

# **Aide à l'interprétation de résultats de puces à ADN par l'utilisation d'une base de connaissances**

Claude Pasquier      Karim Jevardat de Fombelle      Richard Christen

{claude.pasquier, karim.jevardat-de-fombelle, richard.christen}@unice.fr

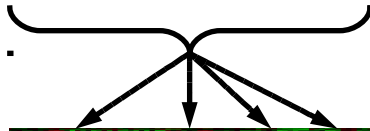
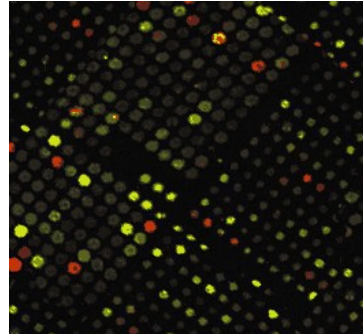
UMR 6078 CNRS-UNSA Laboratoire Jean-Maetz – 06235 Villefranche sur mer

Journées Ouvertes Biologie Informatique Mathématique (JOBIM)  
Saint Malo 10-12 juin 2002

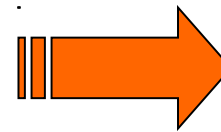
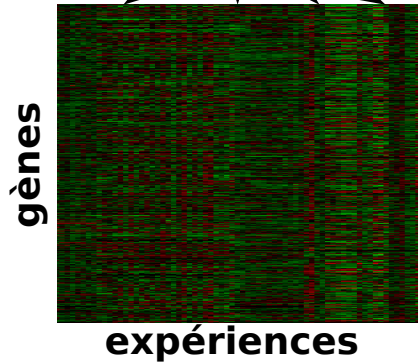
# Un système informatique intégré permettant :

- D'annoter les données issues d'une classification avec des informations biologiques provenant d'une base de connaissances,
- D'enrichir la base de connaissances en fonction des résultats d'expériences

**Expériences effectuées à l'aide de puces à ADN**



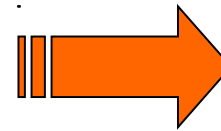
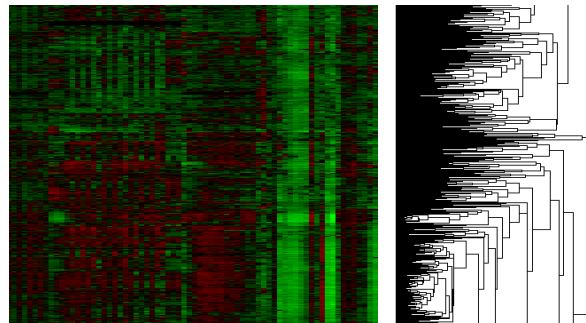
**Collecte des résultats de plusieurs expériences dans un tableau**



**Plusieurs millions de données à analyser**

**Méthode de partitionnement**

**Regroupement des gènes co-exprimés**



**Arbre de partitionnement de plusieurs milliers de feuilles à analyser**

# 3 types d'aides à l'interprétation des résultats

## I. Intégration des bases de données de gènes avec GeneOntology

- Quels transcrits sont caractéristiques d'une ou plusieurs fonctions ?
- Quelles sont les fonctions associées aux produits des gènes ?

## II. Exploration globale du partitionnement

- Les transcrits associés à une fonction donnée sont ils regroupés dans le partitionnement ?
- Des corrélations entre plusieurs fonctions sont elles visibles ?

