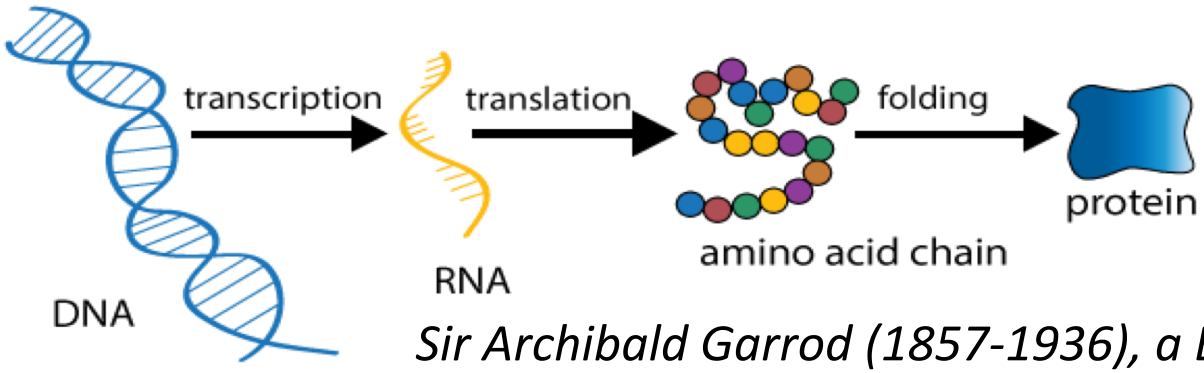
Workshop inter-CATI autour des omics

Transcriptomic Resources
Véronique Brunaud

Réunion inter-CATI du 17 oct 2019

Transcription – Translation

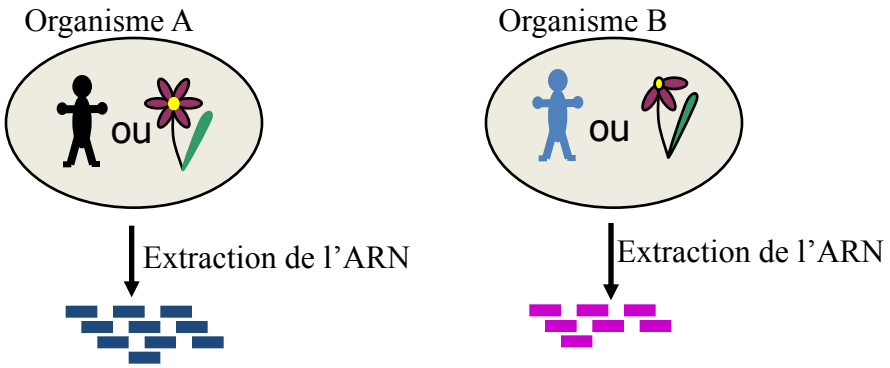


Sir Archibald Garrod (1857-1936), a British medical doctor, was the first to suggest that genes were connected to enzymes : "one gene encode a single enzyme"

Transcriptome = mesurer les ARNs produits par la plante, dans la plante entière, un organe, ou une cellule à un temp t de son développement et dans des conditions définies.

