

# A personalized, integrated system for rice functional genomics

Pierre Larmande

**CIRAD, UMR 1096 DAP, Avenue Agropolis, F-34398 Montpellier 05, France**

# Context

- Our lab is involved in rice functional genomics.
- Three plant databases, namely OryGenesDB, OryzaTagLine and GreenPhylDB.
- OryGenesDB : reverse genetics and genomic DB  
OryzaTagLine : phenotypic mutant lines DB.  
Greenphyl : plant comparative genomics platform.
- DB interconnections needed.

## Passport data

Line: AJP D10

Cocultured callus	Other lines generated from the same callus
Construct	p4979
T2 seed stock	available
Delivered	No
FST	8AK7D03-8AK7D03

## Mutant: APN G08/0

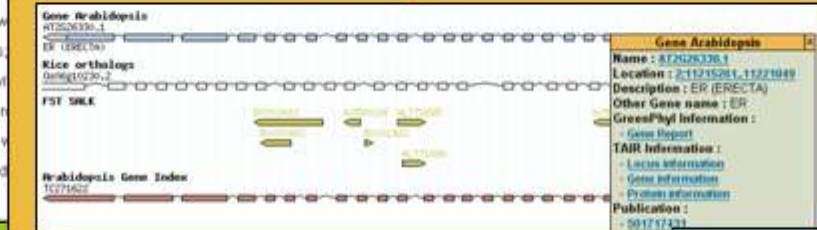
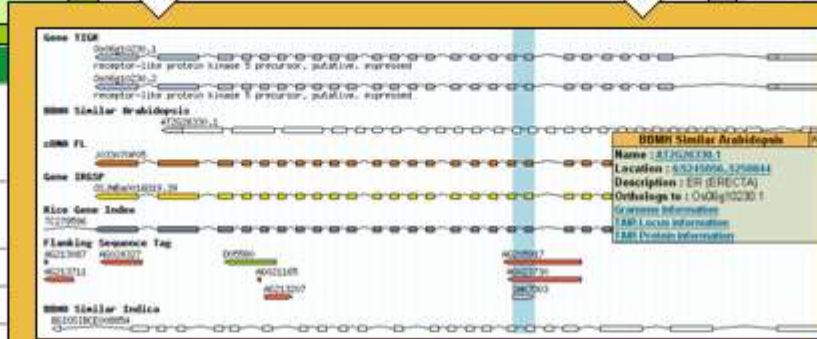
Referenced or designated mutants	dwarf (abbr. Dw)
Phenotypic class	Morphology
Trait	Size
Organ	all organs
Developmental stage	tillering to heading

**Phenotype description**  
Dwarf. Decreased plant size (in proportion), late flowering panicles incompletely exerted and abnormal, few panicle branches; high sterility; abnormal panicles, to dark green, normal to twisted leaves; with or without small (shorter leaves than average size leaves), normal to semi-rolled, narrow leaves; normal to low tillering; without thin tillers; normal to leaf blade with an undulating (wavy) shape; normal to compact plant.



