
Utilisation d'une base de connaissances pour analyser les données des puces à ADN

Claude Pasquier

CNRS (UMR 6078) – Université de Nice

Réunion du groupe Alphy (Alignement et Phylogénie) Lyon 29-30 Avril 2002

Cadre du projet

Equipe Bio-informatique

resp. : Prof. Richard Christen

Analyse des données des puces

resp. : Claude Pasquier
Karim Jevardat de Fombelle (CDD)
Michael Lamarre (Thèse)

Conception des puces

resp. : Prof. Richard Christen
Olivier Croce (Thèse)
Sébastien Flavier (Thèse)

Problème d'interprétation des données

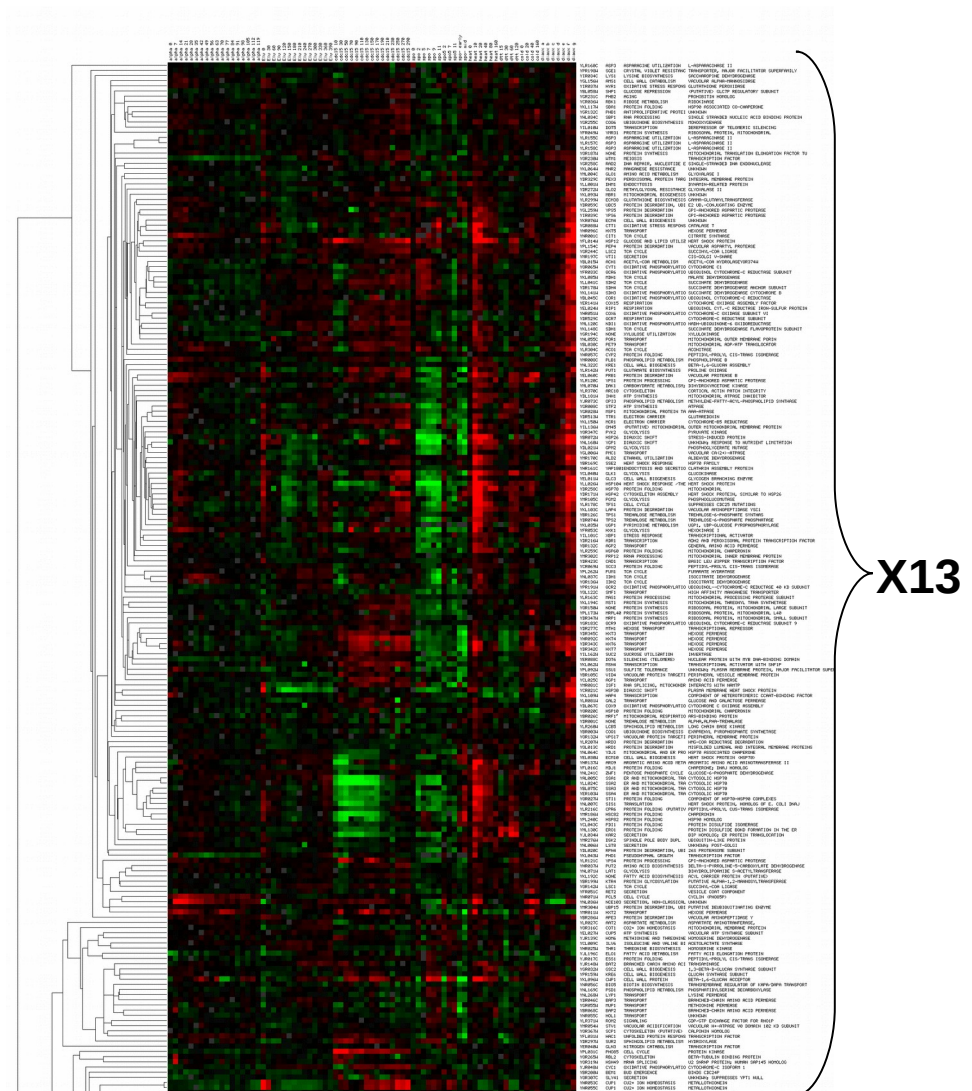
Données extraites de :

Eisen MB, Spellman PT, Brown PO, Botstein D.
Cluster analysis and display of genome-wide
expression patterns. Proc Natl Acad Sci U S A.
1998 Dec 8;95(25):14863-8.

