

# URGV Plant Genomics



6 research groups, ca. 90 personnes

26 researchers and engineers

3 university professors

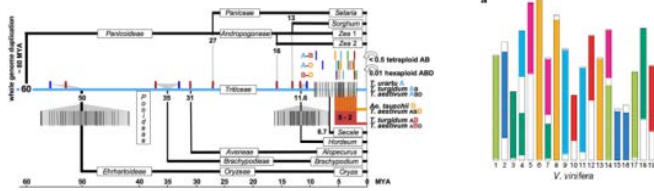
20 technicians and assistant engineers

4 administrative officers

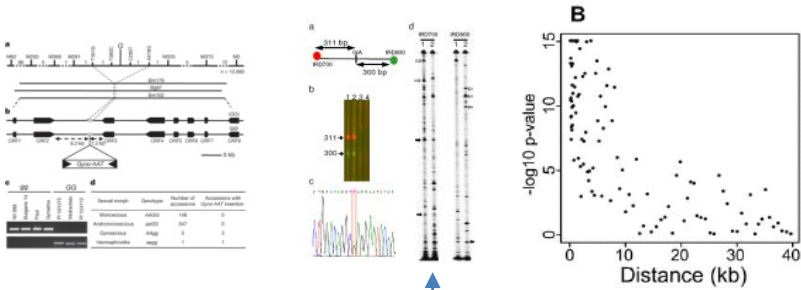


# Genomics

## Comparative and Translational Genomics



## Genetic Determination of Economically Important Traits



## Aims

Understand genomes:

- organisation
- evolution
- functioning (hapmap, comparative genomics, polyploidy)

Predict phenotypes from genotypes

- positional cloning
- TILLING
- association genetics



+ international consortia

Collaborations :

Kick-Off meeting of the Saclay Plant Sciences LabEx

October 13, 2011



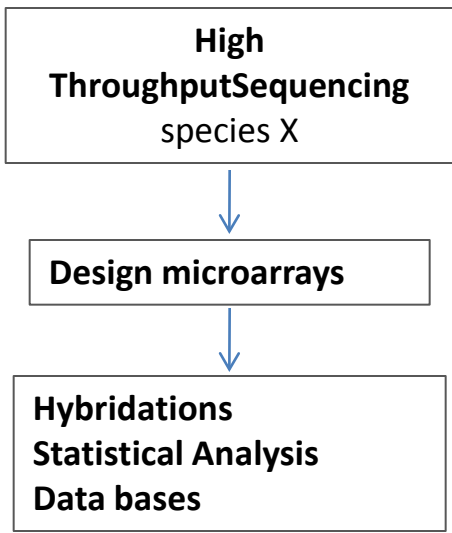
UNIVERSITÉ PARIS-SUD 11

# Transcriptomics

Translation from model species to crops



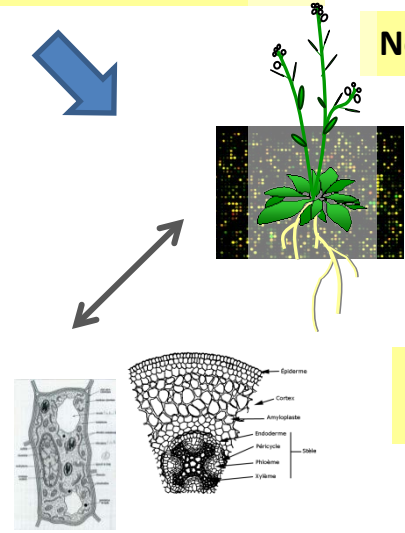
**URGV « pipeline »**



Utilisation of URGV expertise

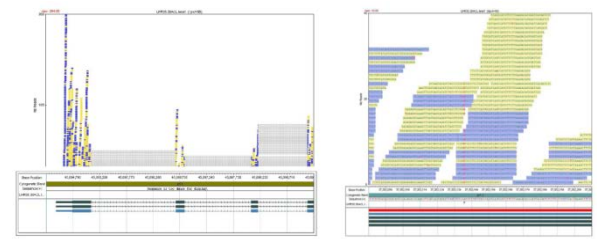
- High throughput sequencing equipment
- Bioinformatic and statistical analysis

New technologies



Transcriptional and post-transcriptional regulation

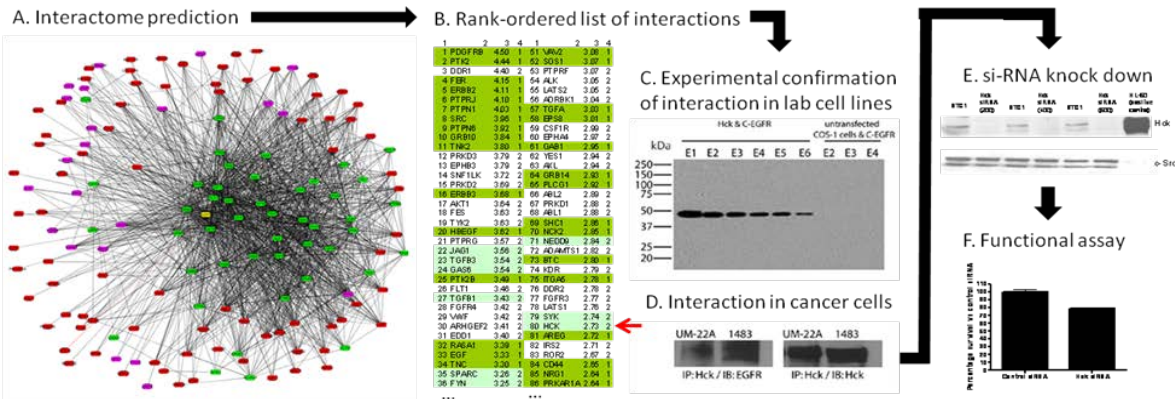
From whole plant/organs to tissues and cells



