

GSBA

BIOINFORMATICS TEAM

C. Landès



September 2016

RATIONALE of the TEAM

(set up end of 2013)



Observation:

- Exponentials needs in data treatment
- Some skills spread in various teams

Aims:

- Anticipate the needs
- Provide solutions for the IRHS' projects
- Optimise skills and create synergy

A deal with the other teams:

- At first provide support to the IRHS projects
- Then develop new projects if human and financial means available

Human ressources

- Researchers : Sébastien Aubourg (DR INRA)
Julie Bourbeillon (MC AGO)
Claudine Landès (PR UA)
Besnik Pumo (PR AGO)
Jean-Pierre Renou (DR INRA)

- Engineers and assistants :
Martial Briand (AI INRA)
Fabrice Dupuis (Tech INRA)
Sylvain Gaillard (IE INRA)
Sandra Pelletier (AI INRA)
Pierre Santagostini (IE AGO)

- PhD students :
Arthur Chambon (LERIA)
Nicolas Daccord (Epicenter)
Marc Legeay (LERIA)

Facilities

Calculation server: 64 processing cores, 1.5 To RAM

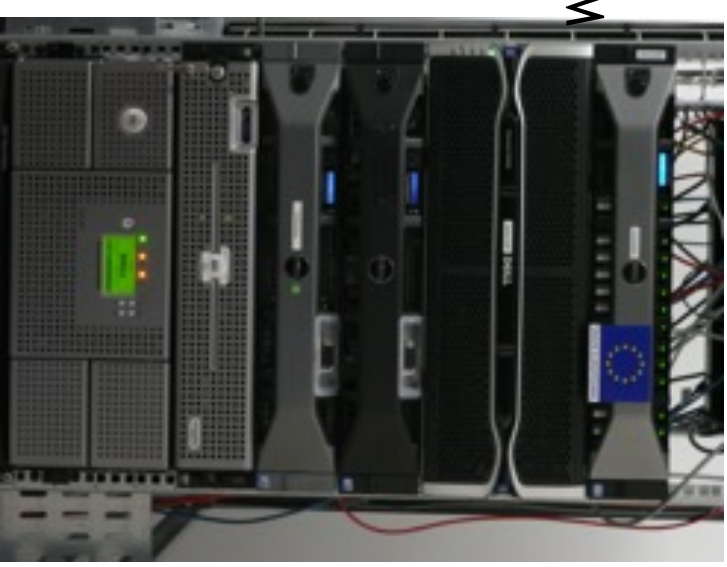
Storage server: 2 X 80 To disks

Objectives: X 3 (PHENOTIC 2 Project)

And

Servers will be hosted by the Angers University Data Center in 2017

IRHS has been seeking to become an operating relay of the
Bioinformatics Platform network of Biogenouest



RUNNING PROJECTS & PARTNERSHIP

- GRIOTE Regional project:
Structuring the bioinformatics community in Pays de Loire for HT-omics and data integration.
 - 9 labs in computer sciences (incl. LERIA/ Univ. Angers) & genomics
 - + BioGenouest Network
 - 3 PhD grants for the team in co-direction.
- GRC Databases (IBISA): Databases for *Rosaceae*, *Apiaceae* and Plant Associated Bacteria GRCs (related IRHS teams)
- ANANdb (SFR QuASaV): Database and analysis tools for the SFR Genomics PF
- PHENOTIC 2: Analysis tools for the Phenotyping PF (GEVES & LARIS/U. Angers)
- Bioinformatic services for IRHS teams (genomics, transcriptomics, imaging)

Organisation



3 main tasks:

Research: Genes annotation and function at the genome level, Networks

Services: DBs, analysis tools, Galaxy instance, servers and system administration...