→ **Comparaisons des transcriptomes**

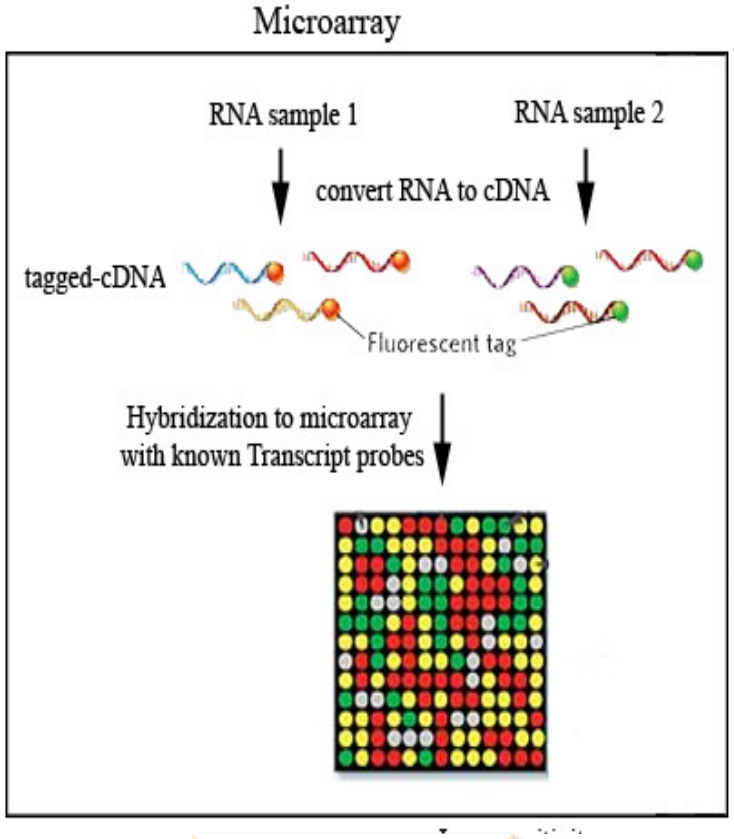
Transcriptomic comparisons

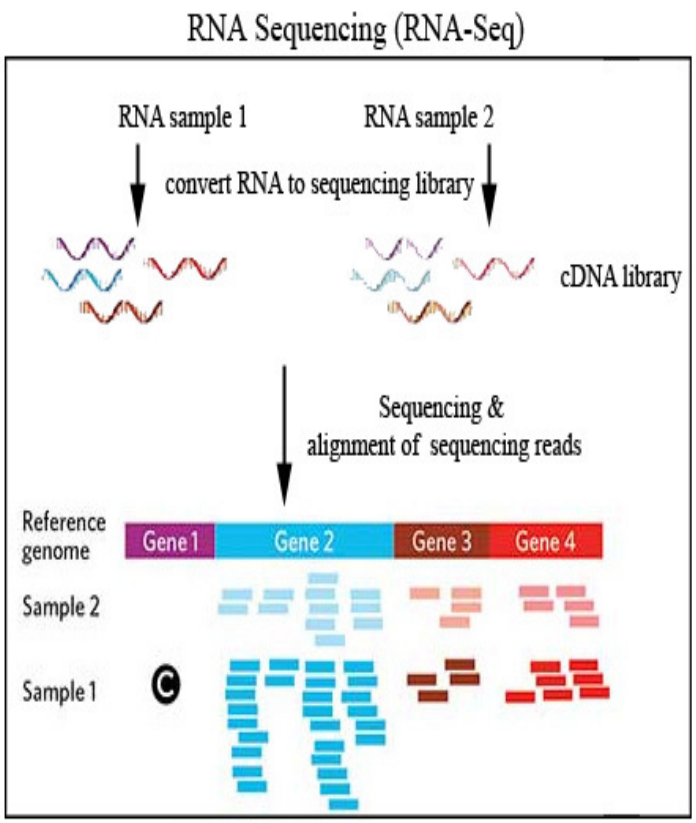


- Comparaison entre des échantillons d'organes différents d'une plante : par exemple fleur et racine.
- Comparaison entre des échantillons de génotypes différents : mutant / sauvage ou l'étude de plusieurs génotypes d'une même espèce.
- Comparaison d'échantillons de plantes cultivées dans 2 conditions : des plantes traitées versus plantes non traitées, effet d'une sécheresse, effet d'un pathogène…
- Comparaison d'échantillons prélevés à des stades de développements différents ou cinétique

….

Micro-Puces & RNA-Seq





Hybridation

Puces d'ADN: 1990

Séquençage

RNA-Seq: 2006

Différences majeures entre RNA-Seq et puces à ADN

- Pas d'obligation d'utiliser une annotation
- Pas besoin d'avoir toutes les comparaisons dès le début (design complet en théorie !)