## III. Focalisation sur l'étude d'un groupe de gènes

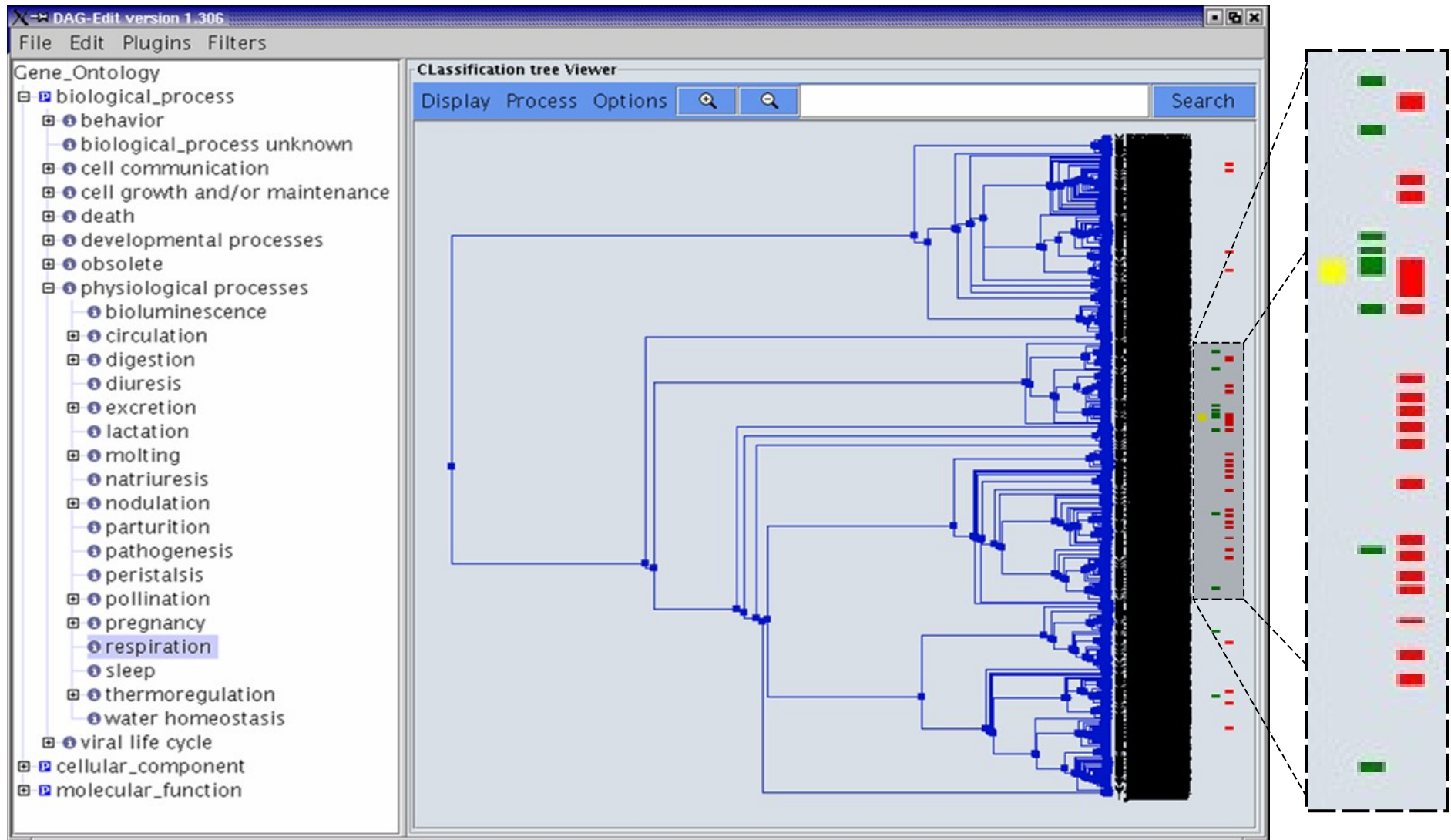
- Peut-on nommer des groupements de gènes déjà connus ?
- Est-il possible d'identifier des co-expressions entre transcrits qui ne reflètent pas la connaissance biologique ?