Training: Programming and statistics, expertise in data analysis

Output: PlantDB a Phenotype database

GRC Rosaceae & Apiaceae (IBISA)

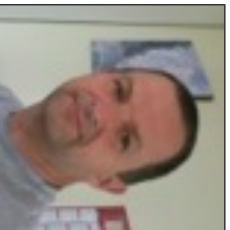


numéro	espèce synonyme	statut	groupes	collection	local_espèce	nombre de variétés	variétés
FR0100 - GRC	Rosa indica 1899	probable	FR01000000	Collection France	FR01000000	2237	FR01000000
FR0100 - VITE	Rosa indica 1899	probable	FR01000000	Collection France	FR01000000	2027	FR01000000
FR0100 - GRC	Rosa indica 1899	probable	FR01000000	Collection France	FR01000000	2014	FR01000000
FR0100 - GRC	Rosa indica 1899	probable	FR01000000	Collection France	FR01000000	2013	FR01000000

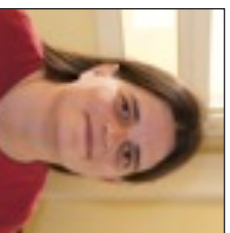
- **PlantDB** is our data management system focused on plants in orchard. This database is our response to the lack of tools to manage plants maintained by vegetative propagation.
- It is built to handle plants in orchard and greenhouse and link them to phenotyping / genotyping data, trace samples and genealogy for breeding projects.
- *It has recently gained the ability to store historical data as well as phenotyping data.*
- This database is also the core of the BRC (Biological Resource Center) management system.



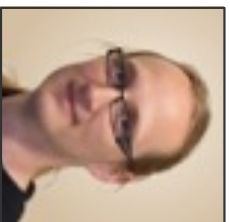
A Lelièvre



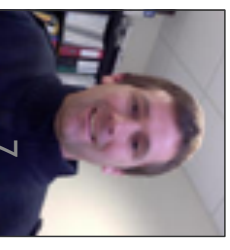
F Dupuis



J Bourbeillon



S Gaillard



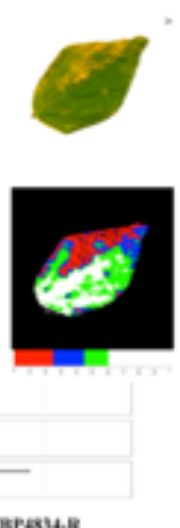
Output: Bioinformatics for Phenotyping

Rousseau *et al.*, Plant Methods 2015 www.phenoplast.org

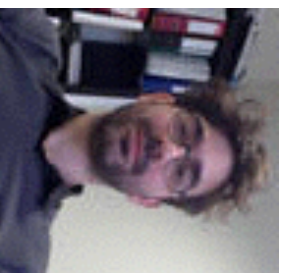
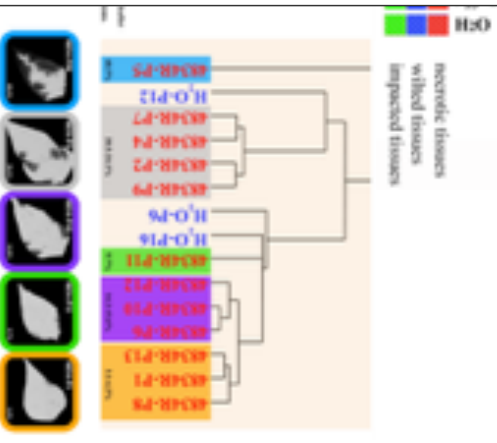
Welcome to PHENOROB/PHENOPLANT

To analyse and compare Phenocam's files (extension: .tif and .tiff), please fill the following form:

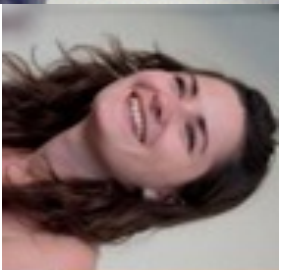
Your email address:



- Phenoplast.org is a web resource developed in collaboration with LERIA (Univ. Angers), dedicated to the phenotyping of the impact of various stresses on plants using image analysis.
- The present resource was initially developed for the analysis of Chlorophyll Fluorescence Imaging data. However, such a resource may be used with any kind of images dataset, if appropriate.
- <http://www.phenoplast.org/> (with EmerSys Team)



