# Evolution of biological networks : Open questions linked to high dimension and complexity

Christine Dillmann, D. de Vienne

UMR Génétique Quantitative et Evolution - Le Moulon

December 3, 2019

# Outline

# Properties of life

**Complexity** : *Juxtaposition* of identical entities, that constitute the parts of more complex entities called macromolecules.
(*DNA, collagene, lignine, ...*)

**Cellular organization** : the cell is the structural, functional and reproductive unit that constitute the parts of all living forms (except viruses).

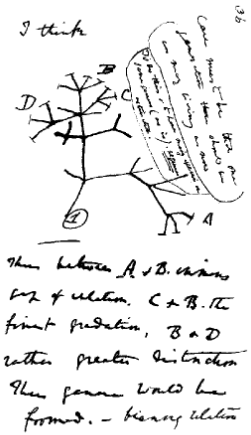
**Metabolism** : chemical reactions that occur within living invividuals, most of them being catalyzed by enzymes.

**Homeostasis** : Invariance of phenotypes with regard to external environment.

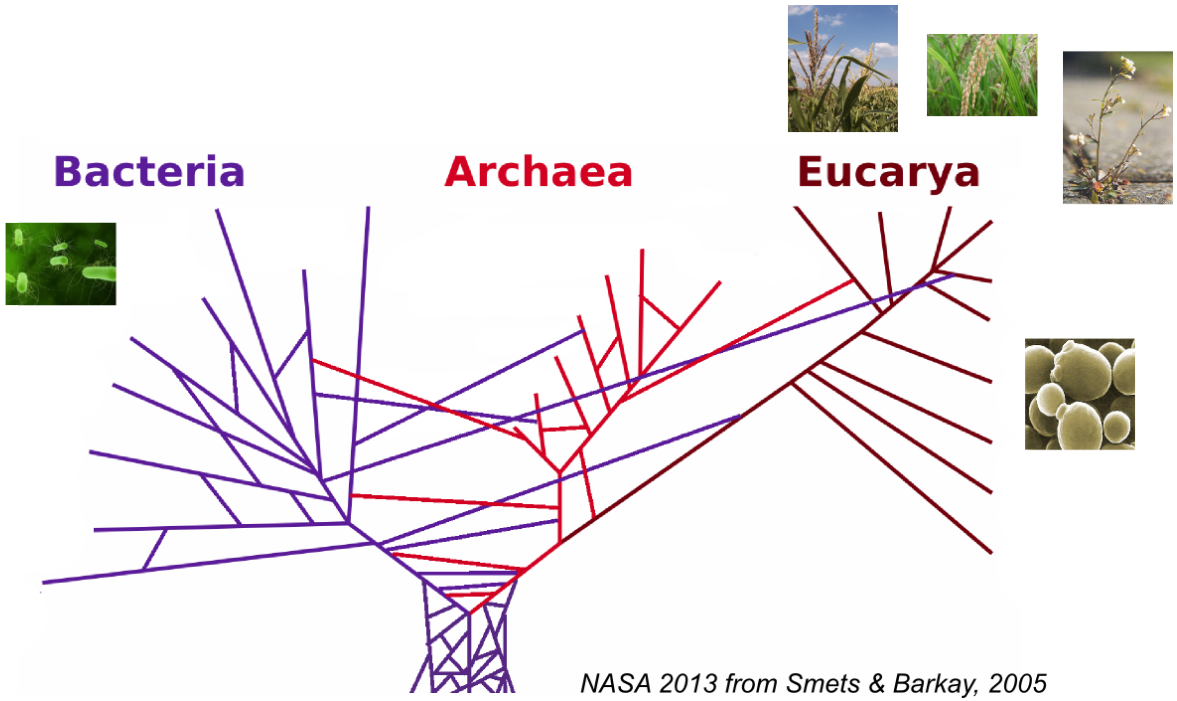
**Evolution** : Set of modifications accumulated through time through

**Reproduction, Heredity, Variability**

# Evolution



Darwin, 1837



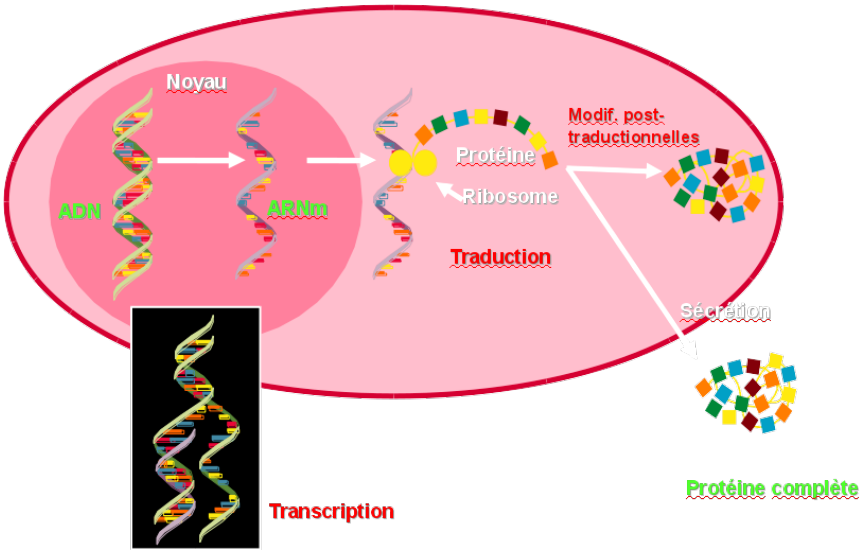

NASA 2013 from Smets & Barkay, 2005

Living species have a complex evolutionary history that result from the combination of evolutionary pressures in changing environments.

| **Variation** | **Transmission** | **Sorting** |
|---|---|---|
| mutations, exchanges | heredity | selection, random drift |

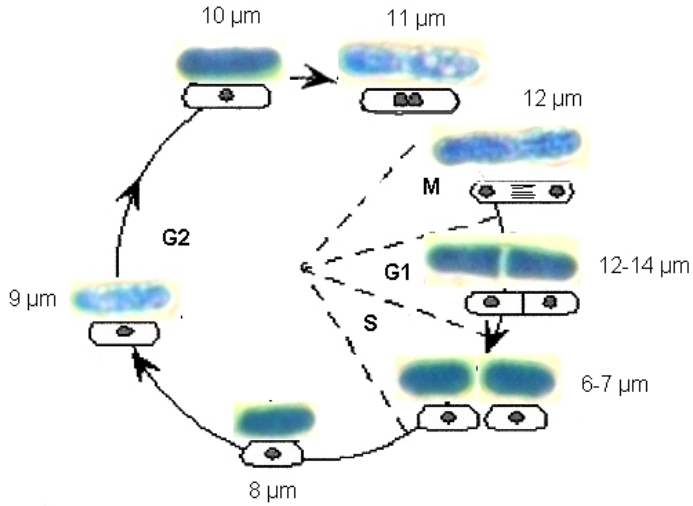
# Heredity and information theory

Set of laws that describe how an organism can be reconstituted from some information, generally encoded in a more compact form.




