

# Bioinformatique

Emploi du temps, groupes de TP, supports de cours et de TP disponibles sur <http://silico.biotoul.fr> sur la page Enseignement

Contrôle avec report : 30%

2 mars 13h30

Contrôle terminal : 70%

30 avril 13h30

Objectifs pédagogiques :

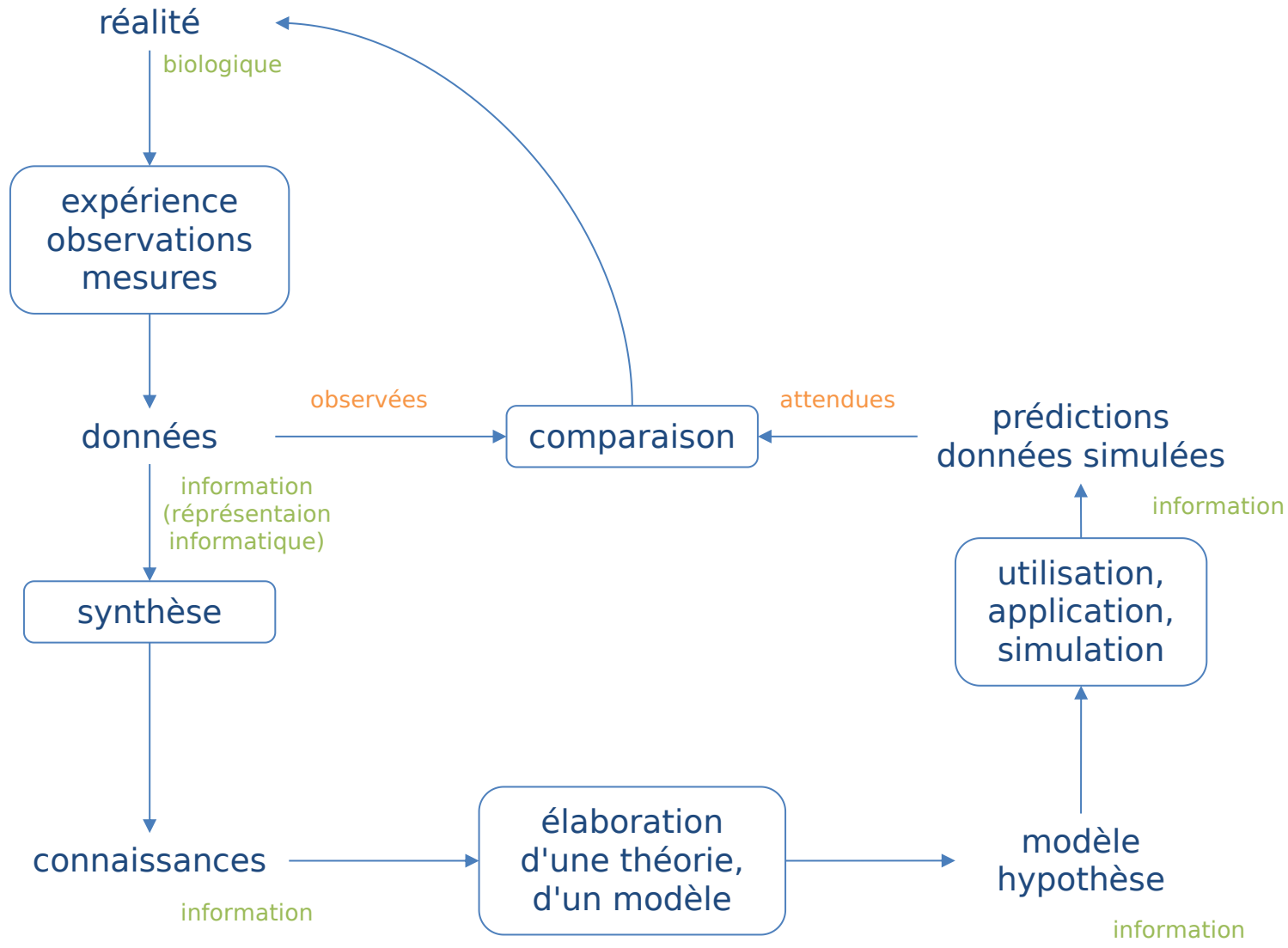
- Aperçu de quelques domaines d'application de la bioinformatique
  - traitement d'image : mesures phénotypiques
  - gestion des données
  - statistiques : corrélation phénotype - génotype
  - banques de données et sites Web publiques
  - modélisation de systèmes biologiques et simulations

Intervenant·e·s

- Roland Barriot (intro, bases de données, ...), Maxime Bonhomme et Raphaël Mourad (génétique, statistiques), Franck Delavoie et Silvia Kocanova (imagerie), Gwennaele Fichant (biologie des systèmes), Elodie Gaulin (bioanalyse)
- contact : [barriot@biotoul.fr](mailto:barriot@biotoul.fr) mais de préférence de vive voix après un cours ou un TD/TP

- Qu'est-ce que la bioinformatique ?
- Plusieurs réponses :
  - pas de l'informatique *bio*
  - traitement de l'information biologique
  - domaine spécialisé de la biologie
  - domaine de recherche multidisciplinaire : biologie, santé, mathématique, statistique, informatique, physique, éthique
- Une définition : modélisation et traitement des informations biologiques par des méthodes informatiques et/ou mathématiques et/ou statistiques.
- Les données et connaissances biologiques posent de nouveaux défis spécifiques pour leur gestion, représentation, et leur analyse/traitement.

