URGI Plant and Fungi Platform: Distributed Resources through GMOD Tools

PAG XXII – GMOD session

January 10 - 15, 2014

joelle.amselem@versailles.inra.fr

URGI: Unité de Recherche en Génomique-Info

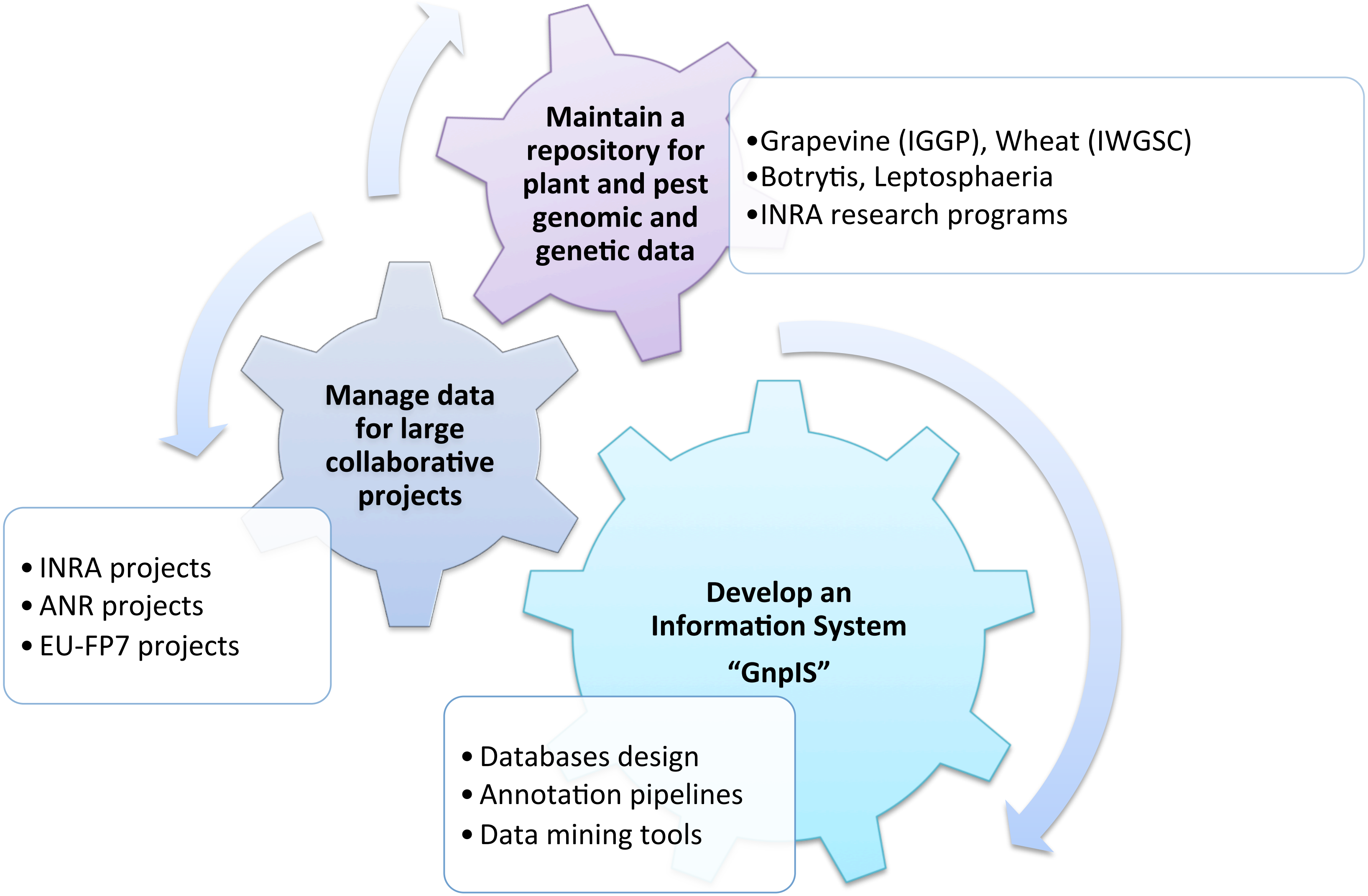
- Research Unit
  - INRA unit (French National Institute for Agricultural Research)
  - Plant breeding and Genetics Department
  - Strong connexions with other plant INRA departments
  - URGI Quality Managment System (QMS) certified iso 9001:2008 since 2012

- Bioinformatic platform
  - Member of the French National Network of Bioinformatic Platforms
  - Member of the French Bioinformatic Infrastructure (IBF)

- Research
  - Data integration
  - Genome structure and dynamics

joelle.amselem@versailles.inra.fr

Bioinformatic platform missions



Maintain a repository for plant and pest genomic and genetic data

- Grapevine (IGGP), Wheat (IWGSC)
- Botrytis, Leptosphaeria
- INRA research programs

Manage data for large collaborative projects

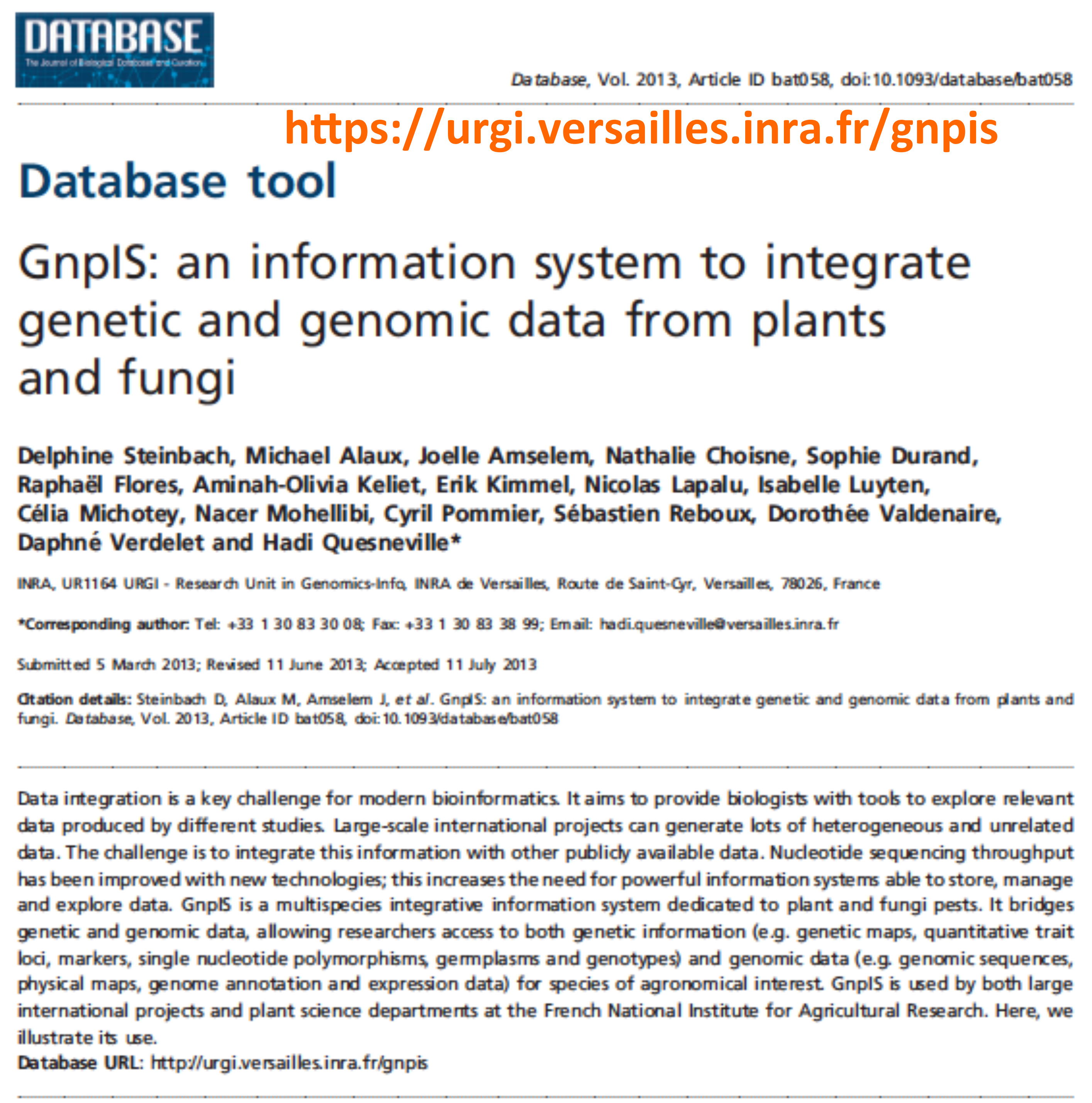
- INRA projects
- ANR projects
- EU-FP7 projects

Develop an Information System "GnpIS"

- Databases design
- Annotation pipelines
- Data mining tools

joelle.amselem@versailles.inra.fr

https://urgi.versailles.inra.fr/gnpis



DATABASE
The Journal of Biological Databases and Curation

Database, Vol. 2013, Article ID bat058, doi:10.1093/database/bat058

Database tool

GnpIS: an information system to integrate genetic and genomic data from plants and fungi

Delphine Steinbach, Michael Alaux, Joelle Amselem, Nathalie Choisne, Sophie Durand, Raphaël Flores, Aminah-Olivia Keliet, Erik Kimmel, Nicolas Lapalu, Isabelle Luyten, Célia Michotey, Nacer Mohellibi, Cyril Pommier, Sébastien Reboux, Dorothée Valdenaire, Daphné Verdelet and Hadi Quesneville*

INRA, UR1164 URGI - Research Unit in Genomics-Info, INRA de Versailles, Route de Saint-Cyr, Versailles, 78026, France

***Corresponding author:** Tel: +33 1 30 83 30 08; Fax: +33 1 30 83 38 99; Email: hadi.quesneville@versailles.inra.fr

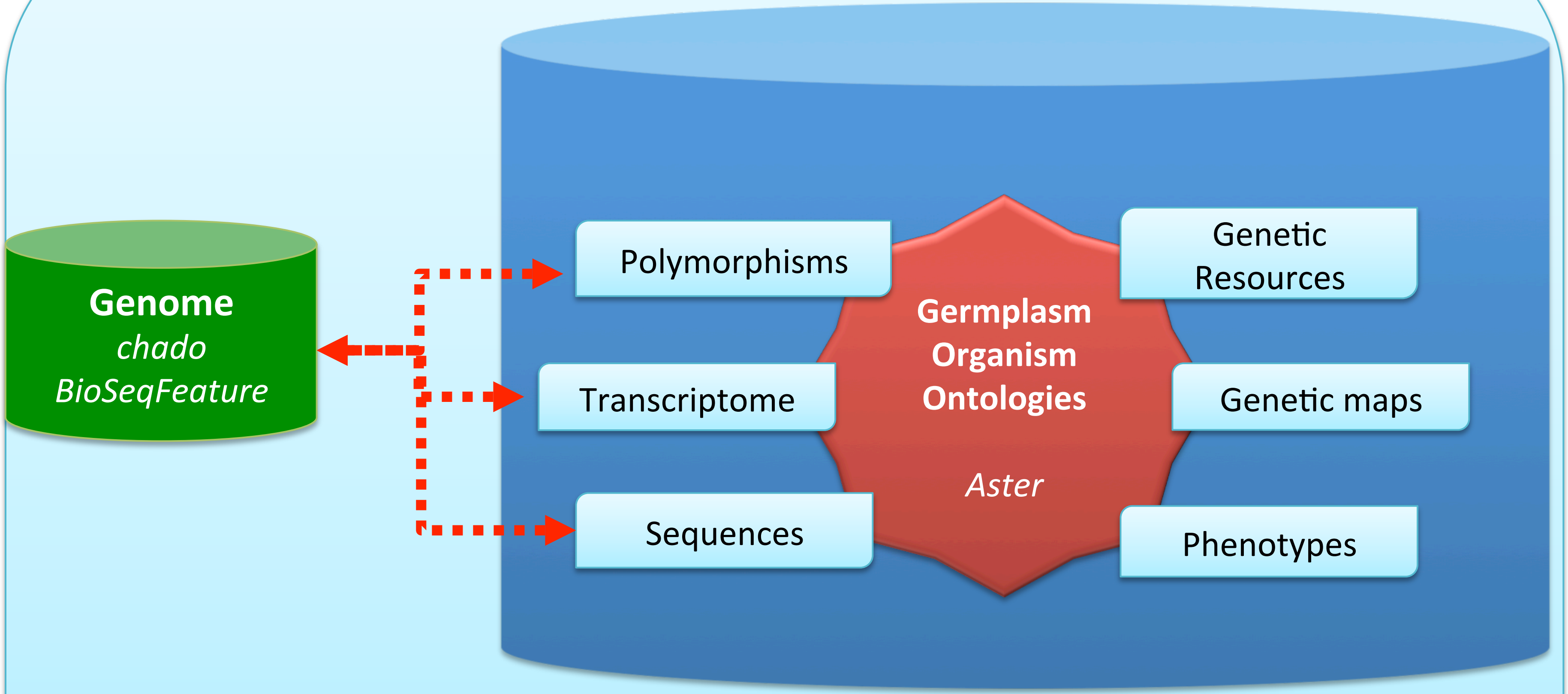
Submitted 5 March 2013; Revised 11 June 2013; Accepted 11 July 2013

Citation details: Steinbach D, Alaux M, Amselem J, *et al*. GnpIS: an information system to integrate genetic and genomic data from plants and fungi. *Database*, Vol. 2013, Article ID bat058, doi:10.1093/database/bat058

Data integration is a key challenge for modern bioinformatics. It aims to provide biologists with tools to explore relevant data produced by different studies. Large-scale international projects can generate lots of heterogeneous and unrelated data. The challenge is to integrate this information with other publicly available data. Nucleotide sequencing throughput has been improved with new technologies; this increases the need for powerful information systems able to store, manage and explore data. GnpIS is a multispecies integrative information system dedicated to plant and fungi pests. It bridges genetic and genomic data, allowing researchers access to both genetic information (e.g. genetic maps, quantitative trait loci, markers, single nucleotide polymorphisms, germplasms and genotypes) and genomic data (e.g. genomic sequences, physical maps, genome annotation and expression data) for species of agronomical interest. GnpIS is used by both large international projects and plant science departments at the French National Institute for Agricultural Research. Here, we illustrate its use.