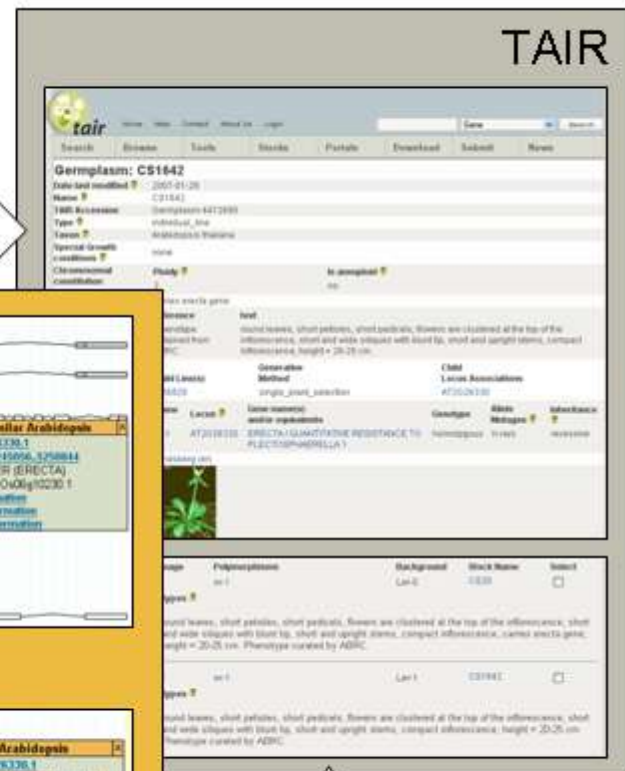
OryzaTagLine

Sequence : phenotype



OryGenesDB

Sequence : sequence



TAIR

Sequence : phenotype



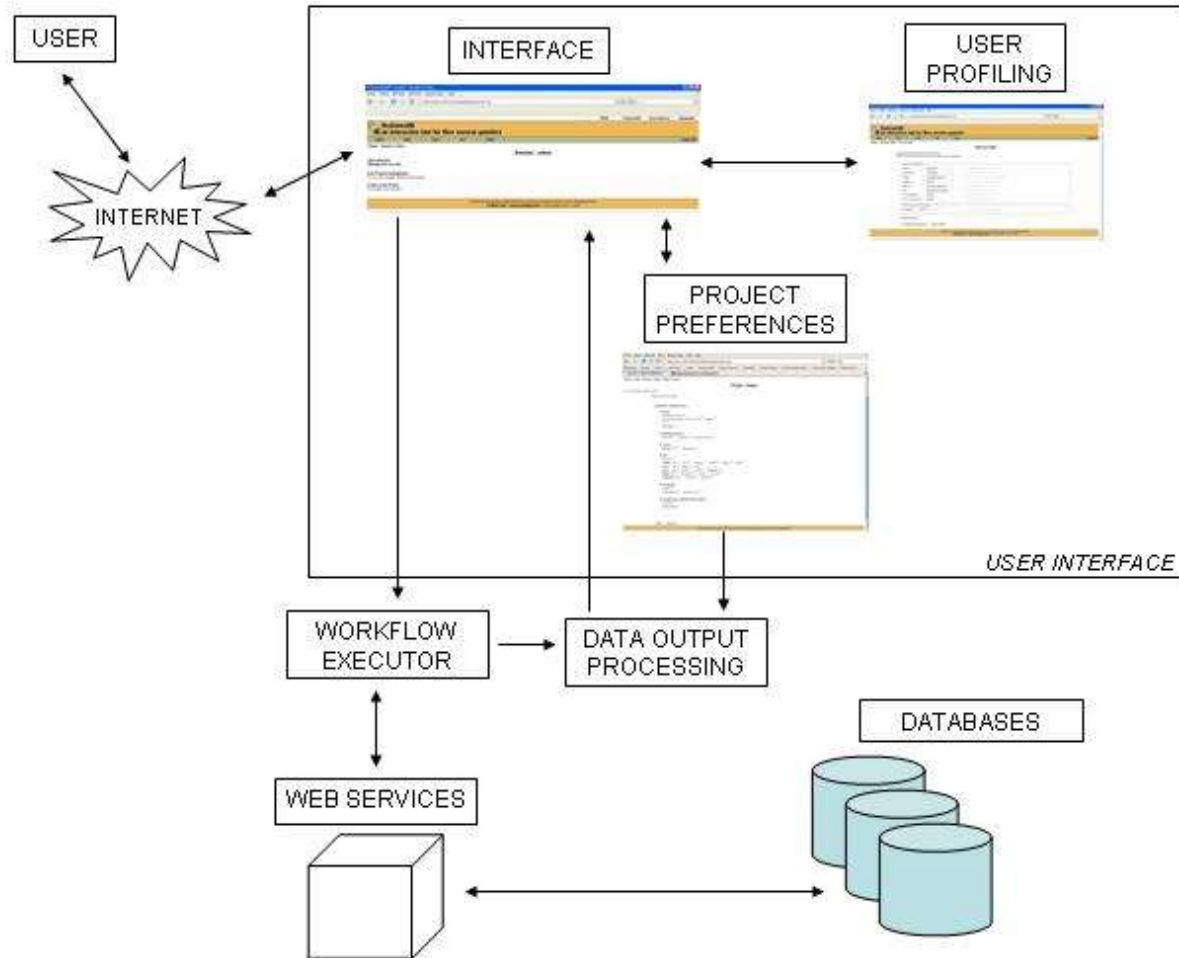
GreenPhyloDB



# Objectives

- Scientists need flexible systems which reproduce automatically their actions and generate synthetic results.
- Our aim is to design and implement a modular system that allows transparent complex queries across these databases using web services.

# The system architecture



# Why using BioMoby ?

- Specific to bioinformatic web services.
- Allows access to hundreds of services.
- Our services are available for a large community and can be discovered with ontology features.
- BioMoby has 3 ontologies : service ontology (e.g. retrieval, analysis) , namespace ontology (e.g. NCBI\_gi, ICIS\_germplasm), data type ontology (e.g. Gene, Dnasequence, etc).

# Data types ontology

- We have used a sub domain developed by the Generation Challenge Programme which provides **genetic, genomic and germplasm data types**. (Bruskiewich et al, 2006)



Registry Browser

- heaven.msri.on.ca
- icapture.mrl.ubc.ca
- icapture.ubc.ca
- imb.uq.edu.au
- inb.bsc.es
- inblosam.com
- llama.med.harvard.edu
- mapman.mpimp-golm.mpg.de
- mips.gsf.de
- mmb.pob.ub.es
- mnon.csic.es
- moby.ucalgary.ca
- NNN.NNN.NNN
- orygenesdb.cirad.fr
  - getFSTByGeneId
  - getGeneIdByFST
  - getGeneIdByLocation
  - getInfoByGeneId
- plantgdb.org
- samples.jmoby.net
- sapporo.genomics.purdue.edu