# Méthode scientifique

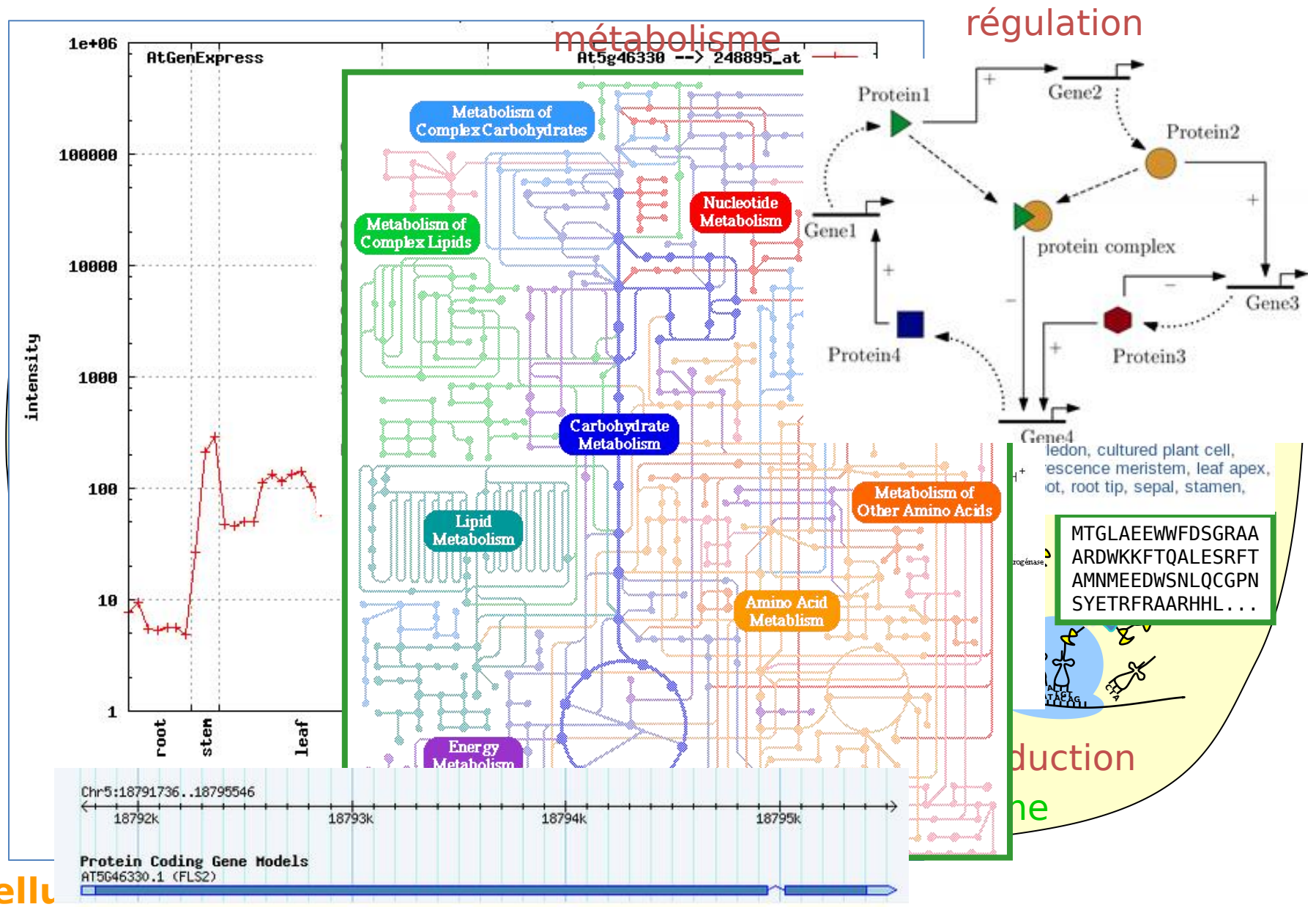


# Historique et domaines d'application liés à la bioinformatique

---

- 1970
- Apparition du terme bioinformatique
  - Disponibilité de séquences de gènes (et du 1<sup>er</sup> génome complet en 1977, bactériophage phi X174)
  - Comparaison de séquences
- 1980
- Alignement de séquences et recherche de séquences similaires
  - Analyse phylogénétique (basée sur l'alignement de séquences)
  - Prédiction de fonction (par similarité de séquence)
  - Banques de données (de séquences)
- 1990
- Débuts de la génomique et génomique comparative ; analyse du contenu d'un génome (1<sup>er</sup> génome complet d'organisme vivant à réplication autonome *Haemophilus influenzae*, 1995)
  - Débuts des analyses globales de l'expression des gènes et des protéines
  - Famille de gènes, de protéines, de domaines
  - Prédiction de la structure 3D des protéines
- 2000
- Généralisation des approches globales
  - Traitement de graphes : interactions protéine-protéine, réseau de régulation de l'expression des gènes, réseau métabolique
  - Apprentissage automatique (*data mining*)
  - Fouille de texte (*text mining*)
  - Biologie des systèmes : modélisation de systèmes biologiques
  - Intégration de données hétérogènes, visualisation
  - Médecine personnalisée
  - NGS/Séquençage très haut débit
  - Traitement d'images liées à la microscopie
- 2010
- Ethique, confidentialité des données, manipulation du génome
  - Modélisation d'une cellule, d'un organisme, d'une population, d'un écosystème
  - Biologie synthétique

# (Quelques) données (mesures et prédictions) et connaissances disponibles



# Données omics

- **Génome**

- séquence(s) nucléique(s) de l'ensemble des chromosomes d'un organisme
- ensemble des gènes d'un organisme

- **Transcriptome**

- ensemble des ARNm ou transcrits présents dans une cellule ou une population de cellules dans des conditions données

- **Protéome**

- ensemble des protéines présentes dans une cellule ou une population de cellules dans des conditions données

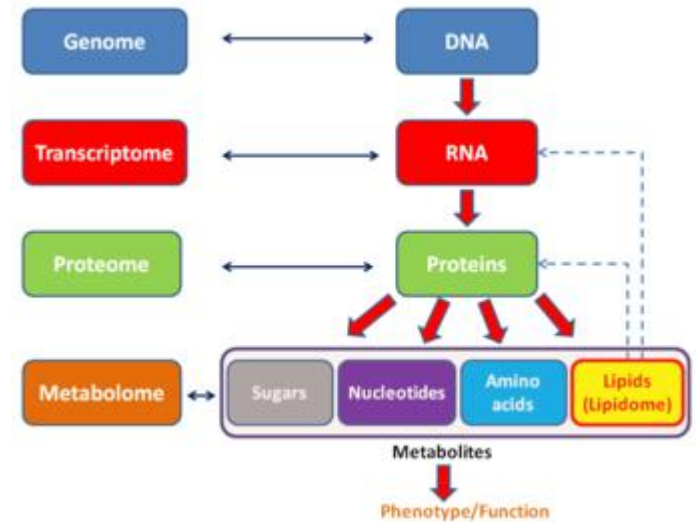
- **Interactome**

- ensemble des interactions moléculaires pouvant survenir *in vivo*
- ensemble des interactions moléculaires dans des conditions données
- ensemble des interactions au sein d'un organisme : moléculaires, physiques, génétiques, fonctionnelles, ...

- **Métabolome**

- ensemble des métabolites présents dans une cellule ou une population de cellules dans des conditions données

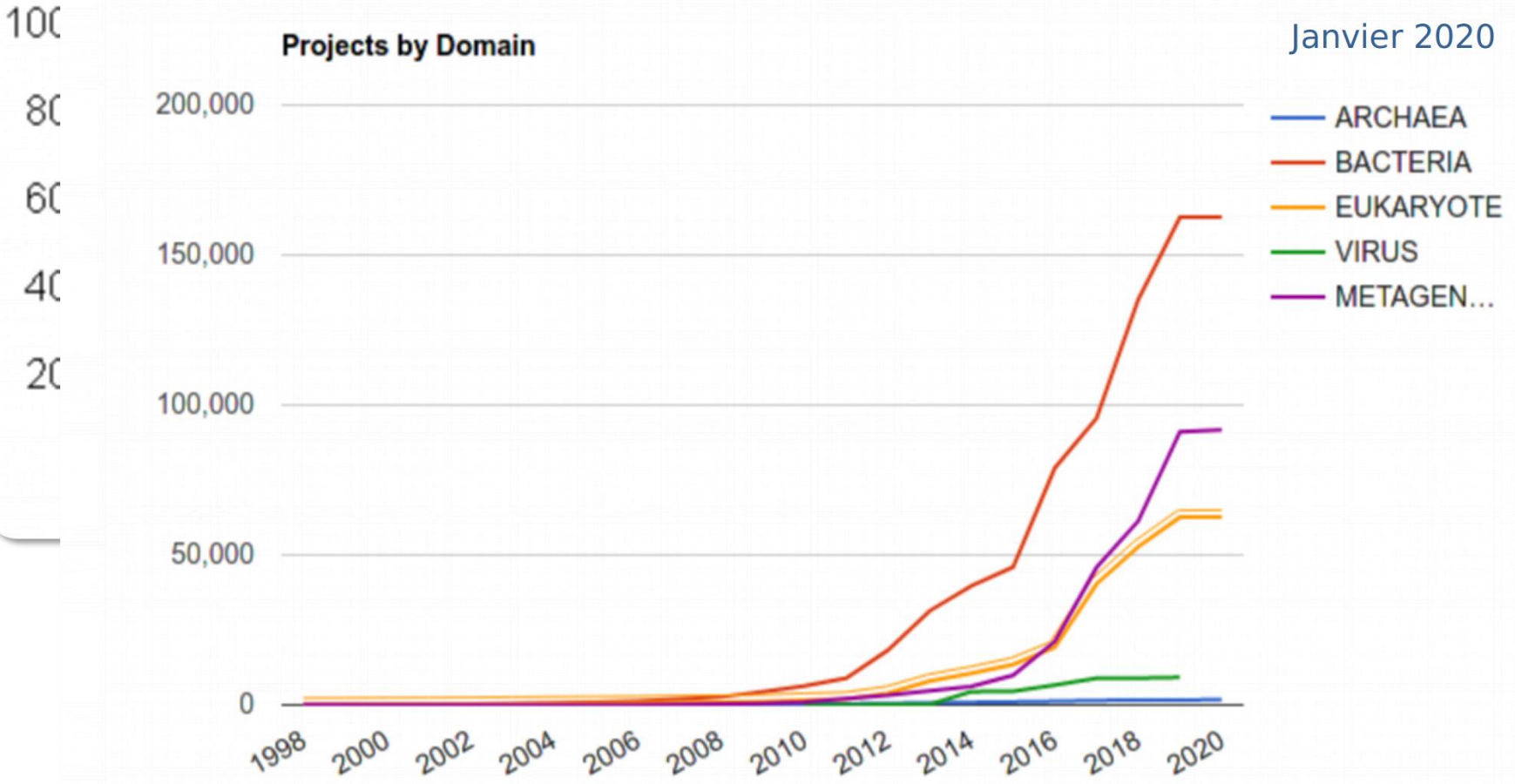
- Exome, lipidome, phénome, régulome, sécrétome, épigénome, méthylome, ...