Database URL: http://urgi.versailles.inra.fr/gnpis

Integrative database

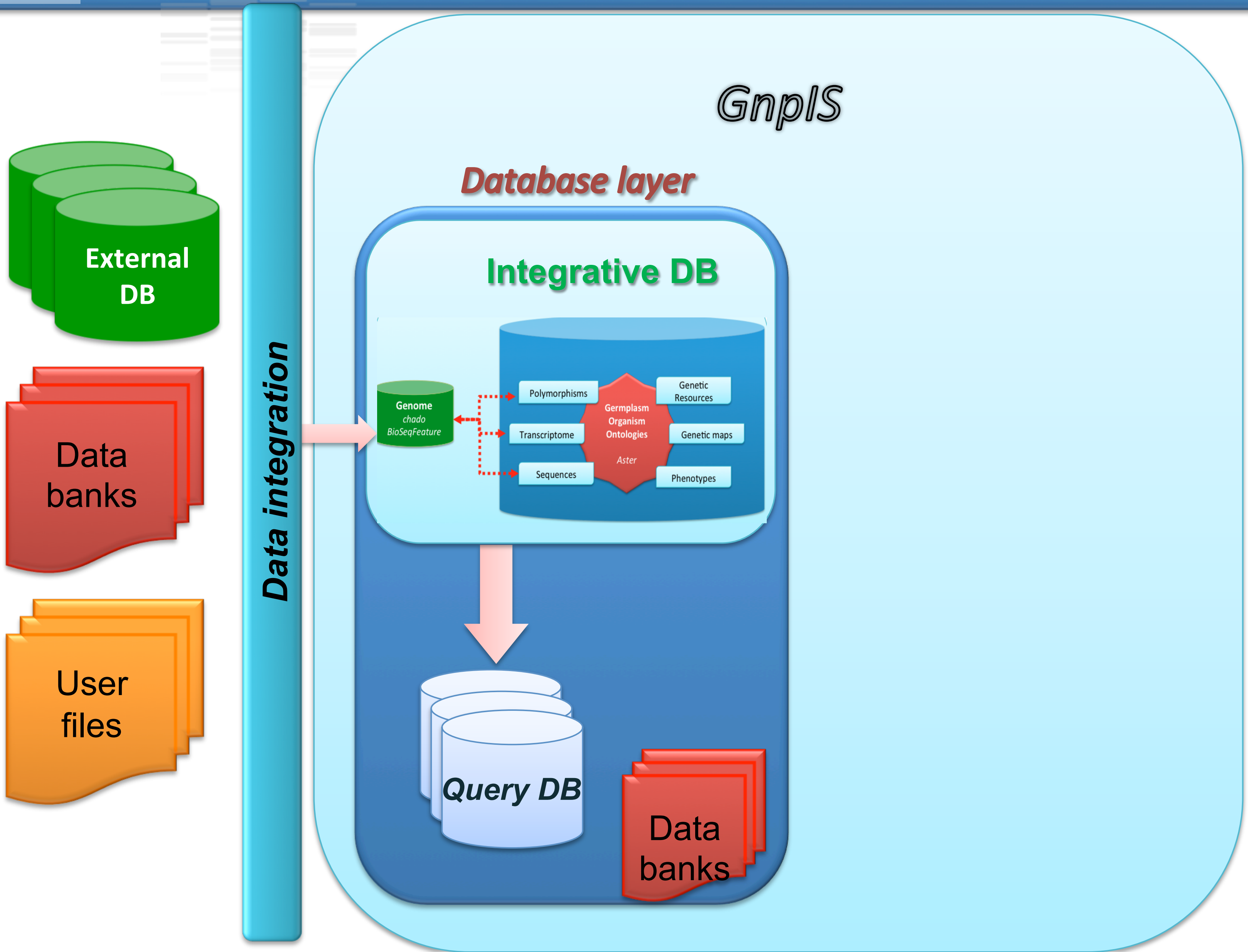


◀···▶ **Xref**

● **Interoperability pivots**

joelle.amselem@versailles.inra.fr

GnpIS architecture



External DB

Data banks

User files

Data integration

GnpIS

Database layer

Integrative DB

Genome
chado
BioSeqFeature

Polymorphisms

Transcriptome

Sequences

Germplasm
Organism
Ontologies
Aster

Genetic Resources

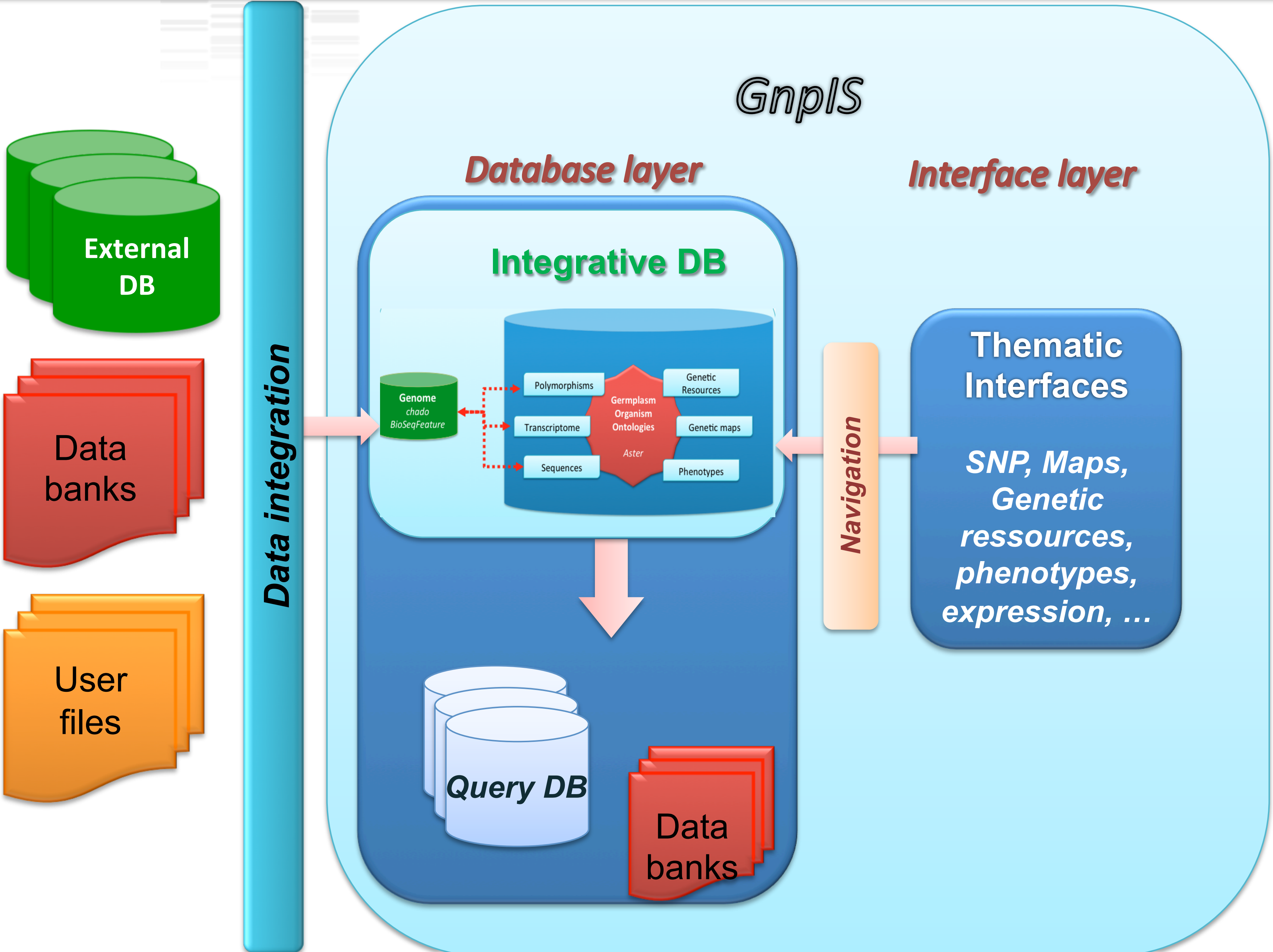
Genetic maps

Phenotypes

Query DB

Data banks

GnpIS architecture



External DB

Data banks

User files

Data integration

GnpIS

Database layer

Interface layer

Integrative DB

Query DB

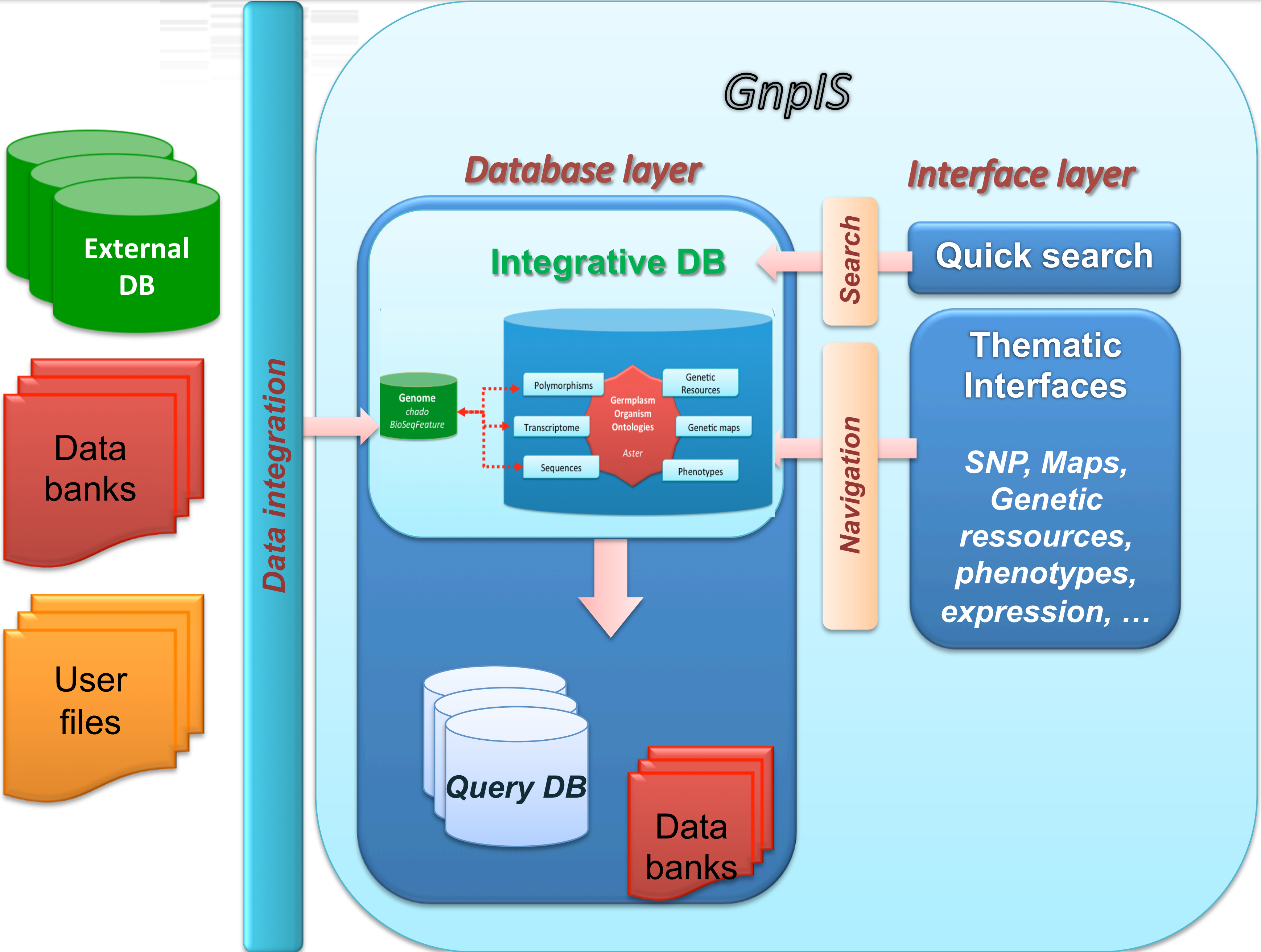
Data banks

Navigation

Thematic Interfaces

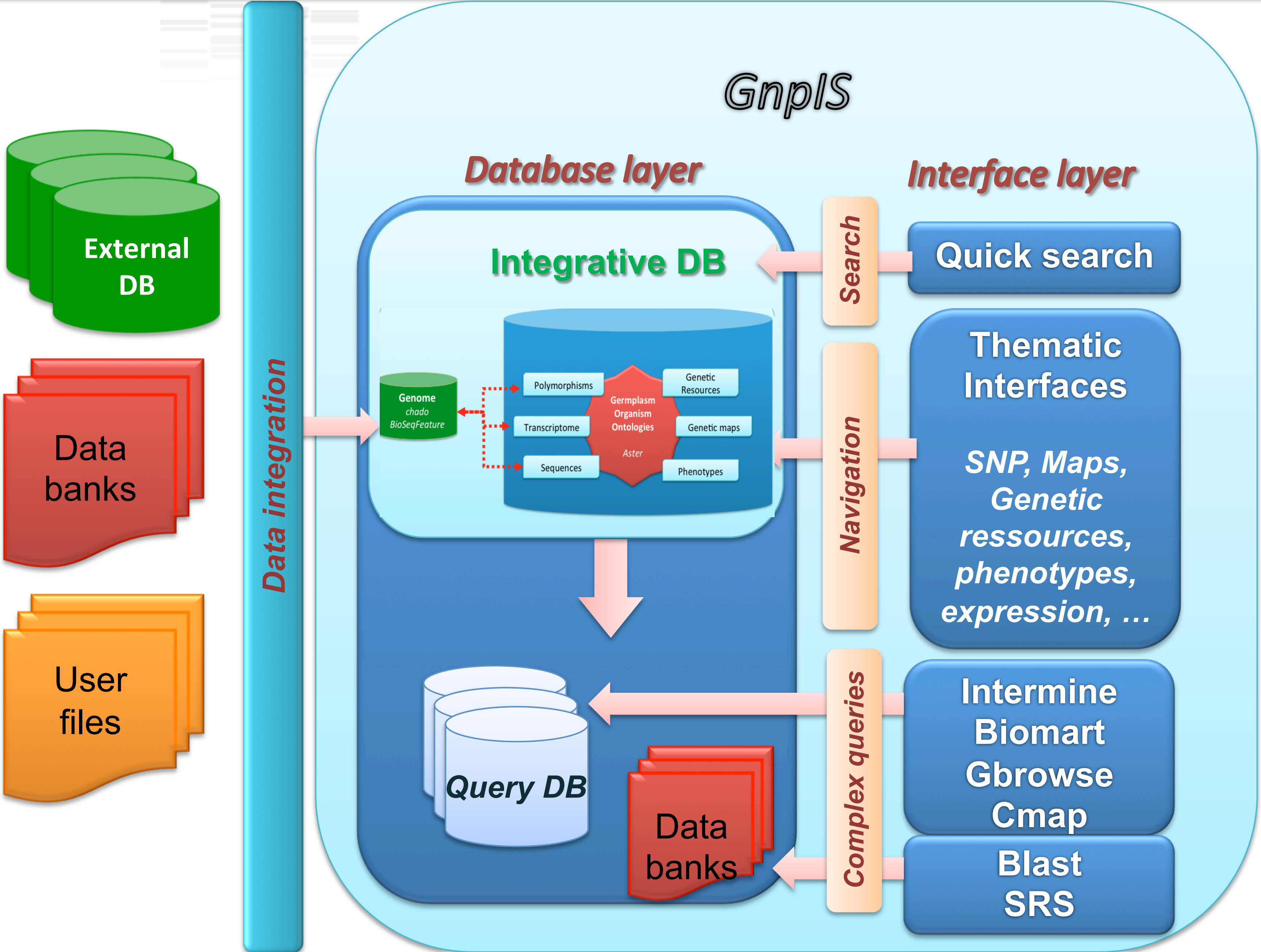
SNP, Maps, Genetic ressources, phenotypes, expression, …

GnpIS architecture



]

GnpIS architecture



External DB

Data banks

User files

Data integration

GnpIS

Database layer

Interface layer

Integrative DB

Genome
chado
BioSeqFeature

Polymorphisms

Transcriptome

Sequences

Germplasm
Organism
Ontologies
Aster

Genetic Resources

Genetic maps

Phenotypes

Query DB

Data banks

Search

Navigation

Complex queries

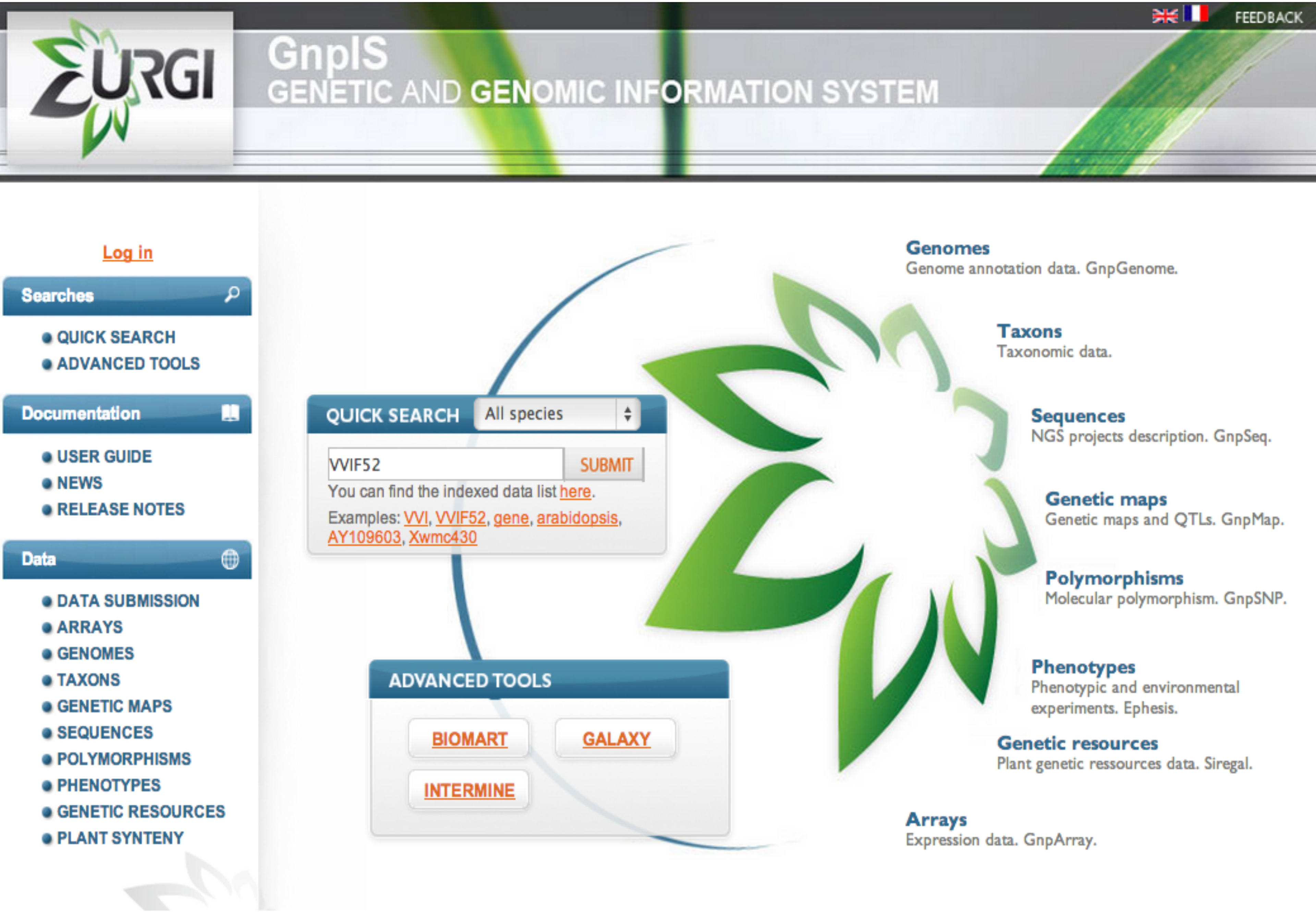
Quick search

Thematic Interfaces

SNP, Maps, Genetic ressources, phenotypes, expression, …

Intermine
Biomart
Gbrowse
Cmap

Blast
SRS



GnpIS
GENETIC AND GENOMIC INFORMATION SYSTEM

You can find the indexed data list here.

Examples: VVI, VVIF52, gene, arabidopsis, AY109603, Xwmc430

Genomes
Genome annotation data. GnpGenome.

Taxons
Taxonomic data.

Sequences
NGS projects description. GnpSeq.

Genetic maps
Genetic maps and QTLs. GnpMap.