# I - Connaissances biologiques et gènes associés

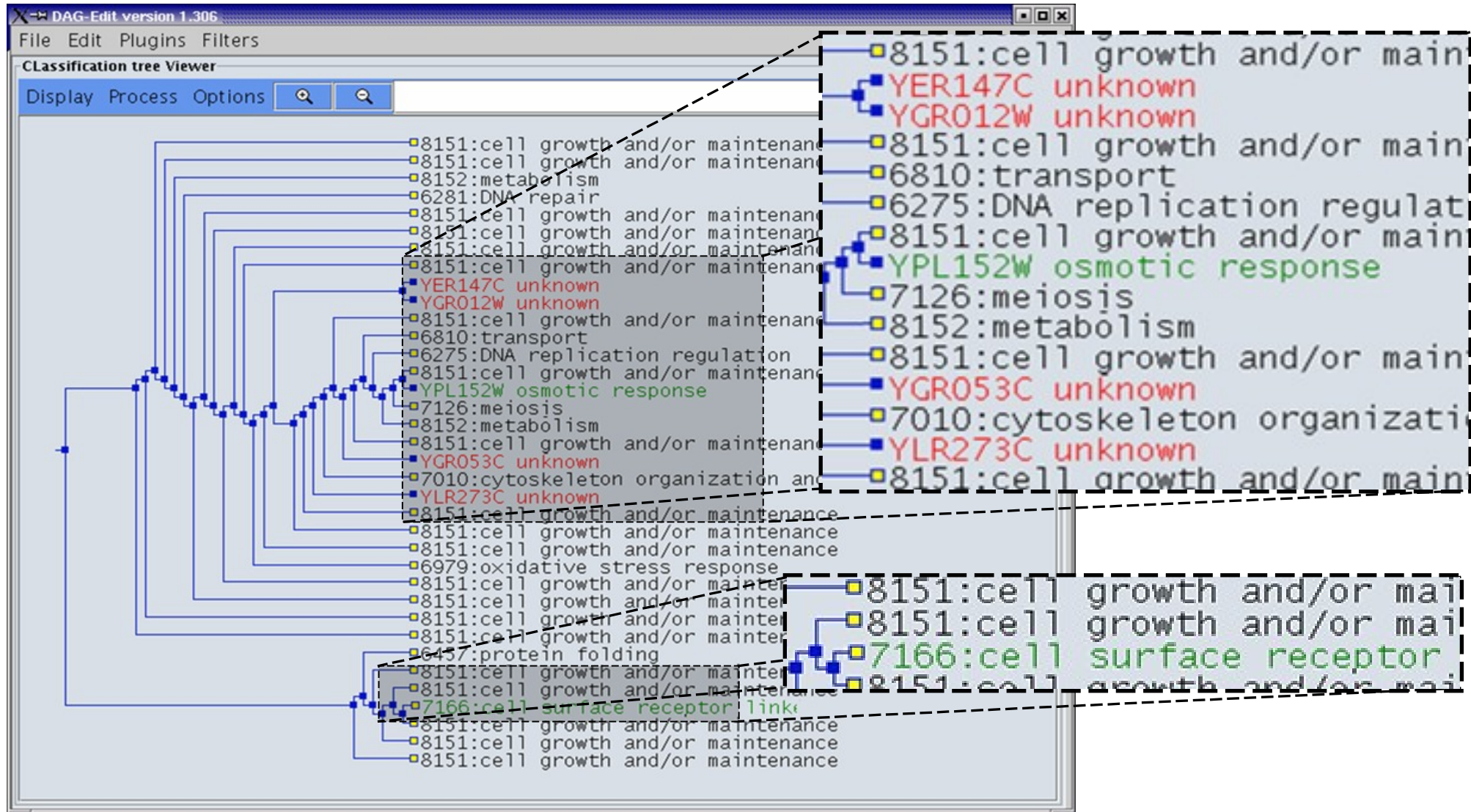
The screenshot displays the DAG-Edit version 1.311 interface, which is used for managing biological data. The main window is divided into several panels:

- Gene\_Ontology:** A hierarchical tree view on the left showing various biological processes. The term "dosage compensation" is selected and highlighted in blue.
- Find terms:** A search panel at the top right with a search bar, dropdown menus for "ID" and "equals", and buttons for "Search all terms", "Search children of selection", "Search selection", and "Case sensitive search". The status "No search performed" is shown.
- DbxrefViewer Plugin:** A panel on the right showing a list of database cross-references (dbxrefs) for the selected term. The list includes identifiers from various databases like SGD, TGI, FB, SP, and TR. Buttons for "Add", "Delete", and "Import" are visible below the list.
- Netscape: Entrez-PubMed:** A browser window in the foreground showing the PubMed search results for the query "10358091". The search results include a link to a paper by Faxiot G, Sergeant A, and Mikaelian I, titled "A new nucleoporin-like protein interacts with both HIV-1 Rev signal and CRM-1".
- GeneDisplayer Plugin:** A panel at the bottom right showing a list of associated genes for the selected term "dosage compensation". The list includes identifiers from various databases. Buttons for "Details" and "All descendant Genes" are visible below the list.