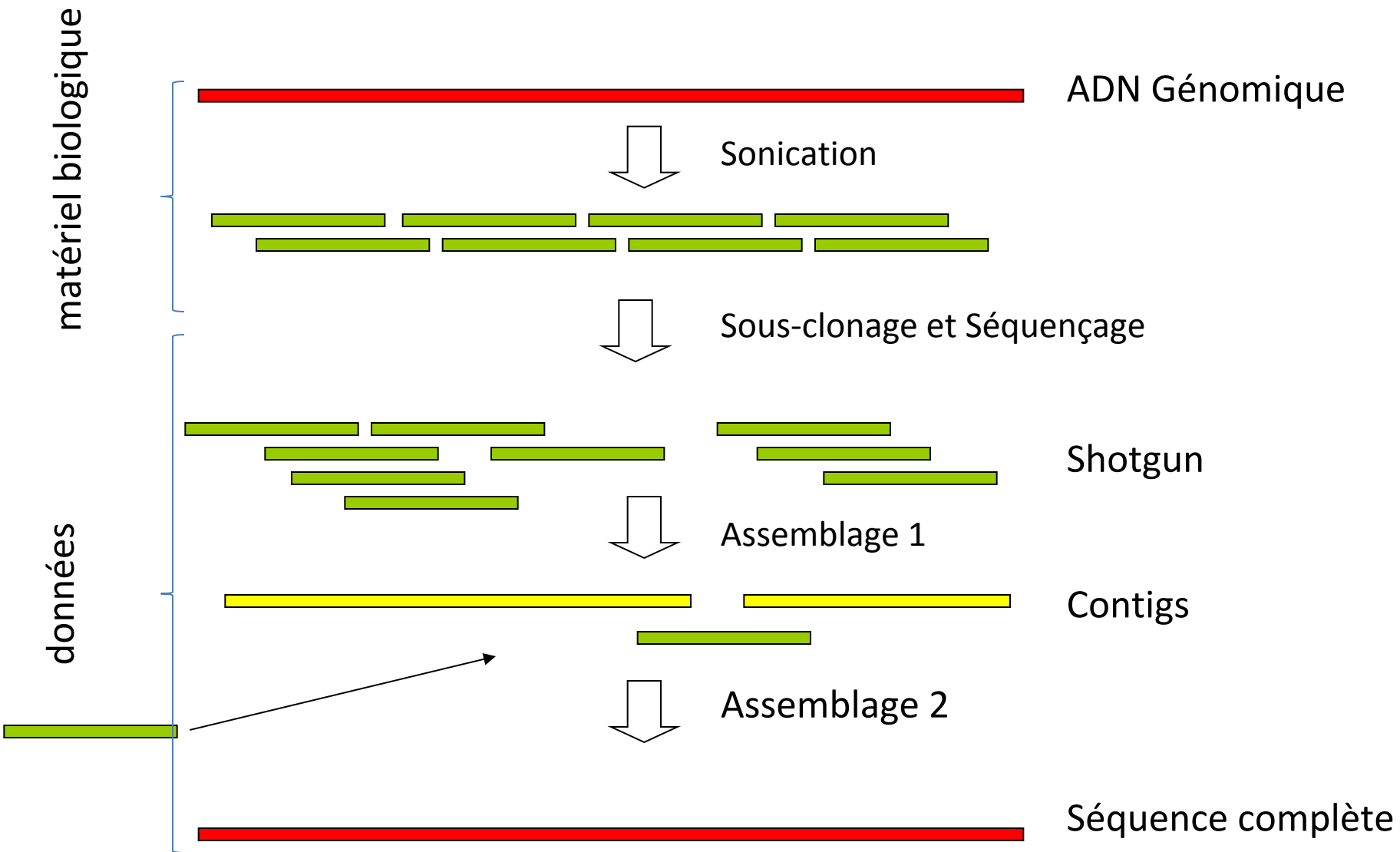
source : Wikipedia

# Séquences disponibles : quelques chiffres

## Completely Sequenced Genomes © January 2009



# Assemblage d'un génome





# L'analyse manuelle d'une séquence peut s'avérer laborieuse

---

TCCTGGCCTACATGTTCTTTGGCAAAGGATCTTCAAATCAACGGCTCCCAGGTGCGGCGATCATCCATTTCTTCGGAGGGATTACGAGATTT  
ACTTCCCCTACATTTCTGATGAAACCTGGCCCTGATTCTCGCAGCCATTGCCGGCGGAGCAAGCGGACTCTAACATTACGATCTTTAATGCCG  
GACTTGTGCGCGCAGCGTCACCGGGAAGCATTATCGCATTGATGGCAATGACGCCAAGAGGAGGCTATTTTCGGCGTATTGGCGGGTGTATTGG  
TCGCTGCAGCTGTATCGTTCATCGTTTTAGCAGTGATCCTGAAATCCTCTAAAGCTAGTGAAGAAGACCTGGCTGCCGCAACAGAAAAATGC  
AGTCCATGAAGGGGAAGAAAAGCCAAGCAGCAGCTGCTTTAGAGGCGGAACAAGCCAAGCAGAGAAGCGTCTGAGCTGTCTCCTGAAAGCGC  
GAACAAAATTATCTTTTCGTGTGATCCGGGATGGGATCAAGTGCCATGGGGGCATCCATCTTAAGAAACAAAGTGAAAAAGCGGAGCTTGACA  
TCAGTGTGACCAACACGGCCATTAACAATCTGCCAAGCGATGCGGATATTGTATCACCCACAAAGATTTAACAGACCGCGCGAAAGCAAAGC  
TGCCGAACGCGACGCACATATCAGTGGATAACTTCTTAAACAGCCGAAATACGACGAGCTGATTGAAAAGCTGAAAAGTAATCTTATAGAAA  
GAGAGTATTGTCATGCAAGTACTCGCAAAGGAAACATTAACCTCAATCAAACGGTATCATCAAAGAAGAGGCTATCAAATTGGCAGGCCAGA  
CGCTGATTGACAACGGCTACGTGACAGAGGATTACATTAGCAAAATGTTTGACCGTGAAGAAACGTTCTTACGTTTTATGGGGAATTTTCATTG  
CCATTCCACACGGCACAGAAGAAGCGAAAAGCGAGGTGCTTCACTCAGGAATTTCAATCATAACAGATTCCAGAGGGCGTTGAGTACGGAGAAG  
GCAACACGGCAAAAGTGGTATTCGGCATTGCGGGTAAAAATAATGAGCATTTAGACATTTTGTCTAACATCGCCATTATCTGTTTCAGAAGAAG  
AAACATTGAACGCCTGATCTCCGCTAAAGCGAAGAAGATTTGATCGCCATTTCAACGAGGTGAACTGACATGATCGCCTTACATTTTCGGTTCG  
GGAAATATCGGGAGAGGATTTATCGGCGCGCTGCTTCACTCCGGCTATGATGTGGTGTTCGCGGATGTGAACGAAACGATGGTCAGCCTC  
CTCAATGAAAAAAGAATAACACAGTGGAACTGGCGGAAGAGGGACGTTTCATCGGAGATCATTGGCCCGGTGAGCGCTATTAACAGCGGCAGT  
CAGACCGAGGAGCTGTACCGGCTGATGAATGAGGCGGCGCTCATCACAACAGCTGTGCGCCCGAATGTCCTGAAGCTGATTGCCCCGTCTATC  
GCAGAAGGTTTAAAGACGAAGAAATACTGCAACACACTGAATATCATTGCCTGCGAAAATATGATTGGCGGAAGCAGCTTCTTAAAGAAAGAA  
ATATACAGCCATTTAACGGAAGCAGAGCAGAAATCCGTGAGTGAACGTTAGGTTTTCCGAATTTCTGCCGTTGACCGGATCGTCCCGATTTCAG  
CATCATGAAGACCCGCTGAAAGTATCGGTTGAACCATTTTTTCGAATGGGTCATTGATGAATCAGGCTTTAAAGGGAAAACACCAGTCATAAAC  
GGCGCACTGTTTTGTTGATGATTTAACGCCGTACATCGAACGGAAGCTGTTTACGGTCAATACCGGACACGCGGTACAGCGTATGTCGGCTAT  
CAGCGCGGACTCAAACGGTCAAAGAAGCAATTGATCATCCGGAAATCCGCCGTGTTGTTTATTTCGGCGCTGCTTGAACTGGTGACTATCTC  
GTCAAATCGTATGGCTTTAAGCAAACCTGAACACGAACAATATATTAATAATCAGCGGTCGCTTTTAAATCCTTTTCAATTTTCGGACGATGTGAC  
CCGCGTAGCGAGGTCACCTCTCAGAAAACCTGGGAGAAAATGTAGACTTGTAGGCCCGGCAAAGAAAATAAAAGAACC GAATGCACTGGCTGAA  
GGAATTGCCGCAGCACTGCGCTTCGATTTACCGGTGACCCTGAAGCGGTTGAACTGCAAGCGCTGATCGAAGAAAAGGATAACAGCGGCGTAC  
TTCAAGAGGTGTGCGGCATTCAGTCCCATGAACCGTTGCACGCCATCATTTTTAAAGAACTTAATCAATAACCGACCACCCGTGACACAATGT  
CACGGGCTTTTTACTATCTCGCAATCTAGTATAATAGAAAGCGCTTACGATAACAGGGGAAGGAGAATGACGATGAAACAATTTGAGATTGCG  
GCAATACCGGGAGACGGAGTAGGAAAGAGGTTGTAGCGGCTGCTGAGAAAGTGCTTACATACAGCGGCTGAGGTACACGGAGGTTTGTCAATCT  
CATTCACAGCTTTTCCATGGAGCTGTGATTATTACTTGGAGCACGGCAAAAATGATGCCCGAAGATGGAATACATACGTTACTCAATTTGAA  
GCAGTTTTTGGGAGCTGTCGAAATCCGAAGCTGGTTCCCGATCATATATCGTTATGGGGCTGCTGCTGAAATCCGGAGGGAGCTTGAGCTTT  
CCATTAATATGAGACCCGCCAAACAAATGGCAGGCATTACGTCGCCGCTTCTGCATCCAAATGATTTTTGACTTCGTGGTGTATTGCGGAGAAC  
AGTGAAGGTGAATACAGTGAAGTTGTGCGGCGCATTACAGAGGCGATGATGAAATCGCCATCCAGAATGCCGTGTTTACGAGAAAAGCGACA  
GAACGTGTCATGCGCTTTGCCTTCGAATT