T. Bourreau
PHENOTIC



C. Rousseau
PHENOTIC



G. Hunault
HIFIH



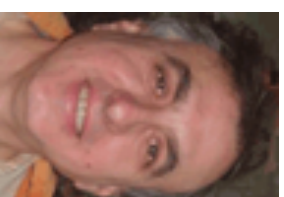
E. Belin
LARIS



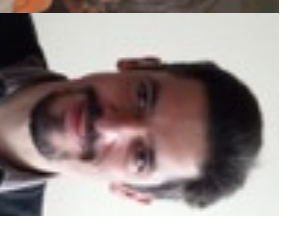
S. Gaillard



J Bourbeillon

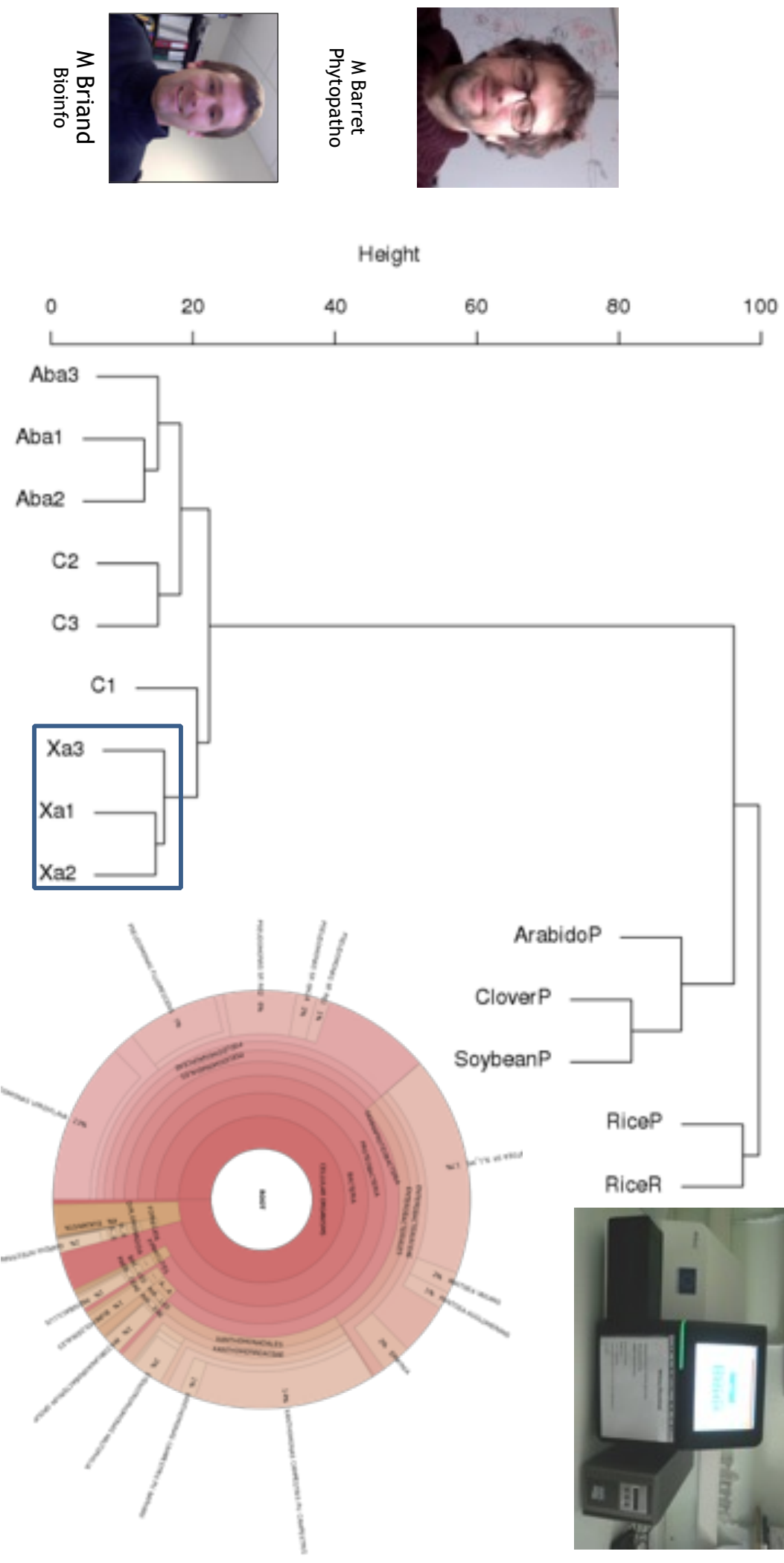


B Pumo

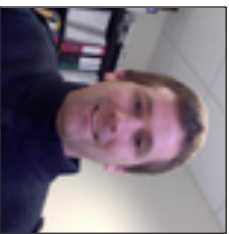


P Santagostini

Output: Clustering of Metagenomic Seed Samples



M Barret
Phytopatho



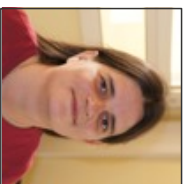
M Briand
Bioinfo

Sequence selection in 30.000 GyrB orthologs to design amplicons

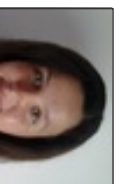
Barret *et al.*, *App. Environ. Microbiol.* 2015

Output: A Data management system for Quasav federation : The ANANdb Project

From Genotypes to Phenotypes



J Bourbeillon



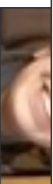
A Khemiri



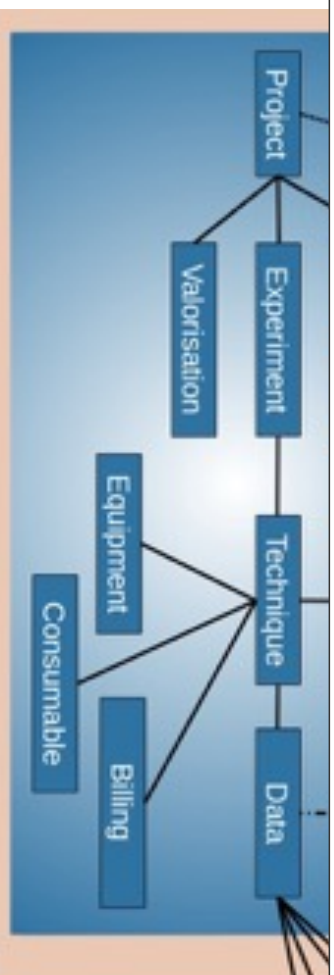
- ANANdb will provide to researchers an easy access to:

- standardized data processing,
- compilation and comparisons,
- traceability of all data,
- easy inventory management and tracking devices,
- whole results and bioinformatics tools developed

- ANANdb will be a precious database that collect all metadata essential for meta-analysis.