Institut de Génomique CNG/CNS

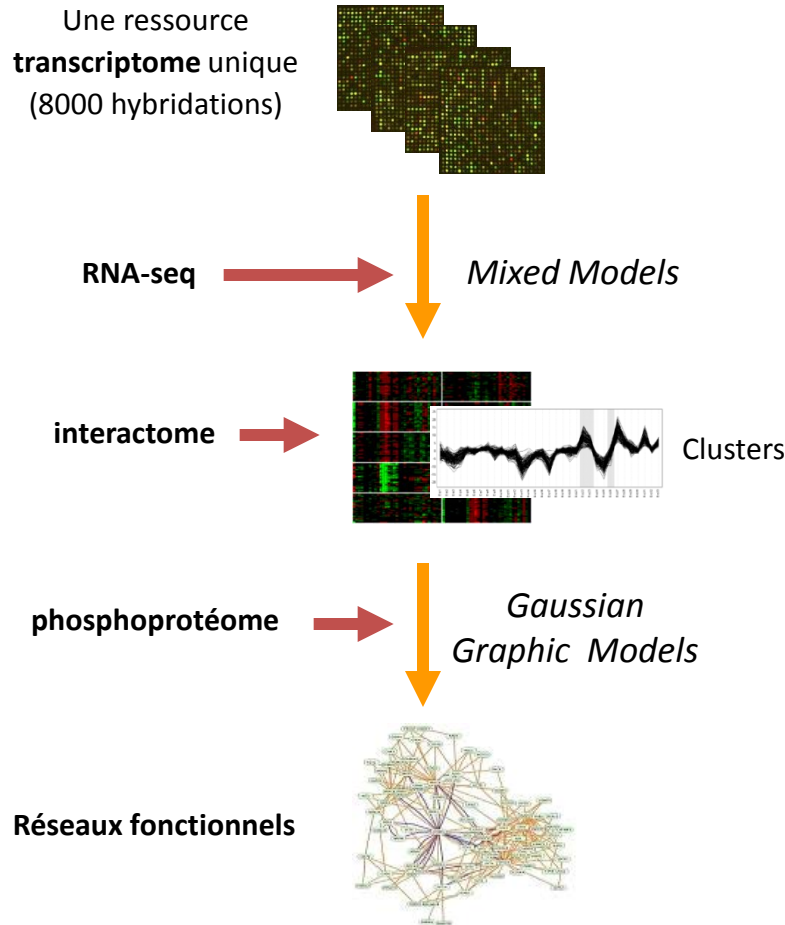
# Proteomics

- Characterise protein complexes of signal transduction by TAP-tag (MAPK modules, signaling dynamics, PTM)
- Identify chromatin-associated proteins and PTMs upon signaling (regulation of the epigenomic machinery)
- Construction of the ORFeome and the protein-protein interactome
- Prediction and functional validation of protein networks



Collaborations :  
 Kick-Off meeting of the Saclay Plant Sciences LabEx  
 October 13, 2011

# Predictive Genomics



- Understand the response of plants to stress and identify the 'master regulators'

- Characterise functional modules of transcription (WIP) and signal transduction (MAPK pathways)

- Explore the orphan gene space to predict their function (stress tolerance)

Compare signaling networks:

- of single and combined stresses (drought and heat)

- of plants and animals to a common pathogen Salmonella (ERANet Systems Biology)



# AERES: Global Analysis URGV – A+

- Development of genomic resources on model species (INRA, CNRS) and cultivated plants (INRA) - 3 platforms: Transcriptomics, TILLING, Positional Cloning
- Reference pole of plant genomics – unique in France and Europe
- Succes in translational research from model species to cultivated species
- Excellent scientific production – doubling research paper output during last 4 years: (93 publications, 73 articles – 2 Science, 1 Nature, 2 Plant Cell, 1 PLoS Genetics, 1 Plos One, 3 Plant J)
- Strong valorisation (6 patents, 3 data bases, 8 industry contracts)

## Recommandations:

- Enlarge international standing (international programmes, foreign students)
- Focus projects due to the low number of researchers, augment student numbers
- Provide office space for every student, post-doc and permanent staff
- Maintain equilibrium between model and cultivated species
- Conserve present leadership by developing new HTP analysis tools
- Develop new bioinformatic approaches and systems biology (personnel?): Sys Bio connects INRA, UEVE and CNRS