## Des séquences... et après ?

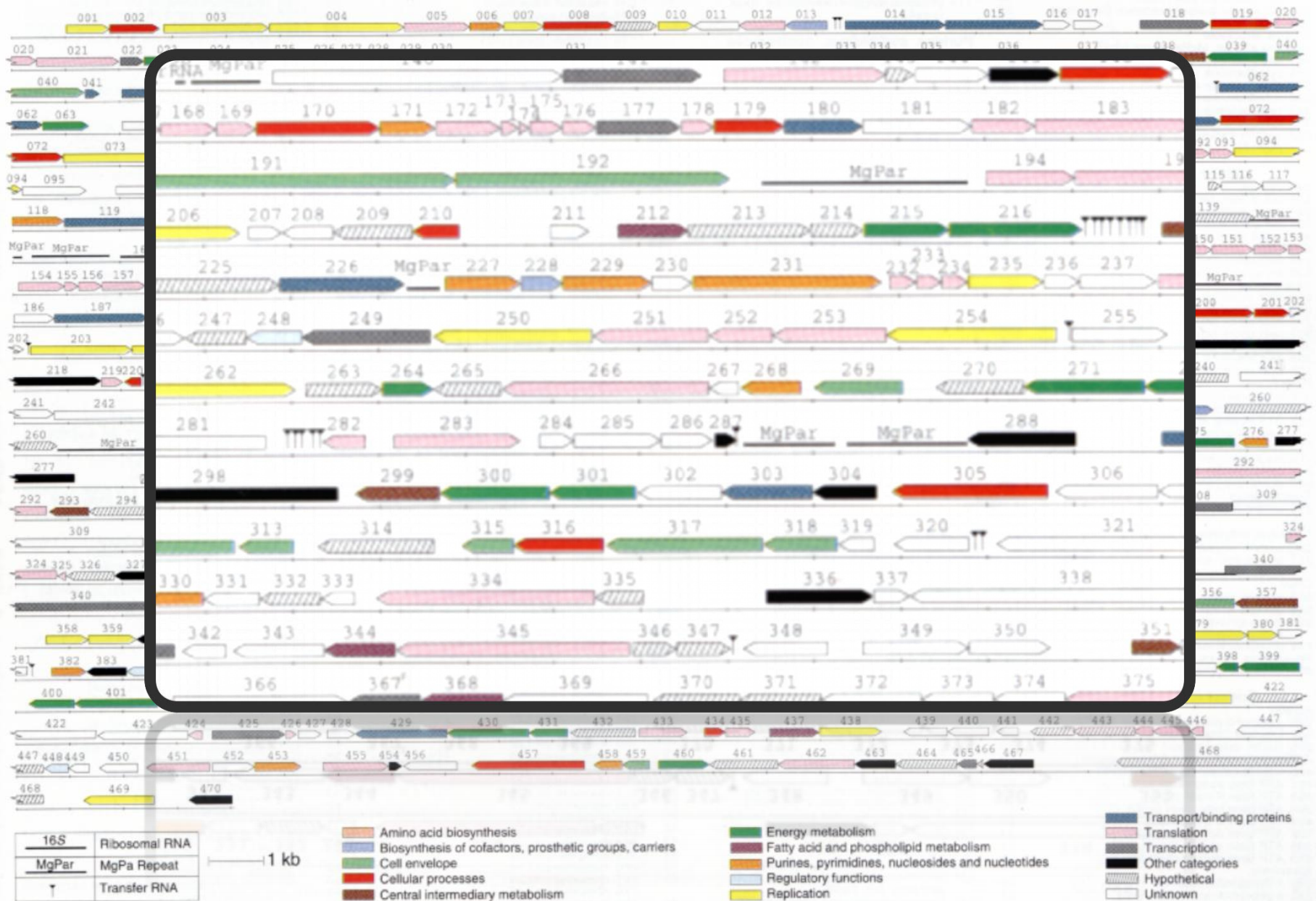
---

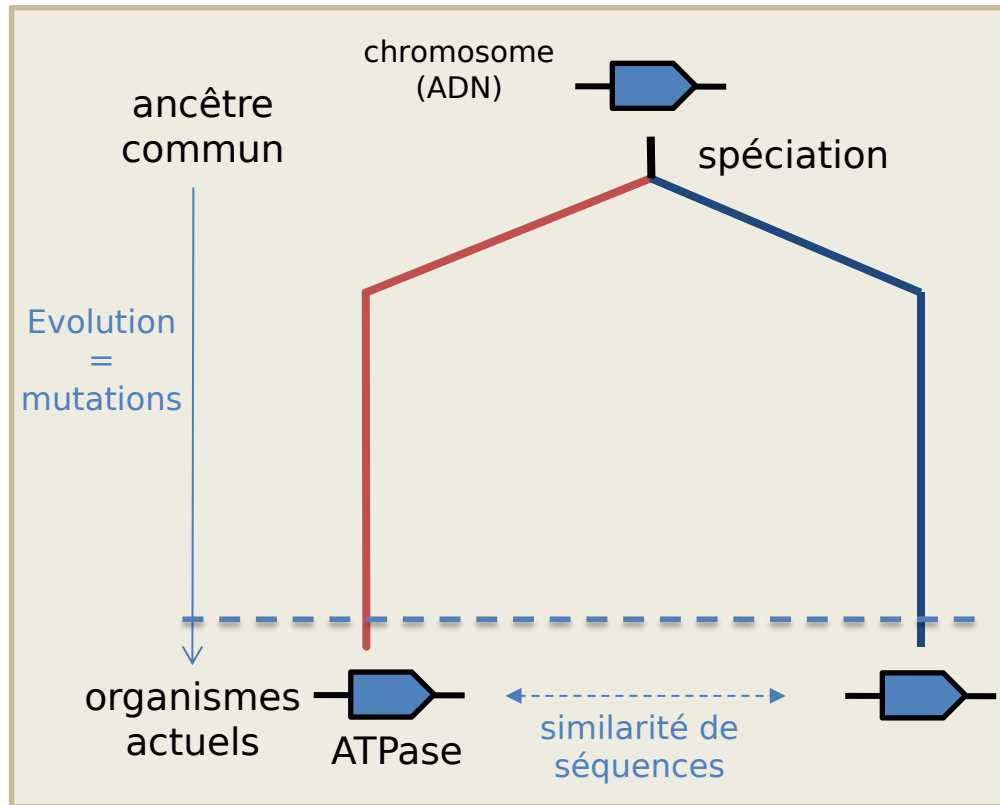
- Annotation
  - régions codantes, régions régulatrice, ...
  - prédiction fonctionnelle
- Reconstruction du réseau métabolique
- Analyse des relations génotype/phénotype
- Analyses évolutives
- Conception de puces d'expression
- Identification de protéine
- Prédiction de structure

- Identification des gènes codant pour :
  - . les ARNr
  - . les ARNt
  - . les protéines
- Identification des unités de transcription (promoteur et terminateur)
- Identification des unités de traduction
- Pour les gènes codant pour les protéines, prédiction fonctionnelle par recherche de similarité de séquences (BLAST) et prédiction de fonction ou classification en grandes classes fonctionnelles (ex: biosynthèse des acides aminés, métabolisme énergétique, ...)

# Génome de *Mycoplasma genitalium*

## Distribution des unités de traduction et classification fonctionnelle





Du temps du séquençage des premiers gènes et génomes

- la conservation/similarité des séquences impliquait des fonctions similaires
- les annotations des gènes caractérisés expérimentalement étaient transférées aux nouveaux gènes/génomes séquencés

## Séquence d'intérêt = Query

Quelles séquences sont proches (suffisamment similaires) et nous indiqueraient la fonction du gène ou de la protéine ?

