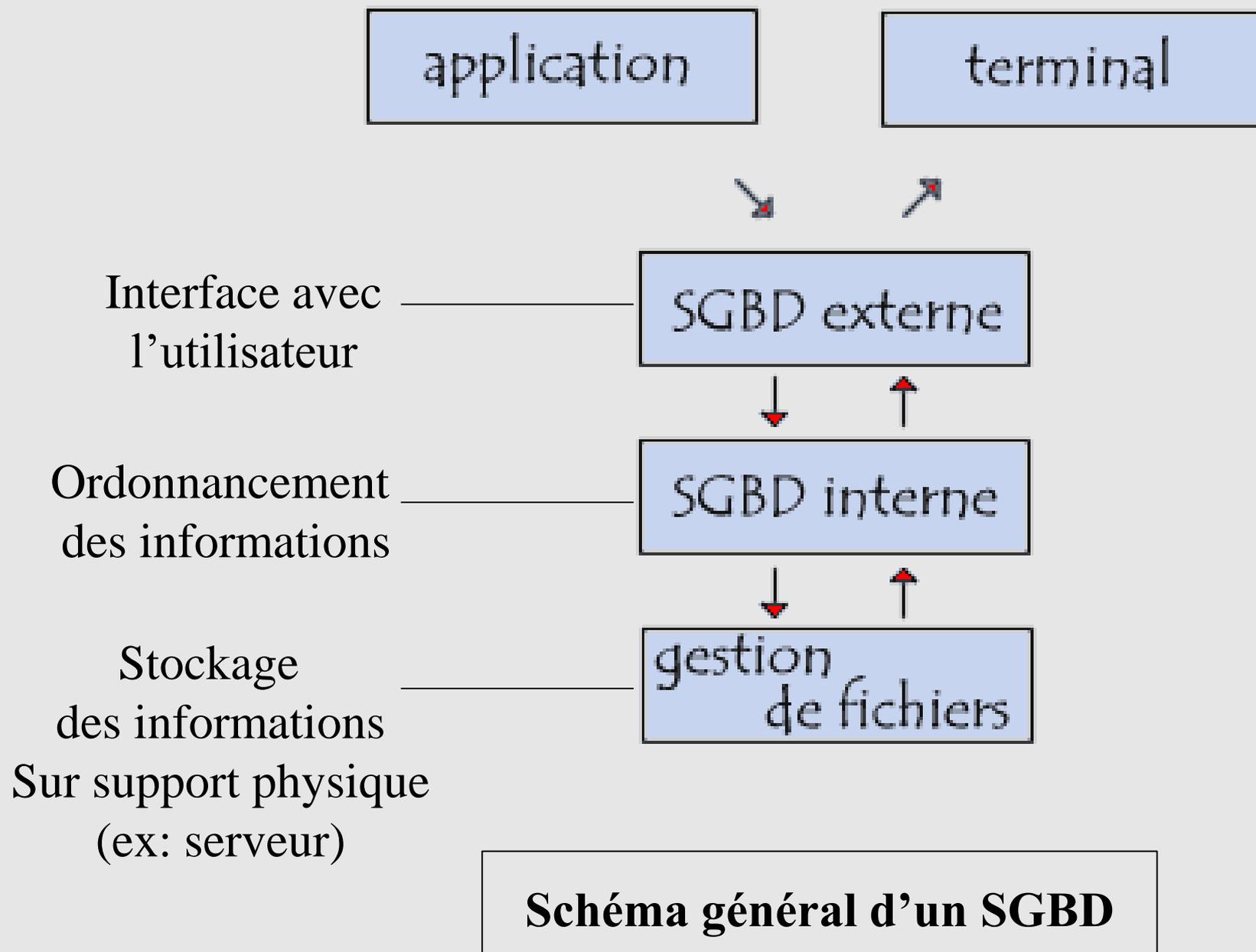
NOTION DE BASE DE DONNÉES

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 (*databank*)** 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

The screenshot displays the Microsoft Access application window. The ribbon at the top includes tabs for File, Home, Create, External Data, Database Tools, Design, Arrange, and Format. The 'Design' tab is active, showing various controls and tools. The main window shows a form titled 'Contact Details' with a 'Main' tab selected. The form contains the following data:

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

The left-hand pane shows 'All Access Objects' with a search bar and categories: Queries (ContactsExtended), Forms (ContactCard, ContactDetails, ContactDS, ContactList, Contacts, DefaultImage, GettingStarted, Main, NameCard, NameList, ReportCenter, ReportCenterDe..., VideoPlayer), and Reports (Comments, ContactAddress..., ContactDetails, ContactList, ContactPhoneB..., Label). The status bar at the bottom indicates 'Record: 1 of 29' and 'No Filter'.

EXEMPLES DE BASES ET DE DONNÉES

Issue Tracking Application - Microsoft Access

Data Insert Page Layout External Data Advanced Tools

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

All Access Objects

Issues by Status Chart Contacts Customers

Customer Name	Product Name	Customer Conta	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

The Google logo is displayed in its characteristic multi-colored font: 'G' in blue, 'o' in red, 'o' in yellow, 'g' in blue, 'l' in green, and 'e' in red.

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

The image shows a screenshot of the IMDb website. At the top, the IMDb logo is on the left, followed by a menu icon and the text 'Menu'. A search bar contains the text 'All' and 'Search IMDb'. To the right of the search bar are links for 'IMDbPro', 'Watchlist', and 'Sign In'. Below the navigation bar, there are two main content areas. The left area is a 'MISSING SNAIL!' poster for the movie 'SpongeBob SquarePants: The Movie'. The poster features a pink snail on a green leaf and text that reads: 'LAST SEEN IN BIKINI BOTTOM', 'HE ANSWERS TO MEOW, GARY AND GARE-BEAR', and 'IN THEATRES MAY 2020 #SAVEGARY'. The right area is a video player for 'Snoop Dogg Alert!' with a duration of 0:31. The video player shows a close-up of SpongeBob's face and a play button icon. Below the video player, the text reads 'Watch the New 'SpongeBob' Trailer'.

<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

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.

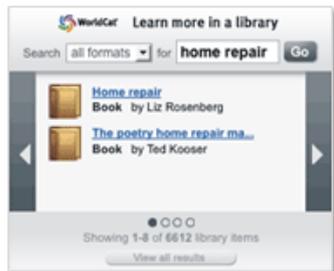
[Mot de passe oublié?](#) | [Créer un compte gratuit](#)

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.