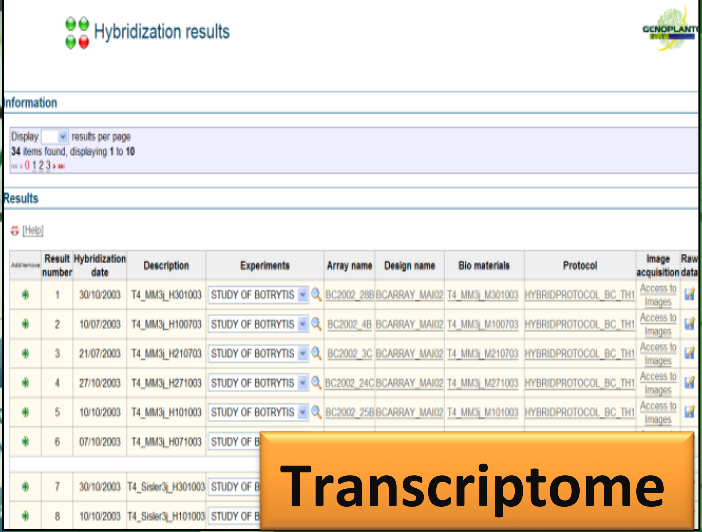
Polymorphisms
Molecular polymorphism. GnpSNP.

Phenotypes
Phenotypic and environmental experiments. Ephesis.

Genetic resources
Plant genetic ressources data. Siregal.

Arrays
Expression data. GnpArray.

https://urgi.versailles.inra.fr/gnpis



Hybridization results

Information

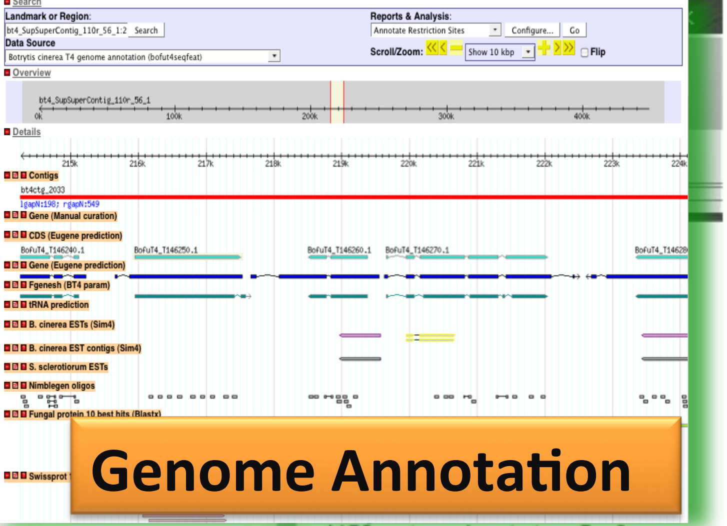
Display results per page
34 items found, displaying **1** to **10**
0 1 2 3

Results

[Help]

Add/remove	Result number	Hybridization date	Description	Experiments	Array name	Design name	Bio materials	Protocol	Image acquisition	Raw data
	1	30/10/2003	T4_MM3j_H301003	STUDY OF BOTRYTIS	BC2002_28B	BCARRAY_MAI02	T4_MM3j_M301003	HYBRIDPROTOCOL_BC_TH1	Access to Images	
	2	10/07/2003	T4_MM3j_H100703	STUDY OF BOTRYTIS	BC2002_4B	BCARRAY_MAI02	T4_MM3j_M100703	HYBRIDPROTOCOL_BC_TH1	Access to Images	
	3	21/07/2003	T4_MM3j_H210703	STUDY OF BOTRYTIS	BC2002_3C	BCARRAY_MAI02	T4_MM3j_M210703	HYBRIDPROTOCOL_BC_TH1	Access to Images	
	4	27/10/2003	T4_MM3j_H271003	STUDY OF BOTRYTIS	BC2002_24C	BCARRAY_MAI02	T4_MM3j_M271003	HYBRIDPROTOCOL_BC_TH1	Access to Images	
	5	10/10/2003	T4_MM3j_H101003	STUDY OF BOTRYTIS	BC2002_25B	BCARRAY_MAI02	T4_MM3j_M101003	HYBRIDPROTOCOL_BC_TH1	Access to Images	
	6	07/10/2003	T4_MM3j_H071003	STUDY OF B						
	7	30/10/2003	T4_Sisler3j_H301003	STUDY OF B						
	8	10/10/2003	T4_Sisler3j_H101003	STUDY OF B						

Transcriptome



Search

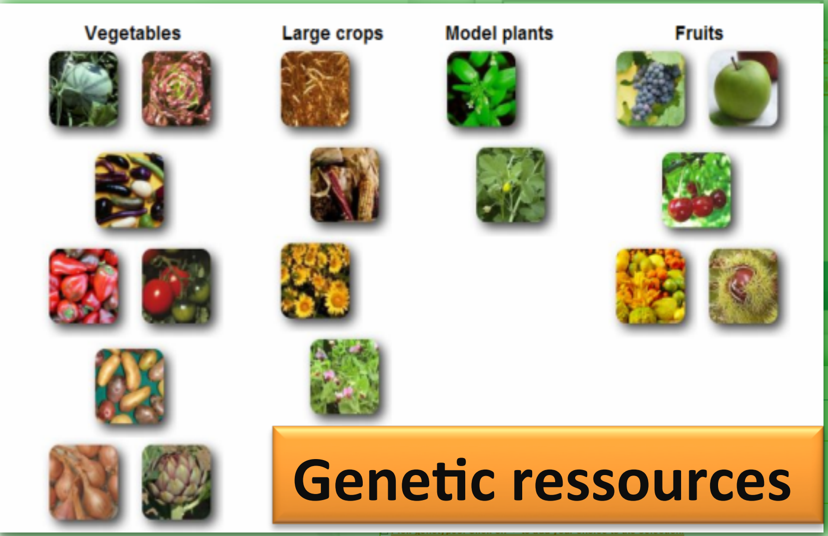
Landmark or Region: bt4_SupSuperContig_110r_56_1:2
Data Source: Botrytis cinerea T4 genome annotation (bofut4seqfeat)
Reports & Analysis: Annotate Restriction Sites
Scroll/Zoom: Show 10 kbp
☐ Flip

Overview

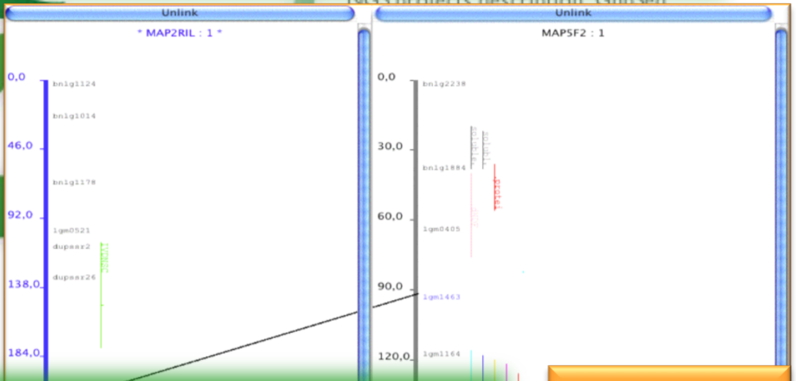
Details

Contigs
Gene (Manual curation)
CDS (Eugene prediction)
Gene (Eugene prediction)
Fgenesh (BT4 param)
tRNA prediction
B. cinerea ESTs (Sim4)
B. cinerea EST contigs (Sim4)
S. sclerotiorum ESTs
Nimblegen oligos
Fungal protein 10 best hits (Blastx)
Swissprot

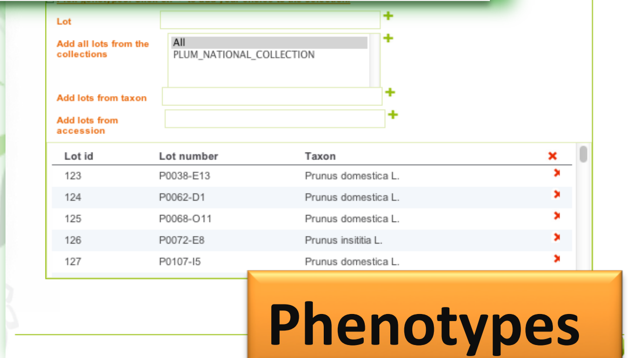
Genome Annotation



Genetic ressources



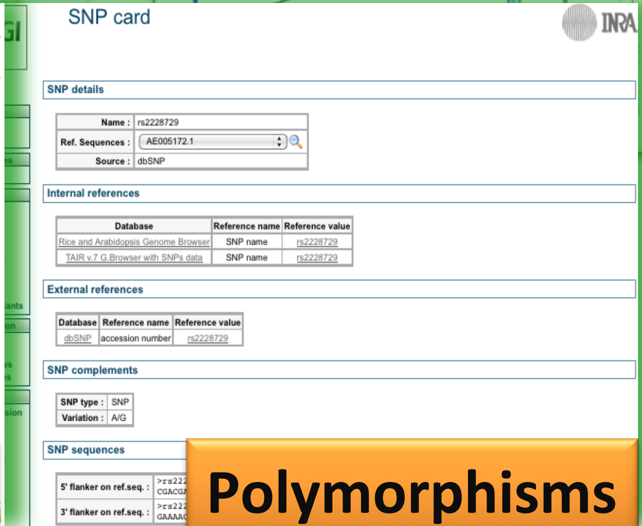
Genetic Maps



Lot	
Add all lots from the collections	All PLUM_NATIONAL_COLLECTION
Add lots from taxon	
Add lots from accession	

Lot id	Lot number	Taxon
123	P0038-E13	Prunus domestica L.
124	P0062-D1	Prunus domestica L.
125	P0068-O11	Prunus domestica L.
126	P0072-E8	Prunus insititia L.
127	P0107-I5	Prunus domestica L.

Phenotypes



SNP card

SNP details

Name :	rs2228729
Ref. Sequences :	AE005172.1
Source :	dbSNP

Internal references

Database	Reference name	Reference value
Rice and Arabidopsis Genome Browser	SNP name	rs2228729
TAIR v.7 G.Browser with SNPs data	SNP name	rs2228729

External references

Database	Reference name	Reference value
dbSNP	accession number	rs2228729

SNP complements

SNP type :	SNP
Variation :	A/G

SNP sequences

5' flanker on ref.seq. :	>rs222 CGACG
3' flanker on ref.seq. :	>rs222 GAAAA

Polymorphisms

GMOD@URGI

GBrowse

Apollo

CHADO

Bio::DB

GBrowse syn

GMOD

CMap

InterMine

biomart

Galaxy

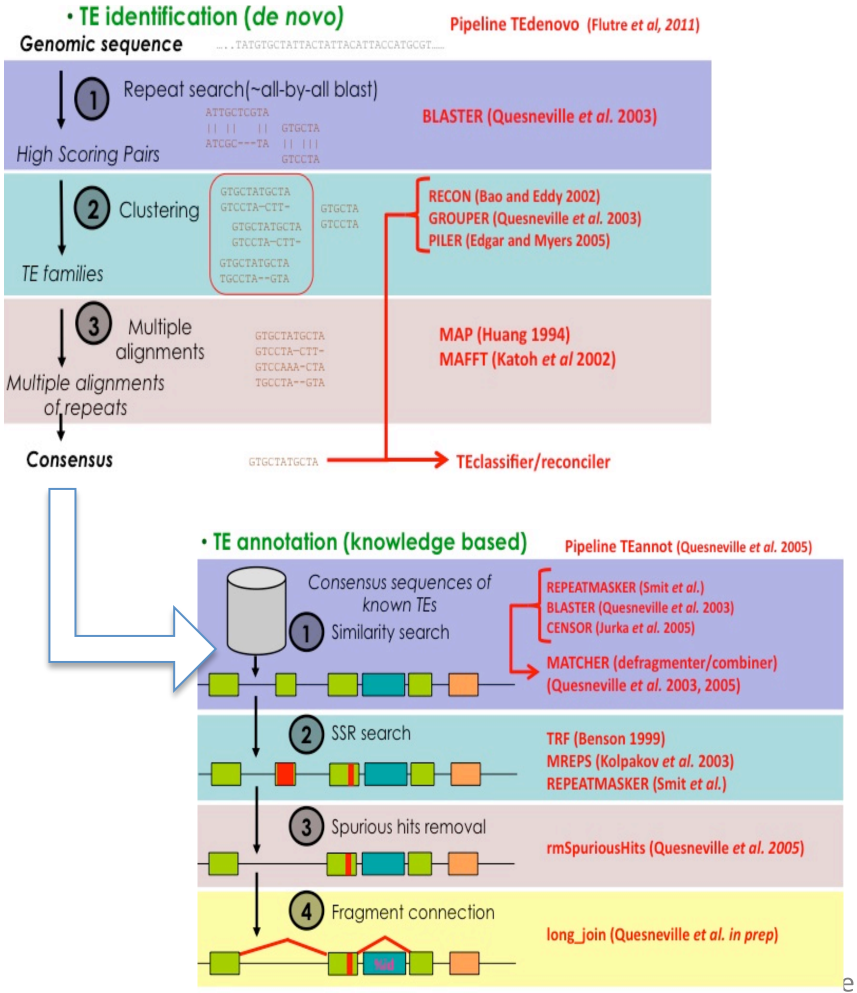
Genome annotation pipelines and data workflows

joelle.amselem@versailles.inra.fr

Genome annotation pipelines

Transposable Elements (TEs) Detection & Annotation Pipelines REPET package

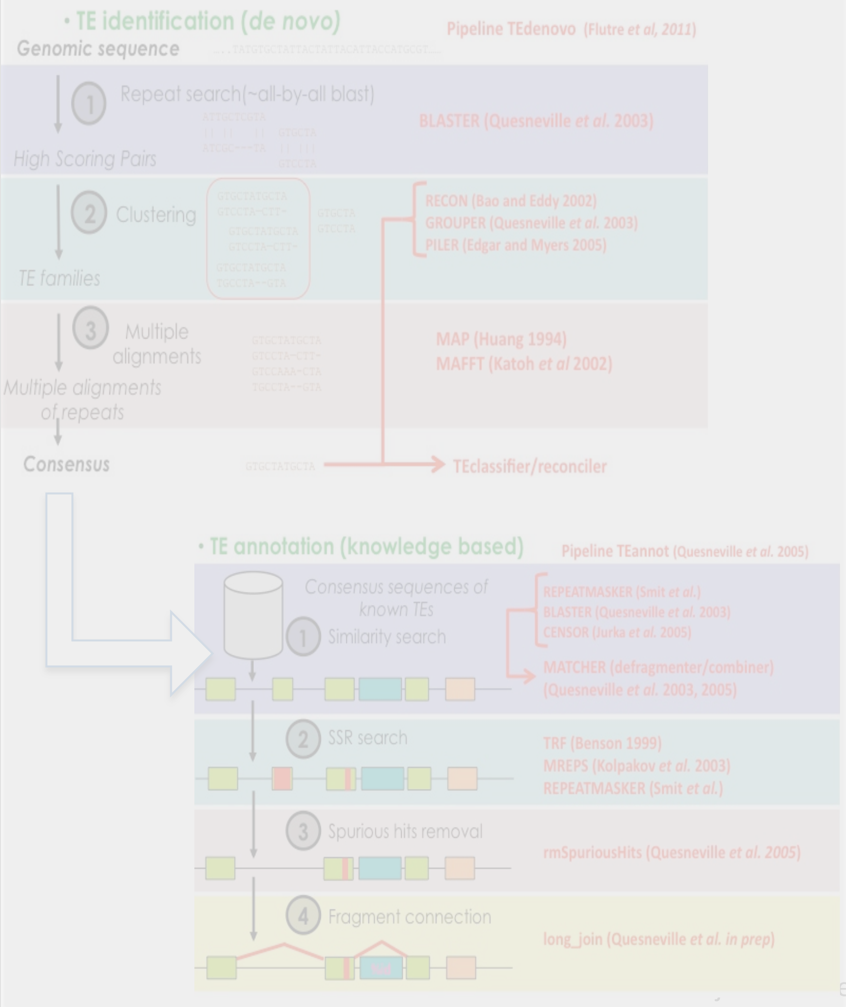
http://urgi.versailles.inra.fr/Tools/REPET



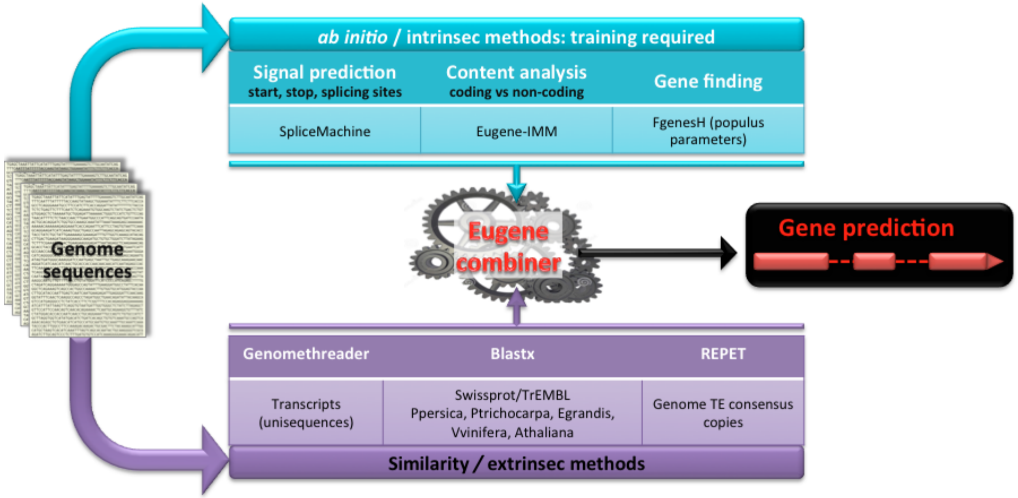
elem@versailles.inra.fr

Genome annotation pipelines

Transposable Elements (TEs)
Detection & Annotation Pipelines
REPET package
http://urgi.versailles.inra.fr/Tools/REPET