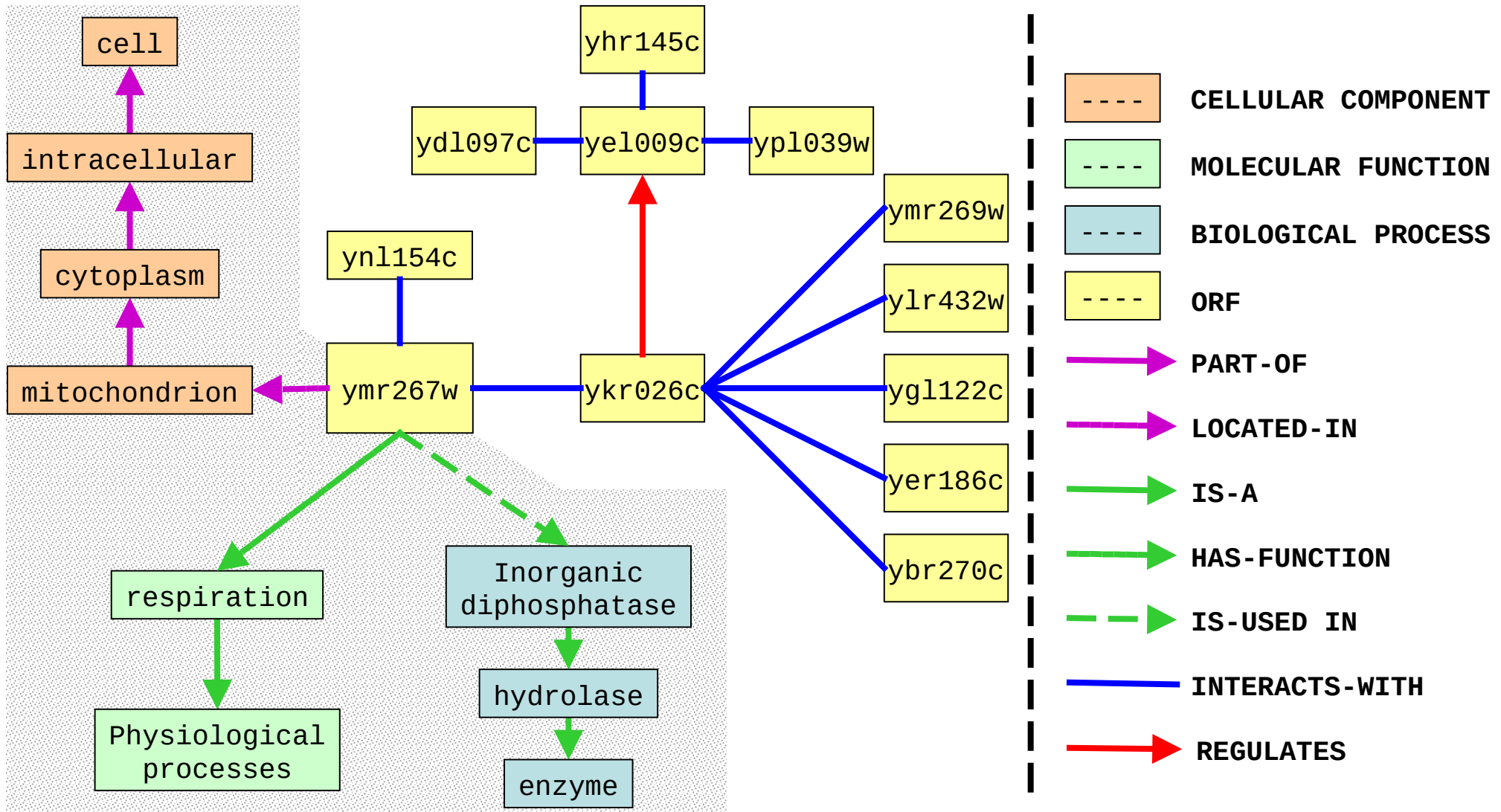
# II - Visualisation de transcrits associés à un terme



# III - Simplification par nommage des branches



# Objectif de la modélisation des connaissances



# Collaborations

## Environnement logiciel

- INRIA – projet OASIS Didier Parigot

## Bases de Connaissances

- INRIA – projet ACACIA Rose Dieng

## Laboratoires partenaires

- CNRS (UMR7637), ESPCI  
Marie-Claude Potier
- CNRS (UMR 6078), UNSA  
Frank Delaunay
- CNRS (UMR 6097 )      Pascal  
Barbry

## Systemes de classification

- Olivier Gascuel à Montpellier
- Alain Guénoche à Marseille

## Manipulation de grands arbres

- IRD - François Chevenet