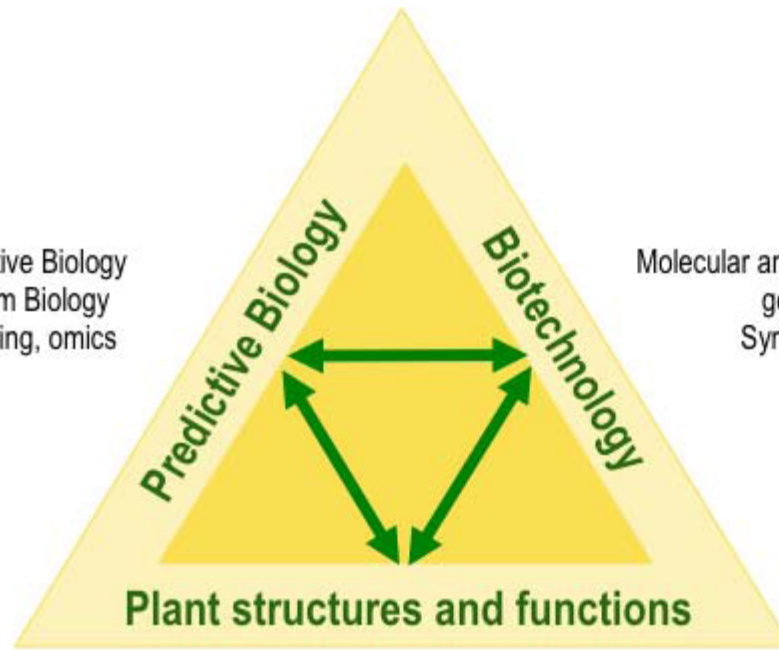


# Integration of LabEx projects into research strategy recommendations

- Enlarge international standing - Install Master of Systems and Synthetic Biology (international programme, many foreign students)
- Augment student numbers – HDRs passed followed by more fellowships
- Provide office space for every student, post-doc and permanent staff – Extension of URGV (from 1000 m<sup>2</sup> (2008) to 1500 m<sup>2</sup> (2011) to 3000 m<sup>2</sup> (2014))
- **Conserve leadership by developing new HTP analysis tools**
  - **Sequencer (partially financed by LabEx) essential for HTP Transcriptomics (Functional Arabidopsis Genomics Group: C. Lurin) and TILLING (Functional Genomics of Crops Group: A. Bendahmane) platforms**
  - **ETD for Orbitrap Mass Spectrometry Proteomics (UEVE Mass Spec Facility Group: J. Tortajada)**
- **Develop new bioinformatic approaches and systems biology**
  - **Recruitment of 1 Post-doc and 1 Engineer as CDDs) – essential for generating HTP kinetic stress data and model development of stress gene regulatory circuits and protein signaling networks (Bioinformatics and Predictive Genomics Group: S. Aubourg; Stress Signaling and Proteomics Group: H. Hirt)**



Integrative Biology  
System Biology  
Modelling, omics



Molecular and cellular engineering,  
genetics, and  
Synthetic biology

Physiology, development, reproduction,  
nutrition, metabolism, and stress responses



# Moving from Descriptive towards Predictive Biology

Spectacular progress in the identification of key genes and molecular mechanisms that control plant development and responses to the environment - **HTP technologies (DNAseq, RNAseq, ChiPseq, proteomics, etc. )**

Inherent complexity of living organisms and the unprecedented increase in analytic tools and data - a major challenge for biology is to use **more mathematics and informatics for handling this complexity**

To understand how genes control organs and interactions with the environment

- we need to develop **quantitative and kinetic data** (gene expression, cell shape, growth, etc.).
- we need to combine the information into functional **interaction networks**
- we need to develop **modelling tools and** statistical methods
- we need to develop **predictive models** for the functioning of environmental adaptation

# Key issues towards Predictive Biology:

- The acquisition, analysis and integration of massive, quantitative and dynamic data
- The development of new formalisms and models to use these data
- The exploration of massive combinatorial correlations between genotypes and phenotypes
- Analysis, reconstruction and simulation of regulatory and interaction networks
- Integrated multi-scale mechanisms from genes to organism
- Interfaces between biology, mathematics, physics or chemistry
- Computer and mathematical tools to analyze data and model systems
- Tools and techniques for phenotyping quantitative, in vivo and dynamic data
- The generic nature of the knowledge obtained for the translation to crops

# Flagship project I : Integrative Analysis of stress responses

Responses at many levels: gene expression, protein regulation and metabolism - quantitative outcome of these responses is the result of the activity of complex multigenic networks.

Adaptations – multigenic (resulting from variation and selection during evolution).

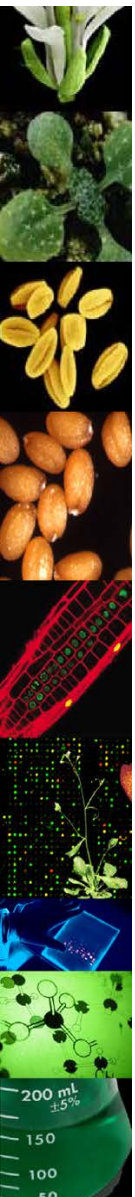
Phenotypic plasticity plays an essential role covering all developmental and morphogenetic modifications that change the growth, the shape or the number of certain organs.