Gene prediction
Eugene (Species training @ URGI)
http://eugene.toulouse.inra.fr/

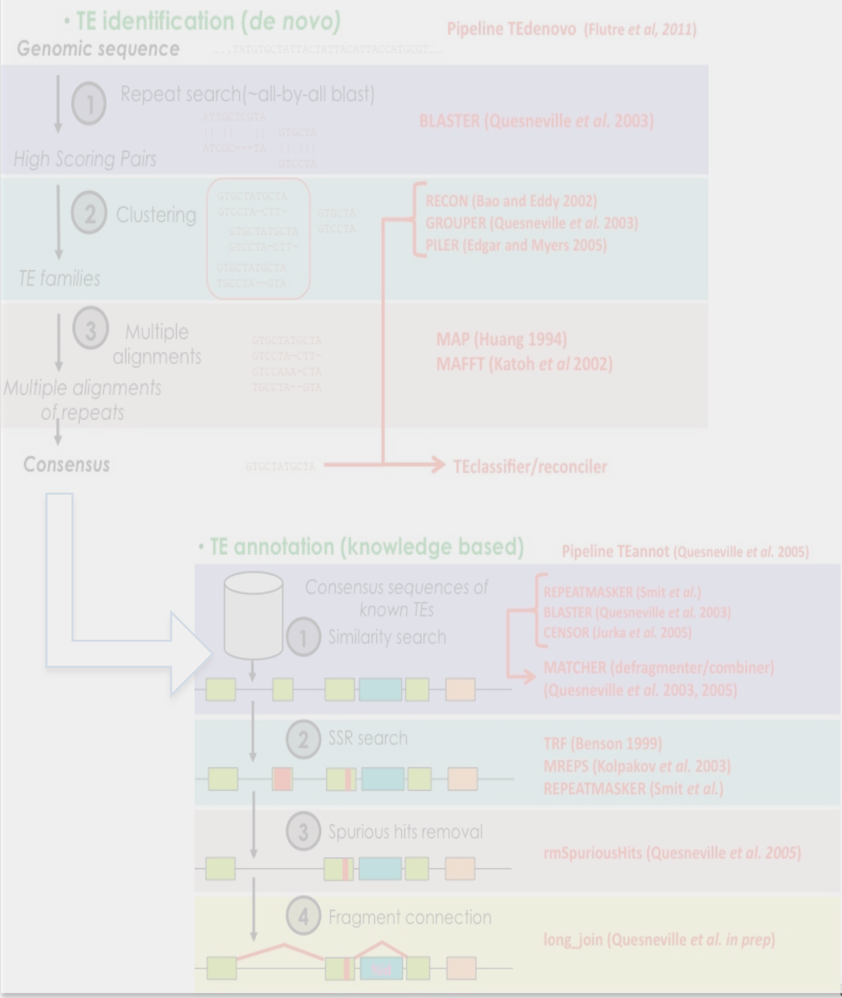


elem@versailles.inra.fr

Genome annotation pipelines

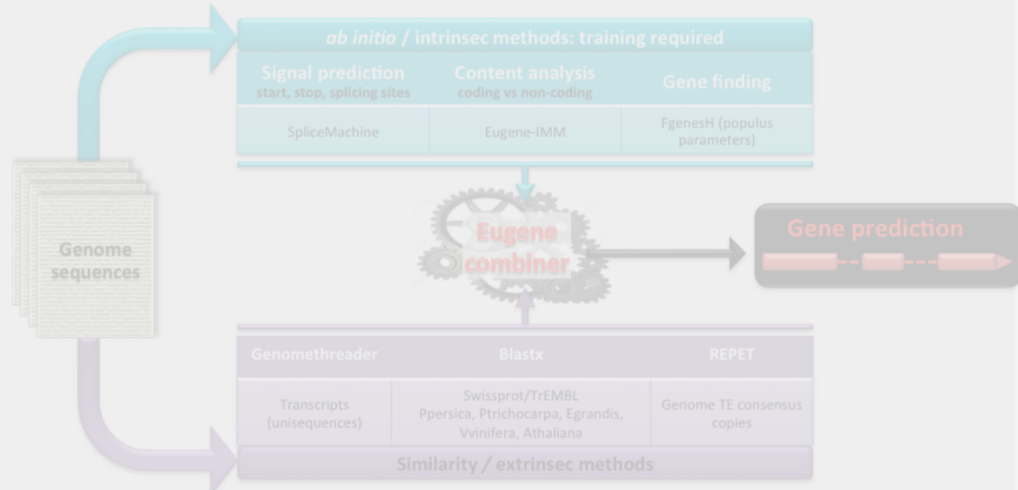
Transposable Elements (TEs) Detection & Annotation Pipelines REPET package

http://urgi.versailles.inra.fr/Tools/REPET

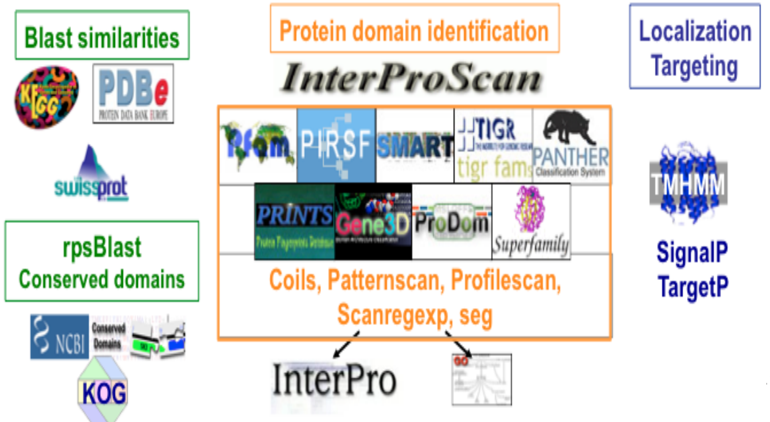


Gene prediction Eugene (Species training @ URGI)

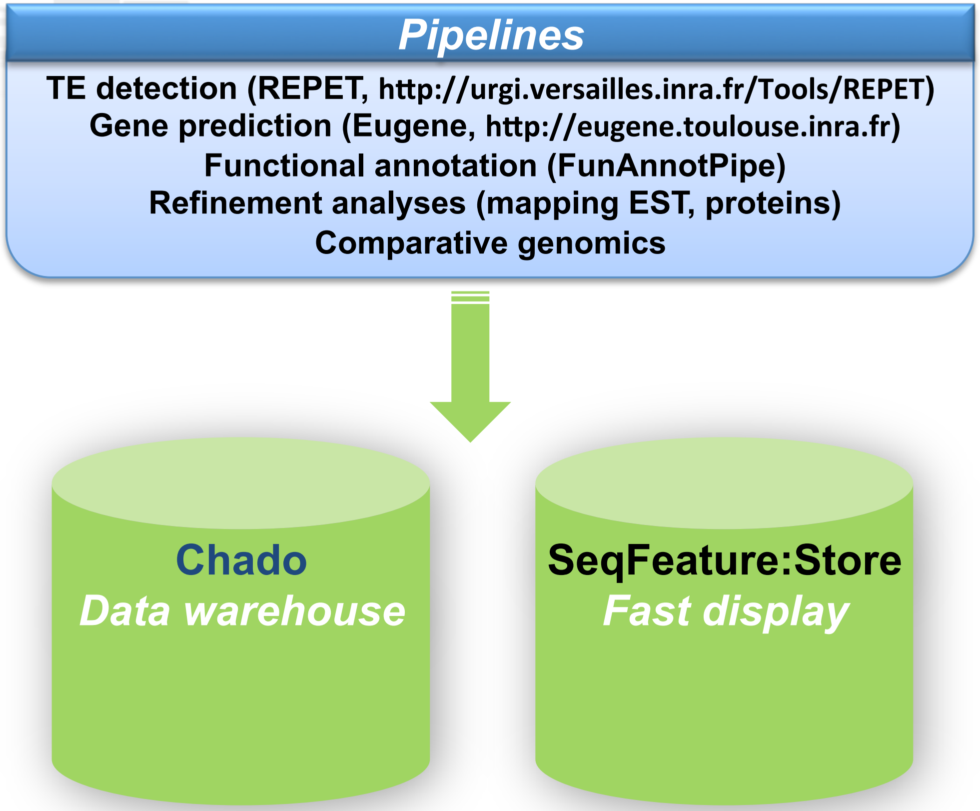
http://eugene.toulouse.inra.fr/



Gene functional annotation



Genome annotation: dataflow and interoperability



Pipelines

TE detection (REPET, http://urgi.versailles.inra.fr/Tools/REPET)
Gene prediction (Eugene, http://eugene.toulouse.inra.fr)
Functional annotation (FunAnnotPipe)
Refinement analyses (mapping EST, proteins)
Comparative genomics

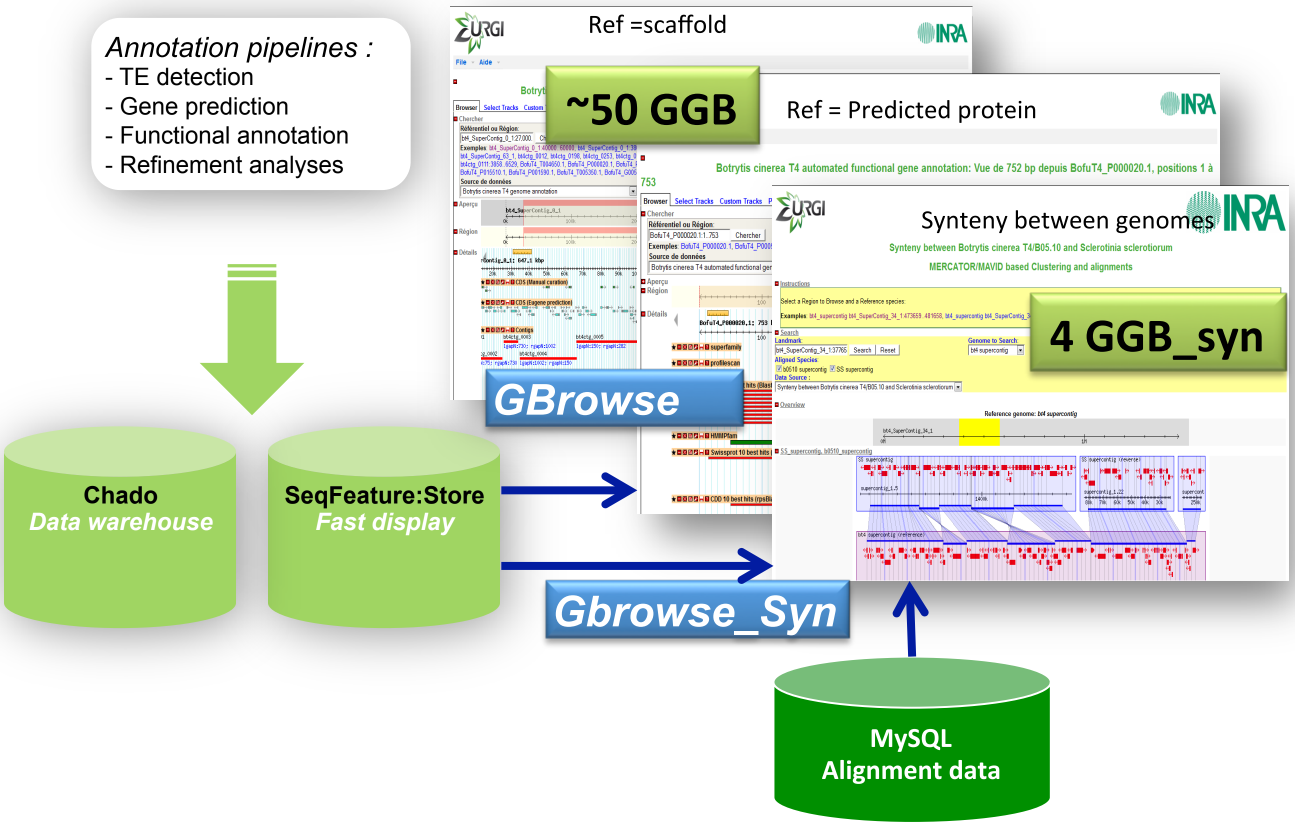
Chado
Data warehouse

SeqFeature:Store
Fast display

joelle.amselem@versailles.inra.fr

Genome annotation: dataflow and interoperability

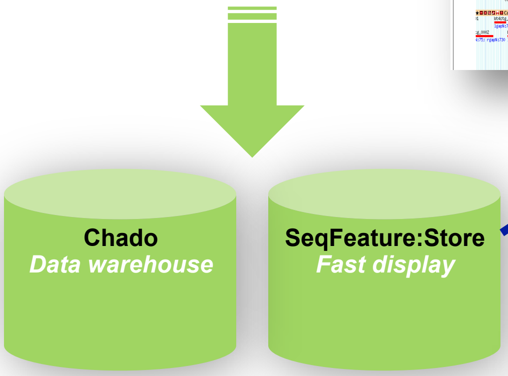
Annotation pipelines :
- TE detection
- Gene prediction
- Functional annotation
- Refinement analyses



joelle.amselem@versailles.inra.fr

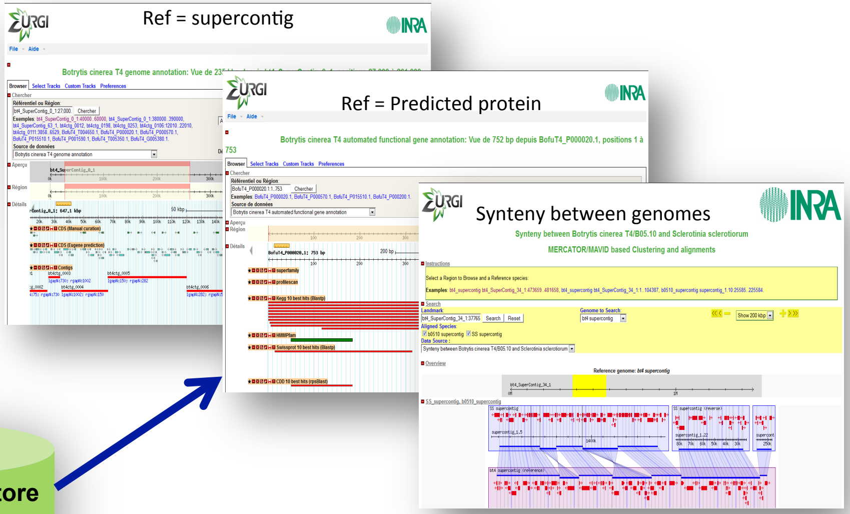
Genome annotation: dataflow and interoperability

Annotation pipelines :
- TE detection
- Gene prediction
- Functional annotation
- Refinement analyses



Chado
Data warehouse

SeqFeature:Store
Fast display



Ref = supercontig

Botrytis cinerea T4 genome annotation

Ref = Predicted protein

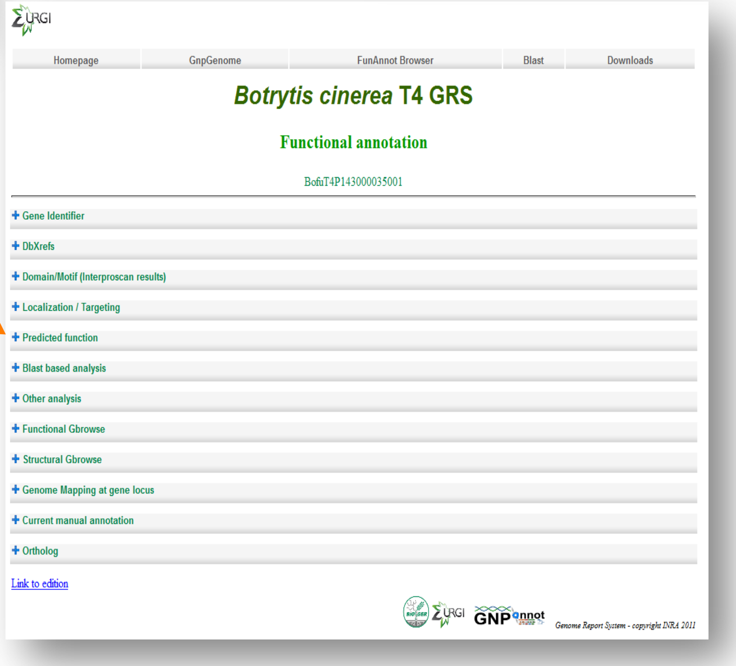
Botrytis cinerea T4 automated functional gene annotation: Vue de 752 bp depuis BofuT4_P000020.1, positions 1 à 753

Synteny between genomes

Synteny between Botrytis cinerea T4/B05.10 and Sclerotinia sclerotiorum

MERCATOR/MAVID based Clustering and alignments

GRS
Genome Report System



Botrytis cinerea T4 GRS

Functional annotation

BofuT4P143000035001

- + Gene Identifier
- + DbXrefs
- + Domain/Motif (Interproscan results)
- + Localization / Targeting
- + Predicted function
- + Blast based analysis
- + Other analysis
- + Functional Gbrowse
- + Structural Gbrowse
- + Genome Mapping at gene locus
- + Current manual annotation
- + Ortholog