*DNA contains the information that can generate the cell machinery.*

*Reproduction ensures the transmission of cells characteristics from one generation to the next.*



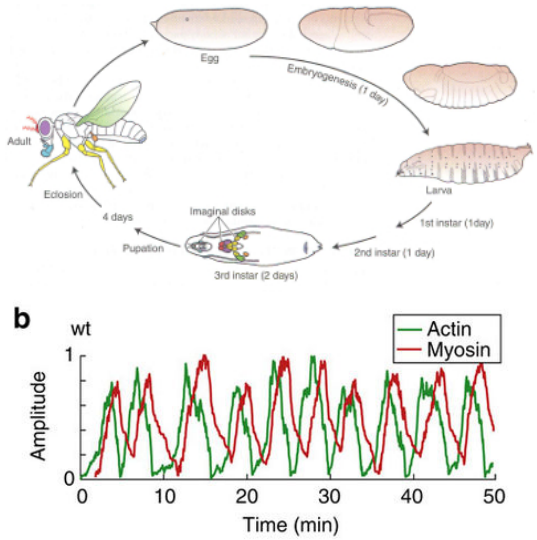

S. pombe, Didier Pol, 2002

Daughter cells inherit from the full DNA content of their mother, but only part of the cytoplasma.

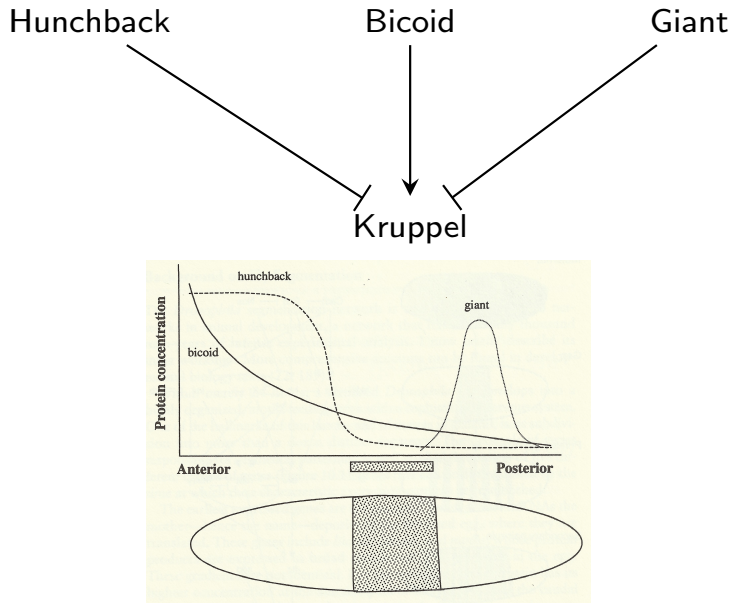
# Developpement

The developmental program is not encoded in the DNA sequence. It results from dynamical molecular processes that take place within cells.

**Example : drosophila**




**Oscillations :** Contractil mechanisms during egg morphogenesis result in the elongation along the antero-posterior axis thanks to asynchronous cell divisions.
`Valencia-Exposito et al. 2016, Nature Com. 7:10746`




**Morhogenetic gradients :** The spatial location of the Kruppel protein is structured along the anterio-posterior axis of the drosoplila egg.

Temporality of molecular interactions during development leads to the dynamical construction of the adult form.

# Outline

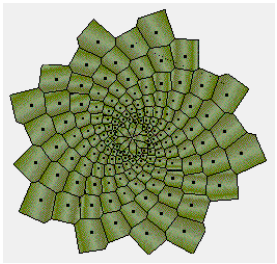
# Emergence

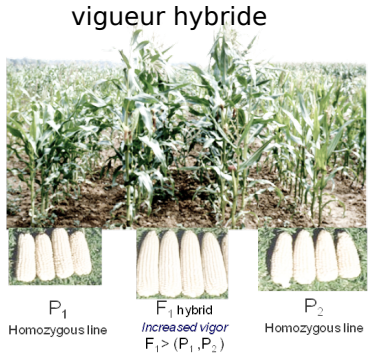
*System's properties cannot be directly predicted from the properties of its parts (DNA)*
The whole is more than the sum of parts.

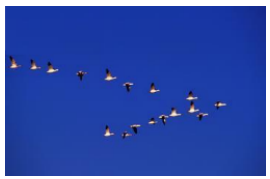
A. Matthies, A. Stephenson, N. Tasker, 2010

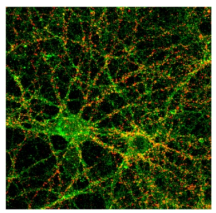
The properties at a level of observation result from the properties of the previous level, but are not reductible to the properties of the previous level. They are often difficult and sometimes impossible to predict.

- High number of parameters
- Entanglement
- Stochasticity


phyllotaxy





social behaviours

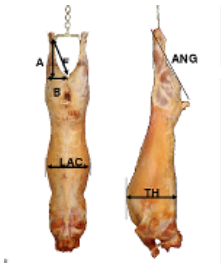

memory


taste

# Genotype and Phenotype

**Genotype :** Set of characteristics of the DNA sequence that identify an individual.

**Phénotype :** Single measurable trait that result of the *expression of numerous genes* in interaction with the *environment*. Can be defined at different observational scales, comprising cellular and molecular scales.

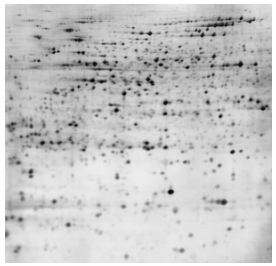



Conformation

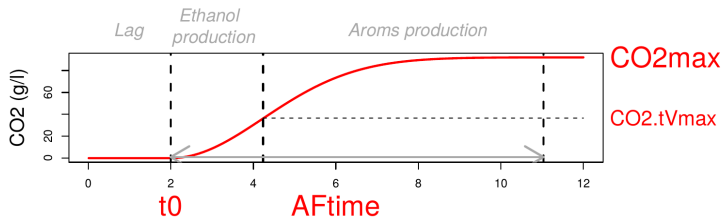
Architecture


Wing number

Weight



Protein abundance




kinetic parameters

One genotype $\leftrightarrow$ several phenotypes (1:n)