Different stress responses share many common features in the perception and signaling mechanisms.

Simultaneous stresses – different responses than to single stresses. Integrated projects on “biotic and abiotic stress - convergences and divergences” are important.

Only validated models of regulatory networks will enable rational plant engineering to produce better crop plants.

# Flagship project I : Integrative Analysis of stress responses

- 
- generate **quantitative and kinetic data** (gene expression, cell shape, growth, etc.) for:
    - **abiotic stresses** (drought, salt, temperature)
    - **biotic stresses** (bacterial, viral or fungal pathogens)
    - **beneficial microbes** (bacterial and fungal microbes)
  - combine the information into functional **interaction networks**
  - develop **modeling tools and statistical methods**
  - develop **predictive models** that undergo of improvement by **reiterative experimental validation**
  - establish **engineering strategies** for improving crop plants

# Flagship project I : Integrative Analysis of stress responses

## Generation of reliable stress kinetic data

Transcriptome  
(RNAseq, ChIPseq,  
Microarray)

Transcriptome (C. Lurin)  
and TILLING (A. Bendahmane)  
Platforms URGV (co-financing  
Illumina sequencer)

Proteome +PTM  
(Mass Spectrometry)

Mass Spec Facility UEVE (J. Tortajada)  
(LC-MS Orbitrap , ANR financed ETD)



## Generation of reliable stress signaling models

Stress Response  
Models  
(Gene and protein  
signaling networks)

Stress signaling & Proteomics Group (H. Hirt)  
Bioinformatics & Predictive Biology Group (S. Aubourg)



Saclay Plant Sciences LabEx  
Kick-Off meeting

Integrative analysis  
of stress responses

S. Aubourg

October 13, 2011

# Plant stress responses : Inference of function by transcript profiling

## Goals :

Explore the orphan gene space (more than 6000 Arabidopsis genes are still without functional annotation) to identify **new candidate genes** involved in defense and adaptation process (of potential interest for crop improvement).

Predict **gene networks** involved in the response to **biotic and abiotic stress** in Arabidopsis.

## Method :

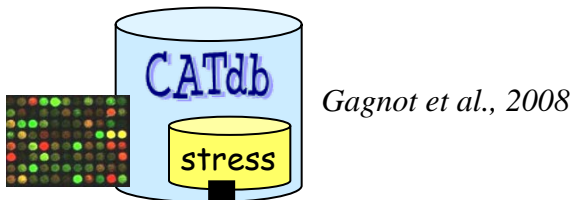
A 'guilt by association' approach based on **meta analysis of transcriptomes** and **data integration** : bioinformatics, statistics and prediction of networks.

## Material :

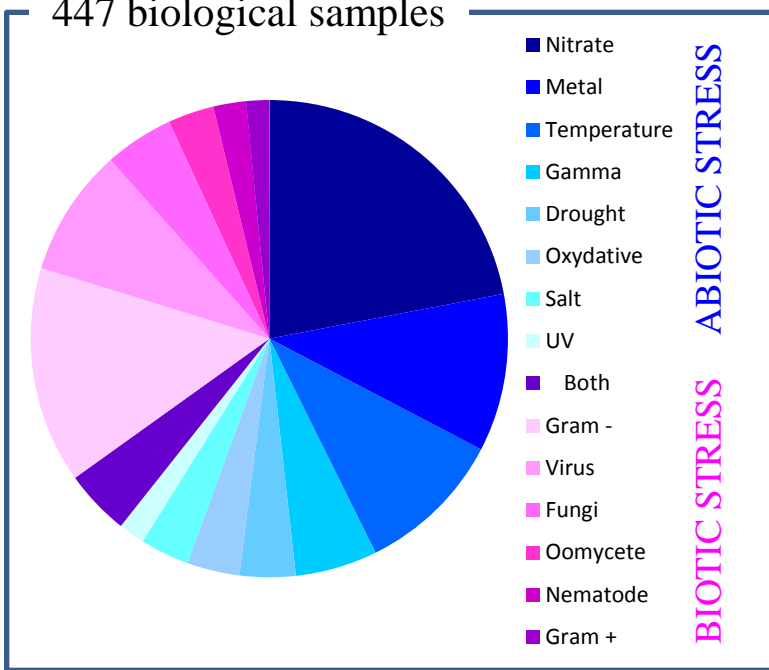
An **original transcriptome resource** containing homogeneous data (CATMA URGV platform, CATdb) with more than 4000 genes not present in the Affymetrix ATH1 chip, and a high diversity of biological samples relative to stress.