Project management for RNA-Seq

Biological question
NEW PROJECT

CATdb

DATABASE

Sample descriptions
Conditions, Treatments

International standard

GEO NCBI Gene Expression Omnibus

RNA extraction
Library construction
Sequencer

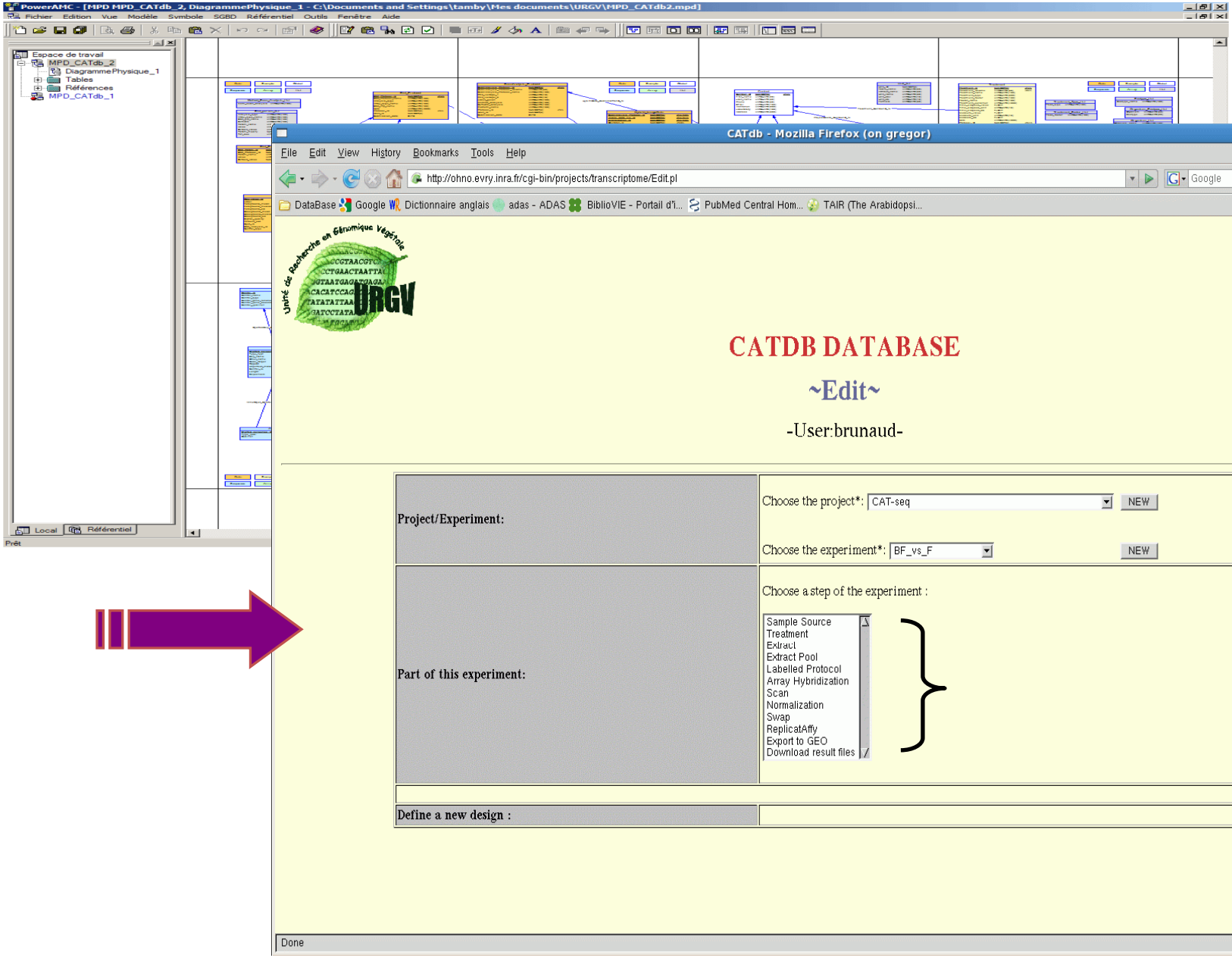
Quality process

Bioinfo and stats analyses

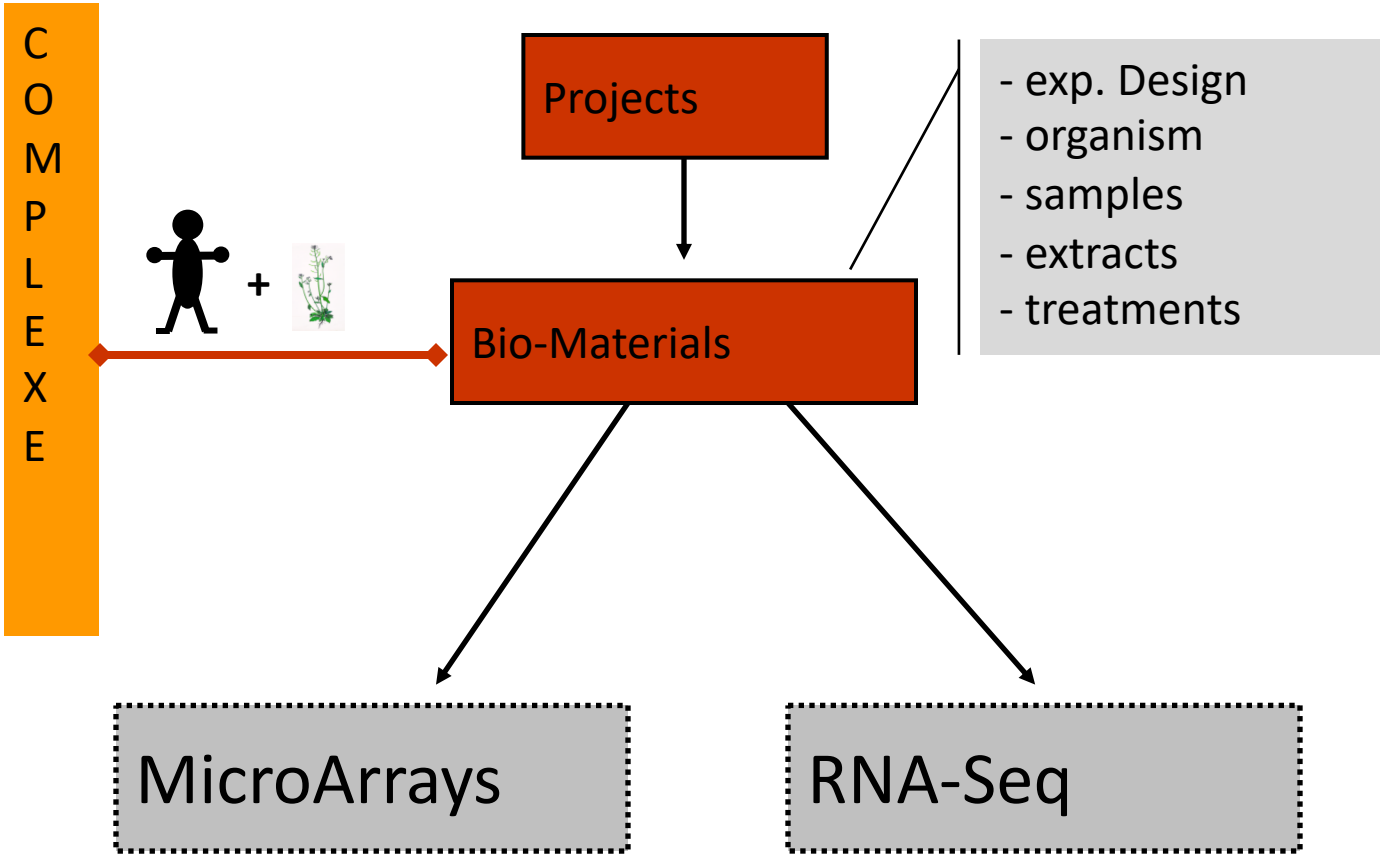
IBiSA

CNRS

CATdb Information system



Experiment and Sample description



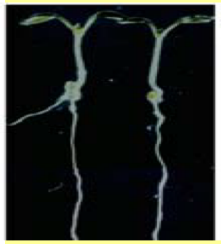
Plan d'expérience : exemple

RA3-4 (INRA2003-04) Elongation cellulaire, paroi primaire

<u>Echantillons</u> :*A. thaliana*, hypocotyles poussés à l'obscurité. Ceux traités à l'isoxabène sont récoltés 3h après le traitement