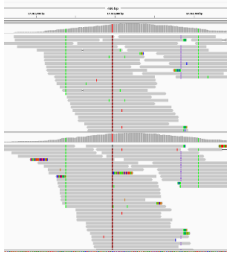
Several genotypes $\leftrightarrow$ same phenotype (n:1)

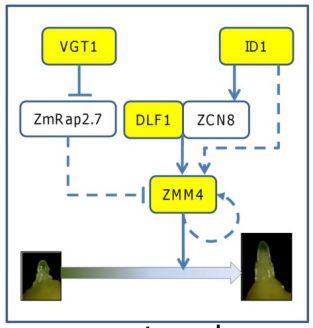
# Outline

# From genotype to phenotype in changing environments

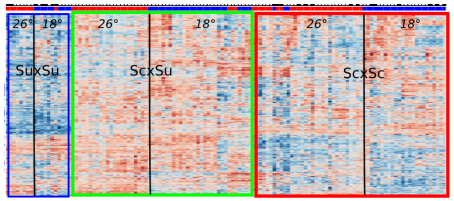
Genetic variation

polymorphisms



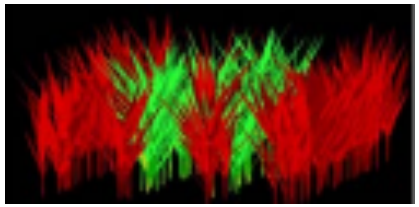


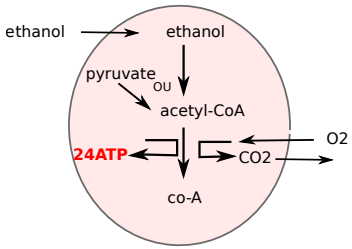

gene networks




molecular phenotypes

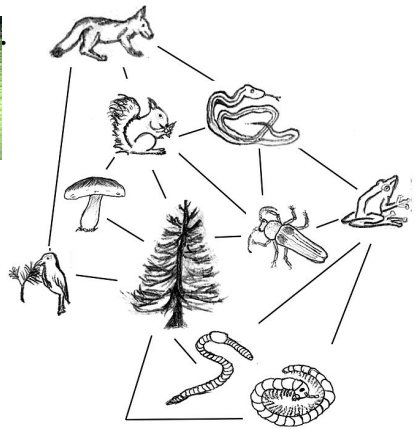
development






metabolic network

biotic environment



functional genomics

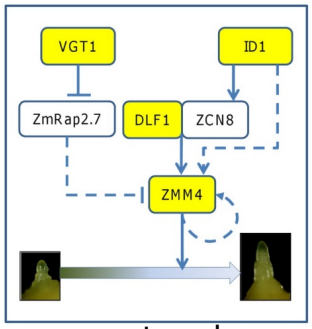
physical environment

trophic network

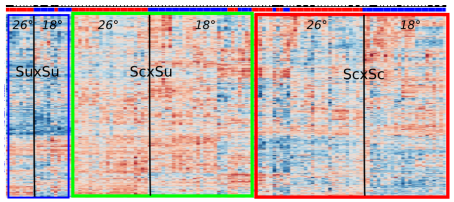
Phenotypic variation

# Dealing with H.SMITH data

Genetic variation

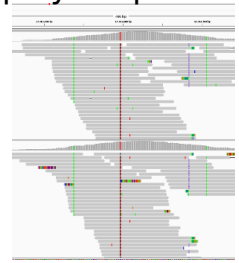


gene networks

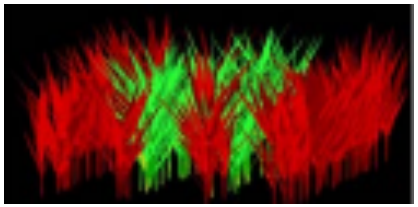


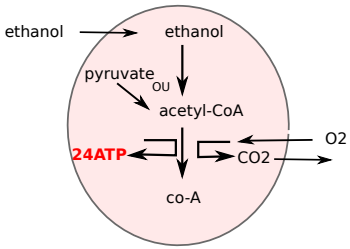
molecular phenotypes

polymorphisms



development





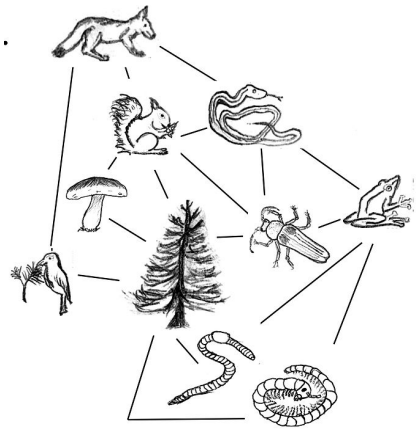
metabolic network

*Heterogeneity*
*Scale entanglement*
*Missing datas*
*Incompleteness*
*Temporality*
*High dimension*