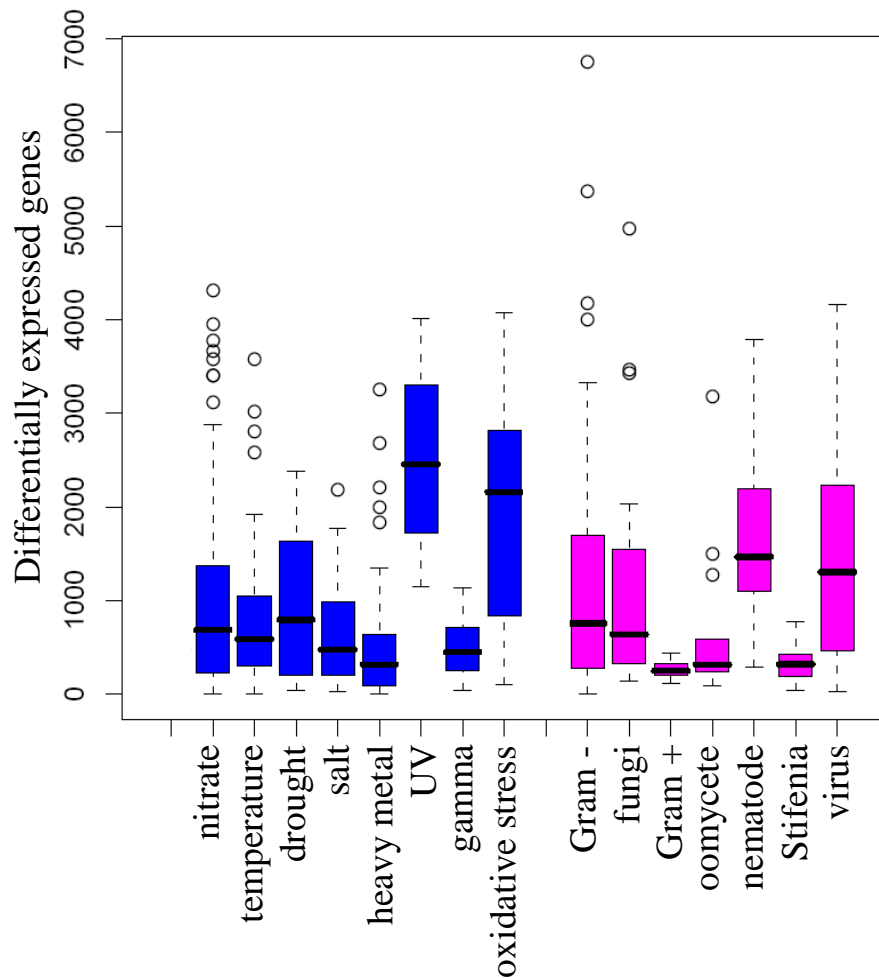
# The 'stress' fraction of CATdb



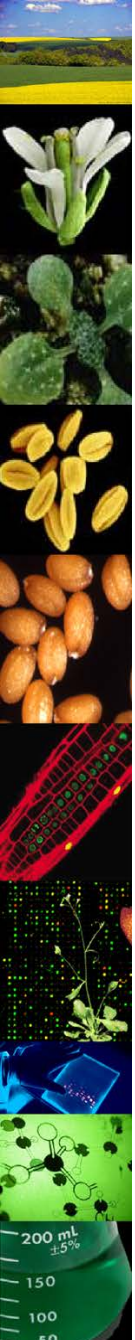
447 biological samples



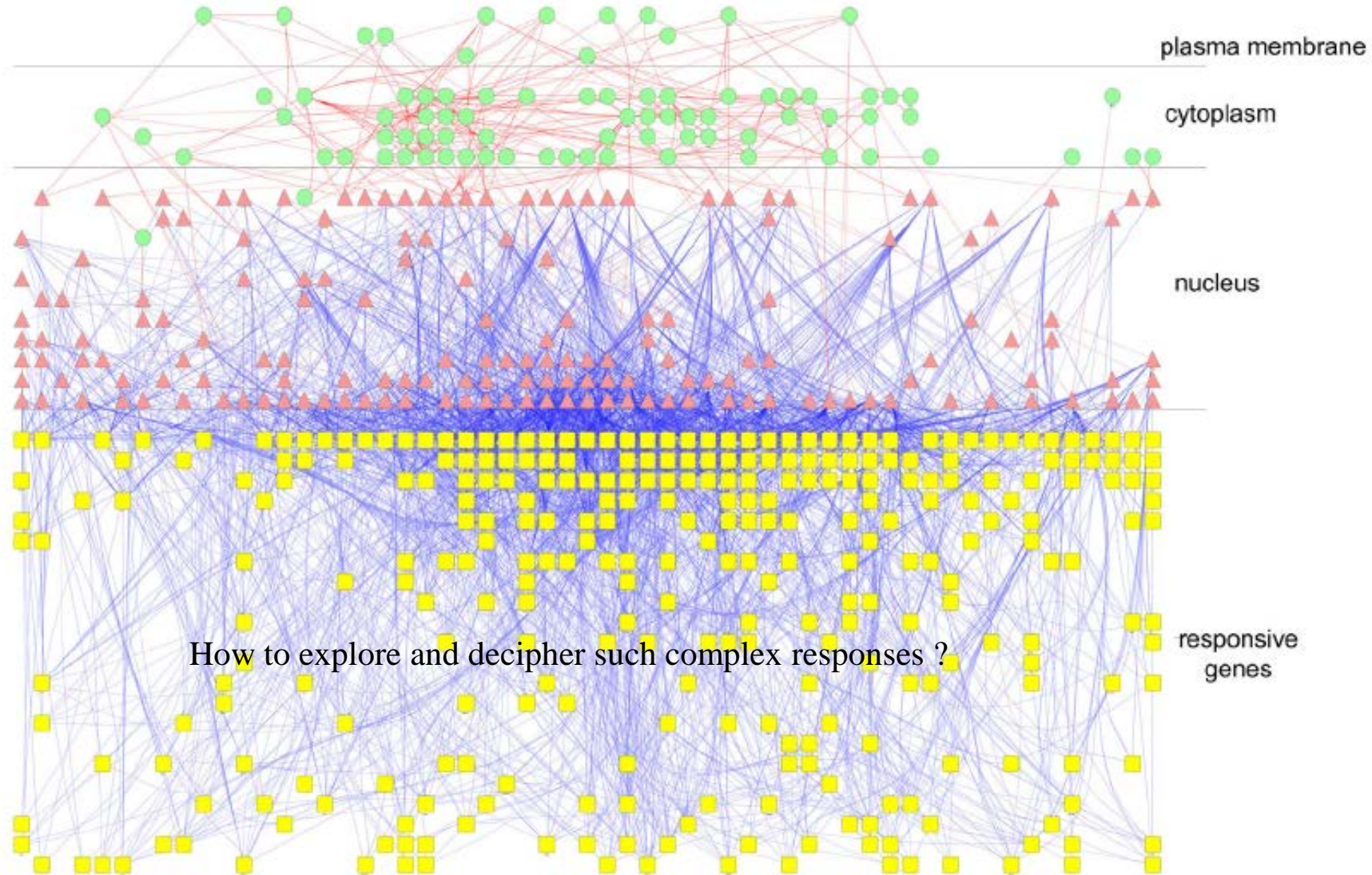
~ 18 000 genes 'impacted' by stress







Stress

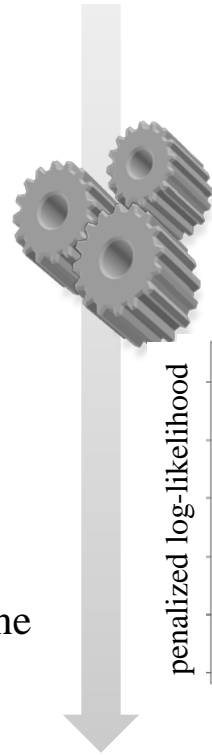


How to explore and decipher such complex responses ?

# Gene clustering based on transcript profiles

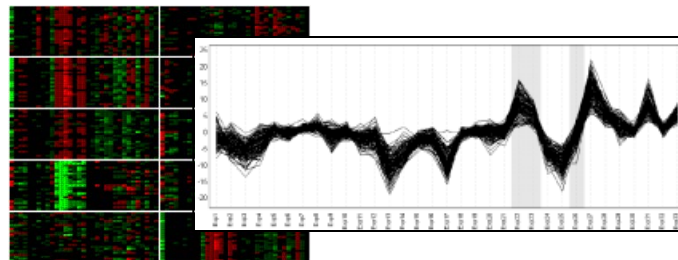
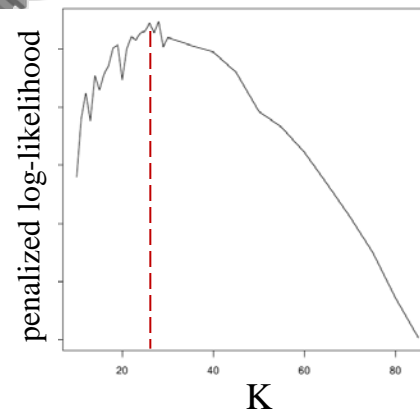
Matrix {DE genes x swaps }

- Gaussian mixture model
- Mathematical criteria to select the cluster number (K)
- General modelling of the variable (experiment) status
- Posterior probability for gene assignment



## SelvarClust SRUW

*Maugis et al., 2009*



Clusters of  
co-regulated genes

	K
Nitrate	in progress
Temperature	in progress
Drought	~35
Salt	in progress
Heavy metal	in progress
UV	in progress
Gamma	~50
Oxydative stress	~55
Flagellin	32
Gram-	~55
Fungi	in progress
Gram+	~15
Oomycete	in progress
Nematode	27
Virus	~45

