GnpIS, the plant information system of INRA URGI unit:
Towards a bridge between genomic and genetic data

Delphine Steinbach, Erik Kimmel, Michael Alaux, Sophie Durand, Cyril Pommier, Isabelle Luyten, Nacer Mohellibi, Daphné Verdelet, Aminah-Olivia Keliet, Sébastien Reboux, Hadi Quesneville
INRA Unité de recherche en Génomique-Info, Centre INRA Versailles, FRANCE
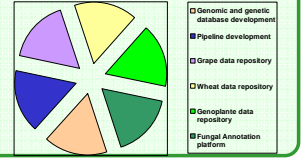
INRA

URGI Génomique-info

Summary: URGI is a bioinformatics research INRA unit (previously only dedicated to Genoplante projects but now open to other projects) . It hosts a bioinformatics platform which offers services (databases, tools, power units..) to biologists. It develops also an information system (I.S.) to store genomic and genetic data for plant, plant models or plant of agronomical interest (Wheat, Grapevine, Maize, Rape...) and its bioagressors (fungi). Our research axis are data integration (in our GnpIS information system) and dynamic genome study based on repeat detection (REPET software). The results presented here are the development of all the I.S., its evolution in time, its databases, data content and the current development concerning the creation of a new portal to query through all the databases of the S.I, using both Lucene (via HibernateSearch) and BIOMART (EBI) technologies.
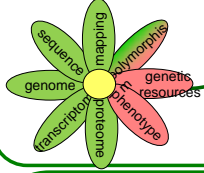
GnpIS description

The GnpIS information system is composed of 8 modules: 5 genomic modules and 3 genetic modules and above all these modules, *a new portal allowing to query through the different modules separately (via the flower) or simultaneously via datamarts or indexes (publication in prep. 2009).*

1) GnpSeq, the EST, mRNA sequence database which contains home-made clusters, contigs and annotations
2) GnpMap, the mapping database
3) GnpGenome, a multispecies database based on Chado/Gbrowse GMOD tools, containing genomic sequences and their annotations
4) GnpArray, the MIAME compliant expression database
5) GnpProt, the proteomic database
6) GnpSNP, the polymorphism (SNP/DIP/STR) database
7) SiReGal, Genetic Resources database for accessions collections
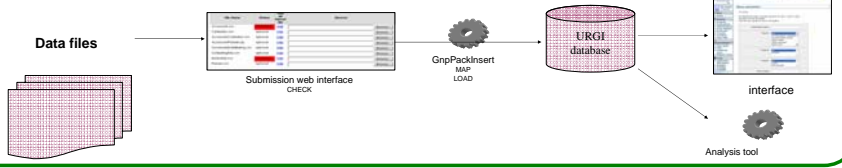8) Ephesis (development in progress), the genotypes, environment and experimentation database

Platform missions

- Genomic and genetic database development
- Pipeline development
- Grape data repository
- Wheat data repository
- Genoplante data repository
- Fungal Annotation platform

GnpIS data flow

Biologist Data

sequence, mapping, genome, transcriptom, proteome, polymorphism, genetic resources, phenotype

Data files

URGI-Extractor EXTRACT TRANSFORM

Submission web interface CHECK

GnpPackInsert MAP LOAD

URGI database

interface

Analysis tool

GnpIS interoperability project development

Project aim: « Bridge Genomics and Genetics »
Build an information system to store and query through all genetic and genomic data in a convivial and powerful way.

Project Organization:
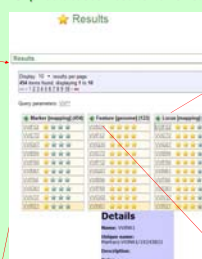**Phase I : 2001-** : building of each database module (Gnp*): « the Flower »
Phase II : 2005- : building of links between Gnp* interfaces : (GnpMap-GnpGenome, GnpSNP-GnpGenome, GnpMap-GnpSeq)

Phase III : 2007- :
Development of a new web query portal: *first version in december 08 with 2 new tools:*
-- a quick search query tool: based on **Lucene (Java/HibernateSearch) and indexes from GnpIS databases**
-- an advanced search query tool using **Biomart (EBI)**, based on GnpIS dedicated marts

GnpGenome-GnpSNP: SNP link

GnpMap-GnpGenome: marker link

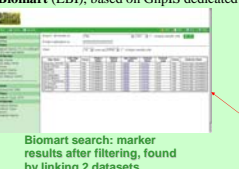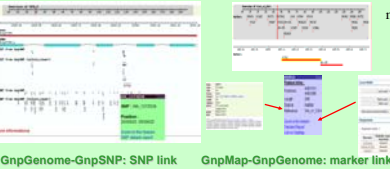Biomart search: marker results after filtering, found by linking 2 datasets GnpMap-GnpGenome

Quick search (Lucene) for VVI * in all Gnp* databases: results items list

VVIF52 locus details in GnpMap

VVIF52 marker details in GnpMap (not shown)
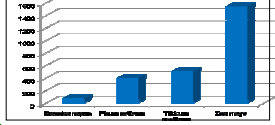
VVIF52 feature details in GnpGenome (not shown)