**6221gènes de *Saccharomyces
cerevisiae***

80 expériences

2467 gènes ont une fonction répertoriée



Modélisation hiérarchique des concepts dans GO

GO:0003673 : Gene Ontology (30674)

- ➔
 - ⊕ **P** [GO:0008150 : biological process \(23487\)](#)
 - ⊕ **I** [GO:0008151 : cell growth and/or maintenance \(15526\)](#)
 - ⊕ **I** [GO:0008152 : metabolism \(10826\)](#)
 - ⊕ **I** [GO:0009058 : biosynthesis \(2327\)](#)
 - ⊕ **I** [GO:0009059 : macromolecule biosynthesis \(1684\)](#)
 - ⊕ **I** [GO:0006412 : protein biosynthesis \(1572\)](#)
 - ⊖ **I** [GO:0006411 : protein metabolism and modification \(4159\)](#)
 - ⊕ **I** [GO:0006412 : protein biosynthesis \(1572\)](#)
 - ⊕ **I** [GO:0006461 : protein complex assembly \(111\)](#)
 - ⊕ **I** [GO:0007582 : physiological processes \(695\)](#)
 - ⊖ **I** [GO:0007585 : respiration \(26\)](#)
- ➔
 - ⊕ **P** [GO:0005575 : cellular component \(14402\)](#)
 - ⊕ **I** [GO:0005623 : cell \(11523\)](#)
 - ⊕ **P** [GO:0005622 : intracellular \(10239\)](#)
 - ⊕ **P** [GO:0005737 : cytoplasm \(5373\)](#)
 - ⊕ **P** [GO:0005739 : mitochondrion \(1433\)](#)
- ➔
 - ⊕ **P** [GO:0003674 : molecular function \(21459\)](#)
 - ⊕ **I** [GO:0005215 : transporter \(1920\)](#)
 - ⊖ **I** [GO:0005372 : water transporter \(5\)](#)