- schematikon.org

Object

- Allele
- AlleleAssociation
- Amigo
- Annotation
- Array
- AvailableMaterial
- BasicClone
- BasicGFF3MultiF
- BasicGFF3Seque
- BasicGFFSequen
- BlastJob
- BlatJob
- Boolean
- BoundingBox
- BriefOccurrenceR
- BriefTaxonConce
- CDNALibrary
- ChinookServiceDI
- Chromatogram\_E
- Collector
- CommentedObjec
- CompartmentPrec


Service Types

- Service
  - AdministrativeService
  - Analysis
  - Bioinformatics
  - Conversion
  - Edit
  - GetFastaConsensiFromGrou
  - hem
  - ObjectHandling
  - Parsing
  - Registration
  - Rendering
  - Resolution
  - Retrieval
    - Searching
    - ServicesAdministrativos
    - ServicesAnnotation
    - ServicesClusterQueries
    - ServicesSequenceQueries
  - SetOperation
  - Shim
  - TestExtractFetcher
  - Testing

Name: getFSTByGeneId  
 Type: Retrieval  
 Category: moby  
 Auth: orygenesdb.cirad.fr  
 Desc: This service takes a gene identifier (ex : 0s03g51030 or 0s03g51030.1) and returns associated F  
 URL: http://sat.cirad.fr/sat/Orygenesdb/services/getFSTByGeneId  
 Contact: orygenesdb@cirad.fr  
 LSID: urn:lsid:biomoby.org:serviceinstance:orygenesdb.cirad.fr,getFSTByGeneId:2007-06-06T14-39-37Z  
 Signature URL:  
 Path to RDF:  
 Primary inputs:  
 Name: gene\_id  
 Data Type:  
 Name: GCP\_SimpleIdentifier  
 Auth:  
 Desc:  
 Contact:  
 LSID: urn:lsid:biomoby.org:objectclass:GCP\_SimpleIdentifier:2006-11-28T07-40-56Z  
 ID: null  
 Namespaces:  
 Name: OryGenesDB\_Orysa  
 Auth:  
 Desc:  
 ID: null