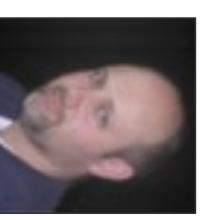
M Briand



S Pelleitier



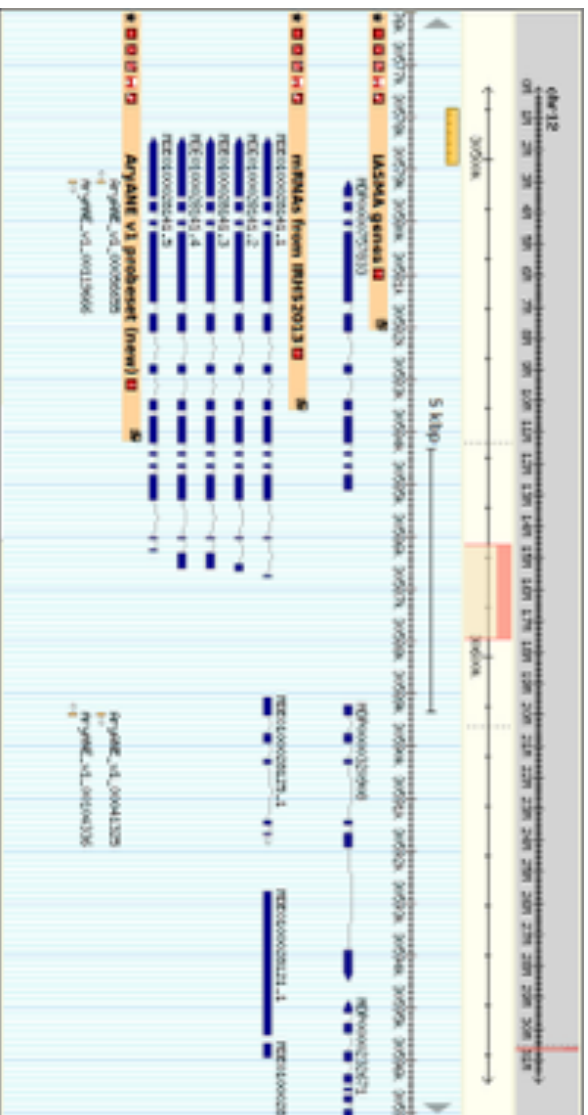
M Legeay



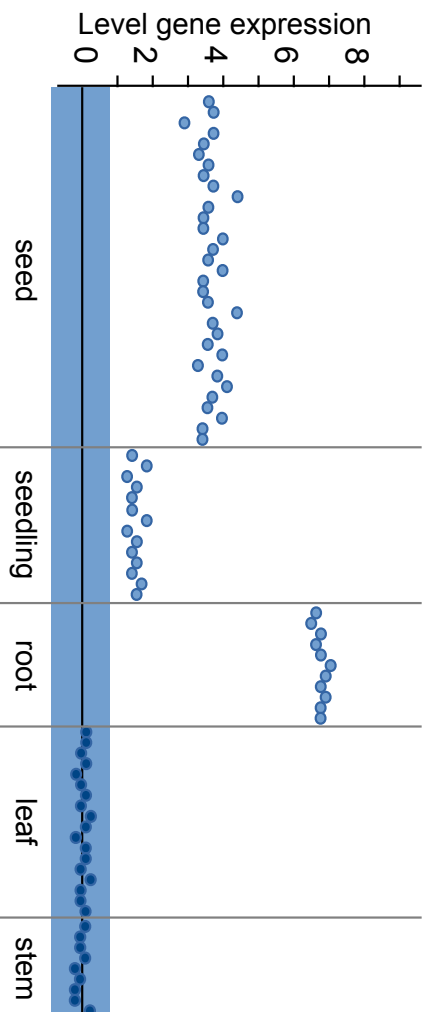
S Aubourg

Output: ANANDb tools

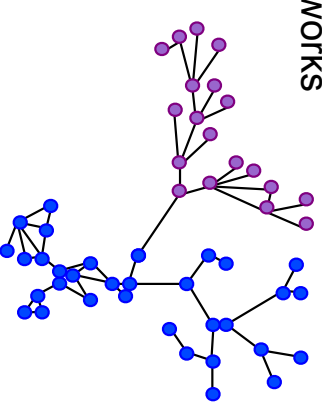
Genome Browsers (apple, rose, various pathogens....)