1

2

3

Code « Evidence » de Gene Ontology

TAS

Traceable Author Statement

IDA

Inferréd from Direct Assay

IMP

Inferréd from Mutant Phenotype

IGI

Inferréd from Genetic Interaction

IPI

Inferréd from Physical Interaction

ISS

Inferréd from Expression Pattern

IEP

Inferréd from Sequéce or structural Similarity

NAS

Non-traceable Author Statement

IEA

Inferréd from Electronic Annotation

NR

Not Recordéd

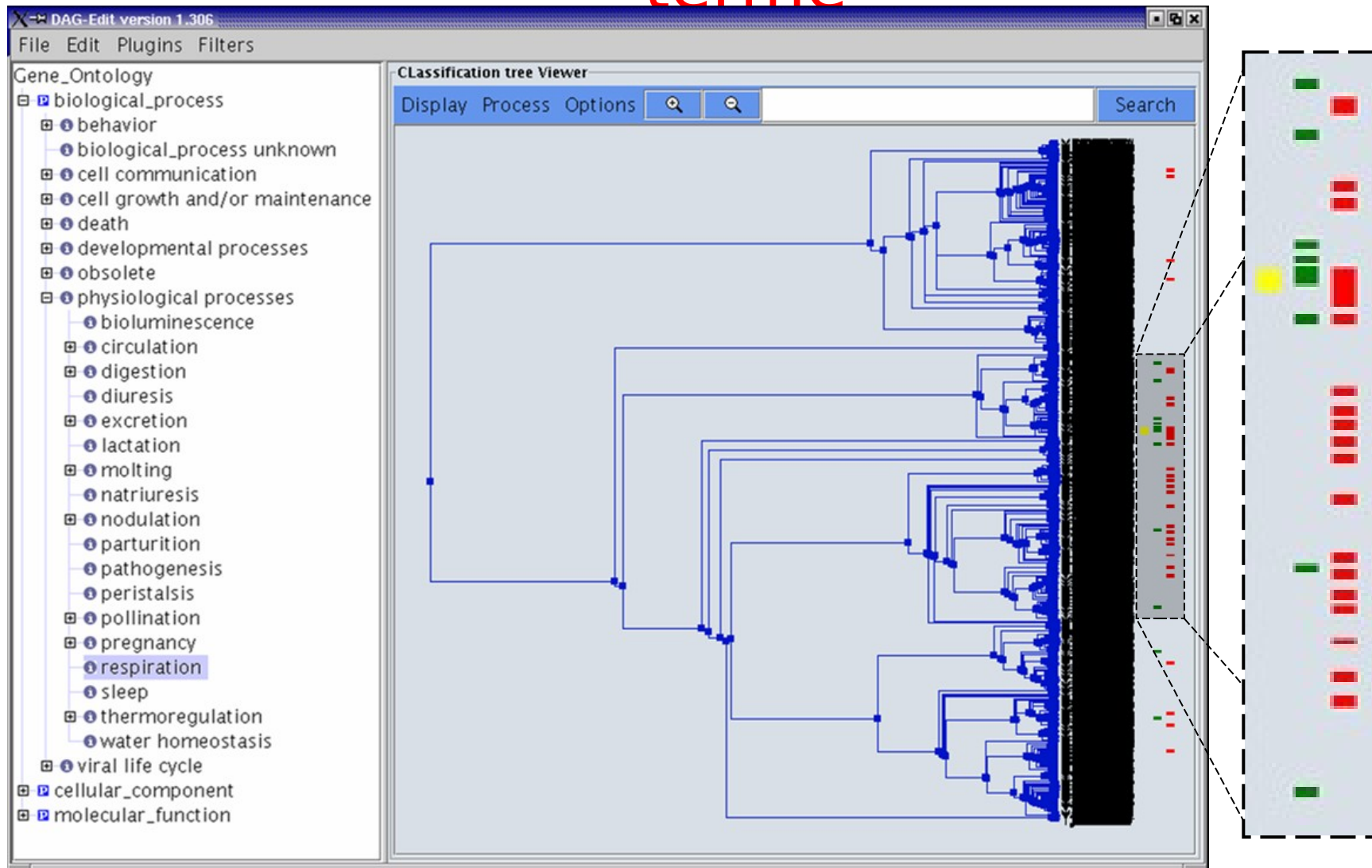
Objectifs du projet

1. utiliser les termes biologiques pour qualifier les résultats de puces à ADN
2. utiliser les résultats des puces pour étendre une base de connaissances
3. exploiter les données recueillies et faire du data-mining

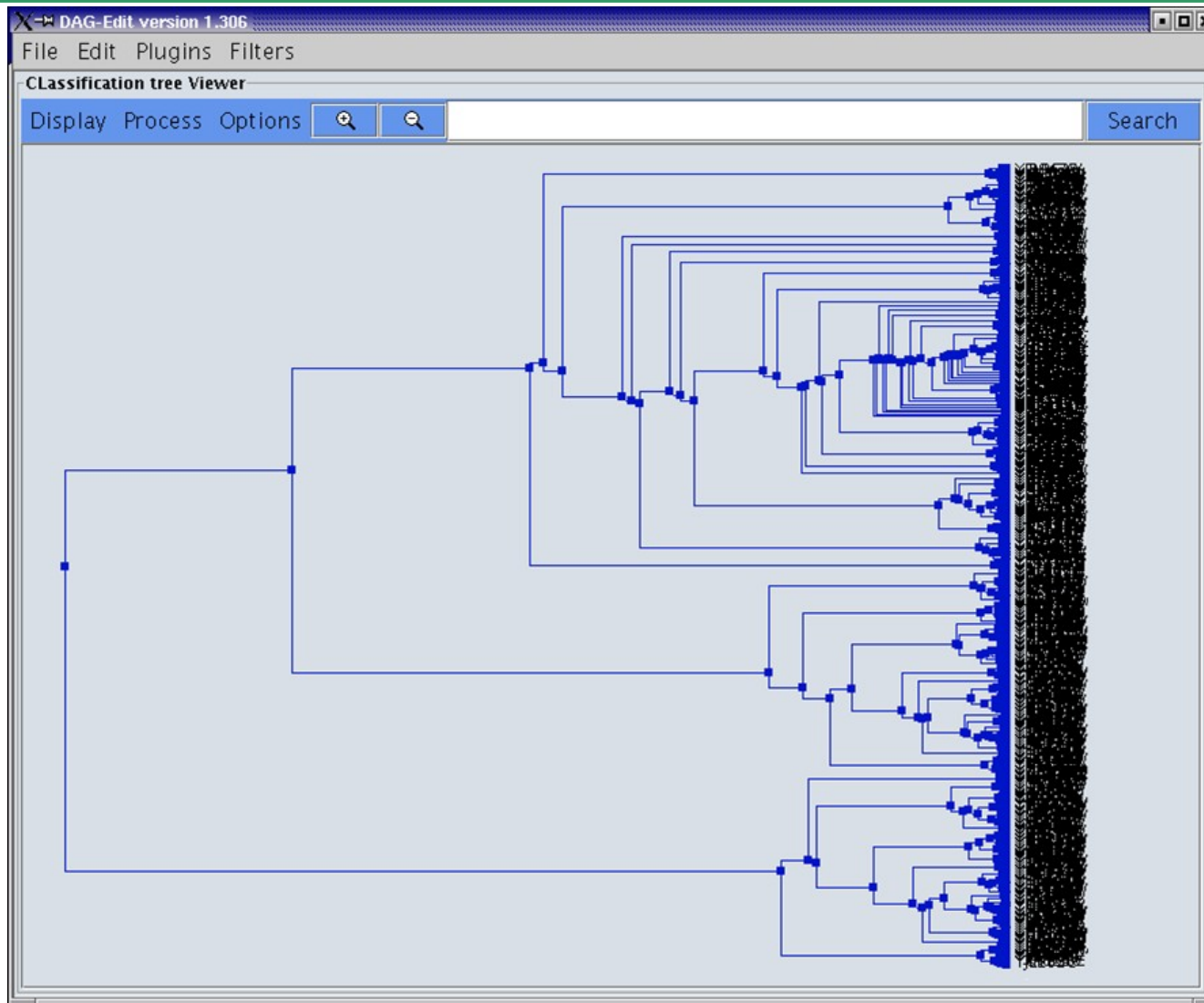
Une des constatations de Eisen et al.