biotic environment

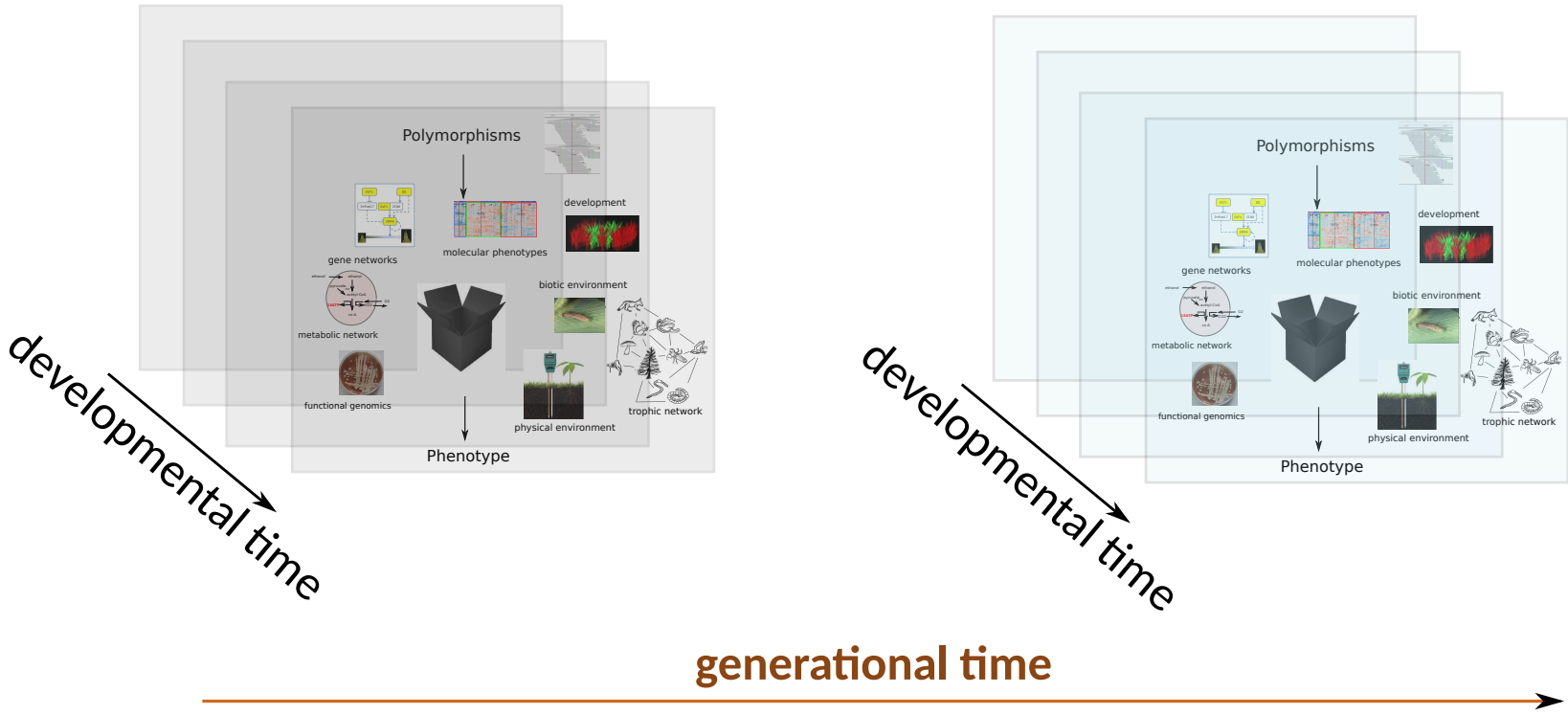

functional genomics


physical environment


trophic network

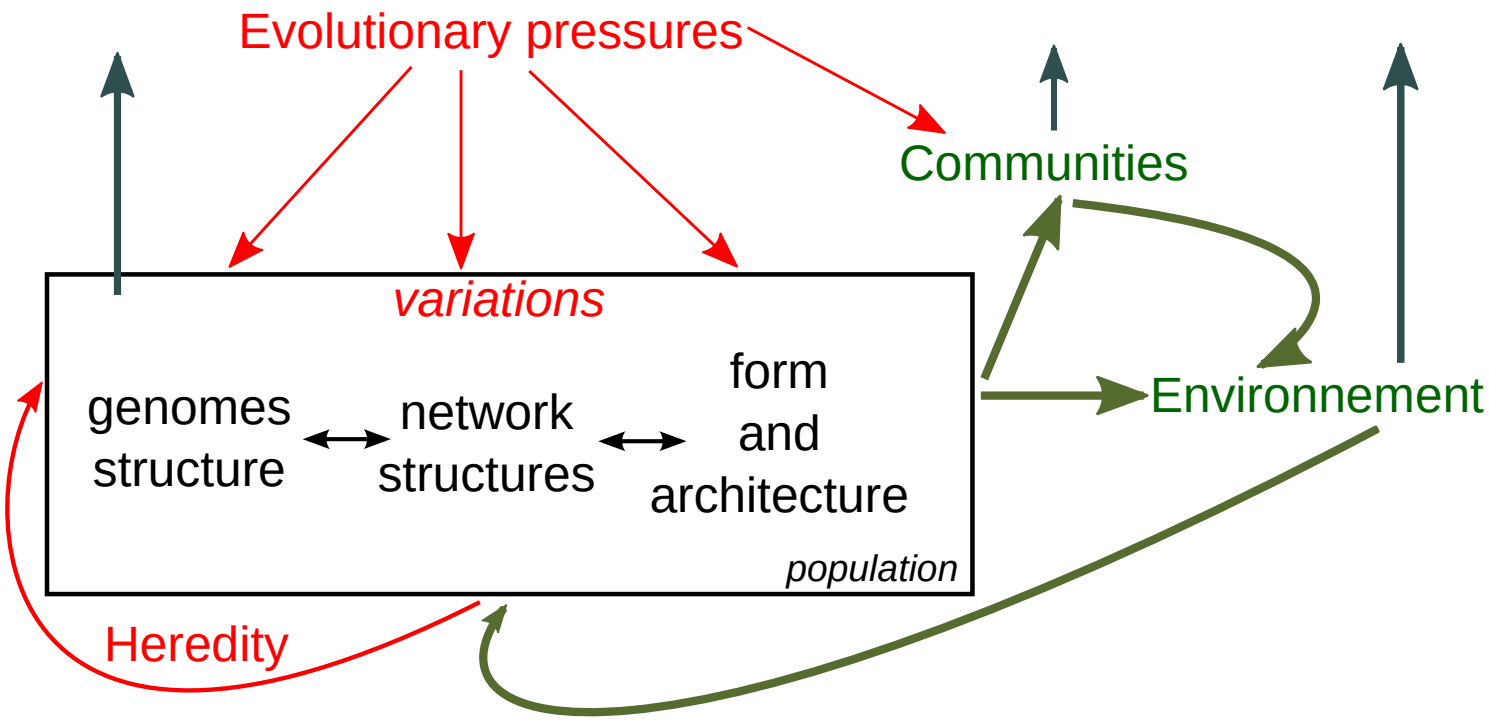
Phenotypic variation

# Temporalities




At the **individual** level, the developmental time conditions the phenotypes. At the **populational** level the generational time conditions genotypic changes. Each living organism has its own temporalities.

# From genotype to phenotype in changing environments



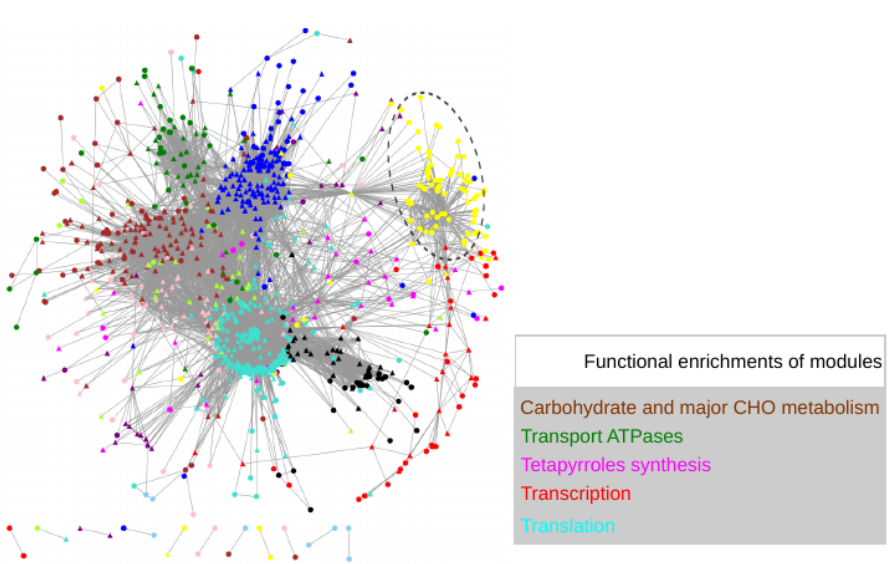

How to predict the phenotype taking into account the complexity, heterogeneity, and temporality of the observations ?