append mode  verbose

## About Dashboard...



# moby Dashboard

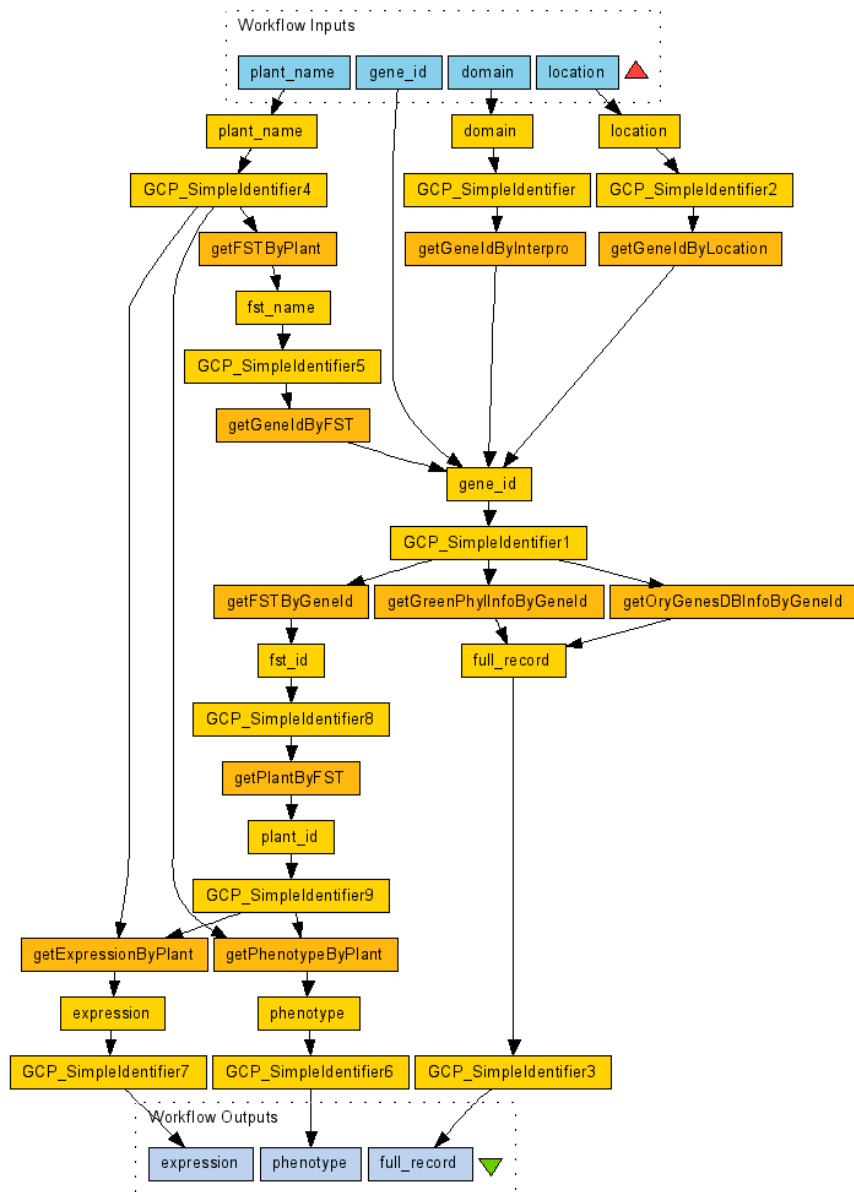
### Dashboard panels

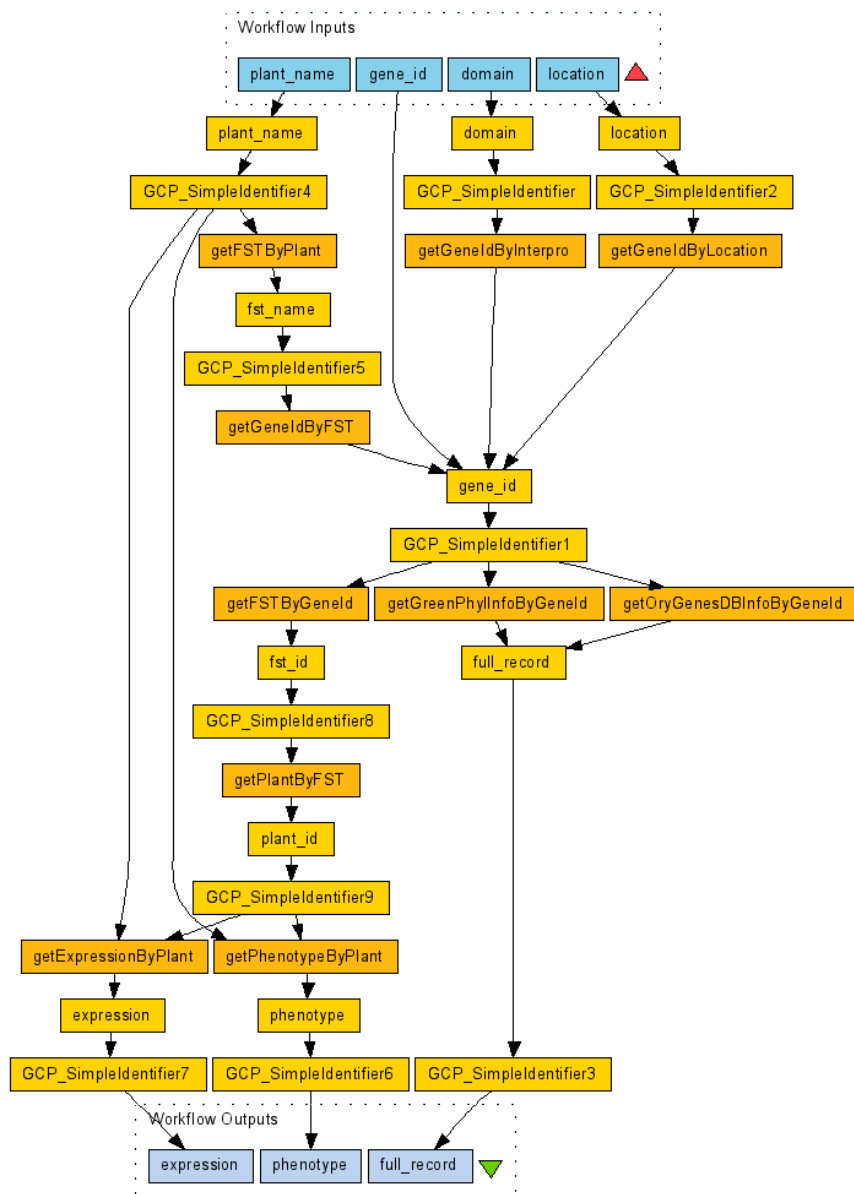
- Registry Browser**  
 A panel showing all Biomoby entities, allowing different sort orders. It also defines which Biomoby registry to use and how and where to cache Biomoby entities locally.
- Registration**  
 A panel allowing to register and unregister any Biomoby entity.
- MoSeS Generator**  
 A panel allowing to generate datatypes and skeletons that can be used by Biomoby service providers to implements their services.
- Simple Client**  
 A panel allowing to create input data and to call an arbitrary Biomoby service. Its purpose is mostly for the service developers to test their new services.

Contact: Martin Senger <martin.senger@gmail.com>



Workflows have been designed with the BioMoby plug-in of Taverna.





- We developed **10 services** organized into **4 workflows**.
- 4 data input could be **location, interpro domain, Gene id, plant name**.
- Each data output participates in a full record summary.

## A personalized integrated system for rice functional genomics

S. FROMENTIN, G. DROC and P. LARMANDE  
 CIRAD, UMR 1095 DAP, Avenue Agropolis,  
 F-34398 Montpellier 05, France  
 Corresponding author: pierre.larmande@cirad.fr

# Acknowledgments: Sébastien Fromentin Gaëtan Droc

### Introduction

We recently developed three plant databases specifically designed for rice functional genomics :

- OryzaGenesDB [1], a reverse genetics and genomic database ,
- OryzaTagLine [2], a phenotypic mutant lines database,
- GreenPhyDB [3], a database for comparative functional genomics to link the two models plant species rice and Arabidopsis through orthologs predictions.