Exp 1

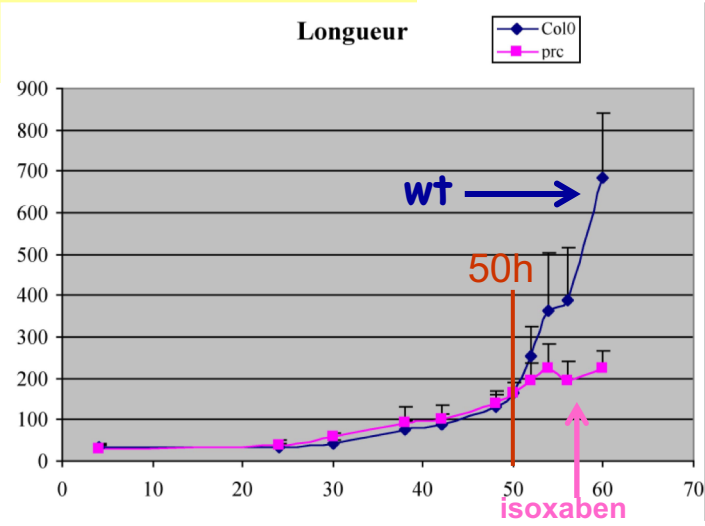
Hyp 45h ⇄ **Hyp 48h** ⇄ **Hyp 52h** ⇄ **Hyp 55h**