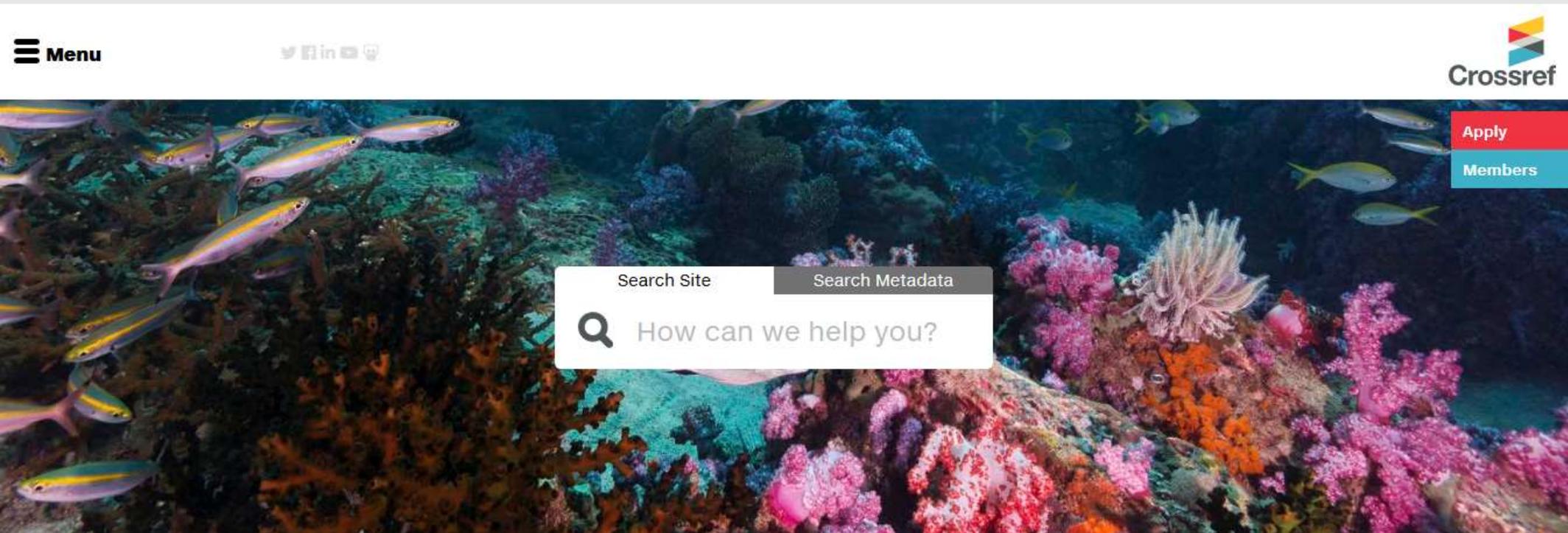
Tutoriel vidéo

[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.

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/>

GenBank

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

GenBank Overview

What is GenBank?

GenBank[®] 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>

EMBL-EBI  Services Research Training About us

 **ENA**
European Nucleotide Archive

Search [Advanced](#) [Sequence](#)

Examples: [BN000065](#), [histone](#)

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](#)

[Advanced search](#)

Sequence Search

Enter or paste a nucleotide sequence or accession number

[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>

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

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

<http://www.uniprot.org>

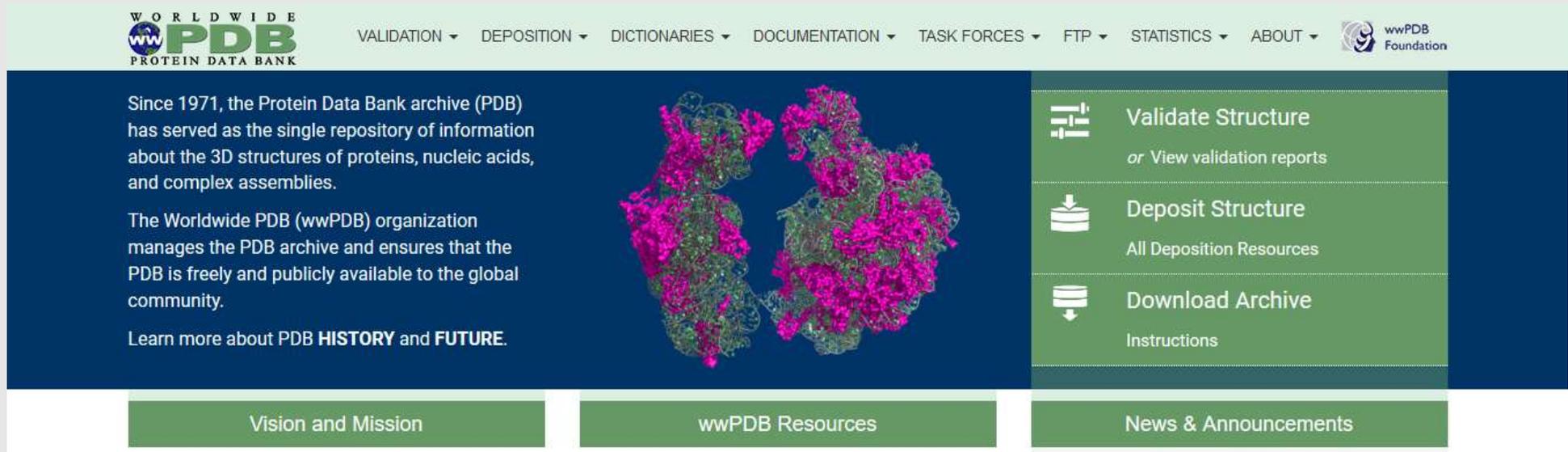
The screenshot shows the UniProt website interface. At the top, there is a navigation bar with the UniProt logo, a search bar containing 'UniProtKB', and a search button. Below the navigation bar, there are links for 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. The main content area features a 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 main database categories: UniProtKB (UniProt Knowledgebase) with 561,568 manually annotated and reviewed records; UniRef (Sequence clusters); UniParc (Sequence archive); and Proteomes (Proteome sets). A 'Supporting data' section includes links for Literature citations, Cross-ref. databases, Taxonomy, Diseases, Subcellular locations, and Keywords. On the right side, there is a 'News' section with social media icons and recent updates like 'UniProt release 2019_11' and 'UniProt release 2019_10'.