« les gènes responsables de la synthèse des protéines de la mitochondrie sont exprimés de manière concordante avec de nombreux gènes impliqués dans la respiration »

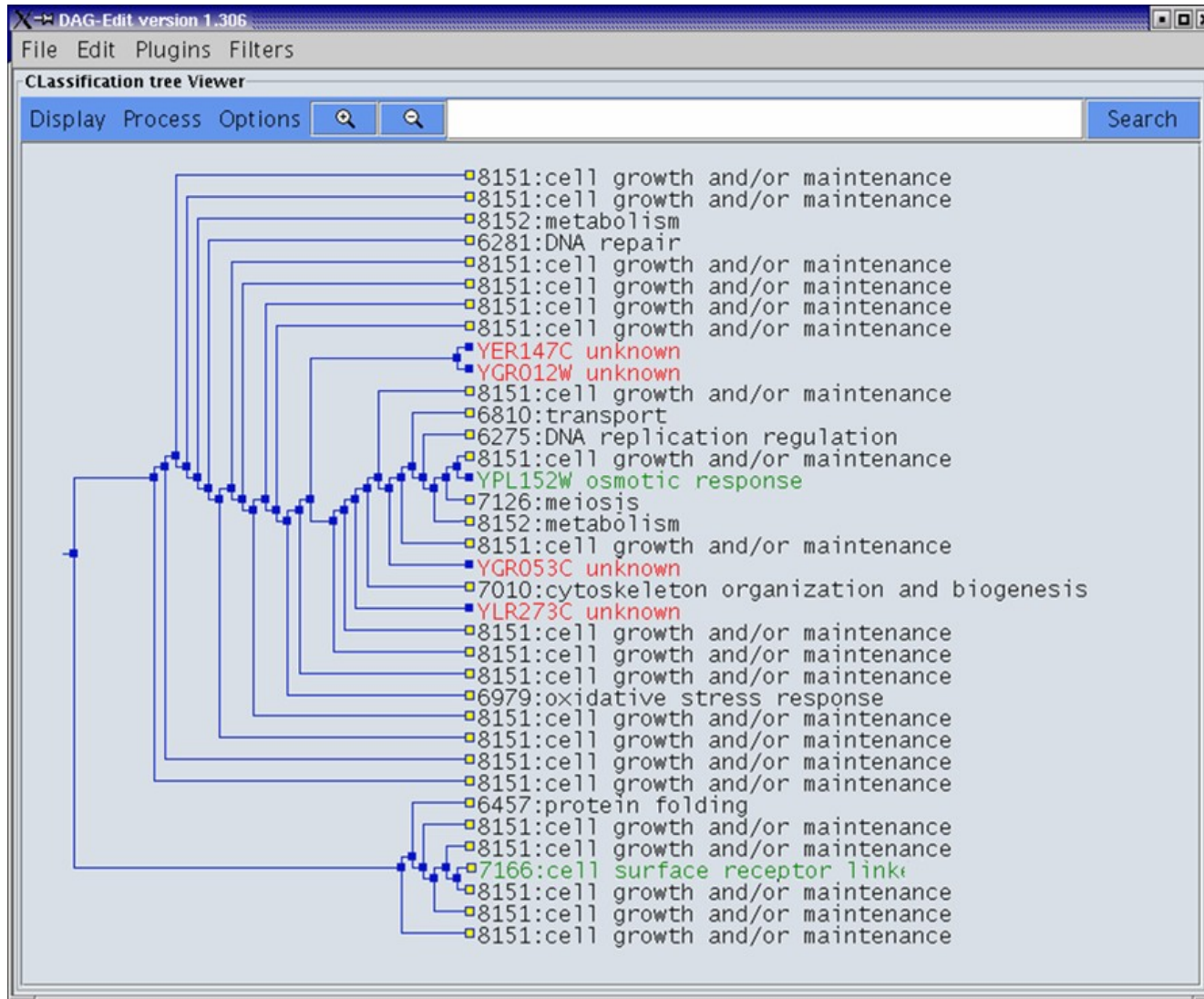
Visualisation de transcrits associés à un terme