# Outline

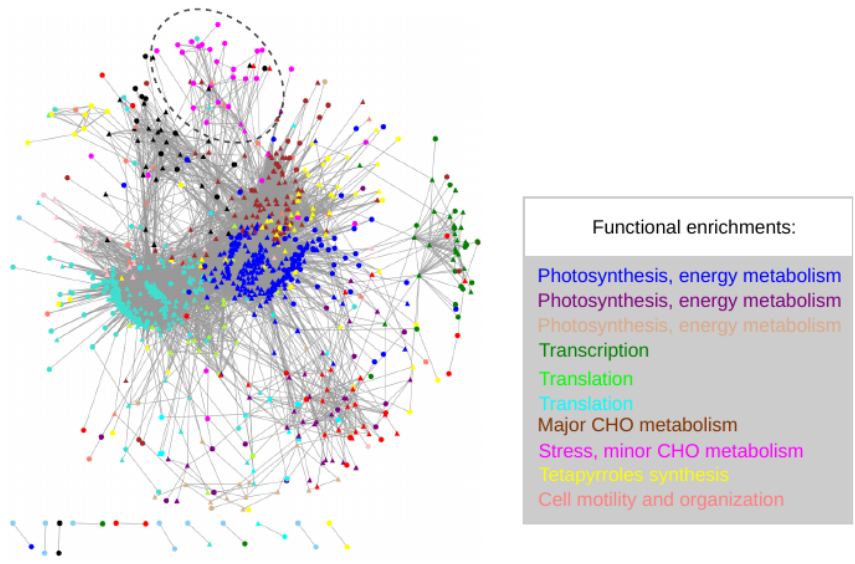
# 01. Maize proteomic response to water-deficit

System's genetics: identification of co-expression modules.

Blein-Nicolas et al, 2019, http://dx.doi.org/10.1101/636514




Control




Water-deficit

One node = one protein. One edge = co-expression in different genotypes. One colour= one **functional cluster**.

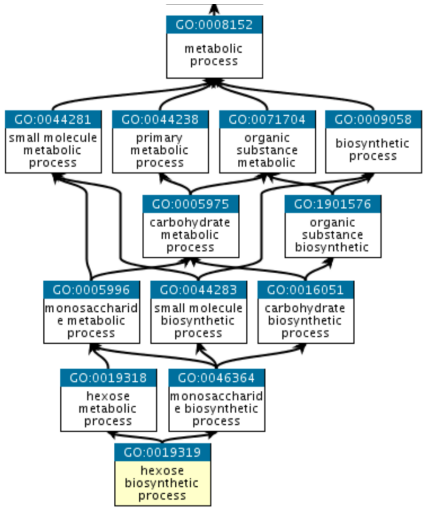
The protein co-expression network changes in response to water deficit.

# Ontologies

## The Gene Ontology

Each gene is attributed

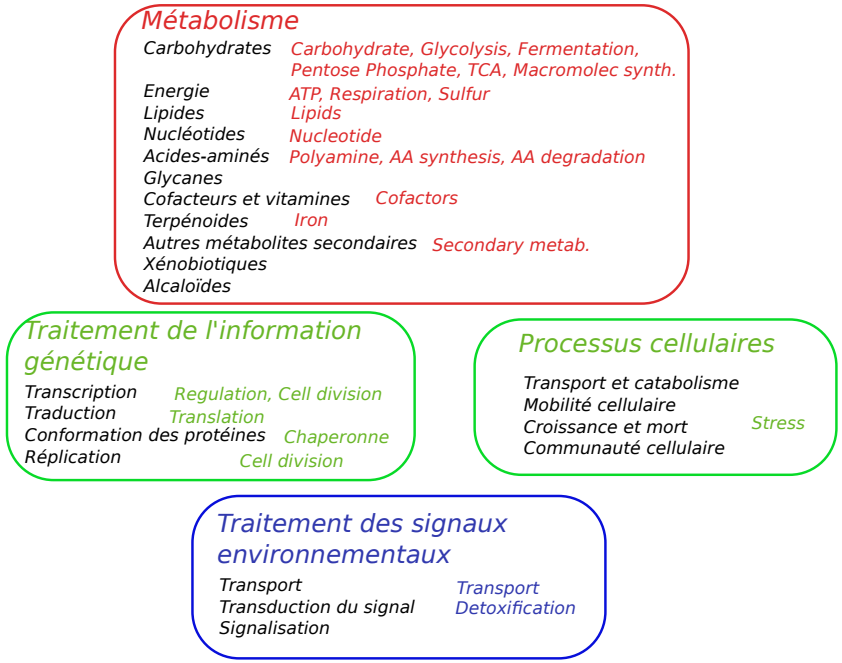
- a molecular function
- a biological process
- a cellular component




Within each domain, terms are ordered in a **loosely hierarchical** manner.