*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/>



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.). 6

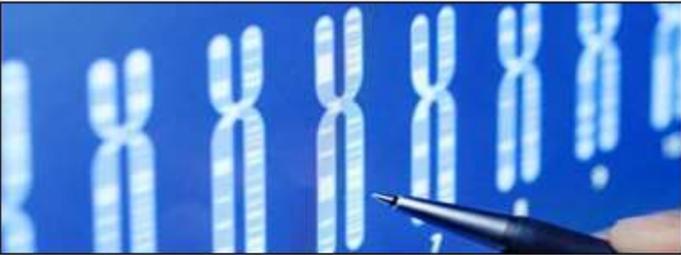
Genome

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 [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/M (Integrated Microbial Genomes and Microbiomes) :

Base de génomes microbiens annotés

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



IMG/M
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.

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](#)

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.

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
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)	CONTACT US
---------------	----------	------------	----------	--	----------------------------

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

FAST-TRACK YOUR PAPER FLYBASE NEWS FLY BOARD
COMMUNITY NEWS MEETINGS COURSES FLYBOOK
Resources for: PUBLIC, TEACHERS, STUDENTS BIOMEDICAL RESEARCH POSITIONS AVAILABLE

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.

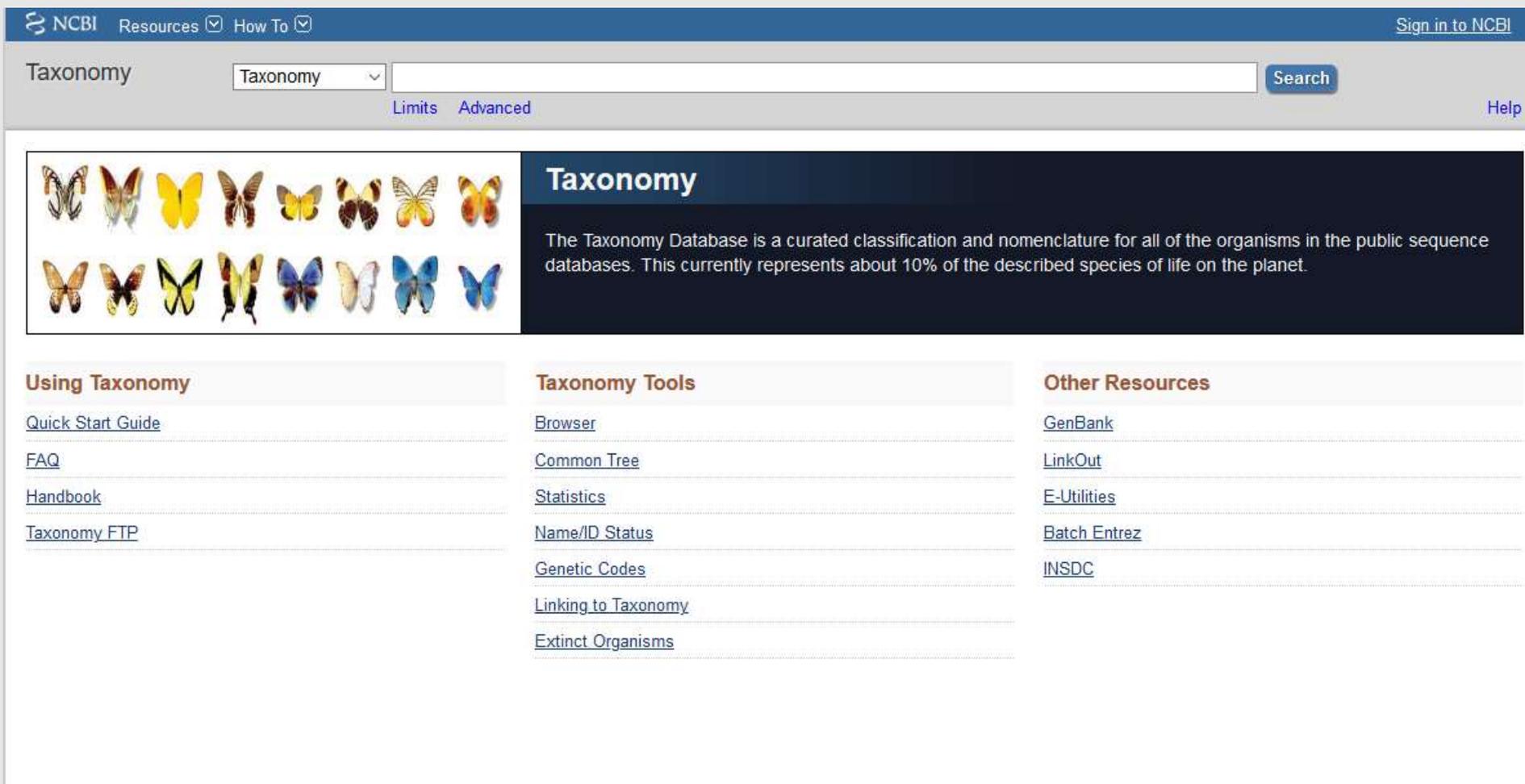
Note: Wild cards (*) can be added to your search term

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>



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](#)

- ✓ **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>

The screenshot shows the PubMed website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' links, and a 'Sign in to NCBI' button. Below this is the 'PubMed' logo and a search bar with a dropdown menu set to 'PubMed' and a 'Search' button. The main content area is divided into three columns:

- Using PubMed:** Includes links for 'PubMed Quick Start Guide', 'Full Text Articles', 'PubMed FAQs', 'PubMed Tutorials', and 'New and Noteworthy'.
- PubMed Tools:** Includes links for 'PubMed Mobile', 'Single Citation Matcher', 'Batch Citation Matcher', 'Clinical Queries', and 'Topic-Specific Queries'.
- More Resources:** Includes links for 'MeSH Database', 'Journals in NCBI Databases', 'Clinical Trials', 'E-Utilities (API)', and 'LinkOut'.