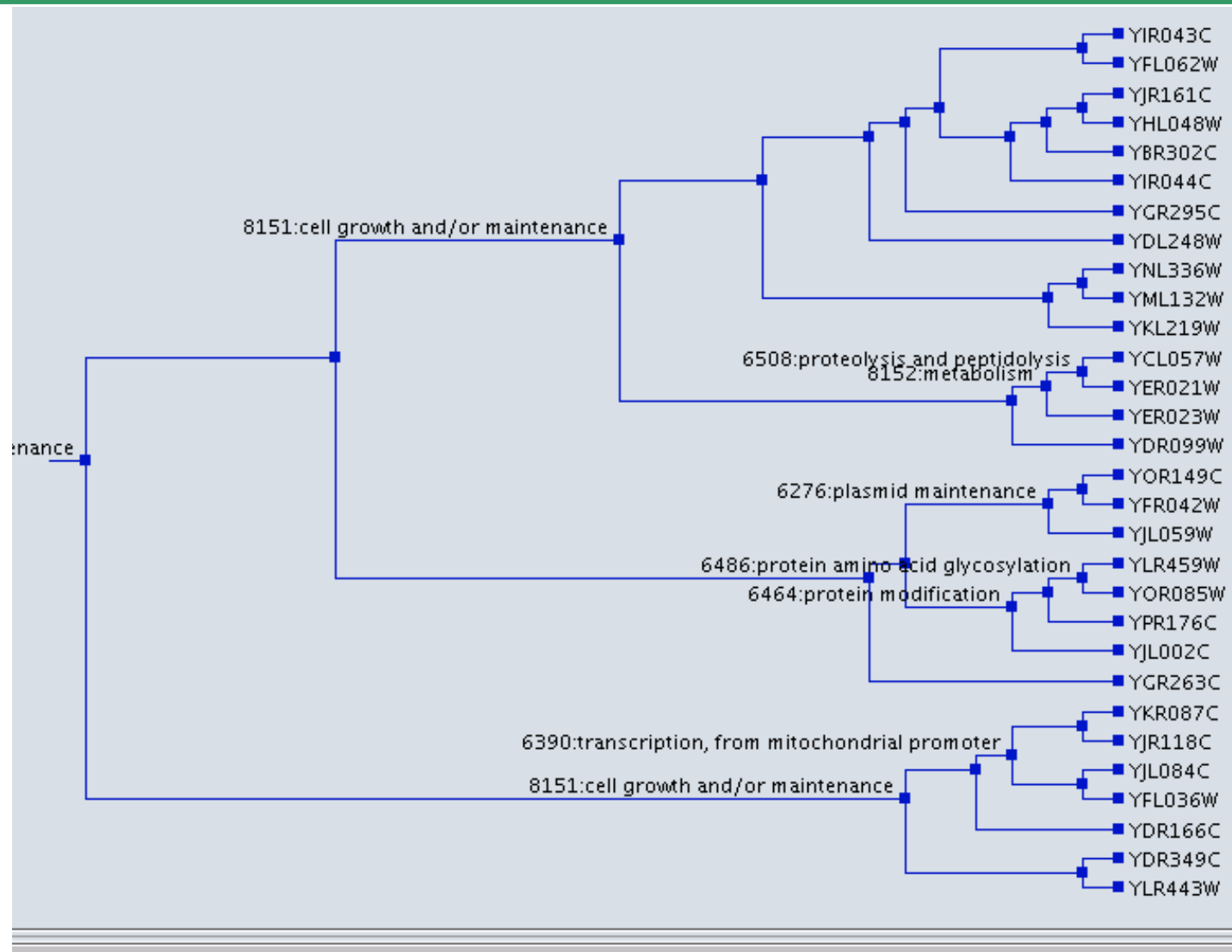
Zoom sur une branche



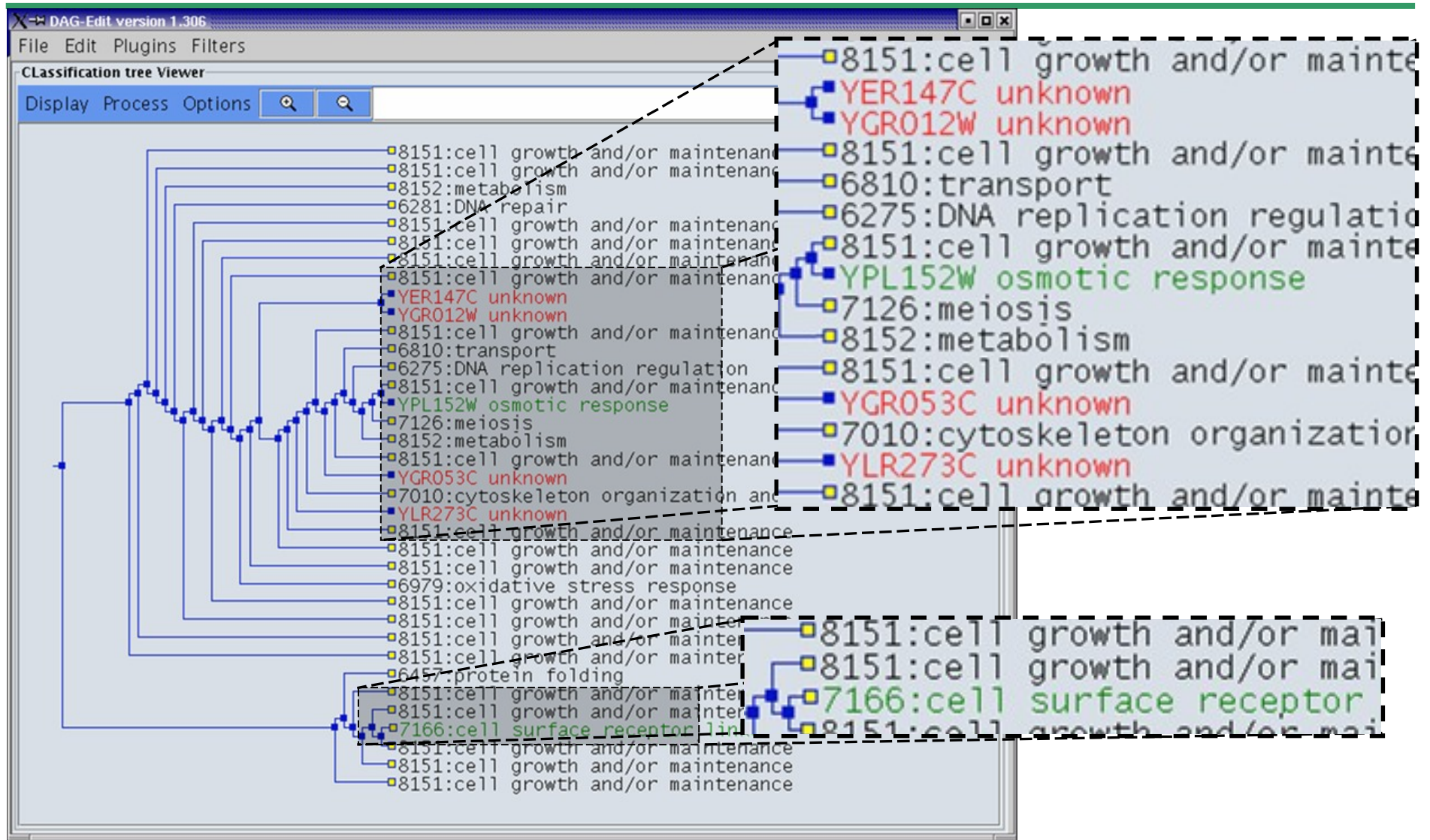
Simplification par nommage des branches



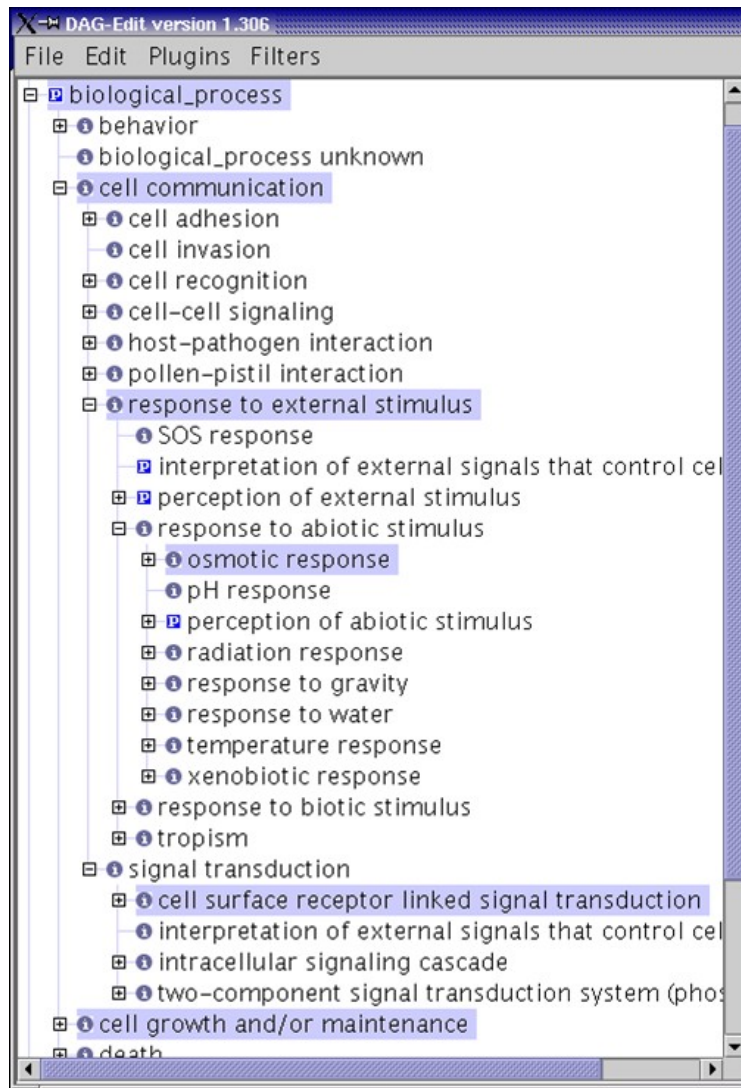
Visualisation d'un groupe de gènes nommés



Simplification par nommage des branches



Localisation des transcrits dans l'ontologie



8151:cell growth and/or maintenance
YER147C unknown
YGR012W unknown
8151:cell growth and/or maintenance
6810:transport
6275:DNA replication regulation
8151:cell growth and/or maintenance
YPL152W osmotic response
7126:meiosis
8152:metabolism
8151:cell growth and/or maintenance
YGR053C unknown
7010:cytoskeleton organization
YLR273C unknown
8151:cell growth and/or maintenance

8151:cell growth and/or maintenance
8151:cell growth and/or maintenance
7166:cell surface receptor
8151:cell growth and/or maintenance