Scientists need flexible systems which reproduce automatically their actions and generate synthetic results.

The aim of the project is to design and implement an architecture which retrieve data from these three remote databases, store queries and results, execute users queries periodically.



### Service definition

We defined several Web Services according to data that are required to :

- extract information from these databases,
- create interconnections between them. (dynamically interconnected database.)

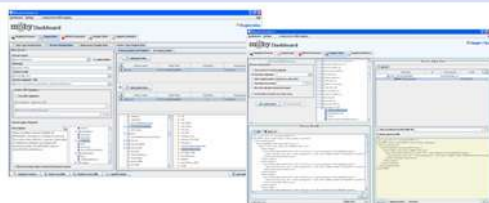
The table lists all Web Services that have been created from our databases. Gene id refers to the Locus Name defined by TIGR and/or TAIR database. Location refers to genomic coordinates. FST id refers to the Flanking Sequence Tag mapped on rice genome and is directly linked to a plant name. Plant name refers to OryzaTagLine.

Number	Web Service Name	Source	Input	Output
1	getOryzaGeneId	OryzaGenesDB	Gene MCO	Full record
2	getOryzaGeneId	OryzaGenesDB	Gene MCO	FST id MCO
3	getOryzaMdbLocusName	OryzaGenesDB	LocusName	Gene MCO
4	getOryzaMdbFST	OryzaGenesDB	FST MCO	Gene MCO
5	getPhenotypeByPlant	OryzaTagLine	Plant name	Full record
6	getExpressionByPlant	OryzaTagLine	Plant name	Full record
7	getFSTByPlant	OryzaTagLine	Plant name	FST id MCO
8	getOryzaMdbInterpro	GreenPhyDB	Interpro domain	Gene MCO
9	getOryzaGeneId	OryzaGenesDB	Gene MCO	Full record
10	getPlantByFST	OryzaTagLine	FST id	Plant name

### Service deployment

Web Services were generated using BioMoby dashboard [4]. First, we registered namespace and services for each database, and we used data types defined by the Generation Challenge Programme [5], like GCP\_SimpleIdentifier or GCP\_Feature.

Next, with the Moses module, we generated data types and skeletons, that we used to implement with Eclipse. Finally, the "Single Client" panel allows us to call our services and display results.

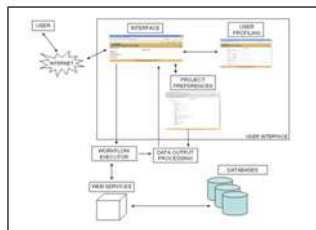


### Results

We developed a web interface to create user accounts and register user profiles. All profiles and preferences are stored in a MySQL data management system and queries are automatically and periodically executed.

With the BioMoby plug-in of Taverna [6], we designed a set of predefined workflows, thus simplifying access to end users. Users can start workflow according to 4 inputs : **location, interpro domain, gene id and plant name**. Entries like "gene\_id" and "fst\_name" make it possible to link our databases.

Workflows have been designed to access and retrieve pertinent information from the different databases (phenotype description, expression, rice Arabidopsis gene orthologs informations...).



### Conclusion

Using this architecture, biologists can automate their web queries therefore reducing the time wasted on several web interfaces to extract the pertinent information. With the stored preferences, users can personalize the system to display synthetic relevant information. They can also be alert when updates are available on the different databases.

References :

- [1] <http://oryzagenesdb.cirad.fr>  
 [2] <http://urgi.versailles.inra.fr/OryzaTagLine/>  
 [3] <http://greenphy.cirad.fr>  
 [4] [http://biomoby.open-bio.org/CVS\\_CONTENT/moby-live/Java/docs/Dashboard.html](http://biomoby.open-bio.org/CVS_CONTENT/moby-live/Java/docs/Dashboard.html)  
 [5] <http://www.generationcp.org/>  
 [6] <http://taverna.sourceforge.net/>



# Poster 15

# Thanks for your attention !