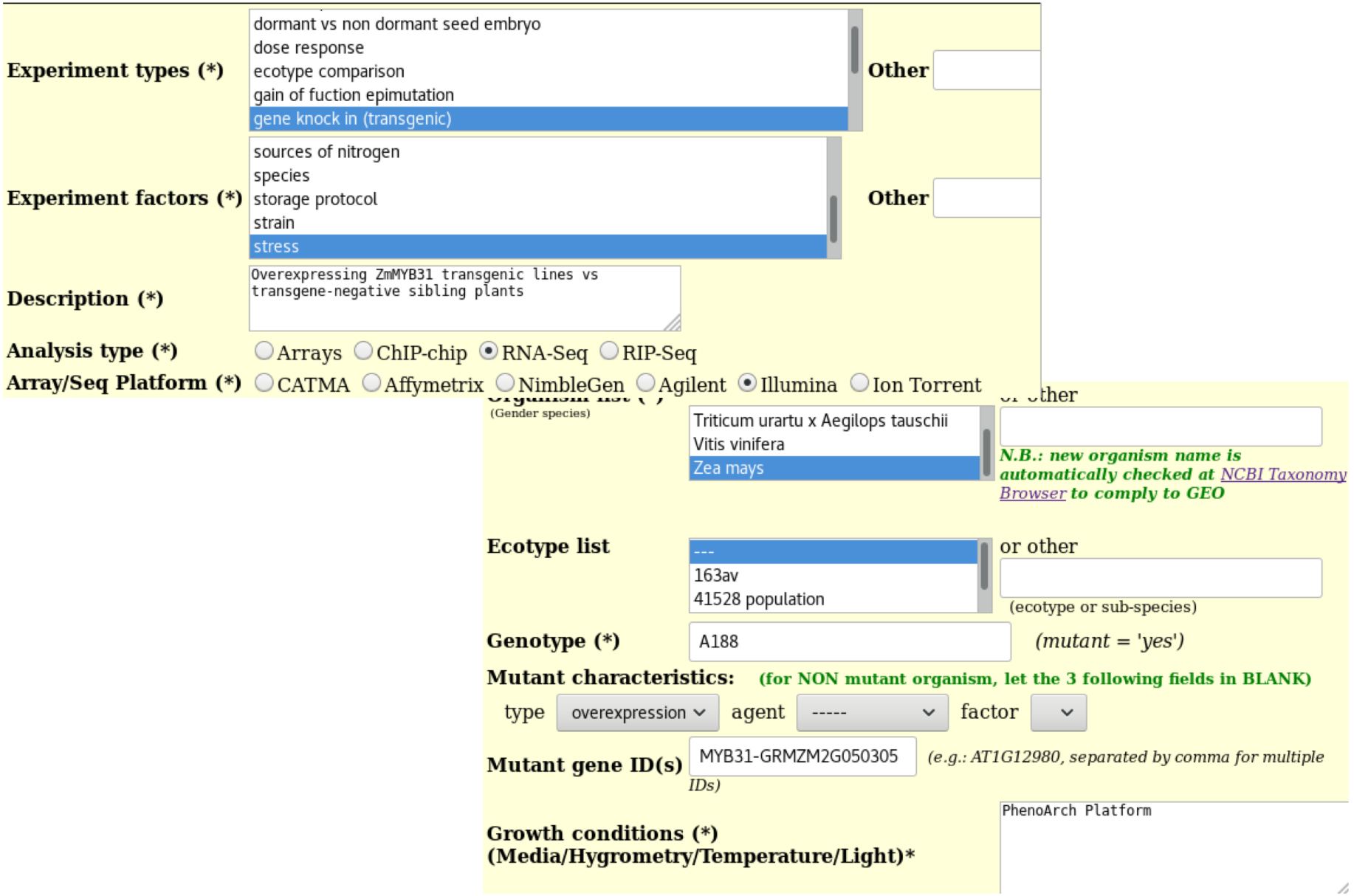
Exp 2

control 45h + 3h ⇄ **Isox 45h + 3h**

control 52h + 3h ⇄ **Isox 52h + 3h**

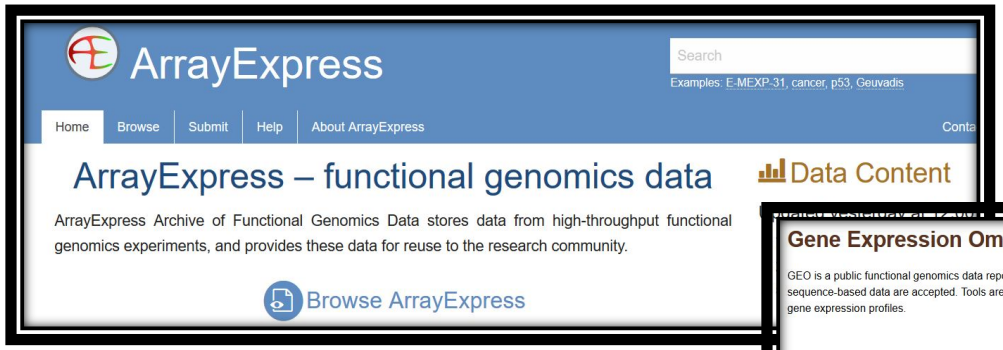


CATdb Information system

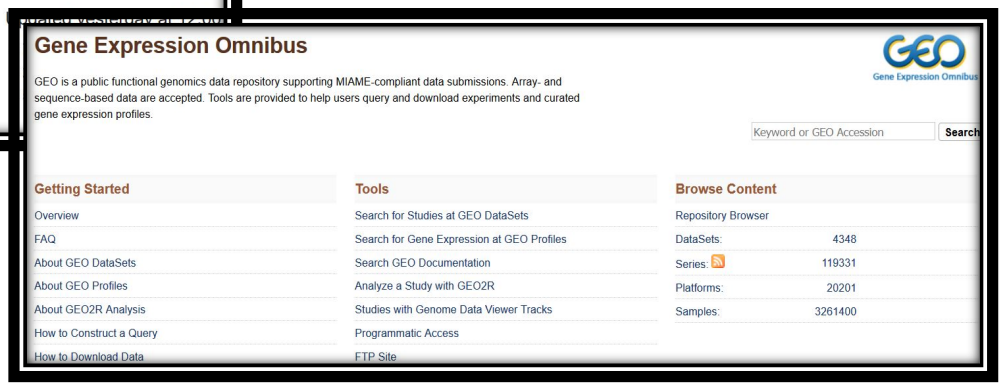


International repository and standards

Resources	Experiments/Samples	Raw data (fastq)
NCBI	GEO (~120000, > 3 million)	SRA
EBI	ArrayExpress (~73000, > 2 million)	ENA
DDBJ		DRA



Collaboration between the 3 databases



The International Nucleotide Sequence Database Collaboration (INSDC)

International repository and standards

MIAME and MINSEQE guidelines

The MIAME (Minimum Information About a Microarray Experiment) and MINSEQE (Minimum Information About a Next-generation Sequencing Experiment) guidelines outline the minimum information that should be included when describing a microarray or sequencing study. Many journals and funding agencies require microarray data to comply with MIAME and MINSEQE standards.