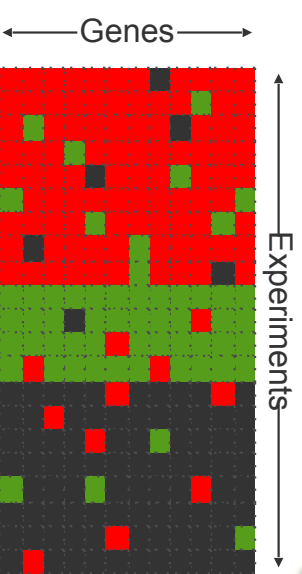
Atlas of expression



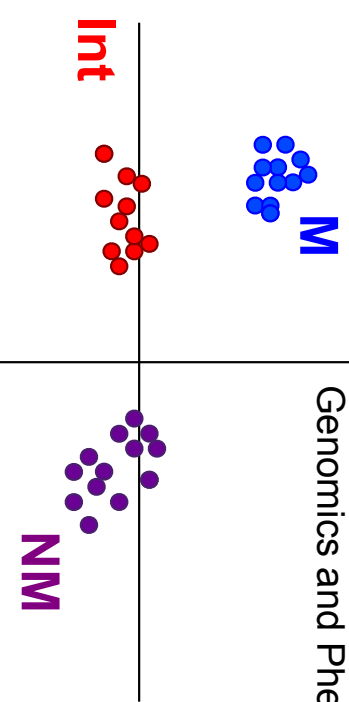
Gene networks



CORGI: Bi-clusters of co-expressed genes



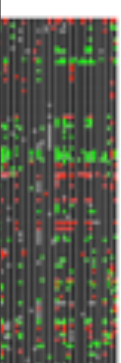
Correlation between Genomics and Phenotype



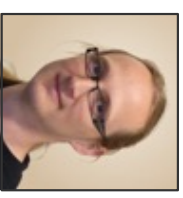
CORGI : a tool for CO-Regulated Genes Identification



Data transformation (-1,0,+1)



Query the database with a gene list (ex: *E. amylovora* response)



- We have developed a gene clustering software combining expression data discretisation and statistics based on the convergence of the binomial law to the normal law, called CORGI : CO-Regulated Genes Identification
- Regarding transcription profiling it allows to identify the cluster of genes having the highest probability to be co-regulated amongst a list of differentially expressed genes, and the experiments which are the most explicative of this cluster.



A simple method to find co-regulated clusters from lists of differentially expressed genes : CO-Regulated Genes Identification (CORGI)



SCIENTIFIC STRATEGY and PROSPECTIVE

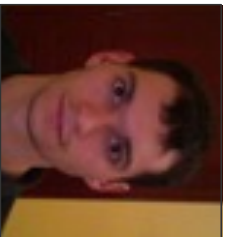
A function inference project:

Finding new defense pathways, key genes
and advanced regulation

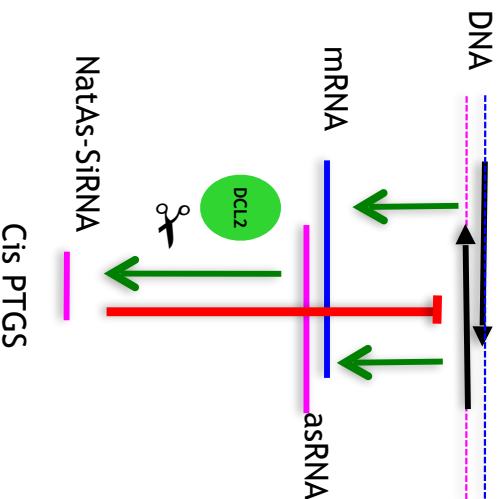
- Understand gene families structures and epigenetic marks in polyploid *Rosaceae* genomes and their contribution to the regulation of expression.
- Define orthologous relationships and propose functional inferences through expert relational and functional annotation and meta-analysis of expression data to find expressologs

Output: Building gene networks including cis-PTGS events?