Below these columns are three sections:

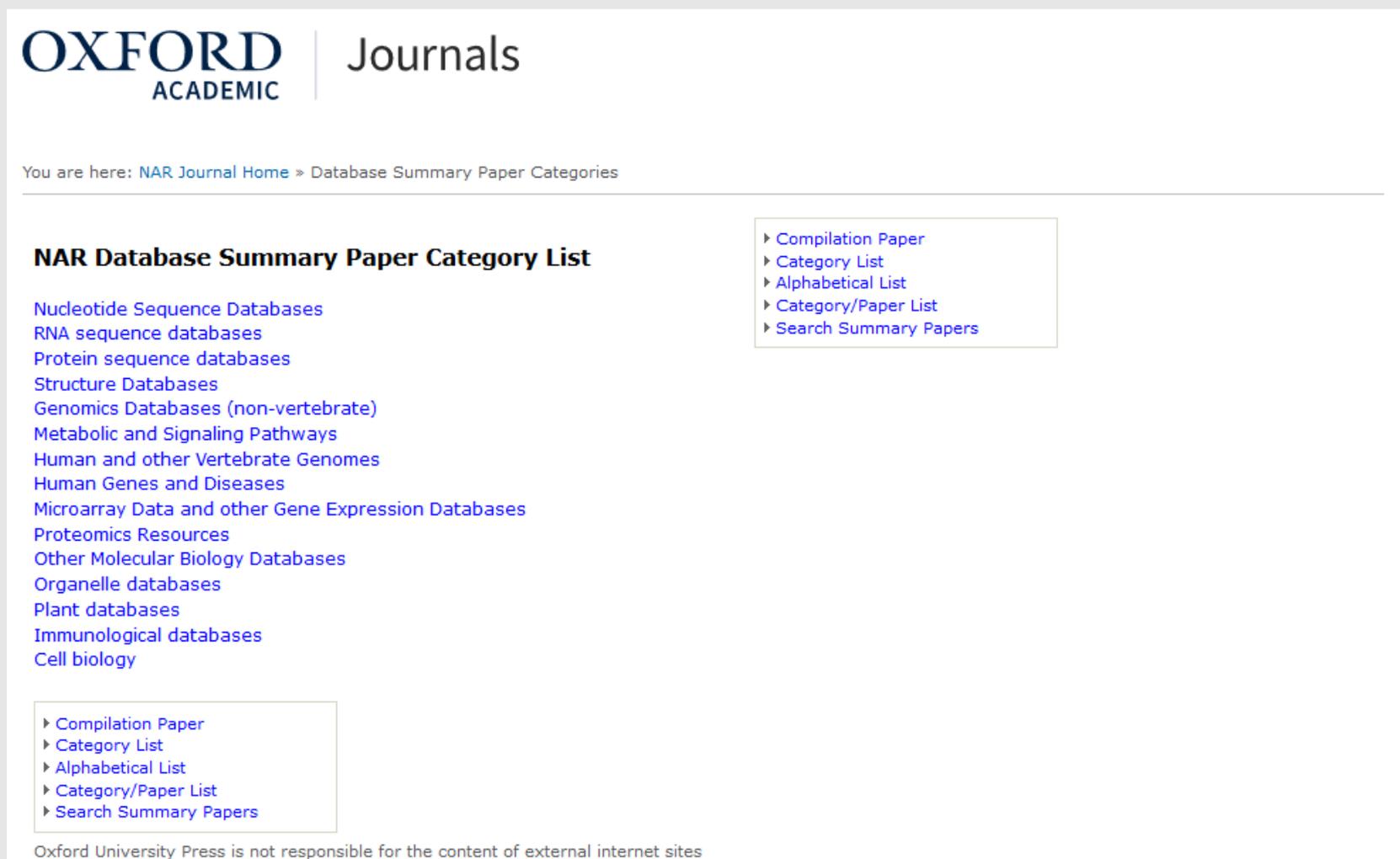
- Latest Literature:** Lists new articles from highly accessed journals, such as 'Am J Clin Nutr (8)', 'Am J Orthod Dentofacial Orthop (35)', 'Cochrane Database Syst Rev (7)', 'J Biol Chem (44)', 'JAMA (1)', 'Lancet (7)', 'N Engl J Med (8)', 'Nature (44)', 'Proc Natl Acad Sci U S A (25)', and 'Science (85)'. It also promotes 'PubMed Journals'.
- Trending Articles:** Lists PubMed records with recent increases in activity, such as 'Clinical outcomes of a scapular-focused treatment in patients with subacromial pain syndrome: a systematic review. Br J Sports Med. 2016.', 'Effectiveness of neuromuscular taping on painful hemiplegic shoulder: a randomised clinical trial. Disabil Rehabil. 2016.', 'Rare and low-frequency coding variants alter human adult height. Nature. 2017.', 'Moconnell's patellar taping does not alter knee and hip muscle activation differences during proprioceptive exercises: A randomized placebo-controlled trial in women with patellofemoral pain syndrome. J Electromyogr Kinesiol. 2016.', and 'Parvovirus B19 during pregnancy: a review. J Prenat Med. 2010.'.
- PubMed Commons:** Lists featured comments, such as 'Circuits in reward & aversion: @gstuber posts journal club review of study identifying neuron populations. bit.ly/2ZG9mN Feb 2', 'Migrating database: @odsouthan provides updated info for finding pharmacological data resource. bit.ly/2ktdWnC Jan 31', 'Genes for lactation persistence in cattle: @Eric_Fauman critiques findings of a genome-wide association study. bit.ly/2RLtsB Jan 30', 'Reviewing replication: R Tibshirani critiques statistics; A Collings cross-posts comment from original study authors bit.ly/24U3H Jan 27', and 'Evaluating connection between food energy supply & obesity: @JamesonVoss discusses ecologic study. bit.ly/2FxpU Jan 29'.