## KEGG pathways

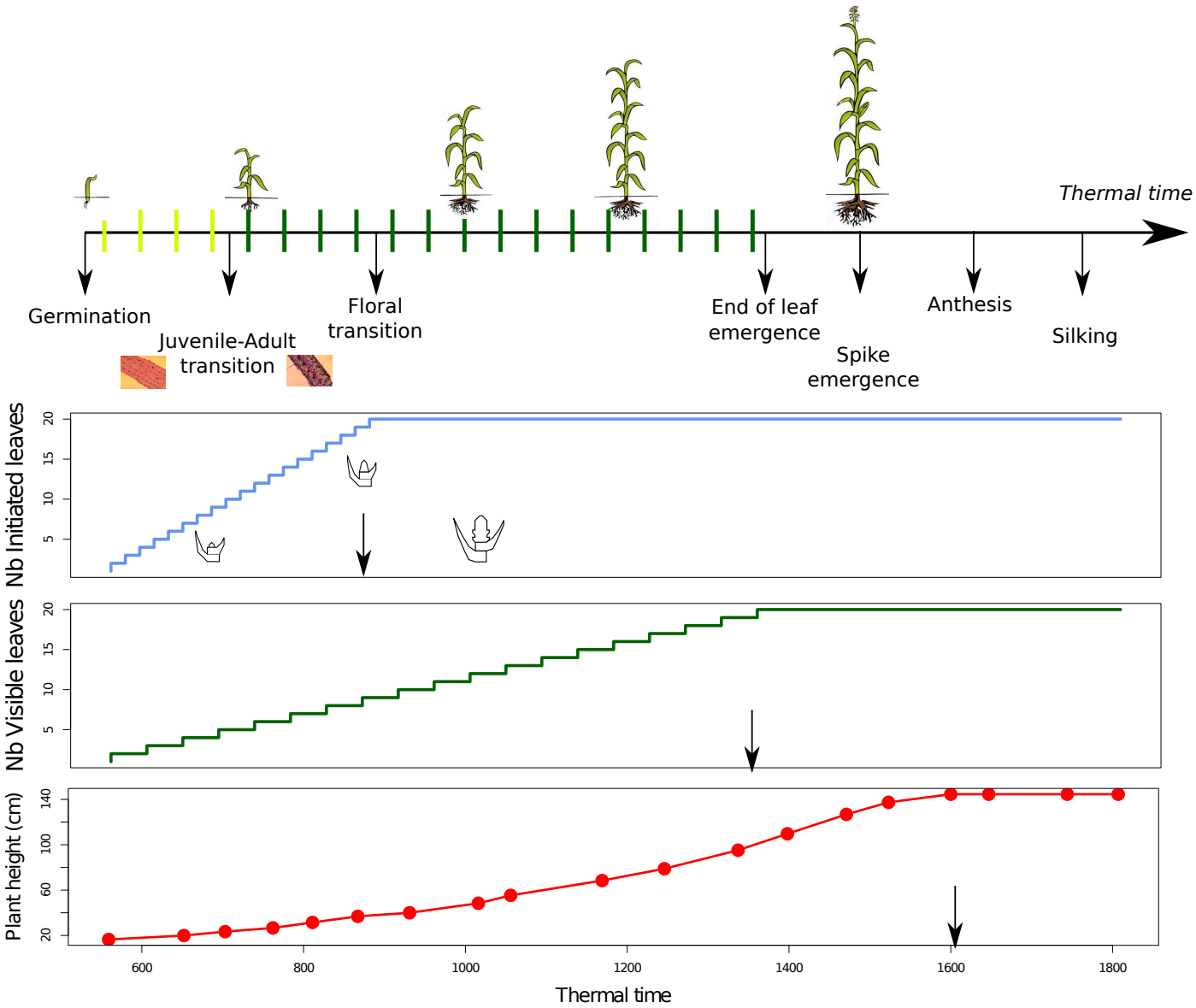
Biological processes are organized into **pathways**




Pathways refer to different hierarchical levels of the Gene Ontology

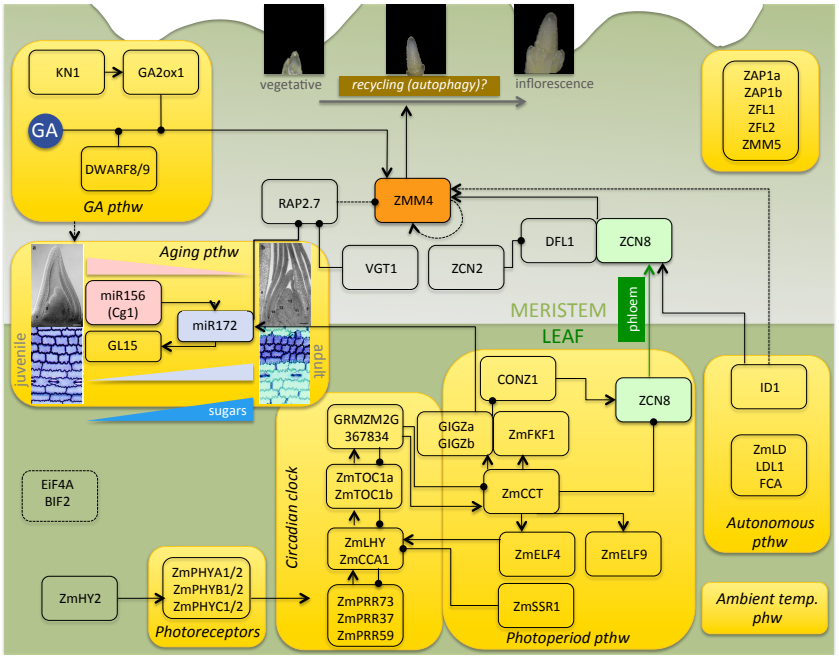
How to organize the information ? How to deal with multifunctionality (one colour = one *overrepresented* function) ?

# 02. Maize developmental transitions



# Gene regulatory network of maize floral transition

## Expert knowledge

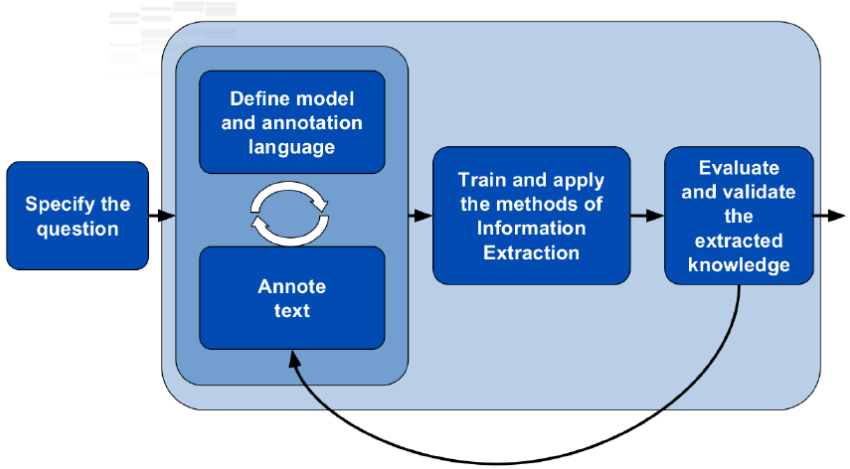



Add information from an automatic exploration of the scientific litterature ?