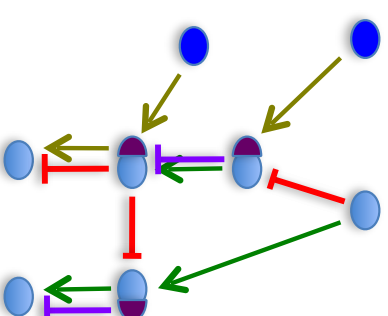
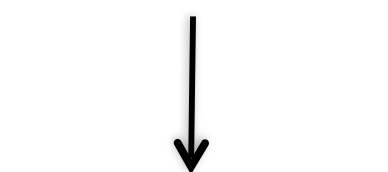
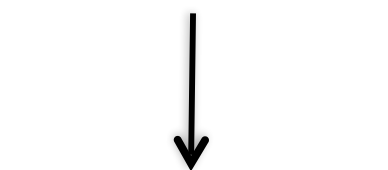
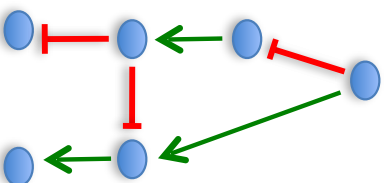
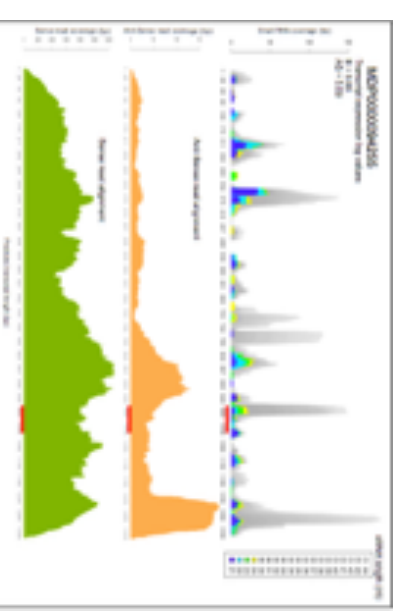
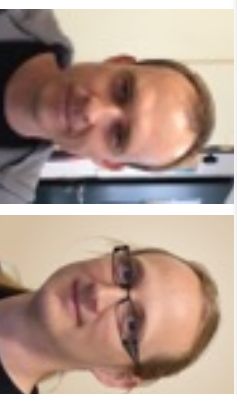
PhD: M. Legéay Dir: B. Duval (*LERIA U. Angers*) & J.P. Renou