BLAST : recherche de séquences similaires (Hits) par alignement local de la séquence query avec les séquences d'une banque

**Basic Local Alignment Search Tool**

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

**NEWS**

**BLAST+ 2.10.0 is released - Improved Composition-based statistics.**

We have updated the BLAST process to improve the stability of BLAST results against changes in the number of results requested.

Mon, 23 Dec 2019 16:00:00 EST [More BLAST news...](#)

**Web BLAST**

**Nucleotide BLAST**  
nucleotide ► nucleotide

**blastx**  
translated nucleotide ► protein

**tblastn**  
protein ► translated nucleotide



**Protein BLAST**  
protein ► protein

# BLAST formulaire de requête

**BLAST** <sup>®</sup> >> blastp suite


blastn **blastp** blastx tblastn tblastx


### Enter Query Sequence


Enter accession number(s), gi(s), or FASTA sequence(s)  [Clear](#) [Query subrange](#) 

>OsProt  
 MERNKFASKMSQHYTKTICIAVVLVAVL.FSLSSAAAAGSGAAVSVQLEALLEFKNGVADD  
 PLGVLGWRVKGSGDGAVRGGALPRHCNWTGVACDGAGOVTSIQLPESKLRGALSPFLGN  
 ISTLQVIDLTSNAFAGGIPPQLGRLGELEQLVVSSNYFAGGIPSSLCNCSAMWALALNVN  
 NLTGAIPTSCIGDLSNLEIFAYLNNLDGELPPSMAKLGIMVVDLSCNQLSGSIPPEIGD  
 LSNLQIQLYENRFSGHIPRELGRCKNLTLNIFSNGETGEIPGELGELTNLEVMRLYKN  
 ALTSETPRSLRRCVSLLNLDLSMNQAGTIPPELGEPLSRLSLHANRLAGTVPASLTLN  
 IVMII TTI ELCEMHI GCDI DACTGCI DNI DD I TVNMML SCOTDACTCM/TQI AMACSEM


From   
 To


Or, upload file  No file selected. 

Job Title   
 Enter a descriptive title for your BLAST search 

Align two or more sequences 

### Choose Search Set

Database  


Organism [Optional](#)   exclude   
 Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. 

Exclude [Optional](#)  Models (XM/XP)  Non-redundant RefSeq proteins (WP)  Uncultured/environmental sample s

### Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterative BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm 

Search database nr using Blastp (protein-protein BLAST)  
 Show results in a new window

# BLAST résultats

**Job Title** OsProt  
**RID** [2ANVU5N101R](#)  
*Search expires on 01-21 14:32 pm*

[Download All](#) ▾

**Program** BLASTP [?](#) [Citation](#) ▾

**Database** nr [See details](#) ▾

**Query ID** lcl|Query\_91880

**Description** OsProt

**Molecule type** amino acid

**Query Length** 1183

**Other reports** [Distance tree of results](#) [Multiple alignment](#)  
[MSA viewer](#) [?](#)

### Filter Results

**Organism** *only top 20 will appear*  exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

---

**Percent Identity**  to  **E value**  to  **Query Coverage**  to

[Filter](#) [Reset](#)

- Descriptions
- Graphic Summary
- Alignments
- Taxonomy

**Sequences producing significant alignments** Download ▾ Manage Columns ▾ Show  [?](#)

select all *100 sequences selected* [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">LRR receptor-like serine/threonine-protein kinase FLS2 [Oryza sativa Japonica Group]</a>	2372	2372	100%	0.0	100.00%	<a href="#">XP_015634951.1</a>
<input checked="" type="checkbox"/>	<a href="#">OSJNBa0058K23.7 [Oryza sativa Japonica Group]</a>	2357	2357	99%	0.0	100.00%	<a href="#">CAE02151.2</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein OsJ_16186 [Oryza sativa Japonica Group]</a>	2352	2352	99%	0.0	99.91%	<a href="#">EAZ32006.1</a>
<input checked="" type="checkbox"/>	<a href="#">H0313F03.16 [Oryza sativa]</a>	2326	2326	99%	0.0	98.89%	<a href="#">CAH68341.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein OsI_17436 [Oryza sativa Indica Group]</a>	2055	2055	99%	0.0	89.59%	<a href="#">EEC78020.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: LRR receptor-like serine/threonine-protein kinase FLS2 [Oryza brachyantha]</a>	1855	1855	96%	0.0	85.69%	<a href="#">XP_015691635.1</a>
<input checked="" type="checkbox"/>	<a href="#">PH01B019A14.19 [Phyllostachys edulis]</a>	1638	1638	99%	0.0	73.86%	<a href="#">CCI55350.1</a>
<input checked="" type="checkbox"/>	<a href="#">unnamed protein product [Triticum turgidum subsp. durum]</a>	1615	1615	96%	0.0	73.02%	<a href="#">VAH36634.1</a>



# Banque de séquences protéiques : UniProt

UniProt  Advanced

BLAST Align Retrieve/ID mapping Peptide search Help Contact

## UniProtKB - Q0JA29 (FLS2\_ORYSJ)

Basket

### Display

BLAST Align Format Add to basket History

Help video Add a publication Feedback

- Entry
- Publications
- Feature viewer
- Feature table

**Protein** | LRR receptor-like serine/threonine-protein kinase FLS2

**Gene** | FLS2

**Organism** | *Oryza sativa subsp. japonica* (Rice)

**Status** | Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level<sup>i</sup>

### Function<sup>i</sup>

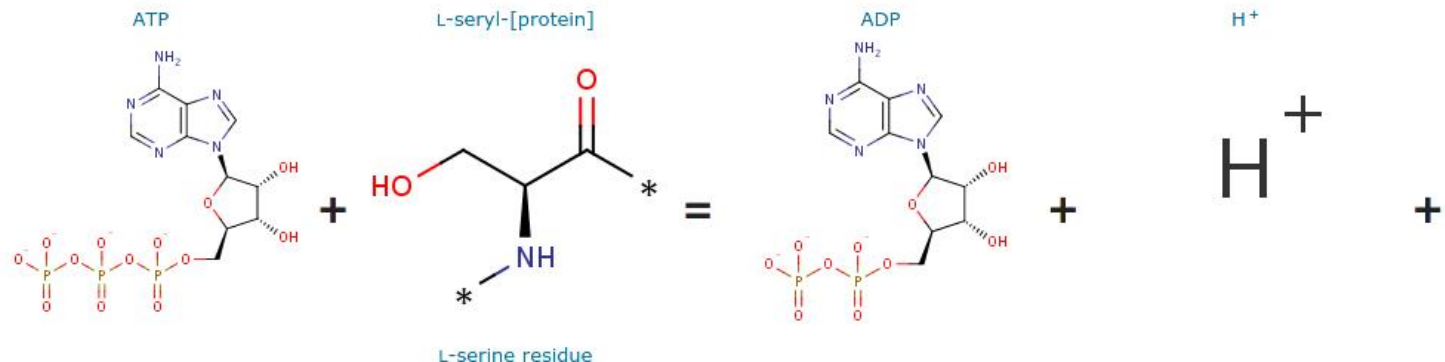
Constitutes the pattern-recognition receptor (PPR) that determines the specific perception of flagellin (flg22), a potent elicitor of the defense response to pathogen-associated molecular patterns (PAMPs). Recognizes flg22 from *Pseudomonas aeruginosa* and *Acidovorax avenae*. flg22 is a peptide derived from the bacterial flagellin N-terminus sequence (PubMed:18986259, PubMed:25617720). Does not recognize flg22 from *Xanthomonas oryzae* pv. *oryzae* (Xoo) or *Xanthomonas oryzae* pv. *oryzicola* (Xoc) (PubMed:25617720). 2 Publications

### Catalytic activity<sup>i</sup>

•  $\text{ATP} + \text{L-seryl-[protein]} = \text{ADP} + \text{H}^+ + \text{O-phospho-L-seryl-[protein]}$

EC:2.7.11.1

Source: Rhea. [« Hide](#)



None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

[▲ Top](#)

- Pression de sélection → conservation de séquence
- Totalité de la séquence mais plus souvent certaines régions → domaines protéiques
  - ex : domaine de liaison à l'ADN, domaine d'interaction avec une autre protéine
- Banques de domaines protéiques
- Méthode de détection des domaines présents sur une séquence donnée

Family: *Pkinase* (PF00069)

 9546 architectures 349448 sequences 73 interactions 7104 species 4704 structures

## Summary

## Domain organisation

## Domain organisation

Below is a listing of the unique domain organisations or architectures in which this domain is found. [More...](#)

## Clan

## Alignments

## HMM logo

## Trees

## Curation &amp; model

## Species

## Interactions

## Structures

Jump to... 

There are **202605** sequences with the following architecture: **Pkinase**

[Z4YHY2\\_DANRE](#) [Danio rerio (Zebrafish) (Brachydanio rerio)] Serine/threonine-protein kinase pim-2 {ECO:0000313|Ensembl:ENSDARP00000076422} (310 residues)



[Show](#) all sequences with this architecture.

