Gene family analysis is an important way to understand complex processes underlying stress response in crops. Several tools exist to study families and propose automatically clustered families or curated published families but automatic clustering is rarely sufficient for precise studies. Biologists need most of the time to manually constitute their families.

We propose to develop an integrative system that will allow to gather sequences and informations from different sources for the analysis and visualisation of a customized family as a synthetic and dynamic view. This work is part of a PhD project and will eventually be integrated in the SouthGreen bioinformatics platform. http://www.southgreen.fr/

**Keywords**: Gene family, data integration, comparative genomics, data visualization

### Family Analysis

The tool integrates several tools for gene family analysis. These tools are GMOD components or are implemented in Galaxy workflow Manager. The tool is also linked to CoGe\(^{(5)}\) the Banana Genome Hub\(^{(2)}\) and the Coffee Genome Hub\(^{(3)}\) in order to get functional information on sequences.

**Selected Family**
- Send the designed family to analysis tools

**Analyse**
- Run Analysis in Galaxy or in CoGe Tools

**Result Visualisation**
- Visualise the results in Galaxy

### Family Constitution

For a customized gene family, the constitution is done in two steps: the constitution of the seed of the family, and the enlargement of the sequence set. The seed set can be selected through gene name, interpro domain, or keywords search. The databases that are queried for the moment are Chado databases\(^{(3)}\) of *Musa acuminata* and *Coffee canephora*. To enlarge the datasets to similar sequences in other species, the tool is linked to HMMer and BLAST workflows in Galaxy workflow manager\(^{(5)}\).

**Constitution of the seed of the family**
- Database
- Gene Name
- Interpro Domain
- Key words
- Personal data
- Sequence Set

**Enlargement of sequence set**
- HMMer
- BLAST

**Constitution of a «cart» of sequences**
- Add sequences
- Remove sequences
- Download Fasta
- At any moment

### Data Visualization (To be done)

Visualization is a crucial step for data understanding. The tool will display an integrative and dynamic view of the results of the gene family analysis and information about the members of the family. The integrative visualization will allow interaction between different views that will contain for example phylogenetic tree, syntenic dotplot, pathway information, genome browser, evidences for stress response...

**Members of the family**
- Display each member of the family and their structural annotation and allow to select one or more sequences and to remove an unwanted element

**Functionnal and relationnal annotation**
- Display informations about selected sequences
- Pathway, Annotation

**Genome Browser**
- Display tracks allowing to visualise gene structure, genomic environment and further informations if available: (SNPs, ESTs...)

**Synteny**
- Graph and dot plot allowing to visualize the synteny of the region around our sequences of interest

**Phylogeny**
- Phylogenetic tree allowing to visualisation of other data available such as expression data, domains, gene structure

### Other Functionalities (To be done)

**History**
- An history of analysis and family modifications will be created so that the user can go back at any moment and keep a track of analysis parameters and why he added or removed a sequence.

**Specific Orientation (stress response)**
- Stress response specific information will be linked to the tool with controlled vocabulary, known annotations, bibliography. These informations will come from stress response specific databases.