GnpIS databases results

Number of new EST/mRNA Sequences per year

- 2003
- 2004
- 2005
- 2006-2007

Pisum sativum, Helianthus, Medicago truncatula, Oryza sativa, Zinnia elegans, Sorghum bicolor, Arabidopsis thaliana, Triticum aestivum, Brassica napus, Vitis vinifera, Zea mays
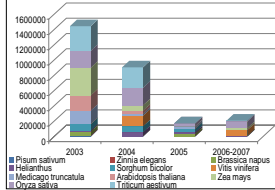
Number of Features by Species

Number of QTL by Species

QTLs map 1   QTLs map 2

URGI Graphical Alignment visualizer

Wheat 3B FPC physical map (AGI)

Vitis vinifera (v.1) Genome Browser

Flower modules: GnpSeq, GnpMap, GnpSNP, SIReGal, Ephesis, GnpProt, GnpArray, GnpGenome — Aster
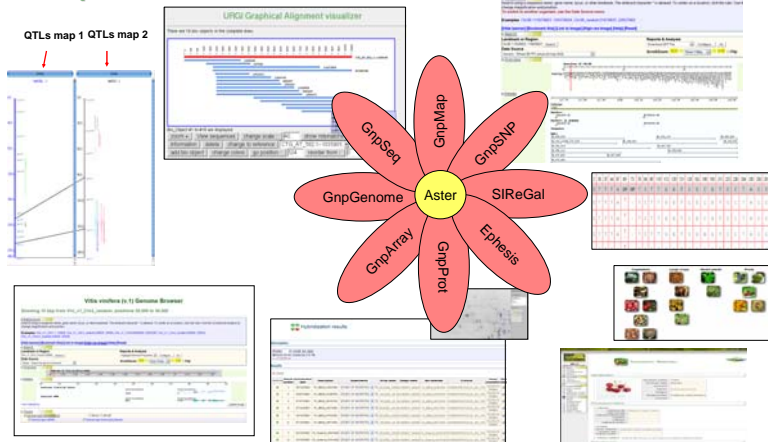
Number of Loci and Markers by Species

Sorghum bicolor, Pisum sativum, Brassica napus, Arabidopsis thaliana, Oryza sativa, Vitis vinifera, Zea mays, Triticum aestivum

Number of Loci, Number of markers
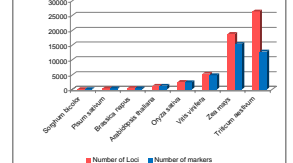
Number of SNP data in 2008

Number of subSNP by species: Oak, Rape, Maritime Pine, Wheat, Arabidopsis, Barley, Rice
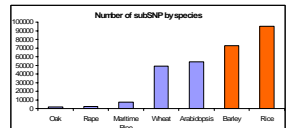
Number of Accessions by Specie: 25, 178, 240, 1100, 1526, 1744
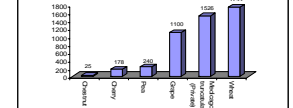
GnpIS time evolution and database interoperability progress
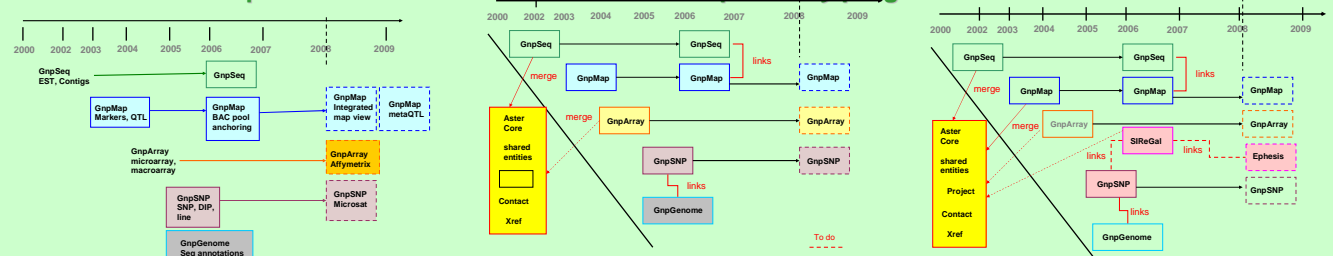
GnpSeq EST, Contigs → GnpSeq
GnpMap Markers, QTL → GnpMap BAC pool anchoring → GnpMap Integrated map view → GnpMap metaQTL
GnpArray microarray, macroarray → GnpArray Affymetrix
GnpSNP SNP, DIP, line → GnpSNP Microsat
GnpGenome Seq annotations

Aster Core shared entities, Contact Xref

merge, links, To do

URGI Contact: urgi-contact@versailles.inra.fr
INRA URGI Web site: http://urgi.versailles.inra.fr
and https://gpi.versailles.inra.fr for projects before publication (account)