There are **11729** sequences with the following architecture: **Pkinase x 2**

[W4GT79\\_9STRA](#) [Aphanomyces astaci] CMGC/CK2 protein kinase {ECO:0000313|EMBL:ETV82199.1} (659 residues)



[Show](#) all sequences with this architecture.

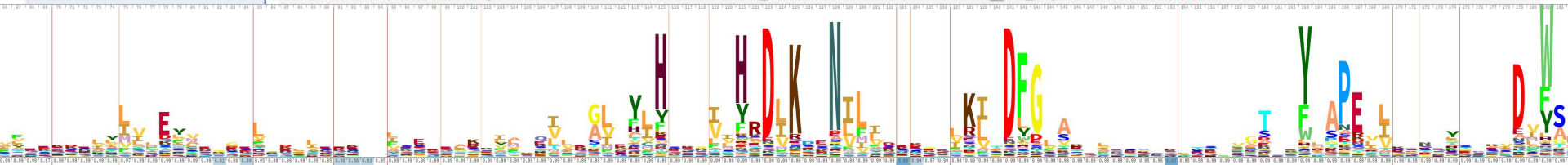
There are **3306** sequences with the following architecture: **Pkinase, Pkinase\_C**

[X0KFA3\\_FUSOX](#) [Fusarium oxysporum f. sp. cubense tropical race 4 54006] Non-specific serine/threonine protein kinase {ECO:0000256|SAAS:SAAS01030073} (635 residues)



[Show](#) all sequences with this architecture.

There are **2855** sequences with the following architecture: **Lectin\_legB, Pkinase**



# Détection de domaines

LRR receptor-like serine/threonine-protein kinase FLS2 (Q0JA29) - protein - InterPro - Mozilla Firefox

https://www.ebi.ac.uk/interpro/protein/UniProt/C... 90% taxid 224324

## InterPro Classification of protein families

Home Search **Browse** Results Release notes Download Help About

### Protein family membership

None predicted

### Entry matches to this protein

Colour By: Accession

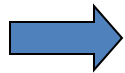
1 100 200 300 400 500 600 700 800

500

IPR000719  
**Protein kinase domain**  
D InterPro domain  
876 - 1180

- Domain**
  - IPR000719 SM00220 PS50011 PF00069
  - IPR013210 PF08263
- Homologous Superfamily**
  - IPR011009 SSF56112
  - IPR032675 G3DSA:3.80.10.10
- Repeat**
  - IPR001611 PF13855
  - IPR003591 SM00369
- Active Site**
  - IPR008271 PS00108
- Unintegrated**
  - G3DSA:1.10.510.10
  - G3DSA:3.30.200.20
  - PRO00019
  - PTHR27000
  - PTHR27000:SF616
  - SSF52047
  - SSF52058
- Other Features**
  - SIGNAL\_PEPTIDE\_N\_REGION
  - SIGNAL\_PEPTIDE
  - SIGNAL\_PEPTIDE\_C\_REGION
  - SIGNAL\_PEPTIDE\_H\_REGION

Transcriptome : ensemble des ARNm ou transcrits présents dans une population de cellules dans des conditions données.



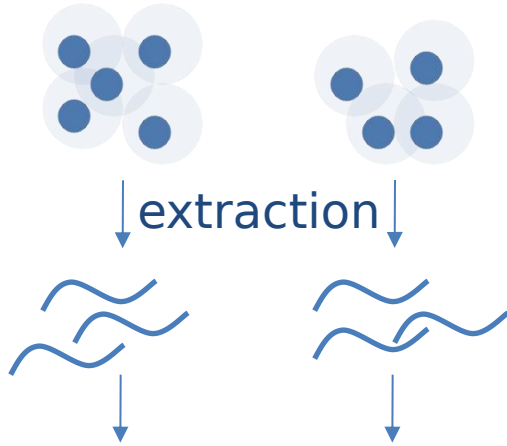
Accès au niveau d'expression de milliers de gènes simultanément (potentiellement l'ensemble des gènes d'un organisme)  
= *instantané* de l'état d'une cellule ou d'une population de cellules

Données d'expression des gènes obtenues par :

- qPCR
- Puces à ADN
- Séquençage ultra-haut débit

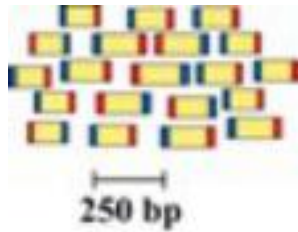
# Transcriptome : acquisition des données (RNAseq)

Échantillon 1    Échantillon 2



ARNm

Préparation de la librairie  
(sélection des fragments,  
ajout des linkers, ...)

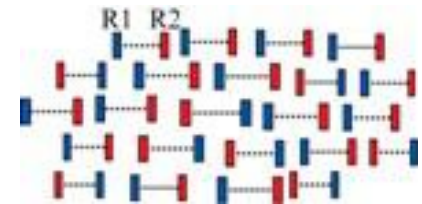
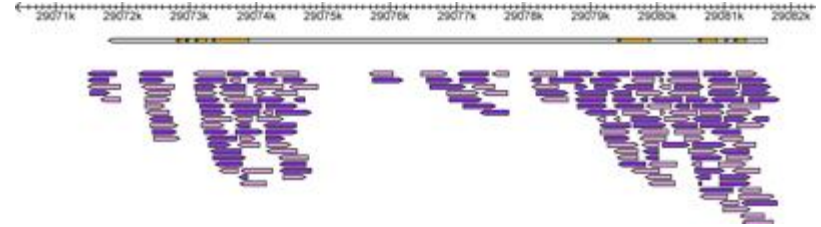


ADNc



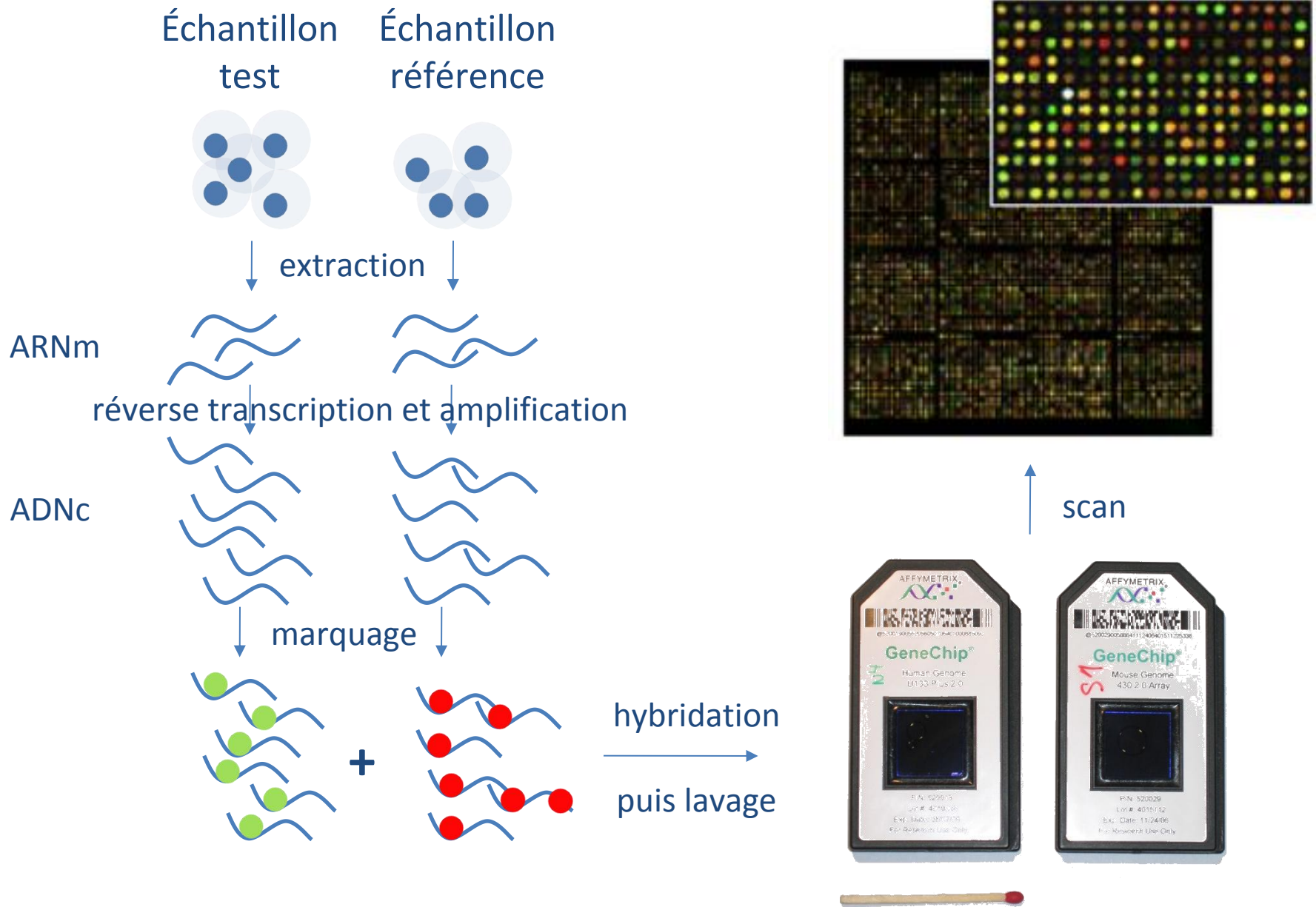
séquençage

Traitement des données  
(assemblage, alignement  
sur le génome,  
normalisation...)