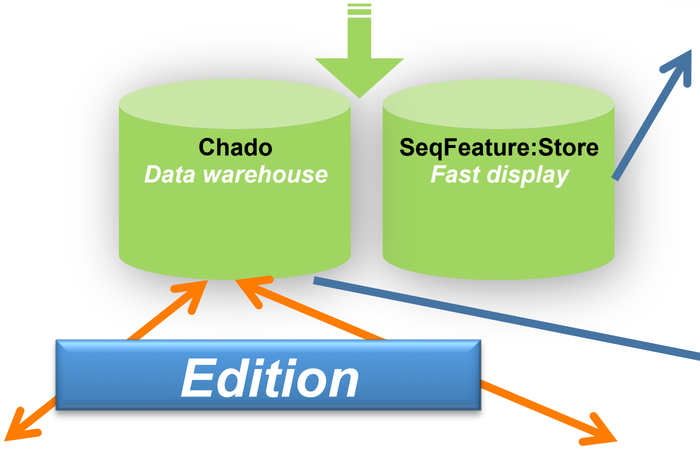
Link to edition

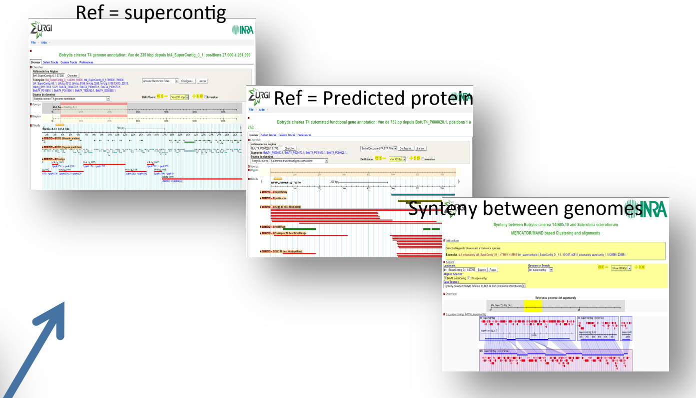
Genome Report System - copyright INRA 2011

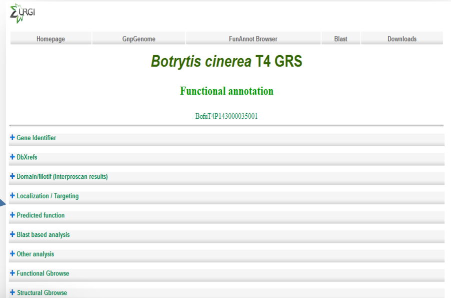
joelle.amselem@versailles.inra.fr

Genome annotation: dataflow and interoperability

Annotation pipelines :
- TE detection
- Gene prediction
- Functional annotation
- Refinement analyses







Botrytis cinerea T4 GRS

Functional annotation

BofuT4P143000035001

- + Gene Identifier
- + DbXrefs
- + Domain/Motif (Interproscan results)
- + Localization / Targeting
- + Predicted function
- + Blast based analysis
- + Other analysis
- + Functional Gbrowse
- + Structural Gbrowse
- Mapping at gene locus
- anual annotation

Genome Report System - copyright INRA 2011

Apollo : manual structural edition

GRE : manual functional edition

Zoom factor = 1.0000 Botrytis fuckeliana:bt4ctg_0006:1-26638

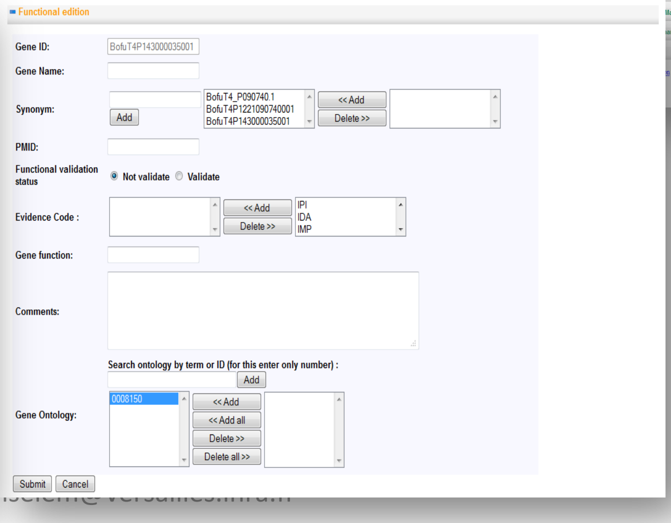
Type	Name	Range	Score
uniprot_trembl	blastx_uniprot_tre...	21835-19697	0.0
uniprot_trembl	blastx_uniprot_tre...	21820-19763	0.0
FNG30_prot	blastx_FNG30_pro...	21613-19760	0.0
FNG30_prot	blastx_FNG30_pro...	21766-19763	0.0
uniprot_trembl	blastx_uniprot_tre...	21613-19763	0.0
FNG30_prot	blastx_FNG30_pro...	21613-19763	0.0

Putative unchara... orum)

Genomic... Geno... Le...
21820-1... 2058

Position 19596 Feature Action

4 Apollo



Functional edition

Gene ID: BofuT4P143000035001

Gene Name:

Synonym: [Add] BofuT4_P090740.1 BofuT4P1221090740001 BofuT4P143000035001 [<< Add] [Delete >>]

PMID:

Functional validation status: ● Not validate ○ Validate

Evidence Code : [<< Add] [Delete >>] IPI IDA IMP

Gene function:

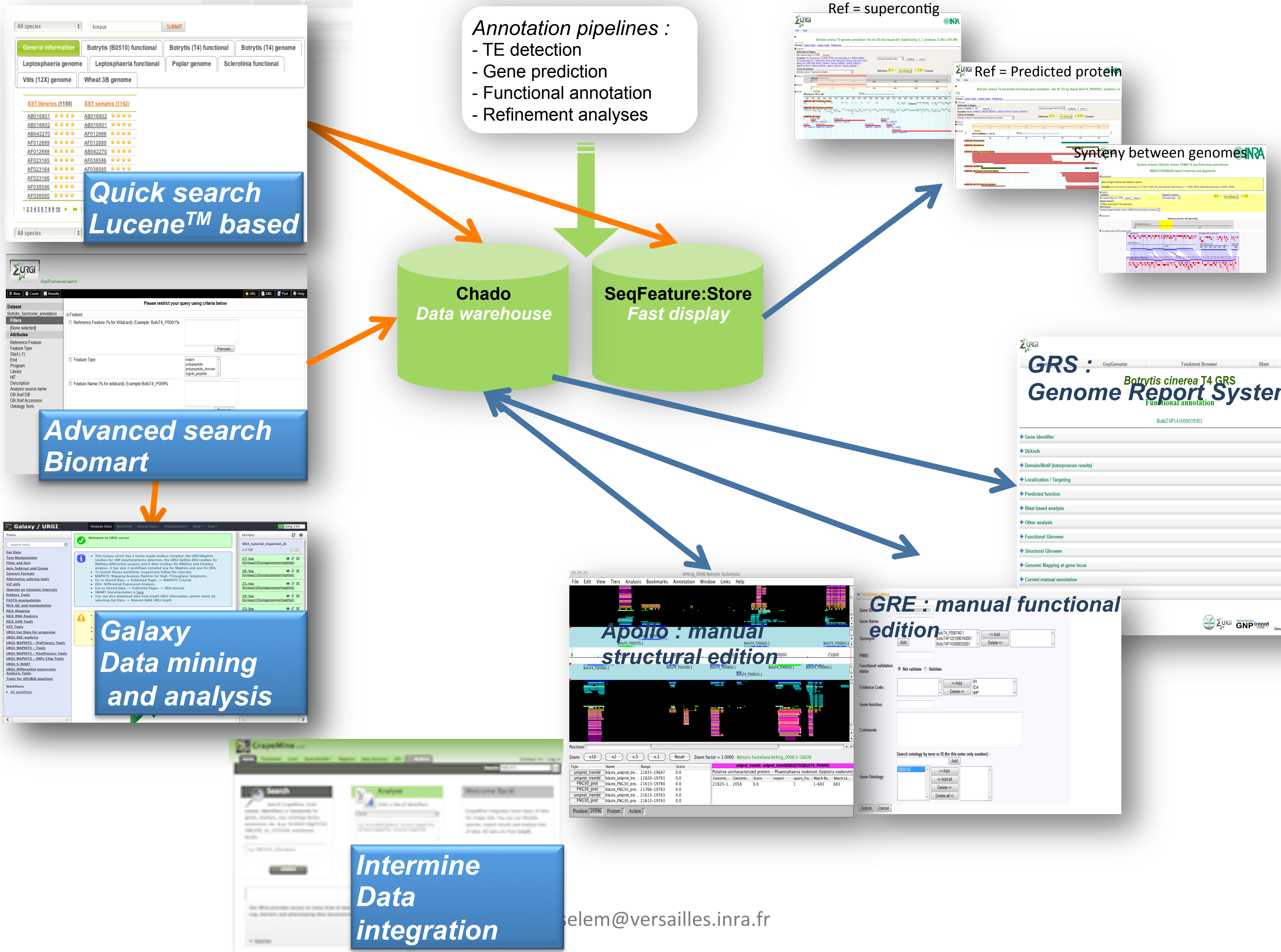
Comments:

Search ontology by term or ID (for this enter only number) : [Add]

Gene Ontology: 0008150 [<< Add] [<< Add all] [Delete >>] [Delete all >>]

[Submit] [Cancel]

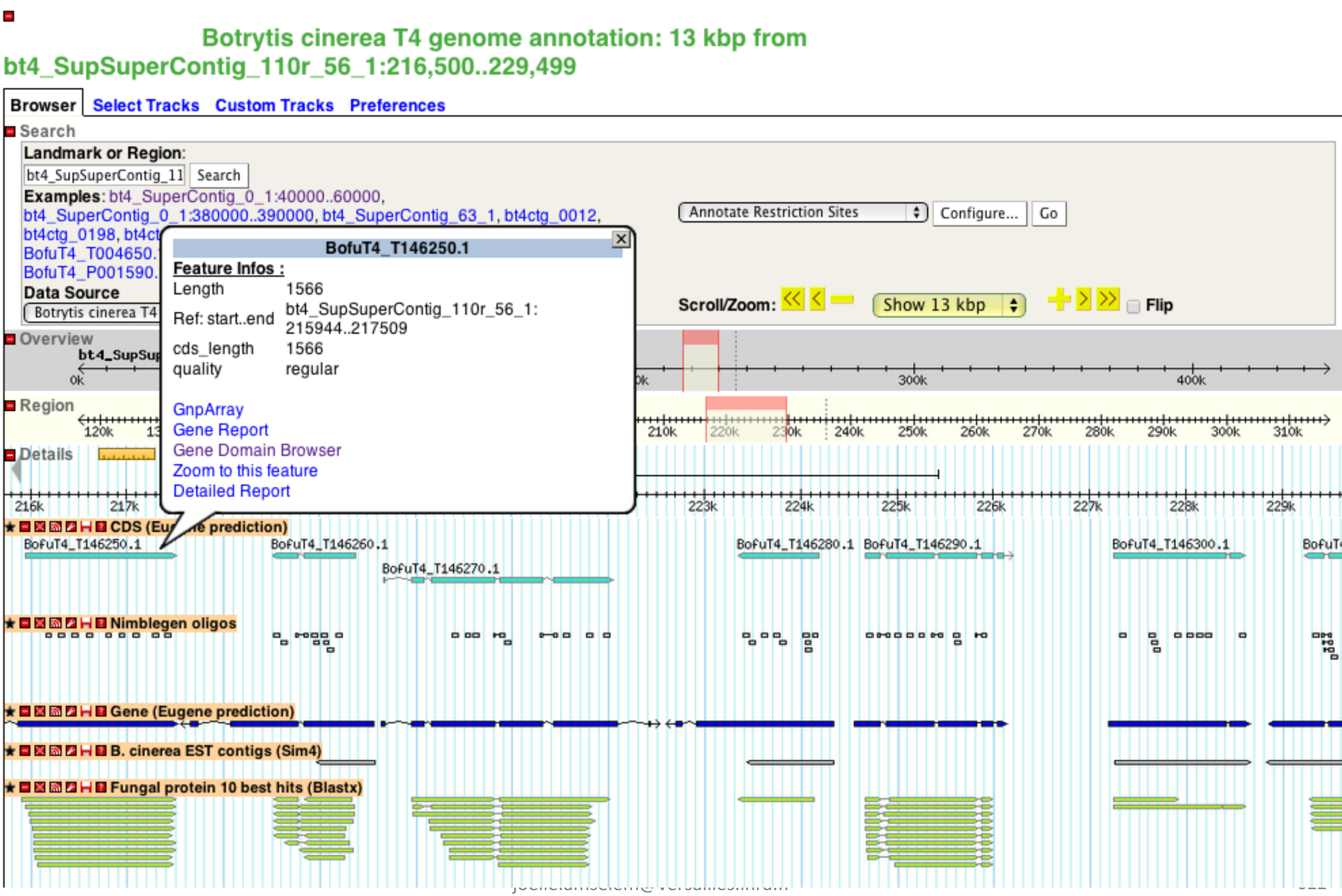
Genome annotation: dataflow and interoperability



Annotation pipelines :
- TE detection
- Gene prediction
- Functional annotation
- Refinement analyses

selem@versailles.inra.fr

From GBrowse scaffold annotation to...



Botrytis cinerea T4 genome annotation: 13 kbp from bt4_SupSuperContig_110r_56_1:216,500..229,499

Landmark or Region:

bt4_SupSuperContig_11 Search

Examples: bt4_SuperContig_0_1:40000..60000, bt4_SuperContig_0_1:380000..390000, bt4_SuperContig_63_1, bt4ctg_0012, bt4ctg_0198, bt4ct… BofuT4_T004650.… BofuT4_P001590.…

Data Source

Botrytis cinerea T4

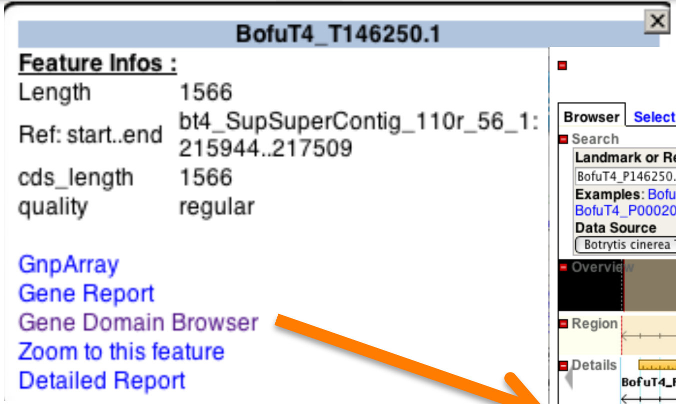
BofuT4_T146250.1

Feature Infos :

Length	1566
Ref: start..end	bt4_SupSuperContig_110r_56_1: 215944..217509
cds_length	1566
quality	regular

GnpArray
Gene Report
Gene Domain Browser
Zoom to this feature
Detailed Report

...to Gene Domain Browser

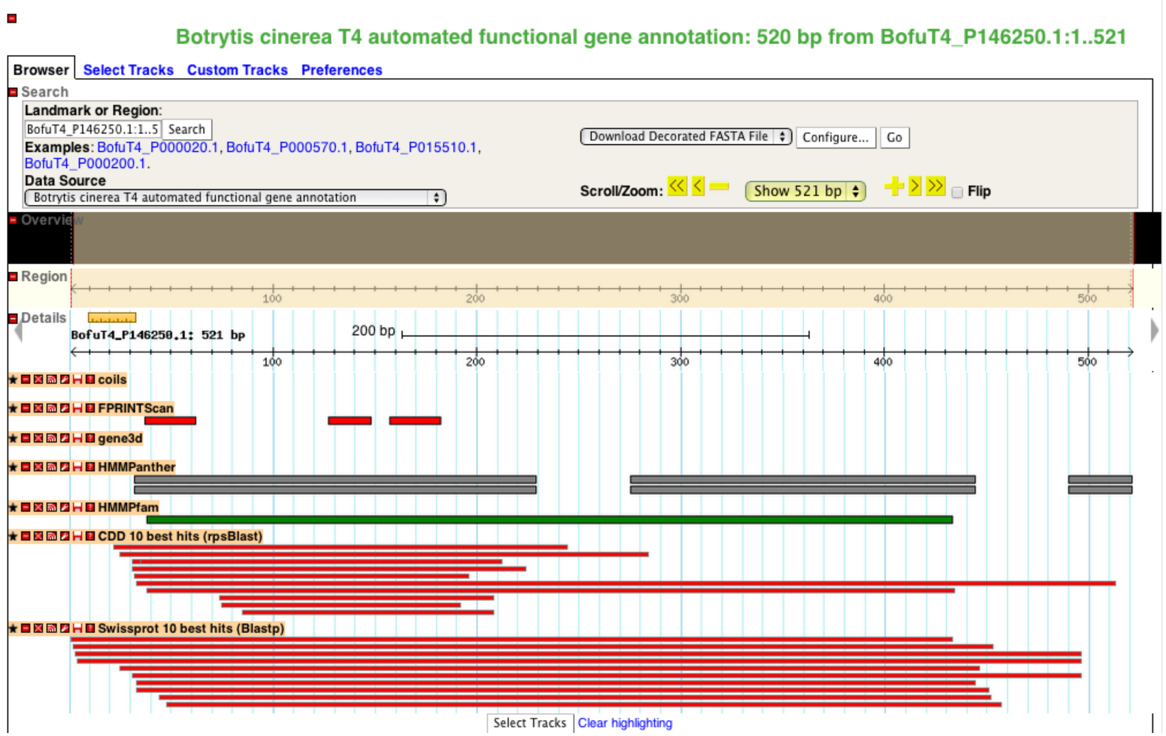


BofuT4_T146250.1

Feature Infos :

Length	1566
Ref: start..end	bt4_SupSuperContig_110r_56_1: 215944..217509
cds_length	1566
quality	regular

GnpArray
Gene Report
Gene Domain Browser
Zoom to this feature
Detailed Report



Botrytis cinerea T4 automated functional gene annotation: 520 bp from BofuT4_P146250.1:1..521

Landmark or Region: BofuT4_P146250.1:1..5

Examples: BofuT4_P000020.1, BofuT4_P000570.1, BofuT4_P015510.1, BofuT4_P000200.1.