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>



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](#)

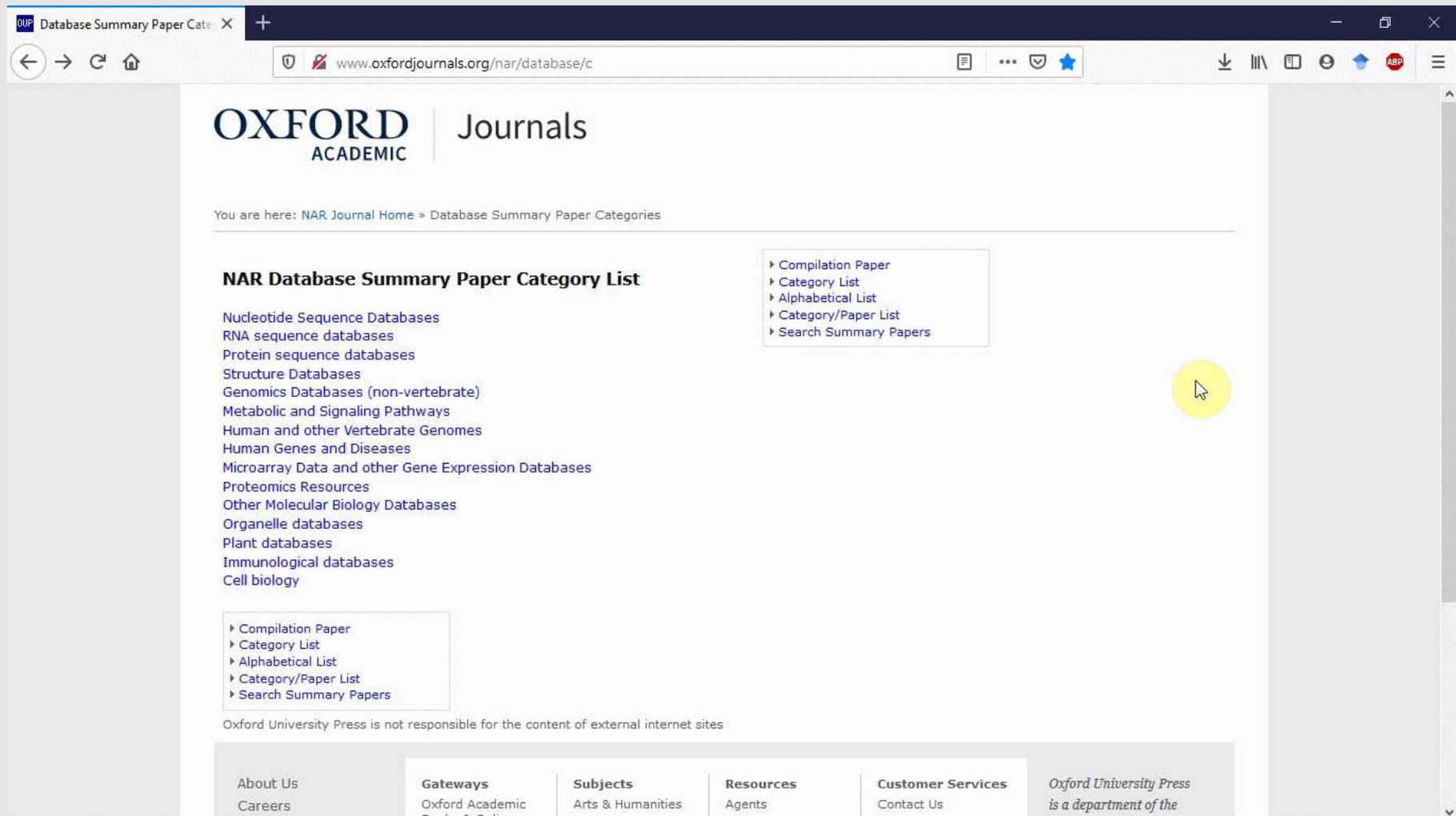
- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

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 displaying the Oxford Academic website. The page title is "Database Summary Paper Categories" and the URL is "www.oxfordjournals.org/nar/database/c". The page features the Oxford Academic logo and a navigation breadcrumb: "You are here: NAR Journal Home » Database Summary Paper Categories".

The main content is titled "NAR Database Summary Paper Category List" and lists various biological databases and resources:

- 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

Two dropdown menus are visible, each containing the following options:

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

At the bottom of the page, there is a footer with navigation links: "About Us", "Careers", "Gateways", "Oxford Academic Books & Online", "Subjects", "Arts & Humanities", "Resources", "Agents", "Customer Services", "Contact Us", and "Oxford University Press is a department of the".

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

Beaucoup de ces systèmes 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 utilisé 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 Send to: Filters: [Manage Filters](#)

Animals (2,835) Customize ...

Molecule types clear Items: 1 to 20 of 2835

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

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](#)