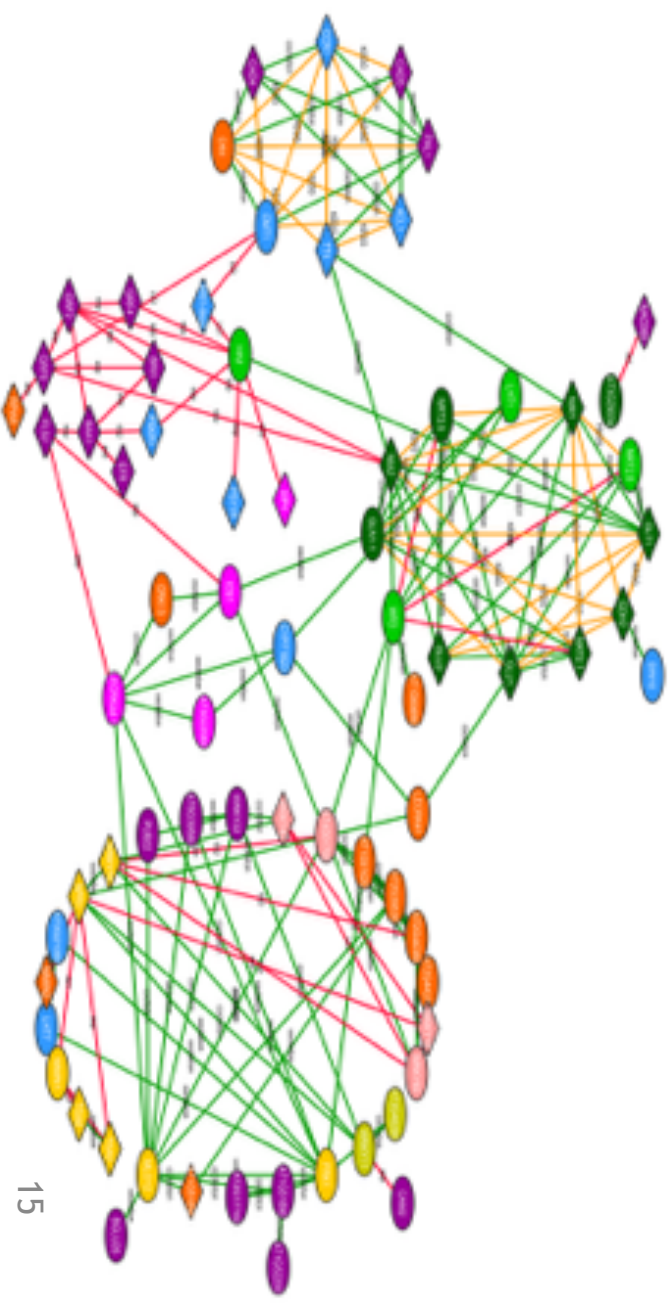
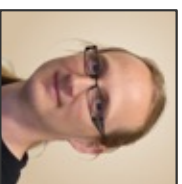
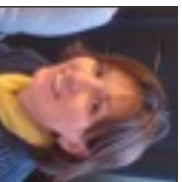
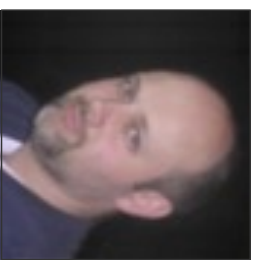
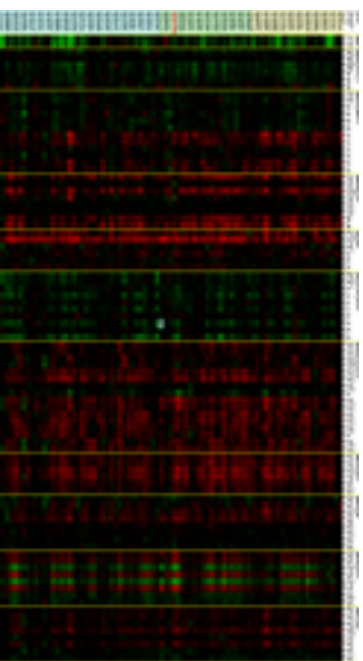
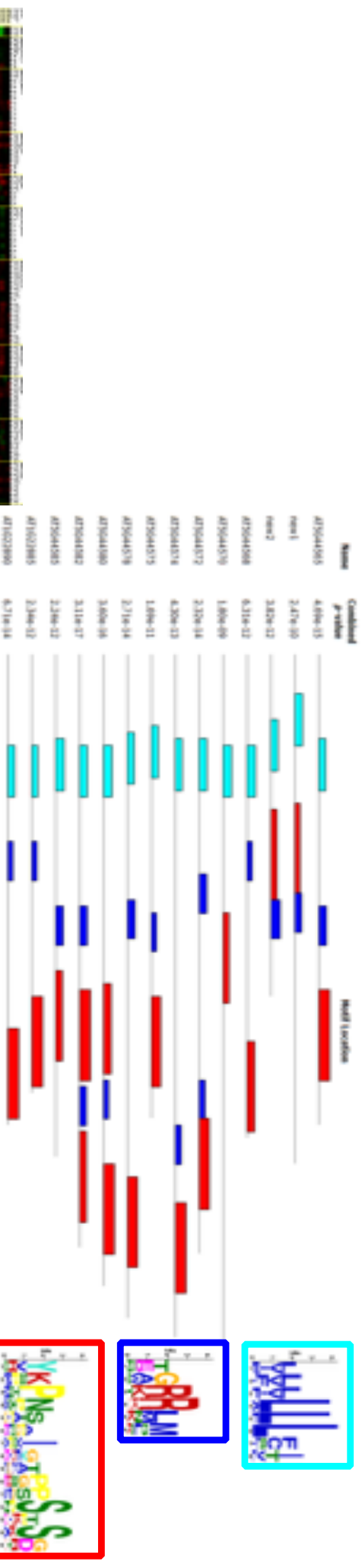
AS transcripts ?



65% of apple genes produce
AS transcripts
(Celton, Gaillard *et al.* *New Phytol*,
2014)



Data integration for functional and relational annotation

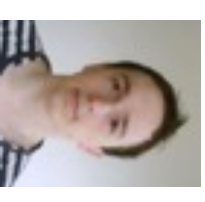


Gene family structure in polyploïds

- Understand interactions between genes structures and epigenetic marks at the genome level and their contribution to the regulation of expression, with a focus on gene families and paralogs in polyploïde genomes of rosaceae (apple and roses).
- Genes duplication is considered to be an important source of genetic diversity implied in organism complexification and/or environment adaptation.
- We will explore interplay between epigenetic marks and sub- or neo-functionalization of duplicated genes according to their type : segmental duplication, whole genome duplication or tandemly arranged genes.



C. Landes



N. Daccord