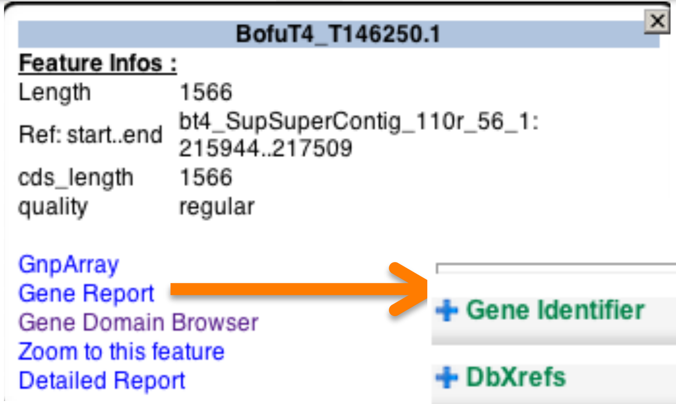
Data Source: Botrytis cinerea T4 automated functional gene annotation

Scroll/Zoom: Show 521 bp ☐ Flip

Gbrowse: functional domains, Motifs

joelle.amselem@versailles.inra.fr

...to gene report

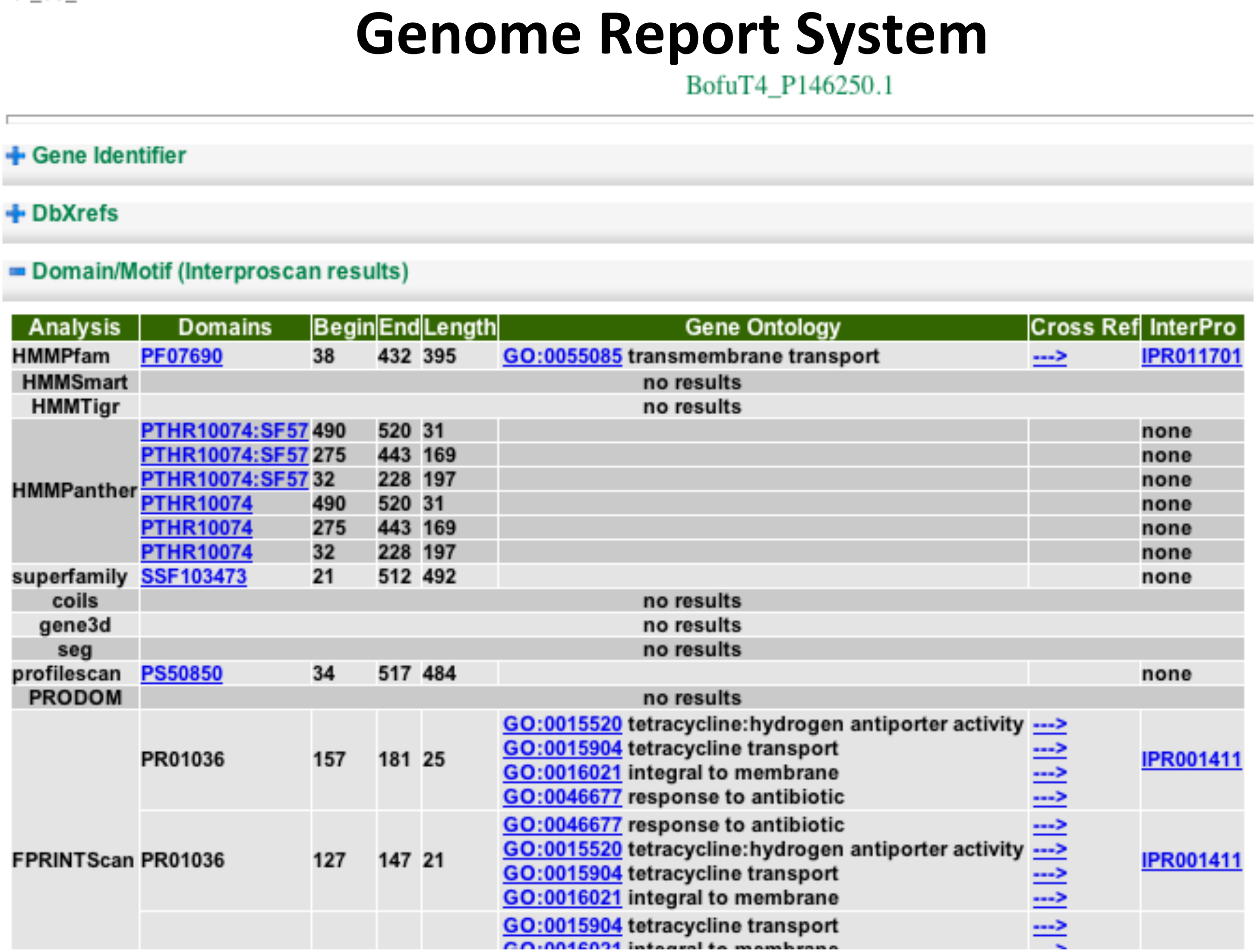


BofuT4_T146250.1

Feature Infos :

Length	1566
Ref: start..end	bt4_SupSuperContig_110r_56_1: 215944..217509
cds_length	1566
quality	regular

GnpArray
Gene Report
Gene Domain Browser
Zoom to this feature
Detailed Report



Genome Report System

BofuT4_P146250.1

+ Gene Identifier

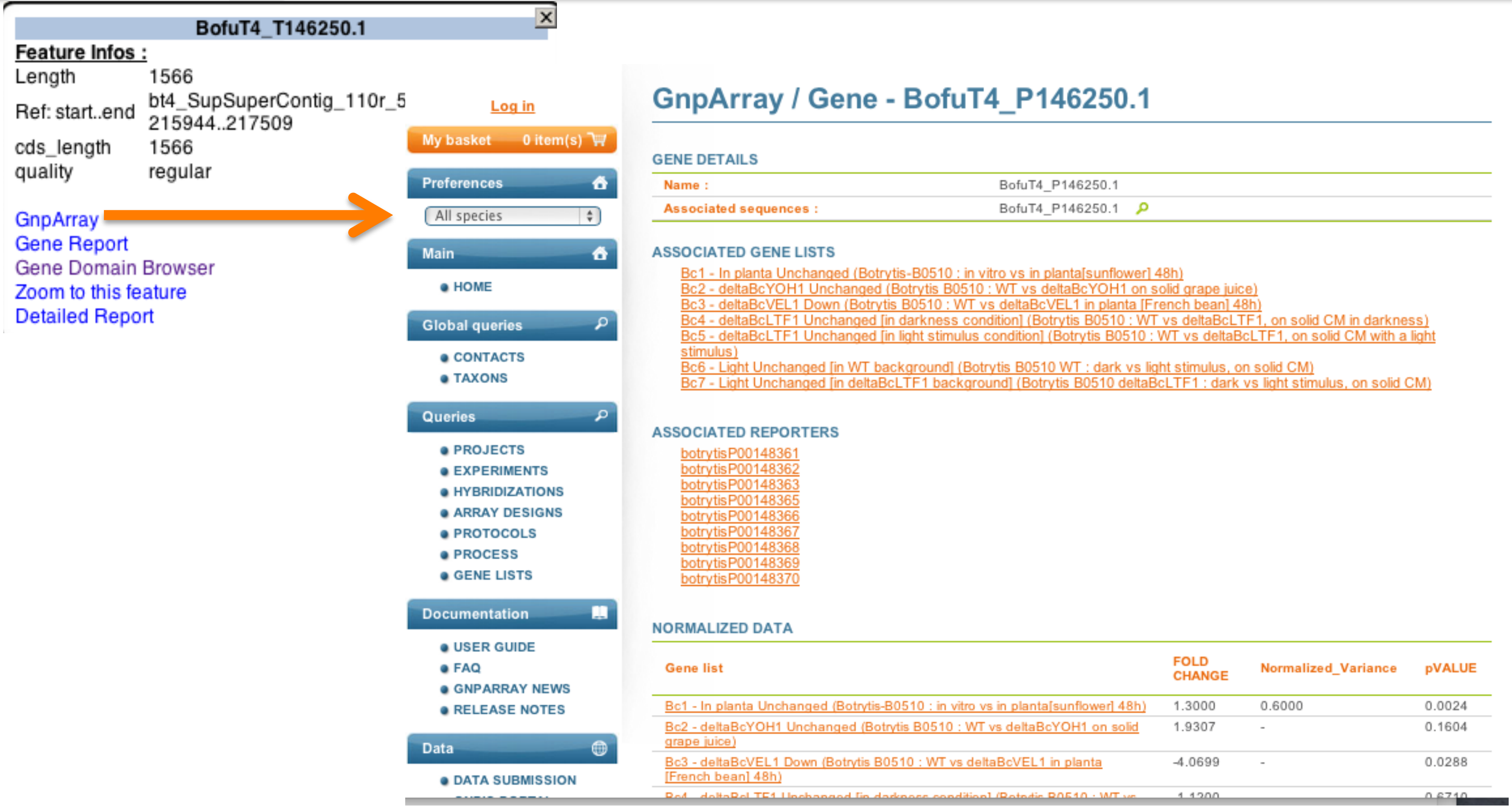
+ DbXrefs

- Domain/Motif (Interproscan results)

Analysis	Domains	Begin	End	Length	Gene Ontology	Cross Ref	InterPro
HMMPfam	PF07690	38	432	395	GO:0055085 transmembrane transport	--->	IPR011701
HMMSmart	no results						
HMMTigr	no results						
HMMPanther	PTHR10074:SF57	490	520	31			none
	PTHR10074:SF57	275	443	169			none
	PTHR10074:SF57	32	228	197			none
	PTHR10074	490	520	31			none
	PTHR10074	275	443	169			none
	PTHR10074	32	228	197			none
superfamily	SSF103473	21	512	492			none
coils	no results						
gene3d	no results						
seg	no results						
profilescan	PS50850	34	517	484			none
PRODOM	no results						
FPRINTScan	PR01036	157	181	25	GO:0015520 tetracycline:hydrogen antiporter activity GO:0015904 tetracycline transport GO:0016021 integral to membrane GO:0046677 response to antibiotic	---> ---> ---> --->	IPR001411
	PR01036	127	147	21	GO:0046677 response to antibiotic GO:0015520 tetracycline:hydrogen antiporter activity GO:0015904 tetracycline transport GO:0016021 integral to membrane	---> ---> ---> --->	IPR001411
					GO:0015904 tetracycline transport	--->	

joelle.amselem@versailles.inra.fr

...to transcriptomic data



BofuT4_T146250.1

Feature Infos :

Length	1566
Ref: start..end	bt4_SupSuperContig_110r_5 215944..217509
cds_length	1566
quality	regular

GnpArray
Gene Report
Gene Domain Browser
Zoom to this feature
Detailed Report

GnpArray / Gene - BofuT4_P146250.1

GENE DETAILS

Name :	BofuT4_P146250.1
Associated sequences :	BofuT4_P146250.1

ASSOCIATED GENE LISTS

Bc1 - In planta Unchanged (Botrytis-B0510 : in vitro vs in planta[sunflower] 48h)
Bc2 - deltaBcYOH1 Unchanged (Botrytis B0510 : WT vs deltaBcYOH1 on solid grape juice)
Bc3 - deltaBcVEL1 Down (Botrytis B0510 : WT vs deltaBcVEL1 in planta [French bean] 48h)
Bc4 - deltaBcLTF1 Unchanged [in darkness condition] (Botrytis B0510 : WT vs deltaBcLTF1, on solid CM in darkness)
Bc5 - deltaBcLTF1 Unchanged [in light stimulus condition] (Botrytis B0510 : WT vs deltaBcLTF1, on solid CM with a light stimulus)
Bc6 - Light Unchanged [in WT background] (Botrytis B0510 WT : dark vs light stimulus, on solid CM)
Bc7 - Light Unchanged [in deltaBcLTF1 background] (Botrytis B0510 deltaBcLTF1 : dark vs light stimulus, on solid CM)

ASSOCIATED REPORTERS

botrytisP00148361
botrytisP00148362
botrytisP00148363
botrytisP00148365
botrytisP00148366
botrytisP00148367
botrytisP00148368
botrytisP00148369
botrytisP00148370

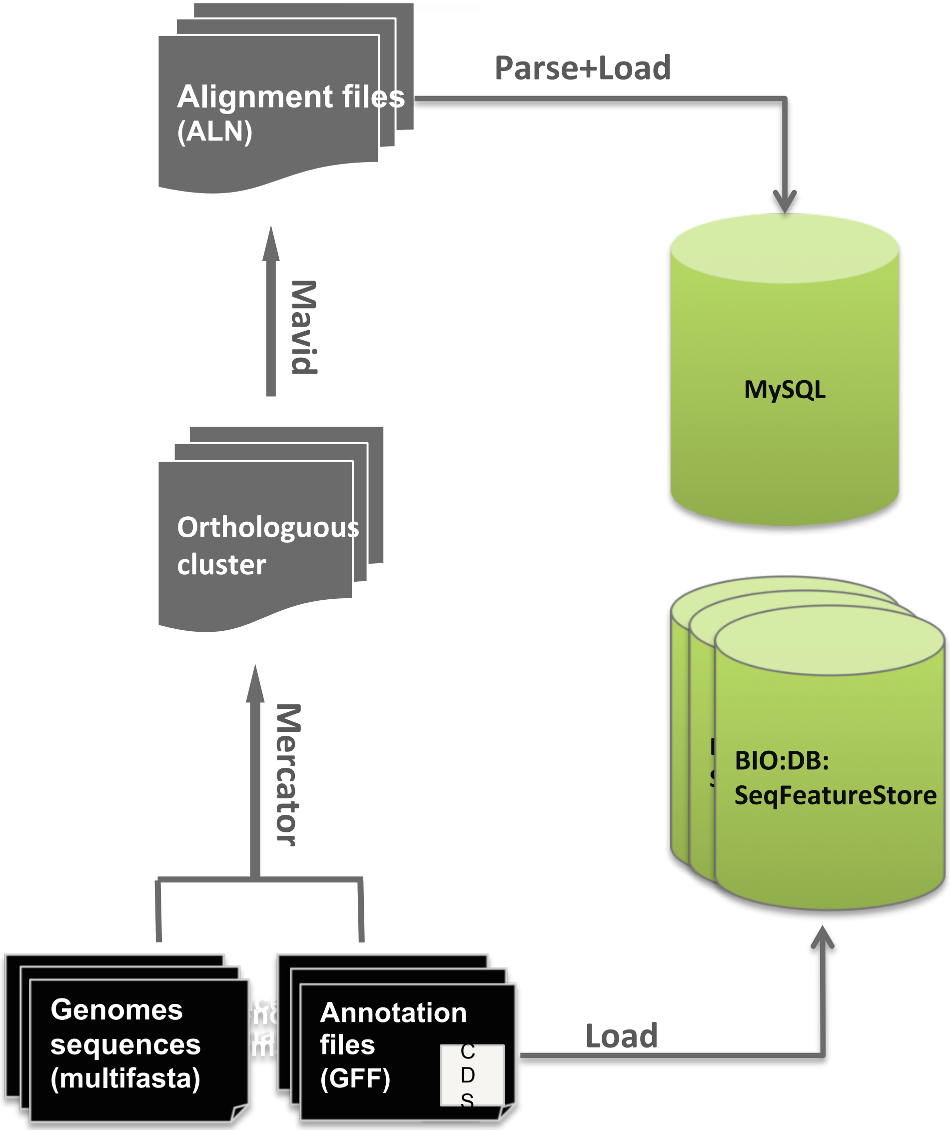
NORMALIZED DATA

Gene list	FOLD CHANGE	Normalized_Variance	pVALUE
Bc1 - In planta Unchanged (Botrytis-B0510 : in vitro vs in planta[sunflower] 48h)	1.3000	0.6000	0.0024
Bc2 - deltaBcYOH1 Unchanged (Botrytis B0510 : WT vs deltaBcYOH1 on solid grape juice)	1.9307	-	0.1604
Bc3 - deltaBcVEL1 Down (Botrytis B0510 : WT vs deltaBcVEL1 in planta [French bean] 48h)	-4.0699	-	0.0288
Bc4 - deltaBcLTF1 Unchanged [in darkness condition] (Botrytis B0510 : WT vs	1.1200	-	0.6710

GnpArray:Gene card

joelle.amselem@versailles.inra.fr

Synteny: Mercator + Mavid based pipeline



Alignment files (ALN)

Parse+Load

MySQL

Mavid

Orthologuous cluster

Mercator

BIO:DB: SeqFeatureStore

Genomes sequences (multifasta)

Annotation files (GFF)