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]) (4253) Nucleotide

Q mouse insulin AND (animals[filter] AND biomol_mrna[PROP]) (4258) Nucleotide

Q mouse insulin AND (biomol_mrna[PROP]) (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

```
ctcggcagc cggaggtcat cctgctagac tcagacctgg atgaacccat agacttgcgc    60
tcggtaaga gccgcagcga ggccggggag ccgccagct cctccaggt gaagcccag    120
acaccggcgt cggcggcggg 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)
FORMED      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             Location/Qualifiers
     source            1..1599
                        /organism="Saccharomyces cerevisiae"
                        /mol_type="genomic DNA"
                        /strain="JZ1C"
                        /db_xref="taxon:4932"
     gene              <1..>1599
                        /gene="SUC2"
     mRNA              <1..>1599
                        /gene="SUC2"
                        /product="invertase"
     CDS               1..1599
                        /gene="SUC2"
                        /codon_start=1
                        /product="invertase"
                        /protein_id="AFN08663.1"
                        /db_xref="GI:393395466"
                        /translation="MLLQAFLELLAGFAAKIASMNTNEDSDNPLVHFTFNEGWMNDPH
                        GLVYDEKDAKWHLYPQYHNDYVMGTPLFVGHATSDDLTNHEDEPIAIAPKKNDSGAF
                        SGSMVVDYNNISGFPNDIIDPQKCVAIWYNTPKSEHQYISYELDGGTFTFYKQNF
                        VLAARSTQVHDPKVPVWYEPSQKWNIAAKSGQYKIEIYSSDOLSKWLESAFAMEGFL
                        GYQTECPGLLIEVPTKQOPSKSYWNFISINPGAPAGGSFNQYFVGSYNGTHFAFDNQ
                        SRVDFGRDYALQTFPNTDPTYGSALGIAWASHWYSAPVFTNPWSSMSLWVPSL
                        NTEYQANPETELINLKAEPILNISNAGPMSRPATMITLRANSYVDLSNSTGLEPE
                        LVYAVNTTQTSKSVFPLSLMFKGLEDPKYLQMGFEASASDFLDRGNSKVFYKE
                        NPYFTNMSVNNQPFKSENLSYKIVYGLLDQNLLELYFNMGDVFSTNTYPMITGNAL
                        GSVHMITGVNLFYIDKPVVBEVK"
```

```
ORIGIN
1 atgcttttgc aagctttcct ttctcttttg gctggttttg cagccaaaat atctgcatca
61 atgcaaaaag aaactagcgc tagcctttg gtcaccltca ccccacaaa gggctggatg
121 aatgaccaca atgggtttgt gtcacgtpaa aaagatgcca atggcctctc gtcctttcaa
181 tcccaaccac atgacacngt atgggttagc ccatgttttt ggggcctgca tactctcngt
241 gatttgactc attggaaga tgaacccttt gctatngctc ccaagcgtaa cgtatcaggt
301 gctttctctc gctccatggt ggttgattac aaccaacaga gttggttttt caatgctact
361 attgatccaa gacaaagatg cgttgcgatt tggacttata acactcctga aagtgaagag
421 caatcacatta gctattctct tgatgggtgt tacactttta ctgaatccca aaagaacctt
481 gttttagctg ccaactccac tcaattcaga gatccaaagg tgttctgcta tgaacctttt
541 caaaaatgga ttatgacgpc tgcacaaatca caagatcca aaattpaat ttactcctrt
601 gatgacttga agtactgcaa gctagaactc gcatllgcta atgaaggttt cttaggctac
661 caatataaat gtccggtttt gatgaaagtc ccaactgagc aagatccttc caatctcat
721 tgggtcatgt ttattcttat caatccaggt gcccctgctg goggttcott caaccetat
781 ttgtttgget ccttcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggtc
841 gattttggtc aggaactacta tgccttgcaa acttttttca acacagaacc aacgtacggt
901 tcaagcttag gtattgctgt ggccttcaaan tgggagtaga gtgcttttgt cccaactaac
961 ccatggagat catccatgta ttigtccgc aagttttctt tgaadactga atatacagct
1021 aatccagaga ctgaattgat caattgaaa gcccaaccac tattgaacat tagtaatgct
1081 gttccctggt ctgcttttgc tactaacaca acttcaacta aggccaattd tctaatgctc
1141 gatttgagca actcagctgg tactctagag tttagattg ttatcgtgtg taacaccaca
1201 caaacctat ccaactcngt ctttcccgag ttctcacttt gttcagggg tttagaagat
1261 cutgagaact atttaagaa yggttttgaa gccagtgctt ctctctttt tttagccgt
1321 ggttaactta aggtcaagtt tgcacagga aaccatatt tcaacaacag aatgtctgtc
1381 aaccaaccac cattcaagtc tgaacacpa ctaagttact ataaagtga cggcctactg
1441 gatcaaaaac tcttgaattt gtacltcaac gatggagatg tggtttctac aatcactac
1501 ttctagacca ccgtaaacgc tctaggatct gtgaacatga ccaactggtg cgtataattg
1561 ttctacattg caaagttcca agtaaggyaa 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:4932"
    gene               <1..>1599
                       /gene="SUC2"
    mRNA              <1..>1599
                       /gene="SUC2"
                       /product="invertase"
    CDS               1..1599
                       /gene="SUC2"
                       /codon_start=1
                       /product="invertase"
                       /protein_id="AFN08663.1"
                       /db_xref="GI:393395466"
                       /translation="MLLQAFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPN
GLWYDEKDAKWHLYFQYNPNDTVWGTPLFWGHATSDDLTHWEDEPIAIAIPKRNDSGAF
SGSMVVDYNNNTSGFFNDTIDPRQRCVAIWTYNTPESEEQYISYSLDGGYTFTEYQKNP
VLAANSTQFRDPKVFVWYEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFL
GYQYECPLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNO
SRVVDVFGKDYALQTFEFDNTDPTYGSAIGIAWASNWEYSAFVPTNPWRSMSLVKRFSL
NTEYQANPETELINLKAEPILNISNAGPWSRFATNTTLTKANSYNVDLSNSTGTLEFE
LVYAVNTTQTISKSVFPDLSLWFKGLEDPEEYLRMGFEASASSFFLDRGNSKVKEVKE
NPYFTNRMSVNNQPFKSENDLSYYKVYGLLDQNIILELYFNDGDVVSTNTYFMTTGNAL
GSVNMTTGVDNLFYIDKFQVREVK"
```

La séquence (format Genbank)