(centaines de)  
millions de séquences  
(reads)

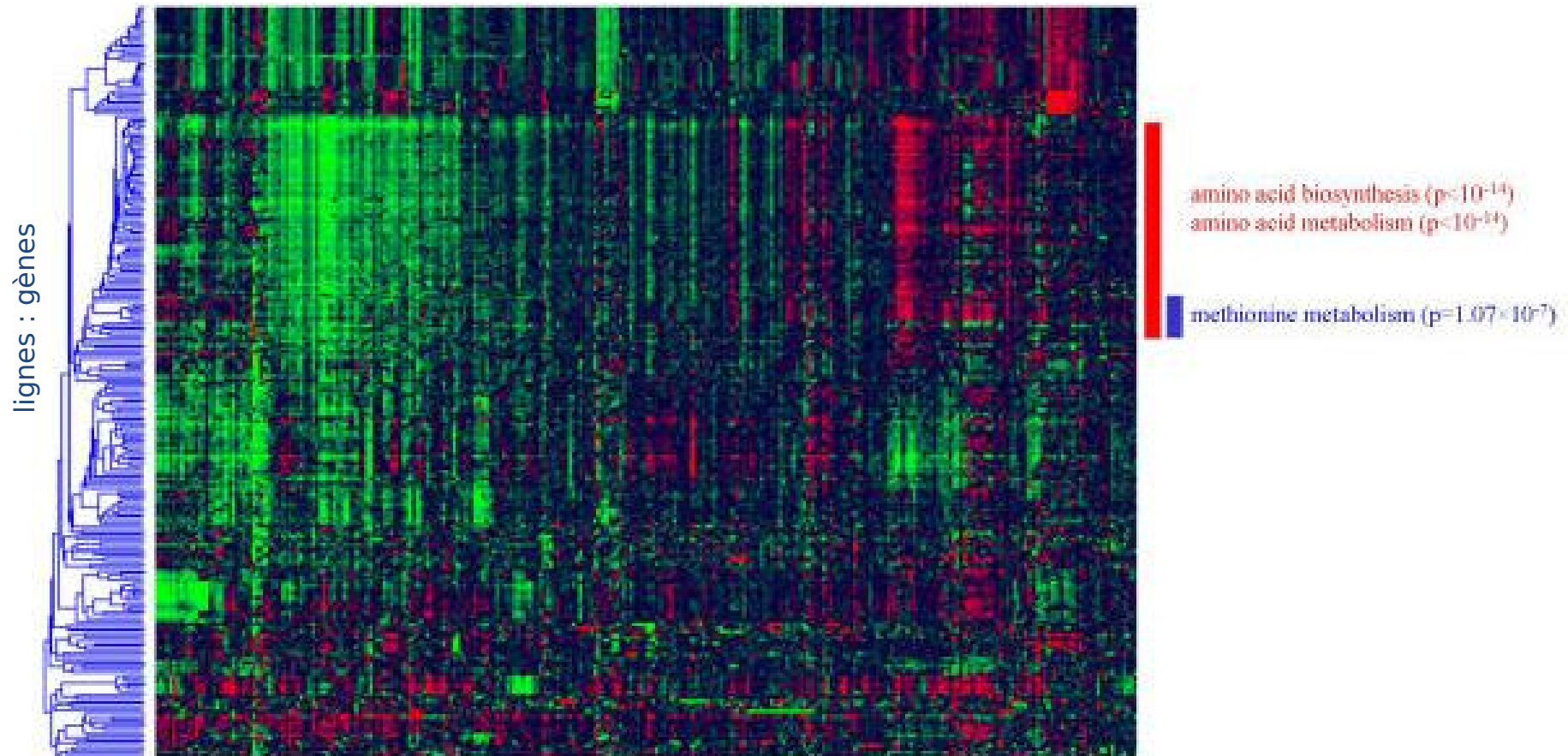
# Transcriptome : acquisition des données (microarray)



# Transcriptome : gènes co-exprimés

- Motivation : les gènes ayant des profils d'expression similaires sont potentiellement co-régulés et participeraient donc à un même processus biologique
- But : regrouper les gènes impliqués dans un même processus biologique

colonnes : condition expérimentales (ex: mutants)



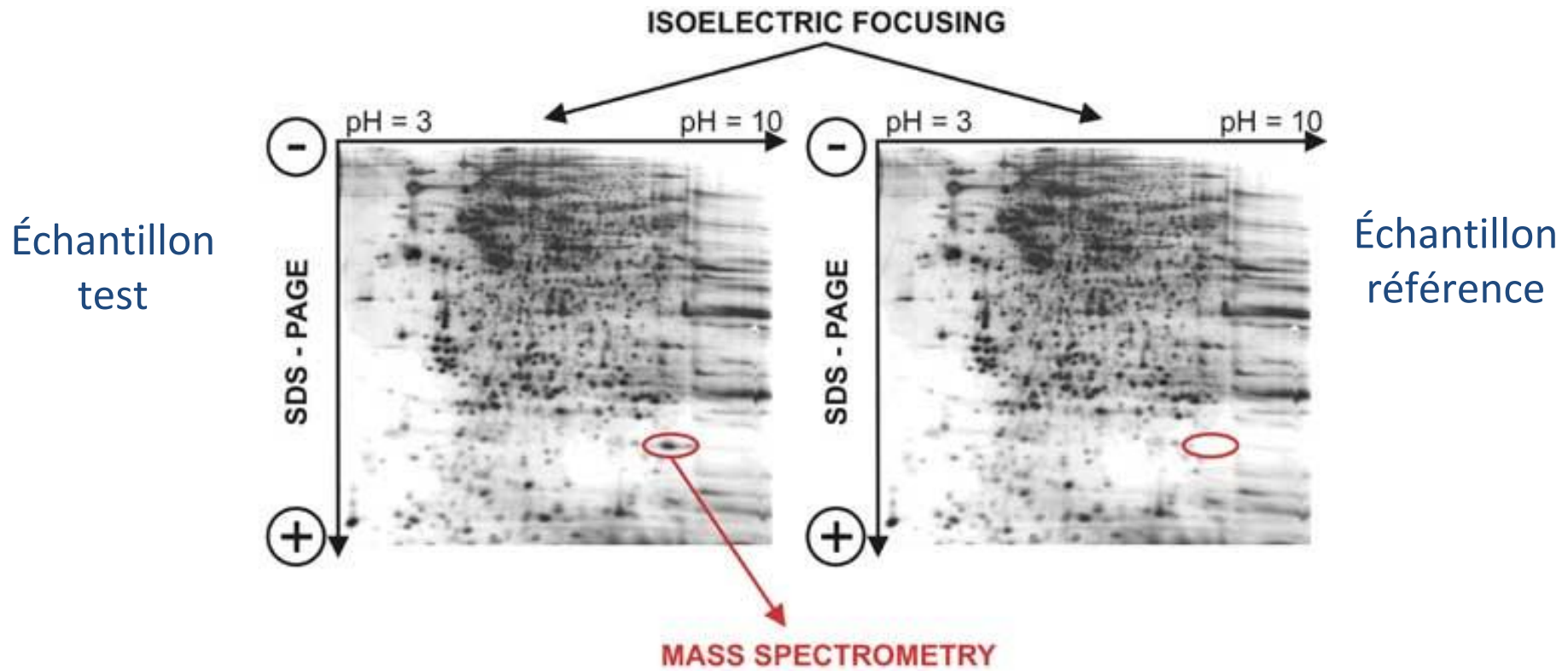


# Protéomique

Protéome : ensemble des protéines exprimées dans une cellule, une partie d'une cellule (membranes, organites) ou un groupe de cellules (organe, organisme, groupe d'organismes) dans des conditions données et à un moment donné.