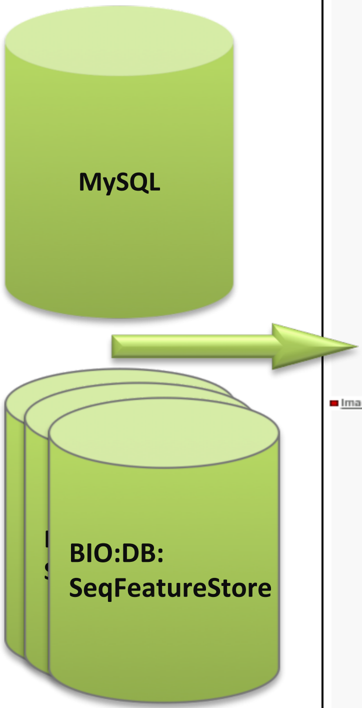
CDS

Load

joelle.amselem@versailles.inra.fr

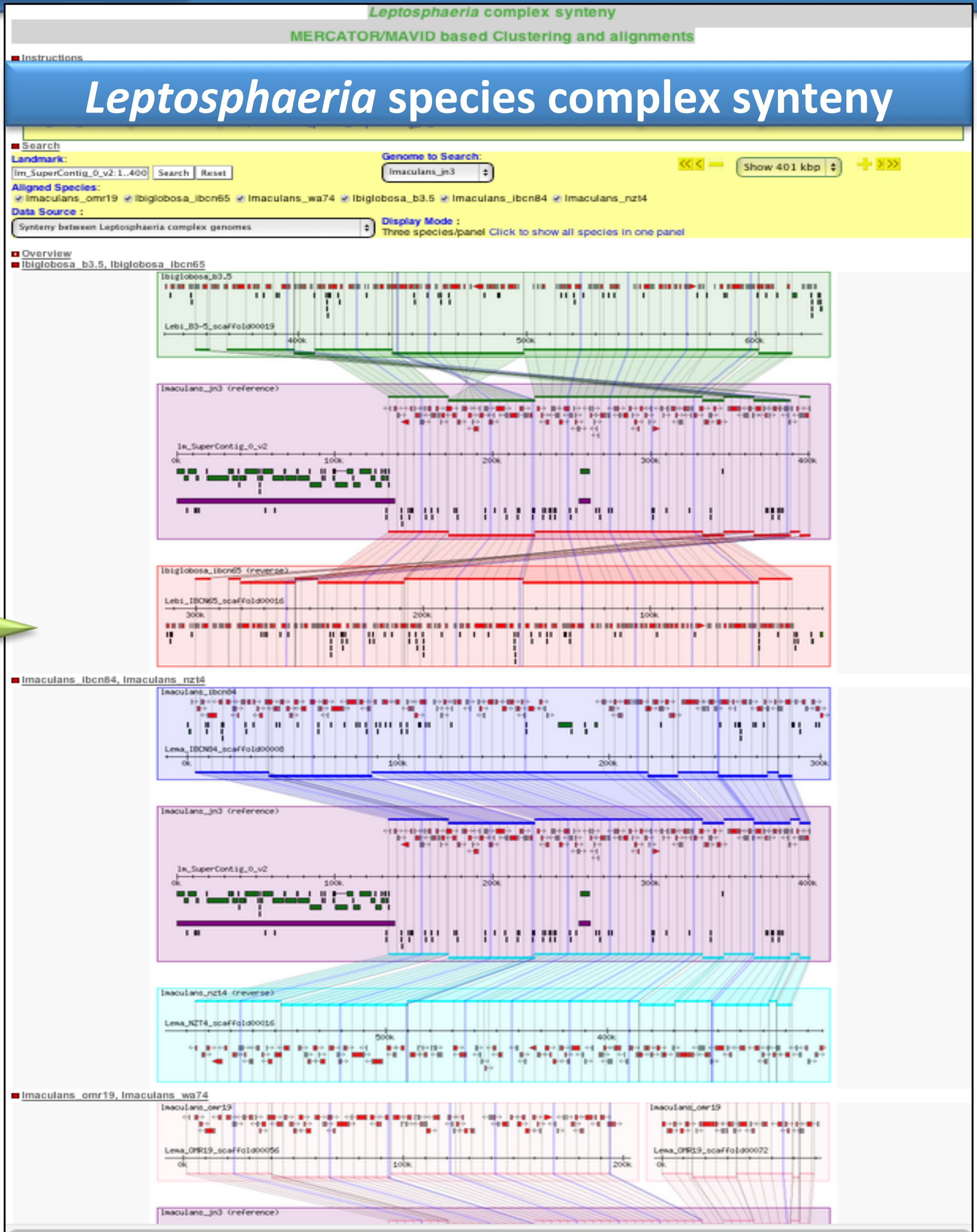
Synteny:

Leptosphaeria species complex synteny



MySQL

BIO:DB: SeqFeatureStore



Leptosphaeria complex synteny

MERCATOR/MAVID based Clustering and alignments

Instructions

Search

Landmark: lm_SuperContig_0_v2:1..400 Search Reset

Genome to Search: lmaculans_jn3

Show 401 kbp

Aligned Species: ☑ lmaculans_omr19 ☑ lbiglobosa_ibcn65 ☑ lmaculans_wa74 ☑ lbiglobosa_b3.5 ☑ lmaculans_ibcn84 ☑ lmaculans_nzt4

Data Source : Synteny between Leptosphaeria complex genomes

Display Mode : Three species/panel Click to show all species in one panel

Overview

lbiglobosa_b3.5, lbiglobosa_ibcn65

lmaculans_ibcn84, lmaculans_nzt4

lmaculans_omr19, lmaculans_wa74

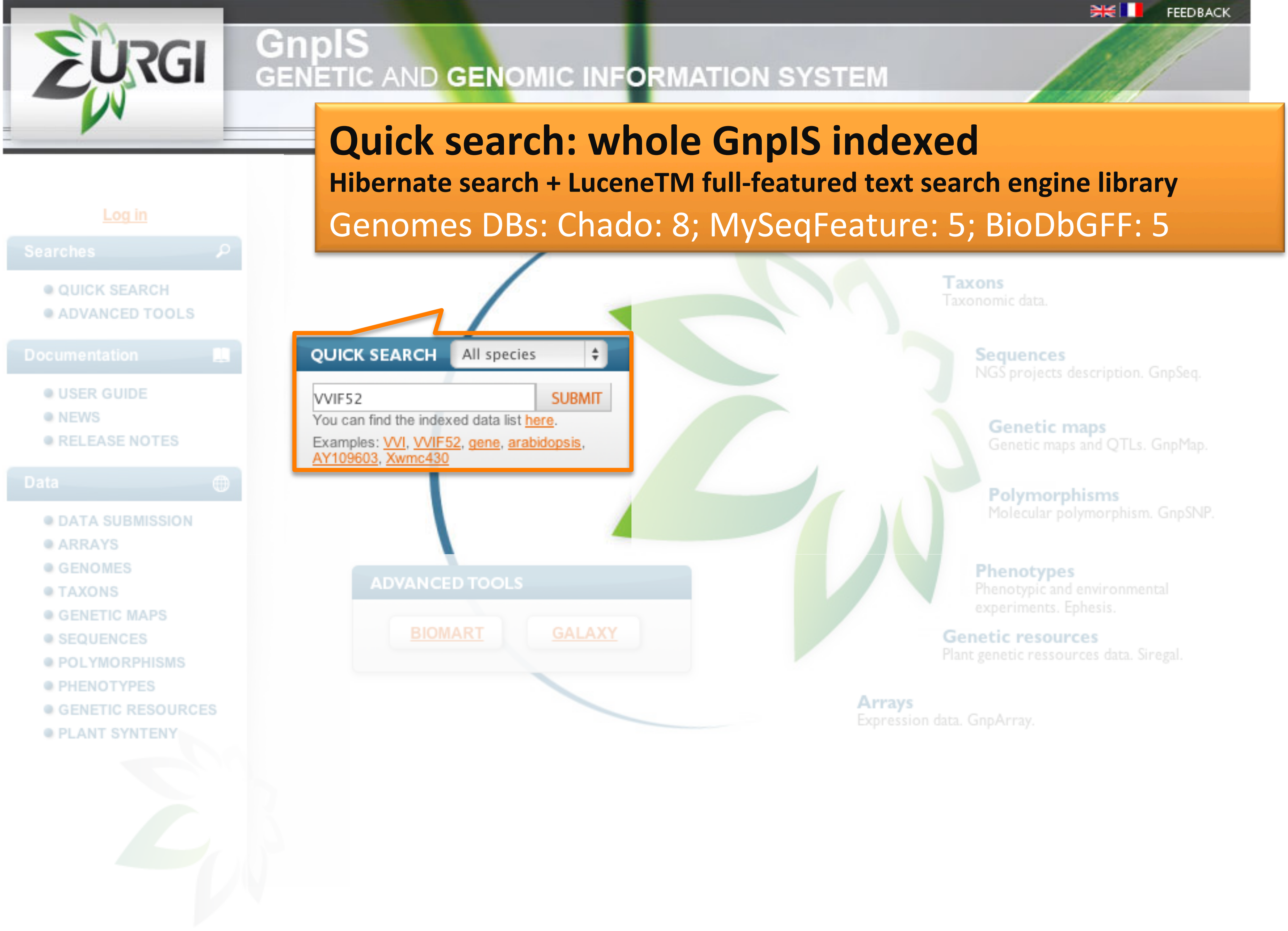
Data querying and browsing system

joelle.amselem@versailles.inra.fr

Quick search: whole GnpIS indexed

Hibernate search + LuceneTM full-featured text search engine library

Genomes DBs: Chado: 8; MySeqFeature: 5; BioDbGFF: 5



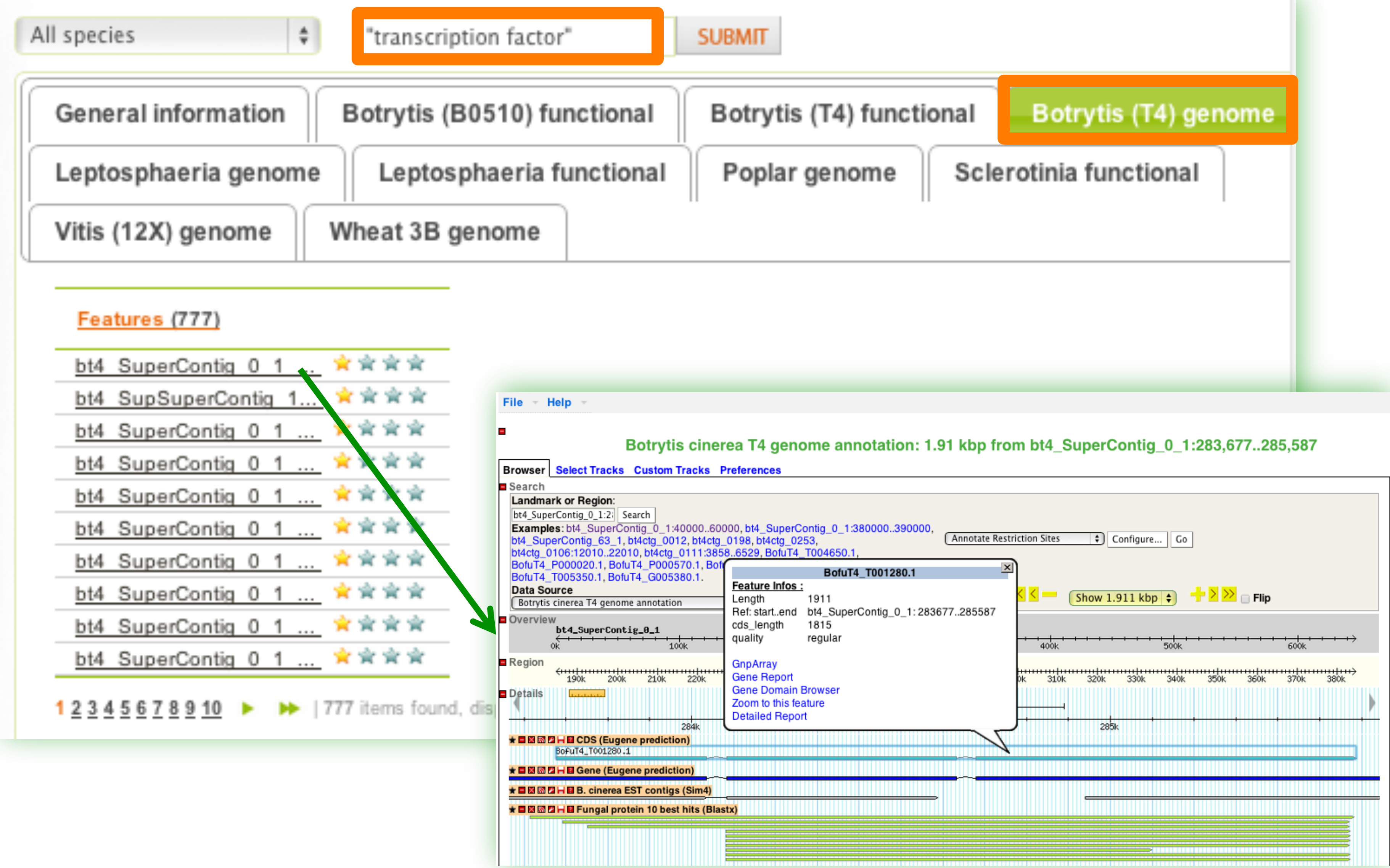
VVIF52

You can find the indexed data list here.

Examples: VVI, VVIF52, gene, arabidopsis, AY109603, Xwmc430

joelle.amselem@versailles.inra.fr

Quick Search Results: structural annotation



Features (777)

- bt4_SuperContig_0_1_...
- bt4_SupSuperContig_1...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...
- bt4_SuperContig_0_1_...

1 2 3 4 5 6 7 8 9 10 ▶ ▶▶ | 777 items found, dis

Botrytis cinerea T4 genome annotation: 1.91 kbp from bt4_SuperContig_0_1:283,677..285,587

Landmark or Region: bt4_SuperContig_0_1:2

Examples: bt4_SuperContig_0_1:40000..60000, bt4_SuperContig_0_1:380000..390000, bt4_SuperContig_63_1, bt4ctg_0012, bt4ctg_0198, bt4ctg_0253, bt4ctg_0106:12010..22010, bt4ctg_0111:3858..6529, BofuT4_T004650.1, BofuT4_P000020.1, BofuT4_P000570.1, Bof BofuT4_T005350.1, BofuT4_G005380.1.

Data Source: Botrytis cinerea T4 genome annotation

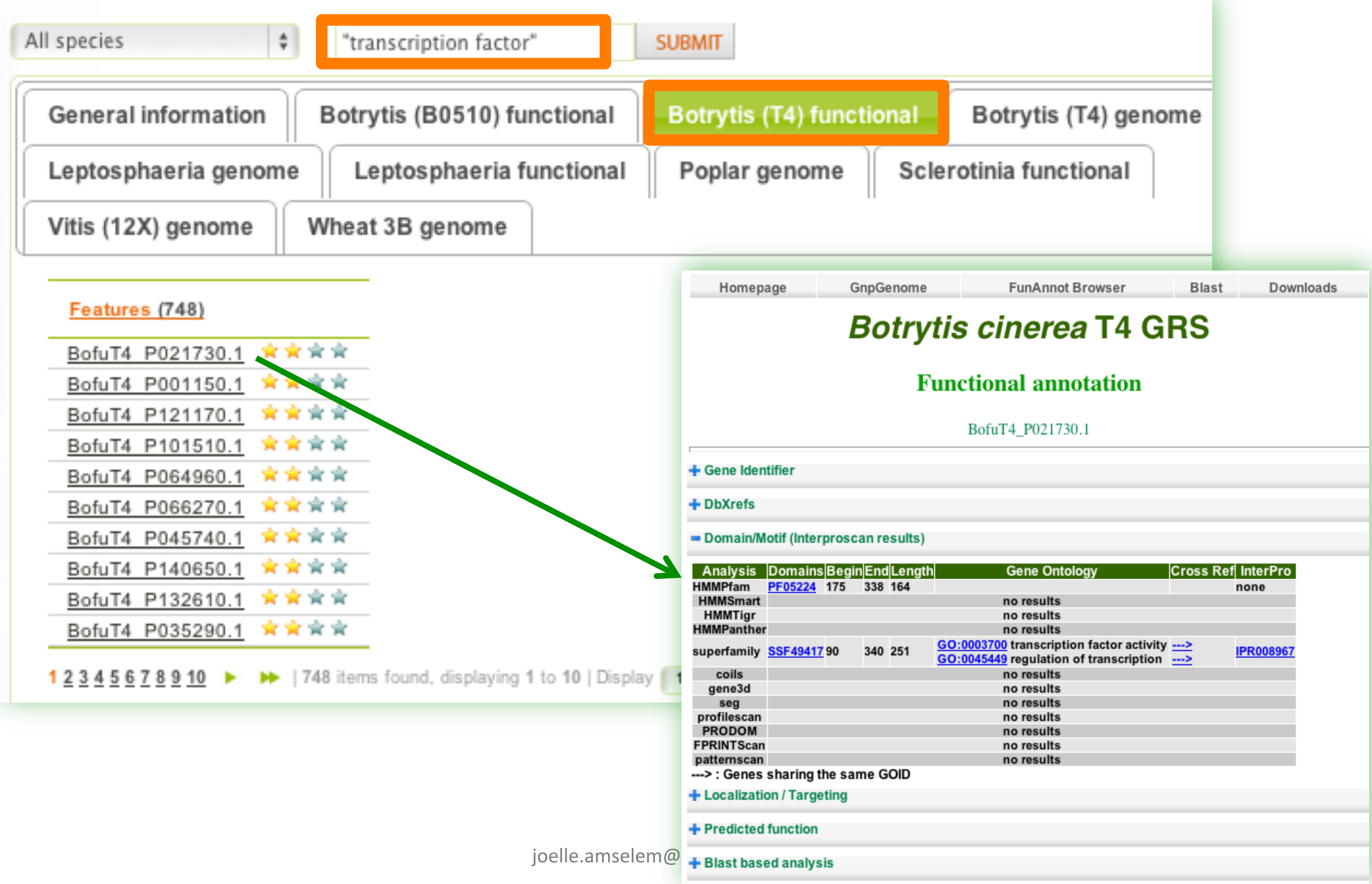
BofuT4_T001280.1

Feature Infos :

Length	1911
Ref: start..end	bt4_SuperContig_0_1: 283677..285587
cds_length	1815
quality	regular

joelle.amselem@versailles.inra.fr

Quick Search Results: functional annotation



All species

"transcription factor"

Features (748)

- BofuT4_P021730.1 ★★☆☆
- BofuT4_P001150.1 ★★☆☆
- BofuT4_P121170.1 ★★☆☆
- BofuT4_P101510.1 ★★☆☆
- BofuT4_P064960.1 ★★☆☆
- BofuT4_P066270.1 ★★☆☆
- BofuT4_P045740.1 ★★☆☆
- BofuT4_P140650.1 ★★☆☆
- BofuT4_P132610.1 ★★☆☆
- BofuT4_P035290.1 ★★☆☆