ORIGIN

```
1 atgcttttgc aagctttcct tttccttttg gctggttttg cagccaaaat atctgcatca
61 atgacaaacg aaactagcga tagacctttg gtccacttca cacccaacaa gggctggatg
121 aatgacccaa atgggttgtg gtacgatgaa aaagatgcca aatggcatct gtactttcaa
181 tacaacccaa atgacaccgt atggggtagc ccattgtttt ggggccatgc tacttccgat
241 gatttgactc attgggaaga tgaaccatt gctatcgctc ccaagcgtaa cgattcaggt
301 gctttctctg gctccatggg ggttgattac aacaacacga gtgggttttt caatgatact
361 attgatccaa gacaaagatg cgttgcgatt tggacttata aactcctga aagtgaagag
421 caatacatta gctattctct tgatgggtgg tacactttta ctgaatacca aaagaaccct
481 gtttttagctg ccaactccac tcaattcaga gatccaaagg tgttctggta tgaaccttct
541 caaaaatgga ttatgacggc tgccaaatca caagactaca aaattgaaat ttactcctct
601 gatgacttga agtcctggaa gctagaatct gcatttgcta atgaaggttt cttaggctac
661 caatatgaat gtccagggtt gattgaagtc ccaactgagc aagatccttc caaatcctat
721 tgggtcatgt ttatttctat caatccaggt gcacctgctg gcggttcctt caaccaatat
781 tttgttggat cettcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggta
841 gattttggta aggactacta tgccttgcaa actttcttca acacagacc aacgtacggg
901 tcagcattag gtattgctg ggcttcaaac tgggagtaca gtgcctttgt cccaactaac
961 ccatggagat catccatgtc tttggtcgcg 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 caaacatat ccaaatccgt ctttcccgcac ttatcacttt ggttcaaggg tttagaagat
1261 cctgaagaat atttaagaat gggttttgaa gccagtgctt cttccttctt tttggaccgt
1321 ggtaactcta aggtcaagtt tgtcaaggag aacctatatt tcacaaacag aatgtctgtc
1381 aacaaccaac cattcaagtc tgagaacgac ctaagttact ataaagtgta cggcctactg
1441 gatcaaaaca tcttggaatt gtacttcaac gatggagatg tggtttctac aaatacctac
1501 ttcatgacca ccggtaacgc tctaggatct gtgaacatga cactgggtgt 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; FNU 1599 bp.
XX
AC: JQ836661.
XX
DT: 06-JUL-2012 (Rel. 11); Created;
DT: 28-DEC-2012 (Rel. 11); Last updated, Version 2;
XX
SR: Saccharomyces cerevisiae strain JZ1C invertase (SUC2) gene, complete cds.
XX
KW: .
XX
OC: Saccharomyces cerevisiae (baker's yeast)
OC: Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes;
OC: Saccharomycetiales; Saccharomycetaceae; Saccharomyces.
XX
RN: [1]
RP: 1-1599
RX: DOI, 10.1101/ADM.02458-12.
RX: FIDMSID, 23104410.
RA: Wang S.-A., Li F.-L.
RT: "Invertase SUC2 Is the Key Hydrolase for Inulin Degradation in
RP: Saccharomyces cerevisiae".
RL: Appl. Environ. Microbiol. 79(1):403-406(2013).
XX
RH: [2]
RP: 1-1599
RA: Wang S.-A., Li F.-L.
RT: .
RL: Submitted (27-MAR-2012) to the EMBL.
RL: Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess
RL: Technology, 189 Songling Road, Qingdao, Shandong 266101, China
XX
DR: EuropePMC; PMC3536098; 23104410.
XX
FH: Key Location/Qualifiers
FF
FT: source 1..1599
FT: /organism="Saccharomyces cerevisiae"
FT: /strain="JZ1C"
FT: /mol_type="genomic DNA"
FT: /db_xref="taxon:4932"
FT
FT: gene 1..1599
FT: /gene="SUC2"
FT
FT: mRNA 1..1599
FT: /gene="SUC2"
FT: /product="invertase"
FT
FT: CDS 1..1599
FT: /codon_start=1
FT: /gene="SUC2"
FT: /product="invertase"
FT: /db_xref="GeneID:16909"
FT: /db_xref="InterPro:IPR001360"
FT: /db_xref="InterPro:IPR009925"
FT: /db_xref="InterPro:IPR01040"
FT: /db_xref="InterPro:IPR01316"
FT: /db_xref="InterPro:IPR01319"
FT: /db_xref="InterPro:IPR014053"
FT: /db_xref="InterPro:IPR020396"
FT: /db_xref="UniProtKB/TrEMBL:I5VAV5"
FT: /protein_id="AF066661.1"
FT: /translation="MLIQAFLLLAGFAKTSAMTSTGSRPLVHTFHKWQWQWQWQ
FT: LNVYKDAWGLLYPQFNQHTYVQVFLFWQWATSOGLTQQRDEPIATAKKSHGGAGSQ
FT: SMVYVYEMTGGPFDITDPRQKQVAMTYNTPSERVTSYSLDGGVTFEYGENPVLA
FT: ANRTQPRQKVPWDEPQKQWMTAARSGYKCEIYVSEELKSMELKAPFANEGFLQVGT
FT: KIQGLIVPFGGPKKSNVWFISINPAKAGGSSNGFVGGPMTHTFAFDWQARVVD
FT: PAKDLYALGTFNDFPTDGLATAMKQWQWELKAPVTPQWQWQWQWQWQWQWQWQWQ
FT: MPTKELTALKAEPILAIQWAGPQKREKATPTTLTKWQWQWQWQWQWQWQWQWQWQ
FT: TQTIKSVVFDLSLWVFLKEDPEKVLKQSPKASASQFLIRGKSKVQKPKVKNFYPTWQW
FT: QVWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQW
FT: QWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQ
FT: WQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQWQW
XX

```

Description générale de la séquence

« Features »
 Description des objets biologiques présents sur la séquence

```

Sequence 1599 bp 45% A; 34% C; 11% G; 8% T; 0 other
atgctctcag aggttctcct tctctcttga gctgctctca cagccaaact atctgcacca 48
atgcaaaag aakctagcga tagactcttg gctcactcca cagccaaact ggcctgagtg 96
aatgaccoca atgggtctgq gtagatgaga aaagatgoca aatggcaatc gactctcaa 144
tacaacocaa atgacacnct atgggtagc coattgtttt ggggccagtc tacttccgat 240
gatttgactc atgggaaga tgaaccactt gctatgcttc ccaagatga agattcaggt 300
gctttctctg gttccatggt gttgactac aacacacaga atgggttttt caatgatatt 360
atgacocaa gatcaagatg cgttggcagt tggacttata aactcctga aagtgaaag 420
caatcacta gctctctctc tgggtggcgt tacccttata ctgactacca aagtaacctc 480
gctttagct cactctccat ctgactacca gttctctga tgaacctctc 540
caaaaatgca ttatgacgq tgcacatca caagatgca aattgcaat ttaactcttc 600
gatgacttga agctcagaa gctagatctc goatttgata atgaggttt cttagactac 660
caatcagat gtcagagctt gattgagtc caactgagc agatcttc caaatcctat 720
tgggcagtc ctattctcat caatccaggt gaactcgtg ggggtctct caaccaatat 780
tttttggtt cttcaatgg tactcaattt gaaggtctg caactcaatc tagagtgta 840
gattctggtt aggtactata tgccttga aacttttca anacagacc caagtagggt 900
tactactag gttactgctc gttctcaac tggagtagaa ttgctctctc cctcaactaa 960
caatgagct cactcaagct ttgtctctctt caactctctt caactctga atctcaagct 1020
aahccagaga ctgacttgat caattgaaa gtcgaaccac tattgacat tagtaactgt 1080
gctctctctg ctgctttctg tactaacaca actcfaatca aggcatttc ttaactctc 1140
gatttgaga actgactcga tactctagc tttgagtgt tttagctct taacacaca 1200
caaacctat ccaactcgtt cttctccagc ttatcacttt gtttcaagg tttgaaagt 1260
cctgagatc atttaagat ggtcttga gcaagtctt cttctctctt tttggaact 1320
ggttaactca agtcaagct tcttaagag aactcaact tcaacaacag aatgctcttc 1380
aacacacac caactcaagc tgaagacgc ctgactctt caaactctc caaactctg 1440
gatcaaaaca tcttgaactt gtaactcaac catggagat tggctctac caactcaact 1500
tcatgacca ccggtcaagc totagatct gtagaacgc caactcgtct agataactg 1560
tctcatattg accagtttca agtcaaggca gtaaatag 1599