YNL323W

Détails de deux transcrits

YPL152W/RRD2

Description : Resistant to Rapamycin Deletion 2

GO term : osmotic response

IGI

Inferred from Genetic Interaction

IMP

Inferred from Mutant Phenotype

YNL323W/LEM3

Description : Product of gene unknown

GO term : cell surface receptor linked signal transduction

IMP

Inferred from Mutant Phenotype

Code « Evidence » de Gene Ontology

TAS

Traceable Author Statement

IDA

Inferréd from Direct Assay

IMP

Inferréd from Mutant Phenotype

IGI

Inferréd from Genetic Interaction

IPI

Inferréd from Physical Interaction

ISS

Inferréd from Expression Pattern

IEP

Inferréd from Sequéce or structural Similarity

NAS

Non-traceable Author Statement

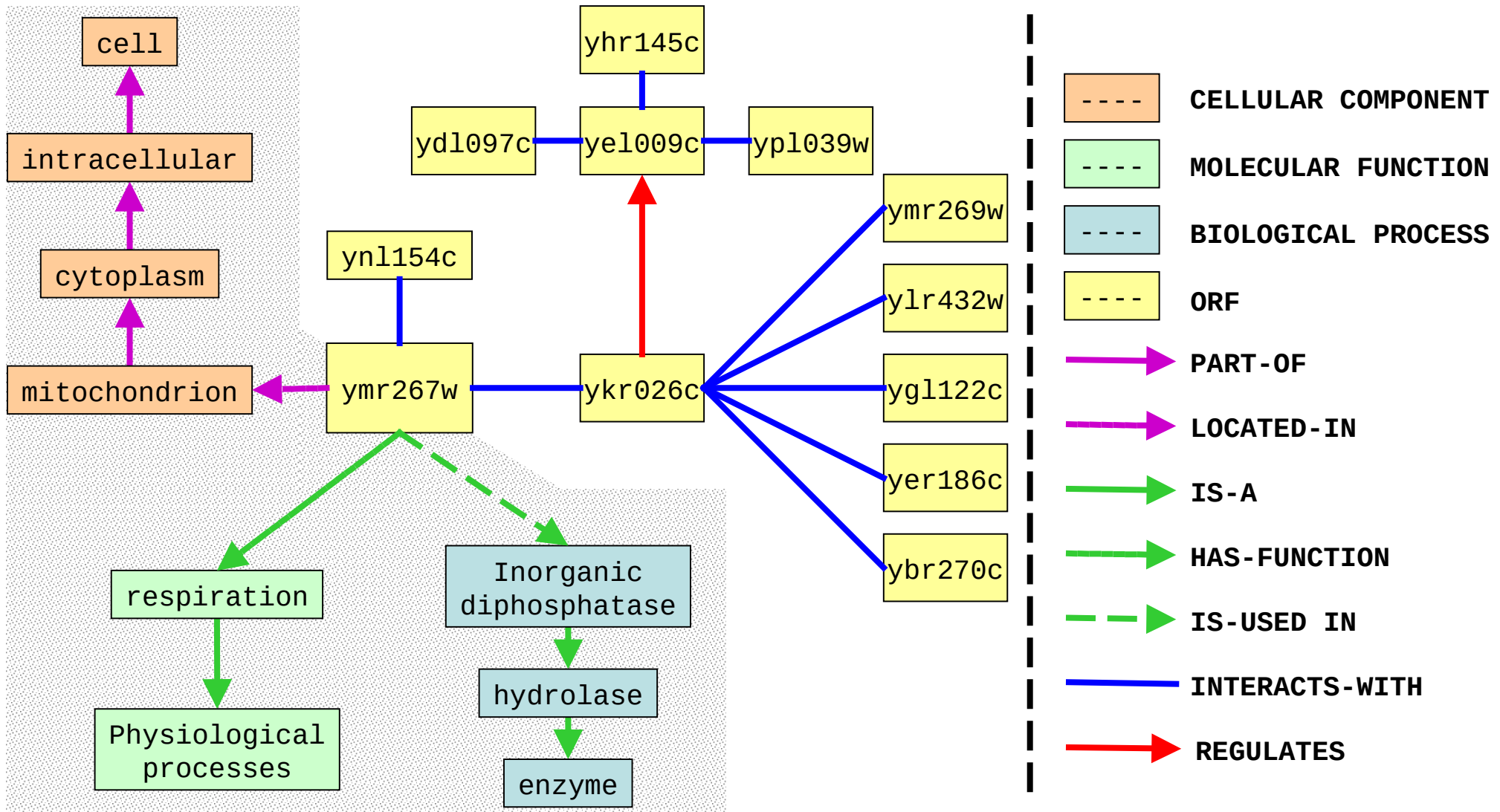
IEA

Inferréd from Electronic Annotation

NR

Not Recordéd

Modélisation des connaissances biologiques



Evolution de Gene Ontologie

GONG = Gene Ontology New Generation

Représentation de la connaissance par logique de description

Basé sur DAML+OIL

- ➔ Liaison avec XML/XML schemas
- ➔ Rapprochement avec RDF/RDFS et le Web sémantique

Axes de recherche et Collaborations

Environnement logiciel

- INRIA – projet OASIS
Didier Parigot

Laboratoires partenaires

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

Data-mining

Bases de Connaissances

- INRIA – projet ACACIA
Rose Dieng

Systèmes de classification

- Olivier Gascuel à Montpellier
- Alain Guénoche à Marseille

Web sémantique