## Itemaize-Bibliome project

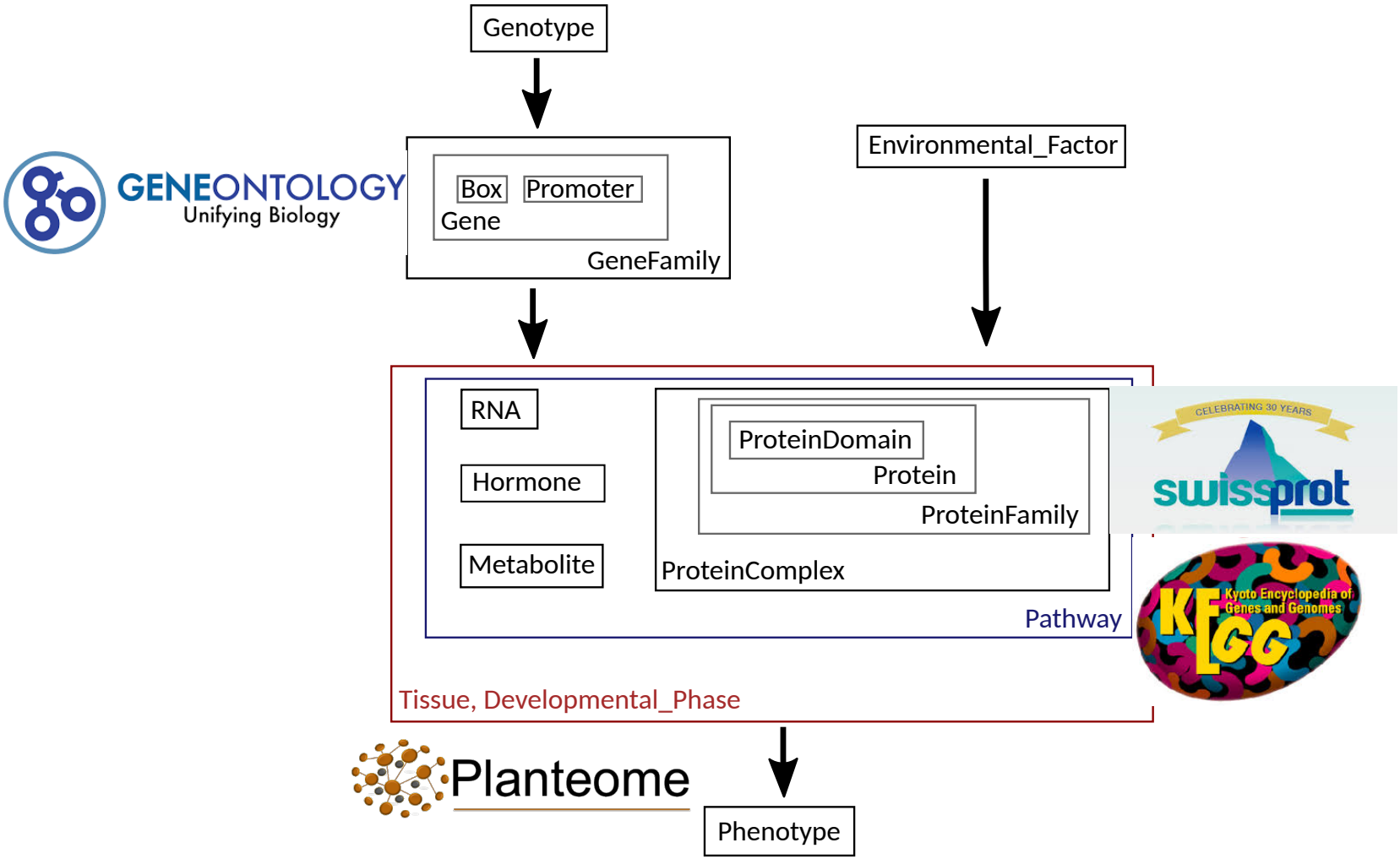
MAiAGE, GQE-Le Moulon, IJPB

Build-up from the Seedev project and transpose to maize floral transition ?



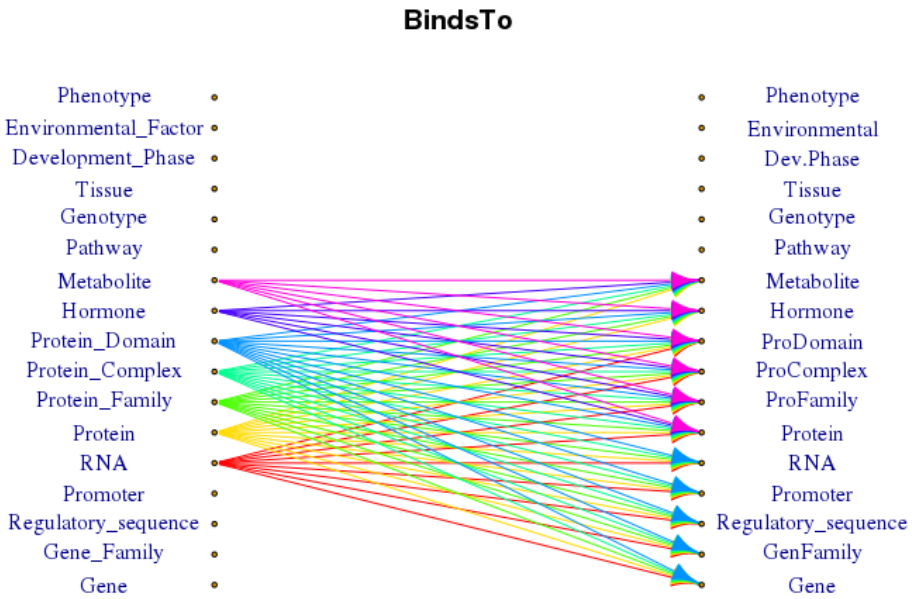
# Knowledge Graph : Entities

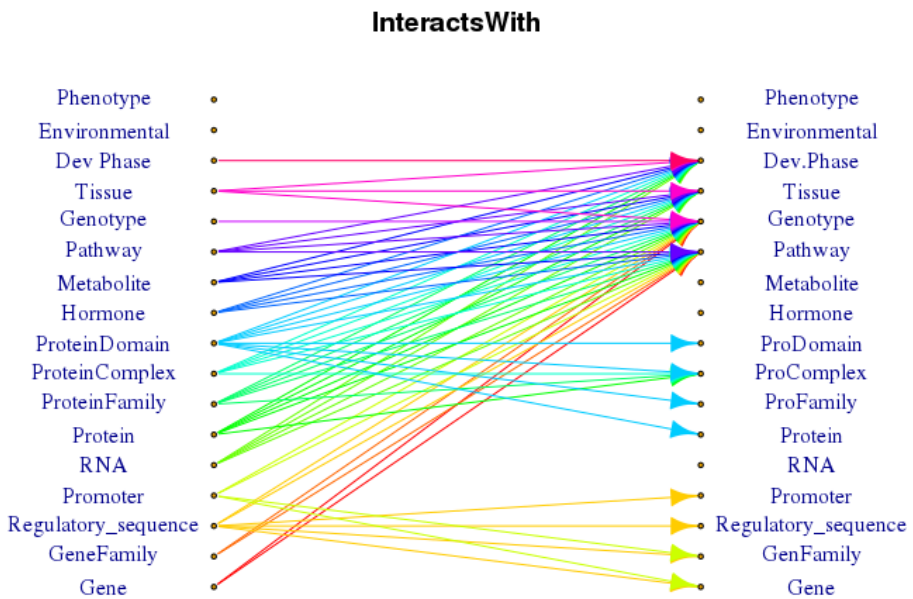
17 different entities were defined. For each entity, possible instances were related to existing ontologies. Hierarchization based on expert knowledge helped to define relations between entities.