# Cluster annotation, data integration



Functional classification  
1 146 stress genes

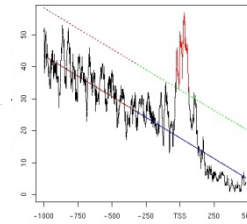


*Derozier et al., 2011*

Subcellular localization  
2 210 TF - 76 families  
Promoter sequences



TFBS



*Bernard et al., 2010*

AI-1

6 475 Y2H interactions

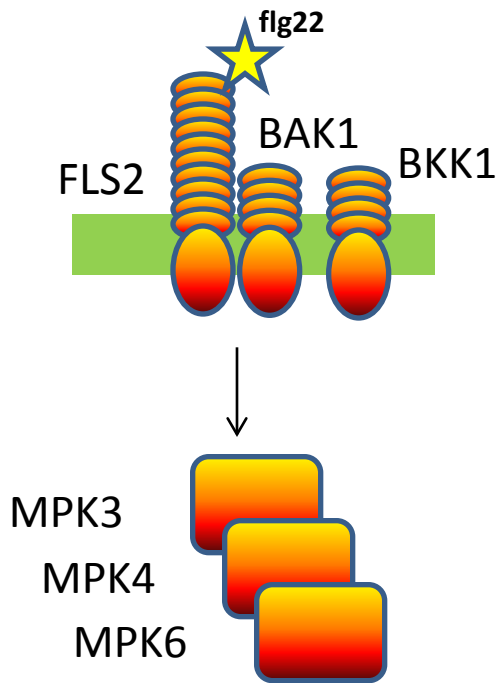


7 945 interactions



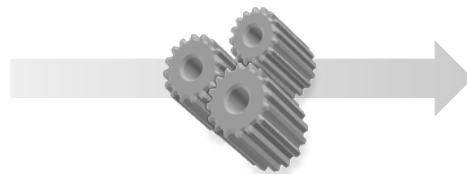
145 494 predictions

# First results



<i>bak1 bkk1</i>	vs	Col
<i>mpk3</i>	vs	Col
<i>mpk4</i>	vs	Col
<i>mpk6</i>	vs	Col
<i>bak1 bkk1 + flg22</i>	vs	<i>bak1 bkk1</i>
<i>mpk3 + flg22</i>	vs	<i>mpk3</i>
<i>mpk4 + flg22</i>	vs	<i>mpk4</i>
<i>mpk6 + flg22</i>	vs	<i>mpk6</i>
Col + flg22	vs	Col

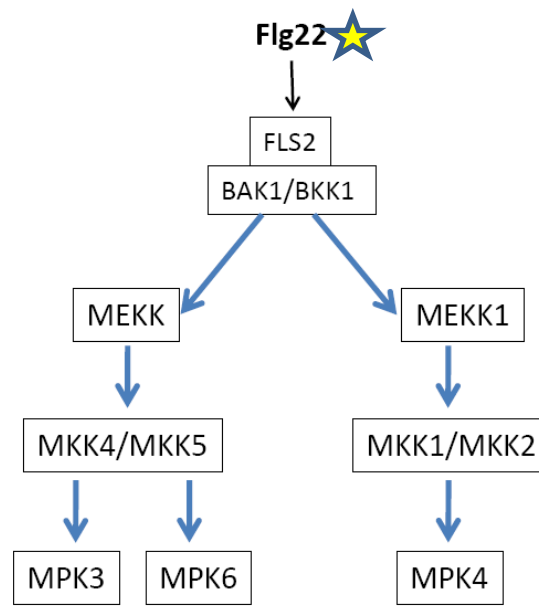
SelvarClust  
SRUW



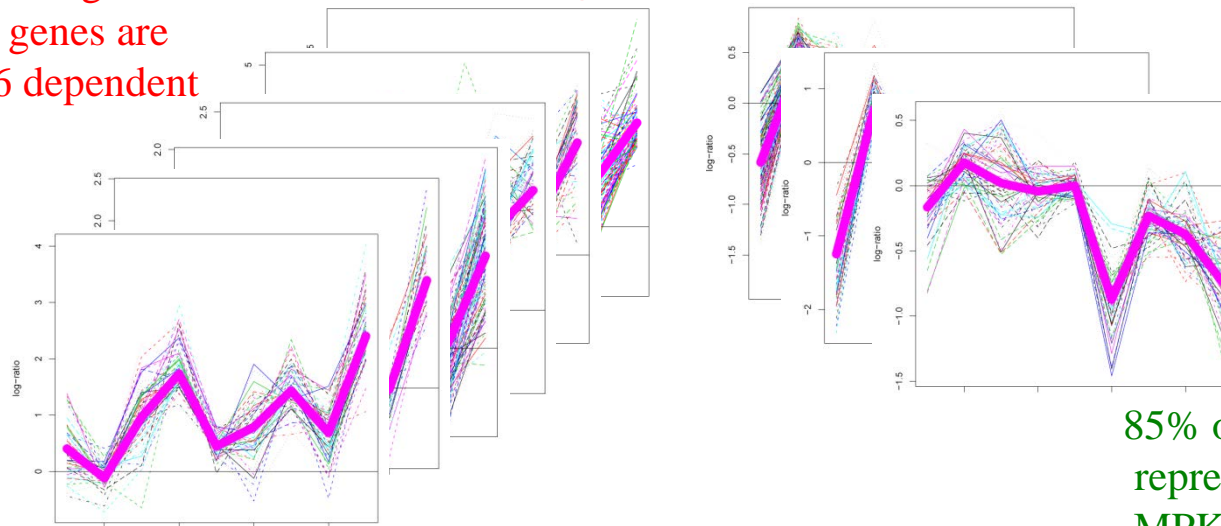
4406 genes found to be differentially expressed at least once among the 9 comparisons