= *instantané* de l'état d'une cellule ou d'une population de cellules

Séparation des protéines par gels d'électrophorèse (1D, 2D) puis identification des spots par spectrométrie de masse

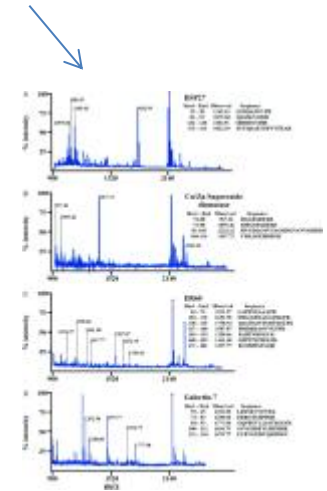
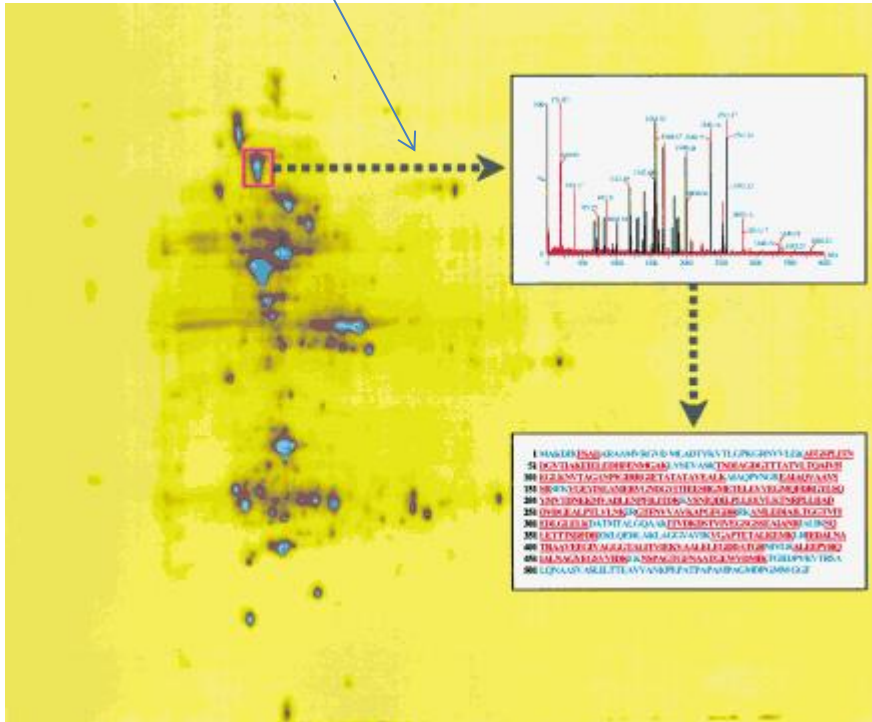


# Protéomique : Identification de protéine

Digestion du spot par une enzyme (ex: trypsine) et mesure du poids des peptides obtenus

Digestion *in silico* du protéome

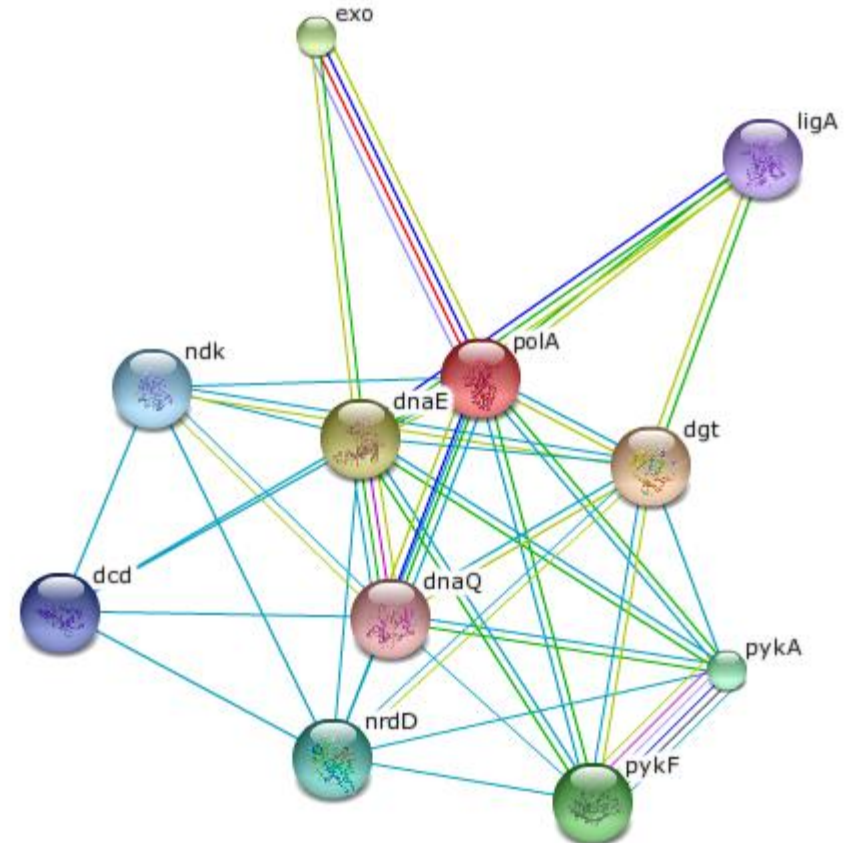
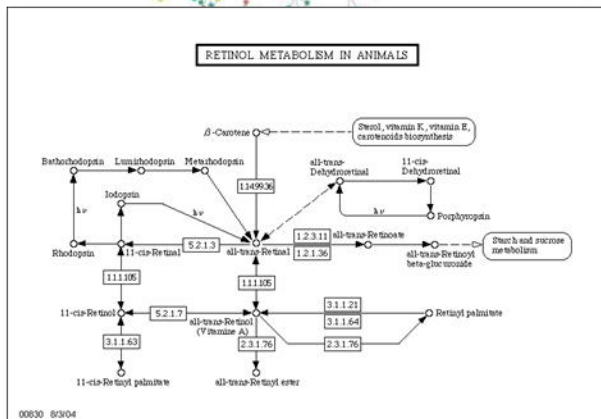
Recherche des protéines correspondant au profil observé



# Réseaux de gènes et de protéines

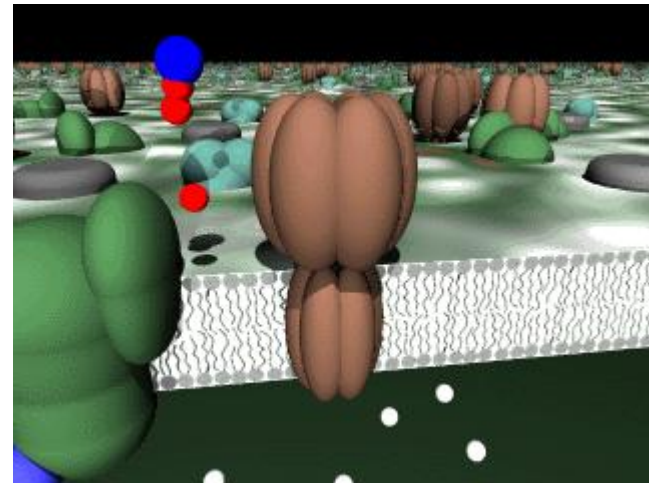
Réseaux :

- d'interactions protéine - protéine, génétiques, fonctionnelles, ...
- de régulation des gènes
- métabolisme (enzymes – substrats)
- transduction du signal



## Intégration et synthèse des connaissances

- modélisation d'un système
  - processus biologique (respiration)
  - organite (mitochondrie)
  - cellule
  - population
  - écosystème



A terme : simulation d'une cellule/organisme/population/écosystème et prédiction de son comportement

## Des observations au modèle

- observations → mesures
- stockage → élaboration d'un schéma de base de données et création de la base de données
- analyses statistiques + recherche d'informations complémentaires (Web)
  - interprétation
- modélisation → modèle mathématique (réseau de pétri)
- hypothèses et simulations

# Concepts et méthodes abordés dans cette UE

- Traitement d'images
  - introduction au traitement d'images
  - segmentation
  - **macros**
- Bases de données
  - introduction aux bases de données
  - conception d'un schéma de base de données
  - utilisation (**requêtes SQL**)
- Traitement de données
  - des statistiques pour des questions biologiques
  - environnement **R**
  - graphiques - corrélation - tests statistiques
- Bioanalyse
  - aperçu de ressources disponibles
  - **banques** de données publiques (ex: UniProt)
- + serveurs d'**analyses** (ex: BLAST)
  - annotations existantes, recherche de domaines sur une séquence, recherche de séquences similaires
  - synthèse des connaissances et observations → esprit critique
- Modélisation
  - introduction à la biologie des systèmes et à la modélisation
  - validation ou remise en question d'une hypothèse / prédiction
  - propriétés émergentes
  - **réseau de pétri**