MIAME compliance is not related to the submission format or route, but rather to the content provided

GEO deposit procedures enable and encourage submitters to supply MIAME and MINSEQE compliant data. All GEO submission procedures are designed to closely follow the MIAME and MINSEQE checklists; if you provide all requested information, your submission will be compliant.

The six most critical elements contributing towards MIAME are:

- Raw data for each assay (e.g., CEL or FASTQ files)
- Final processed (normalized) data for the set of assays in the study (e.g., the gene expression data count matrix used to draw the conclusions in the study)
- Essential sample annotation (e.g., tissue, sex and age) and the experimental factors and their values (e.g., compound and dose in a dose response study)
- Experimental design including sample data relationships (e.g., which raw data file relates to which sample, which assays are technical, which are biological replicates)
- Sufficient annotation of the array or sequence features examines (e.g., gene identifiers, genomic coordinates)
- Essential laboratory and data processing protocols (e.g., what normalization method has been used to obtain the final processed data)

International repository and standards

Home | About Us | Mission | Projects | Conferences

MINSEQE

Minimum Information about a high-throughput SEQuencing Experiment

MINSEQE describes the **Minimum Information about a high-throughput nucleotide SEQuencing Experiment** that is needed to enable the unambiguous interpretation and facilitate reproduction of the results of the experiment. By analogy to the MIAME guidelines for microarray experiments, adherence to the MINSEQE guidelines will improve integration of multiple experiments across different modalities, thereby maximising the value of high-throughput research. [MINSEQE version 1.0 (pdf), FGED Society (June 2012)]

!Sample_source_name_ch[n]	1 per channel	any	Briefly identify the biological material and the experimental variable(s), e.g., vastus lateralis muscle, exercised, 60 min.
!Sample_organism_ch[n]	1 or more	use standard NCBI Taxonomy nomenclature	Identify the organism(s) from which the biological material was derived.
!Sample_characteristics_ch[n]	1 or more	'Tag: Value' format	Describe all available characteristics of the biological source, including factors not necessarily under investigation. Provide in 'Tag: Value' format, where 'Tag' is a type of characteristic (e.g. "gender", "strain", "tissue", "developmental stage", "tumor stage", etc), and 'Value' is the value for each tag (e.g. "female", "129SV", "brain", "embryo", etc). Include as many characteristics fields as necessary to thoroughly describe your Samples.
!Sample_biomaterial_provider_ch[n]	0 or more	any	Specify the name of the company, laboratory or person that provided the biological material.
!Sample_treatment_protocol_ch[n]	0 or more	any	Describe any treatments applied to the biological material prior to extract preparation. You can include as much text as you need to thoroughly describe the protocol; it is strongly recommended that complete protocol descriptions are provided within your submission.
!Sample_growth_protocol_ch[n]	0 or more	any	Describe the conditions that were used to grow or maintain organisms or cells prior to extract preparation. You can include as much text as you need to thoroughly describe the protocol; it is strongly recommended that complete protocol descriptions are provided within your submission.
!Sample_molecule_ch[n]	1 per channel	total RNA, polyA RNA, cytoplasmic RNA, nuclear RNA, genomic DNA, protein, or other	Specify the type of molecule that was extracted from the biological material.

Limites par rapport aux normes ou manques de normes
Usage coté données Transcriptomes

Types de données : description des exp, description des échantillons, données générées (RAW data and analyzed)

BD internationales (GEO/ArrayExpress) et normes (Minimum d'informations MIAME,MINSEQE)

Données: raw data dans BD internationales, RNA-Seq analyses: tables de comptages par gènes ou liste de gènes exprimés différentiellement.

Normes permettent :

→ Description des échantillons avec des consignes minimum

→ Liste des champs obligatoires

→ Plus riche dans catdb mais normal puisque sert de LIMS à la PF

ex: protocoles pour toutes les étapes, cahier de laboratoire

Pb rencontrés :

- Pas de mots clés imposés pour décrire le type de comparaisons ou d'expériences.
- Des ontologies pourraient être imposées ex: Plant Ontology

  ex: dans CATdb 33 tissus ou organes contre 2021 termes de la plante ontolgie
- Versioning et normalisation des noms de gènes, très dependant de l'espèce (vérification Taxonomique mais pas des listes de gènes, juste donner la référence ou source)

Plant Ontology