**31 clusters**

# First results

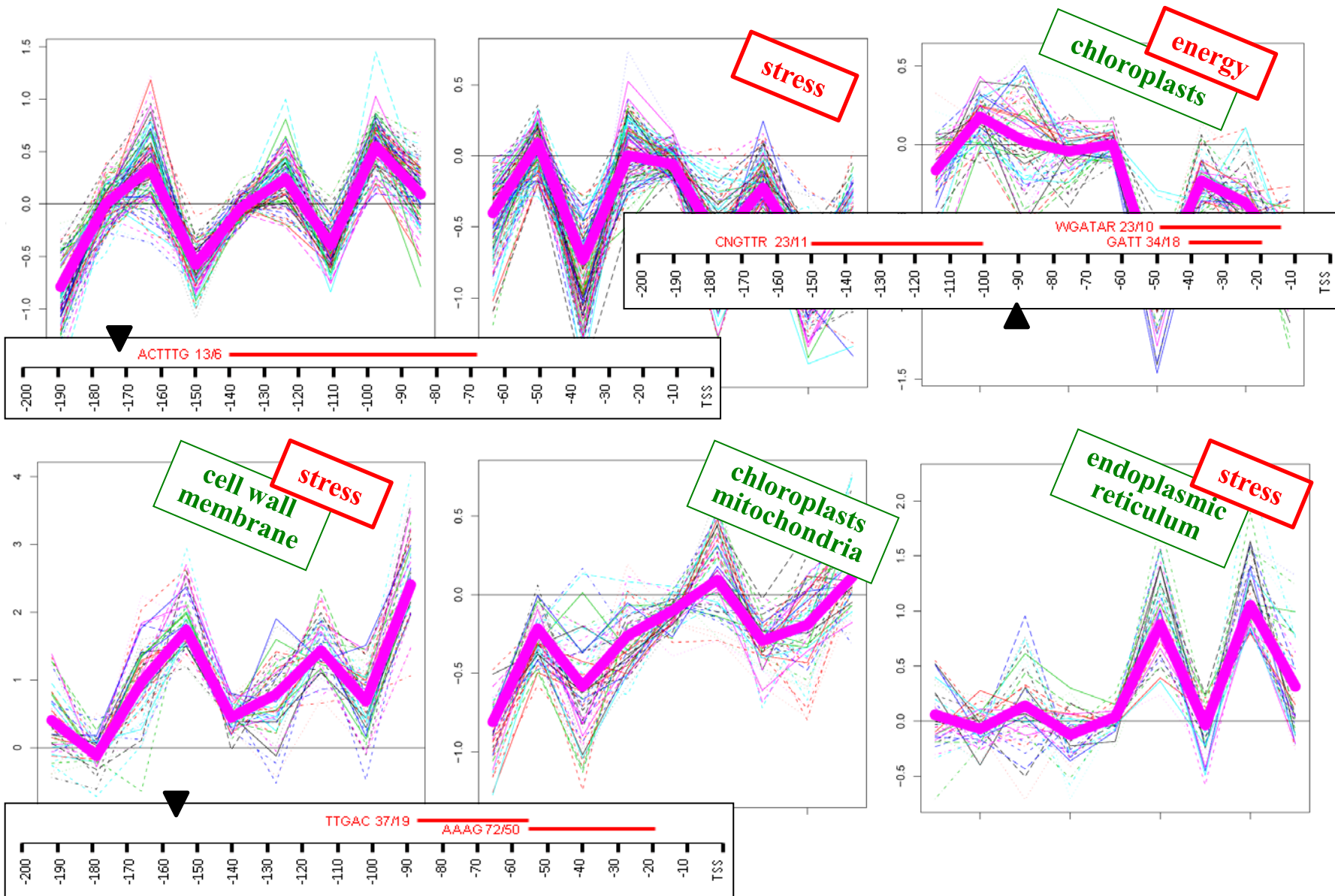


77% of the flagellin-induced genes are MPK3/4/6 dependent



85% of the flagellin-repressed genes are MPK4/6 dependent

# First results



# Beyond the clusters

Comparisons (overlapping genes) of all clusters generated with the different types of stress : ubiquitous or specific responses ?

Modelisation of heterogeneous data.

Experimental validation and cluster enrichment: Transcriptomes of mutants (putative regulators and TF will be targeted in a selection of clusters) and integration of interactome data. (*URGV platform, Claire's group*)

Prediction of gene networks using Gaussian Graphical Model.

Translational research



Crops

- Comparative genomics for prediction of orthologs
- TILLing (*URGV platform, Abdel's group*)



## Bioinformatics

M.-L. Martin-Magniette  
V. Brunaud  
J.-P. Tamby  
A. Lecharny

P. Papastomoulis  
R. Zaag  
A. Cauchard

## Signal Transduction

H. Hirt  
E. Bueso  
N. Frei



G. Celeux (Orsay)



C. Maugis (Toulouse)



S. Huet (Jouy)  
N. Verzelen (Montpellier)  
M. Koskas (Paris)



C. Giraud (Palaiseau)



T. Mary-Huard (Paris)



C. Keribin (Orsay)

SONATA-stats funding:



GAP  
BV  
MIA