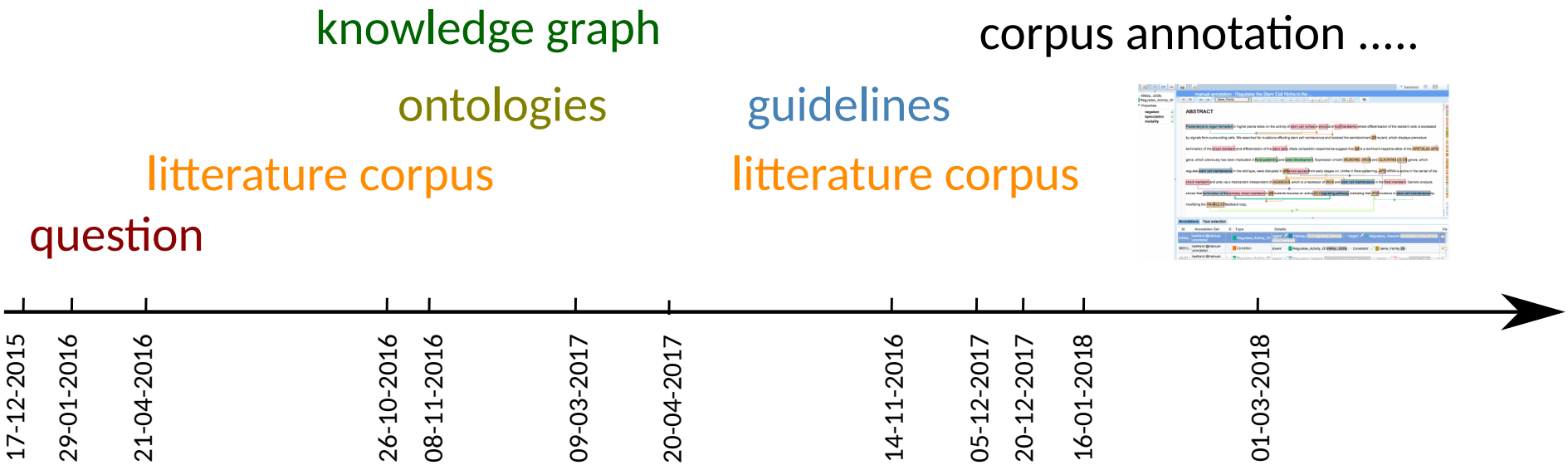
# Knowledge Graph : Relations

9 relations between entities were defined, as well as a comparison operator between entities.



# Itemaize-bibliome : caveats of interdisciplinary work




A side-project (sparse funding, PIA Amaizing) involving eight biologists and five informaticians during two years.

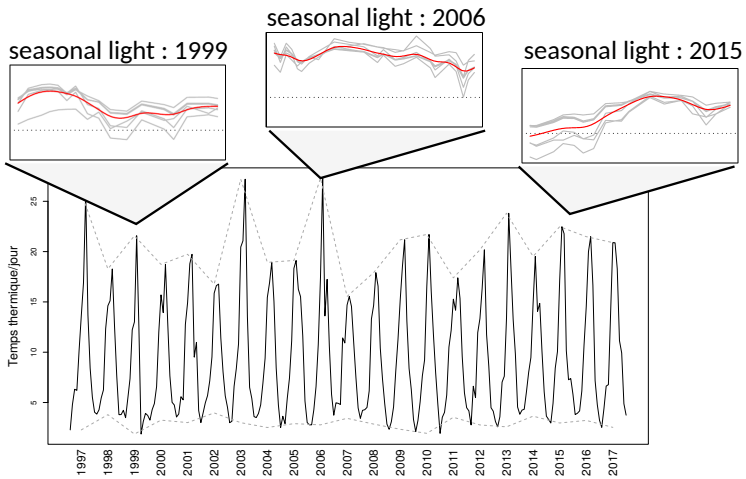
- Corpus annotation is time-demanding and not immediately rewarding ?
- Which kind of funding for such projects ?

# 03. DATAIA Warm-rules project

Predict phenotypic variation from climatic series ?

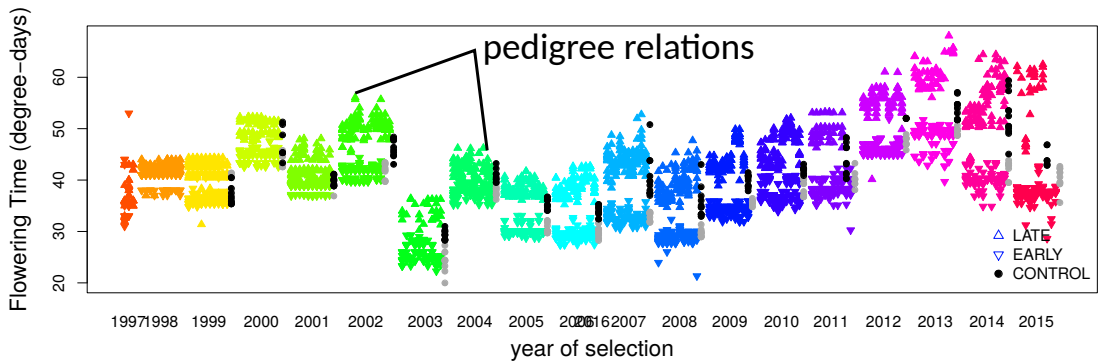
Juliette Debie (AgroParisTech), Fatiha Saïs (LRI), Elodie Marchadier (GQE-Le Moulon), ...

## Climatic series




## Phenotypic series

Saclay's divergent selection for flowering time in maize




(Pid1 /\ Pid2 /\ ...) /\ (Cvar1 /\ Cvar2 /\ Gvar3 ...) => (Pvar1 /\ Pvar2 /\ ...)

contextual identity — cause — phenotype

*climatic variables*
*genotypic variables*

*flowering time*
*plant height*

**Mobilizing** Population genetics, ecophysiology, statistical modeling, AI

# Take-Home messages

- In living organisms, phenotypes have emergent properties that results from a multitude of causes. Interactions are the rule.
  - Organizing knowledge $\longrightarrow$ Knowledge graphs
  - Observation-based clustering
  - Inferring causal relationships

- Which kind of funding and rewards for the necessarily interdisciplinary (and stimulating) work ?