```

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="MLLQAFLEFLLAGFAAKISASMTNETSDRPLVHFTPNKGWMNDPNG
FT                       LWYDEKDAKWHLYFQYNPNDTVWGTPLFWGHATSDDLTHWEDEPIAIAPKRNDSGAFSG
FT                       SMVVDYNNISGFFNDTIDPRQRCVAIWYNTPESEEQYISYSLDGGYTFTEYQKNPVL
FT                       ANSTQFRDPKVFWEPSQKWIMTAAKSQDYKIEIYSSDDLKSWKLESAFANEGFLGYQY
FT                       ECPGLIEVPTEQDPSKSYWVMFISINPGAPAGGSFNQYFVGSFNGTHFEAFDNQSRVVD
FT                       FGKDYALQTFNTDPTYGSALGIAWASNWEYSAFVPTNPWRSSMSLVRKFSLNTEYQA
FT                       NPETELINLKAEPILNISNAGPWSRFATNTLLTKANSYNVDLSNSTGTLEFELVYAVNT
FT                       TQTISKSVFPDLSLWFKGLEDPEEYLRLMGFEASASSFFLDRGNSKVKFVKENPYFTNRM
FT                       SVNQPFSKSENDLSYYKVYGLLDQNIILELYFNDGDVVSTNTYFMTTGNALGSVNMTTGV
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 cacccaacaa gggctggatg      120
aatgacccaa atgggttgtg gtacgatgaa aaagatgcca aatggcatct gtactttcaa      180
tacaacccaa atgacaccgt atggggtaag ccattgtttt ggggccatgc tacttccgat      240
gatttgactc attgggaaga tgaaccatt gctatcctc ccaagcgtaa cgattcaggt      300
gctttctctg gctccatggg ggttgattac aacaacacga gtgggttttt caatgatact      360
attgatccaa gacaaagatg cgttgcgatt tggacttata acactcctga aagtgaagag      420
caatacatta gctattctct tgatgggtgg tactacttta ctgaatacca aaagaaccct      480
gttttagctg ccaactccac tcaattcaga gatccaaagg tgttctggta tgaaccttct      540
caaaaatgga ttatgacggc tgccaaatca caagactaca aaattgaaat ttactcctct      600
gatgacttga agtcctggaa gctagaatct gcatttgcta atgaaggttt cttagggtac      660
caatatgaat gtccaggttt gattgaagtc ccaactgagc aagatccttc caaatcctat      720
tgggtcatgt ttatttctat caatccaggt gcacctgctg gcggttcctt caaccaatat      780
tttgttggat ccttcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggta      840
gattttggta aggactacta tgccttgcaa actttcttca acacagacc aacgtacggt      900
tcagcattag gtattgcctg ggcttcaaac tgggagtaca gtgcctttgt cccaactaac      960
ccatggagat catccatgtc tttggtcctc aagttttctt tgaacactga atatcaagct     1020
aatccagaga ctgaattgat caatttgaaa gccgaaccaa tattgaacat tagtaatgct     1080
ggtcctctgg ctctgttttg tactaacaca actctaacta aggccaattc ttacaatgtc     1140
gatttgagca actcgactgg taccctagag tttgagttgg ttacgctgt taacaccaca     1200
caaaccatat ccaaatcctg ctttcccgac ttatcacttt ggttcaaggg tttagaagat     1260
cctgaagaat atttaagaat gggttttgaa gccagtgett cttecttctt tttggaccgt     1320
ggtaactcta aggtcaagtt tgtcaaggag aaccatatt tcacaaacag aatgtctgtc     1380
aacaaccaac cattcaagtc tgagaacgac ctaagttact ataaagtgta cggcctactg     1440
gatcaaaaca tcttgggaat gtacttcaac gatggagatg tggtttctac aaatacctac     1500
ttcatgacca ccggtaacgc tctaggatct gtgaacatga ccactgggtg cgataatttg     1560
ttctacattg acaagttcca agtaagggaa gtaaaaatag                                     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'accension
 - 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;
 MLLQAFLEFL AGFAAKISAS MTNETSDRPL VHFTPNKGWM NDPNGLWYDE KDAKWHLYFQ
 YNPNDTVWGT PLFWGHATSD DLTHWEDEPI AIAPKRNDSD AFSGSMVVDY NNTSGFFNDT
 IDPRQRCVAI WTYNTPESSE QYISYSLDGG YTFTEYQRNP VLAANSTQFR DPKVFWYEPS
 QKWIMTAAKS QDYKIEIYSS DDLKSWKLES AFANEGFLGY QYECPLIEV PTEQDPSKSY
 WVMFISINPG APAGGSFNQY FVGSFNGTHF EAFDNQSRVV DFGKDYALQ TFFNTDPTYG
 SALGIAWASN WEYSAFVPTN PWRSSMSLVR KFSLNTEYQA NPETELINLK AEPILNISNA
 GPWSRFATNT TLTKANSYNV DLSNSTGTLE FELVYAVNTT QTISKSVFPD LSLWFKGLED
 PEYLRMGFE ASASSFFLDR GNSKVKFVKE NPYFTNRMSV NNQPFKSEND LSYYKVVYGLL
 DQNIILELYFN DGDVVSTNTY FMTTGNALGS VNMTTGVDNL FYIDKQVRE 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  
ctccggcagcccgaggtcatcctgctagactcagacctggatgaacccat  
agacttgcgctcgggtcaagagccgcagcgcaggccgggggagccgcccagct  
ccctccaggtgaagcccgagacaccggcgctcggcgggcggtggcggtggcg  
gcggcagcggcaccaccacgacggcgggagagaagcctccagccgaggcca  
ggacgaacctgcagagtcgctgagcgcagttcaagcccttctttgggaata  
taattatcaccgacgtcaccgcgaactgcctcaccgttactttcaaggag  
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 .

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	