1 2 3 4 5 6 7 8 9 10 ▶ ▶▶ | 748 items found, displaying 1 to 10 | Display

Botrytis cinerea T4 GRS

Functional annotation

BofuT4_P021730.1

+ Gene Identifier

+ DbXrefs

− Domain/Motif (Interproscan results)

Analysis	Domains	Begin	End	Length	Gene Ontology	Cross Ref	InterPro
HMMPfam	PF05224	175	338	164			none
HMMSmart	no results						
HMMTigr	no results						
HMMPanther	no results						
superfamily	SSF49417	90	340	251	GO:0003700 transcription factor activity GO:0045449 regulation of transcription	---> --->	IPR008967
coils	no results						
gene3d	no results						
seg	no results						
profilescan	no results						
PRODOM	no results						
FPRINTScan	no results						
patternscan	no results						

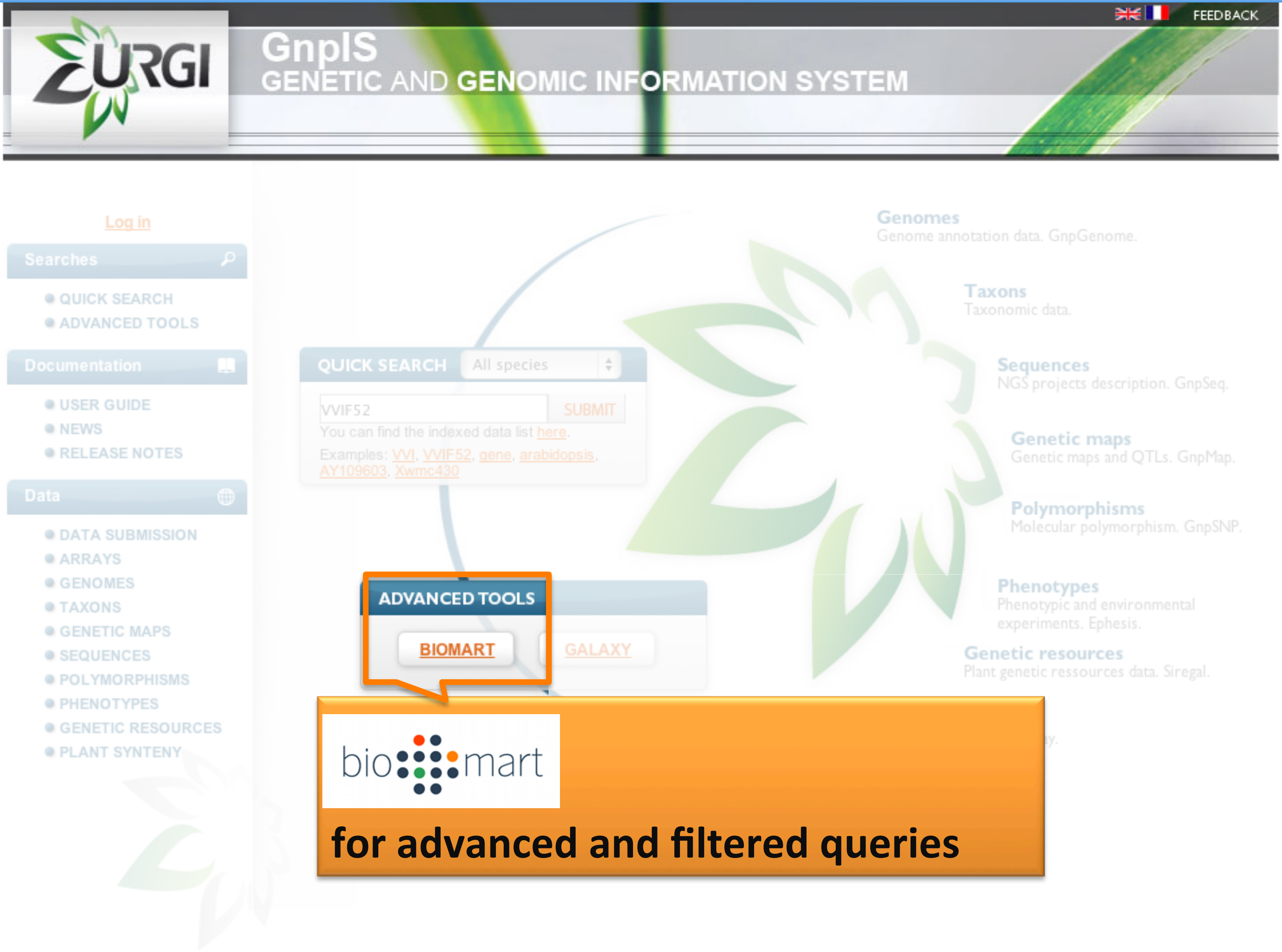
---> : Genes sharing the same GOID

+ Localization / Targeting

+ Predicted function

+ Blast based analysis

joelle.amselem@



VVIF52

You can find the indexed data list here.

Examples: VVI, VVIF52, gene, arabidopsis, AY109603, Xwmc430

Genomes
Genome annotation data. GnpGenome.

Taxons
Taxonomic data.

Sequences
NGS projects description. GnpSeq.

Genetic maps
Genetic maps and QTLs. GnpMap.

Polymorphisms
Molecular polymorphism. GnpSNP.

Phenotypes
Phenotypic and environmental experiments. Ephesis.

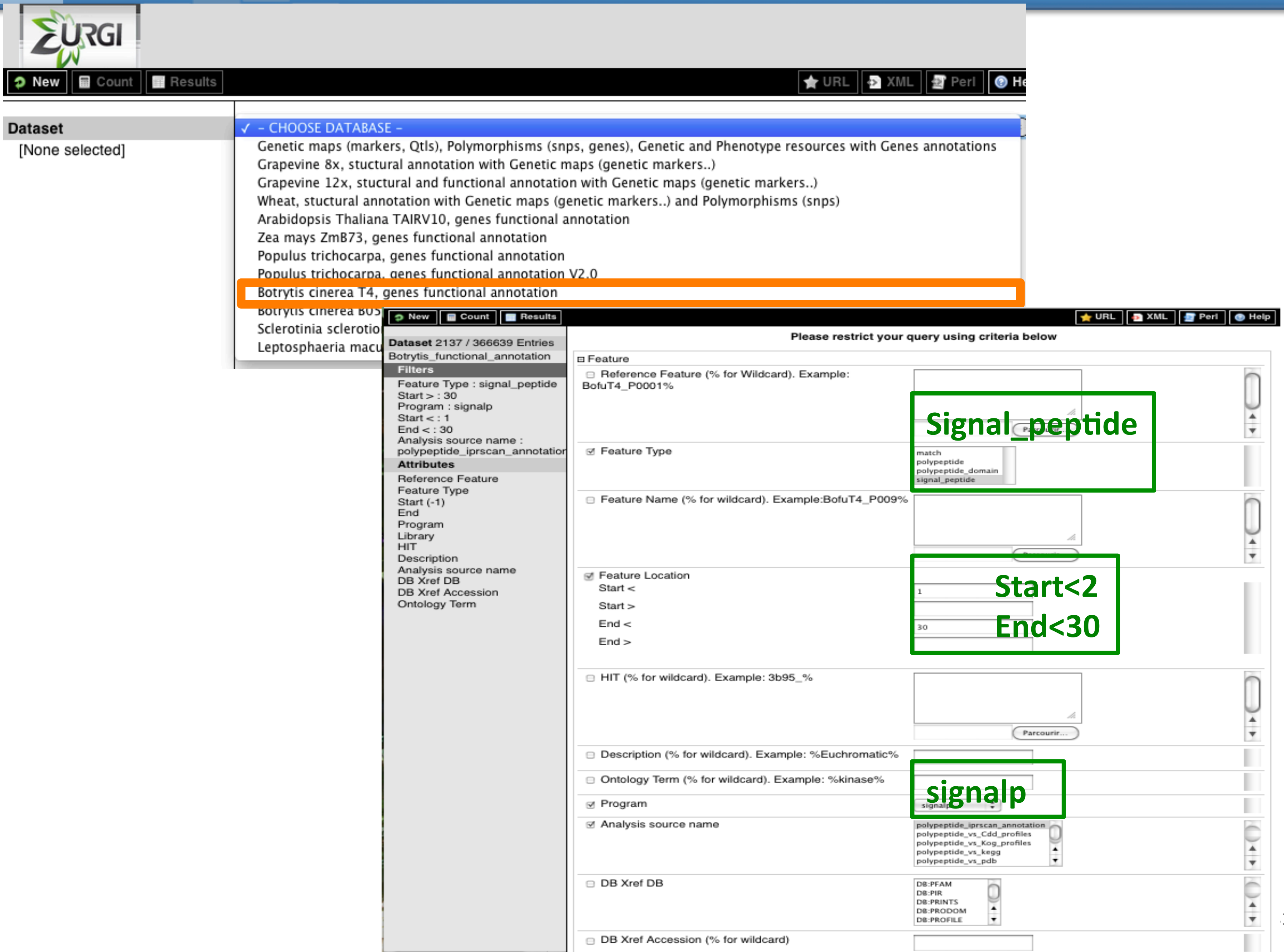
Genetic resources
Plant genetic ressources data. Siregal.

biomart

for advanced and filtered queries

joelle.amselem@versailles.inra.fr

biomart for advanced and filtered queries



Dataset

[None selected]

✓ – CHOOSE DATABASE –

Genetic maps (markers, Qtls), Polymorphisms (snps, genes), Genetic and Phenotype resources with Genes annotations
Grapevine 8x, stuctural annotation with Genetic maps (genetic markers..)
Grapevine 12x, stuctural and functional annotation with Genetic maps (genetic markers..)
Wheat, stuctural annotation with Genetic maps (genetic markers..) and Polymorphisms (snps)
Arabidopsis Thaliana TAIRV10, genes functional annotation
Zea mays ZmB73, genes functional annotation
Populus trichocarpa, genes functional annotation
Populus trichocarpa, genes functional annotation V2.0
Botrytis cinerea T4, genes functional annotation
Botrytis cinerea B05
Sclerotinia sclerotio
Leptosphaeria macu

Dataset 2137 / 366639 Entries
Botrytis_functional_annotation

Filters
Feature Type : signal_peptide
Start > : 30
Program : signalp
Start < : 1
End < : 30
Analysis source name : polypeptide_iprscan_annotation

Attributes
Reference Feature
Feature Type
Start (-1)
End
Program
Library
HIT
Description
Analysis source name
DB Xref DB
DB Xref Accession
Ontology Term

Please restrict your query using criteria below

⊟ Feature

☐ Reference Feature (% for Wildcard). Example: BofuT4_P0001%

☑ Feature Type: match, polypeptide, polypeptide_domain, signal_peptide

☐ Feature Name (% for wildcard). Example:BofuT4_P009%

☑ Feature Location
Start < 1
Start >
End < 30
End >

☐ HIT (% for wildcard). Example: 3b95_%

☐ Description (% for wildcard). Example: %Euchromatic%

☐ Ontology Term (% for wildcard). Example: %kinase%

☑ Program: signalp

☑ Analysis source name: polypeptide_iprscan_annotation, polypeptide_vs_Cdd_profiles, polypeptide_vs_Kog_profiles, polypeptide_vs_kegg, polypeptide_vs_pdb

☐ DB Xref DB: DB:PFAM, DB:PIR, DB:PRINTS, DB:PRODOM, DB:PROFILE

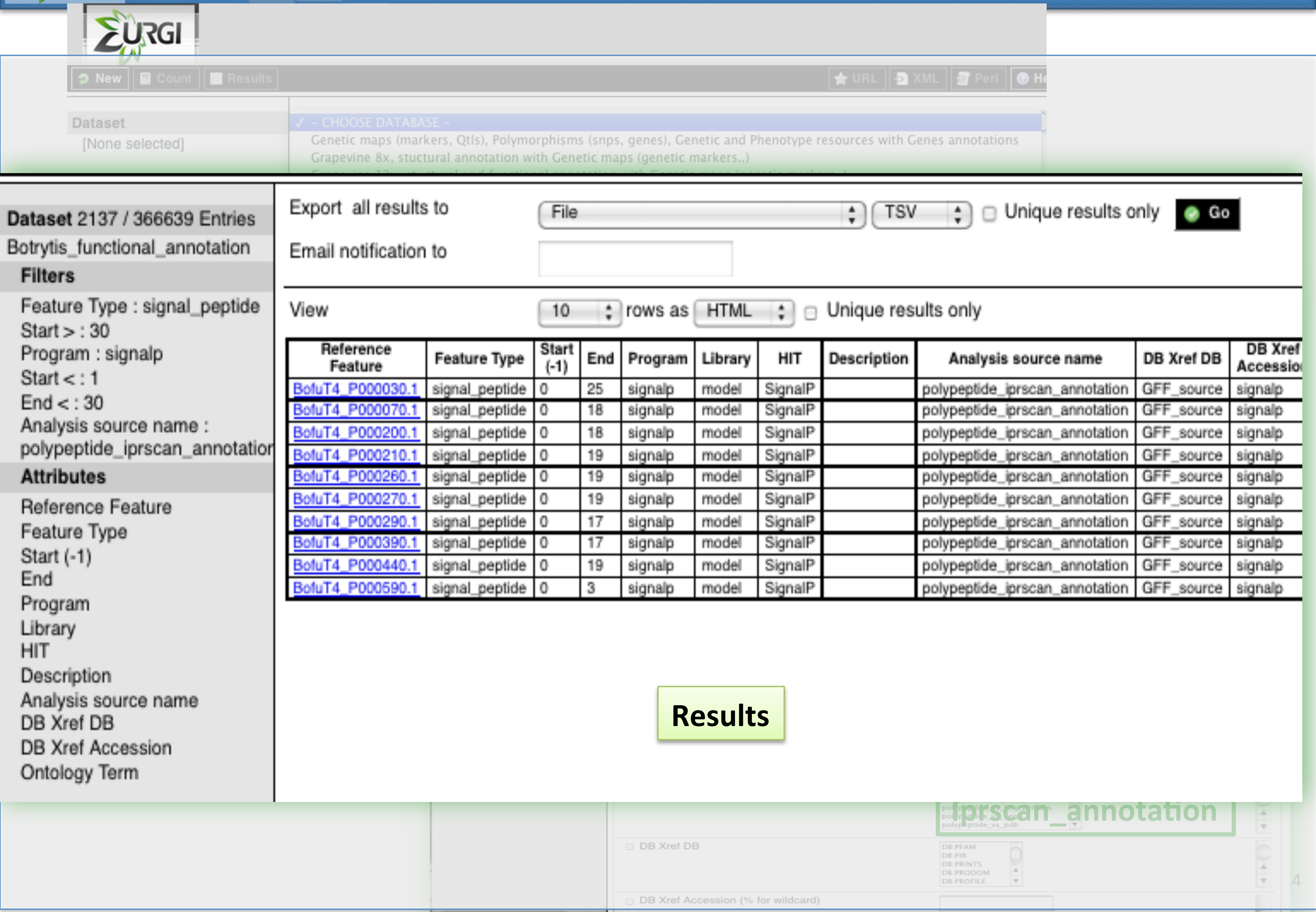
☐ DB Xref Accession (% for wildcard)

Signal_peptide

Start<2
End<30

signalp

to query on functional annotation



Dataset
[None selected]

✓ - CHOOSE DATABASE -
Genetic maps (markers, Qtls), Polymorphisms (snps, genes), Genetic and Phenotype resources with Genes annotations
Grapevine 8x, stuctural annotation with Genetic maps (genetic markers..)

Dataset 2137 / 366639 Entries
Botrytis_functional_annotation

Filters
Feature Type : signal_peptide
Start > : 30
Program : signalp
Start < : 1
End < : 30
Analysis source name : polypeptide_iprscan_annotation

Attributes
Reference Feature
Feature Type
Start (-1)
End
Program
Library
HIT
Description
Analysis source name
DB Xref DB
DB Xref Accession
Ontology Term

Export all results to: File, TSV, ☐ Unique results only, Go
Email notification to

View 10 rows as HTML ☐ Unique results only

Reference Feature	Feature Type	Start (-1)	End	Program	Library	HIT	Description	Analysis source name	DB Xref DB	DB Xref Accessio
BofuT4_P000030.1	signal_peptide	0	25	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000070.1	signal_peptide	0	18	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000200.1	signal_peptide	0	18	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000210.1	signal_peptide	0	19	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000260.1	signal_peptide	0	19	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000270.1	signal_peptide	0	19	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000290.1	signal_peptide	0	17	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000390.1	signal_peptide	0	17	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000440.1	signal_peptide	0	19	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp
BofuT4_P000590.1	signal_peptide	0	3	signalp	model	SignalP		polypeptide_iprscan_annotation	GFF_source	signalp

Results

iprscan_annotation

☐ DB Xref DB
DB:PFAM
DB:PIR
DB:PRINTS
DB:PRODOM
DB:PROFILE

☐ DB Xref Accession (% for wildcard)

to query on functional annotation

Dataset 2137 / 366639 Entries

Botrytis_functional_annotation

Filters

Feature Type : signal_peptide
Start > : 30
Program : signalp
Start < : 1
End < : 30
Analysis source name : polypeptide_iprscan_annotation

Attributes

Reference Feature
Feature Type
Start (-1)
End
Program
Library
HIT
Description
Analysis source name
DB Xref DB
DB Xref Accession
Ontology Term

Reference Feature
BofuT4_P000030.1
BofuT4_P000070.1
BofuT4_P000200.1
BofuT4_P000210.1
BofuT4_P000390.1
BofuT4_P000440.1
BofuT4_P000590.1

Link to GRS

Botrytis cinerea T4 GRS

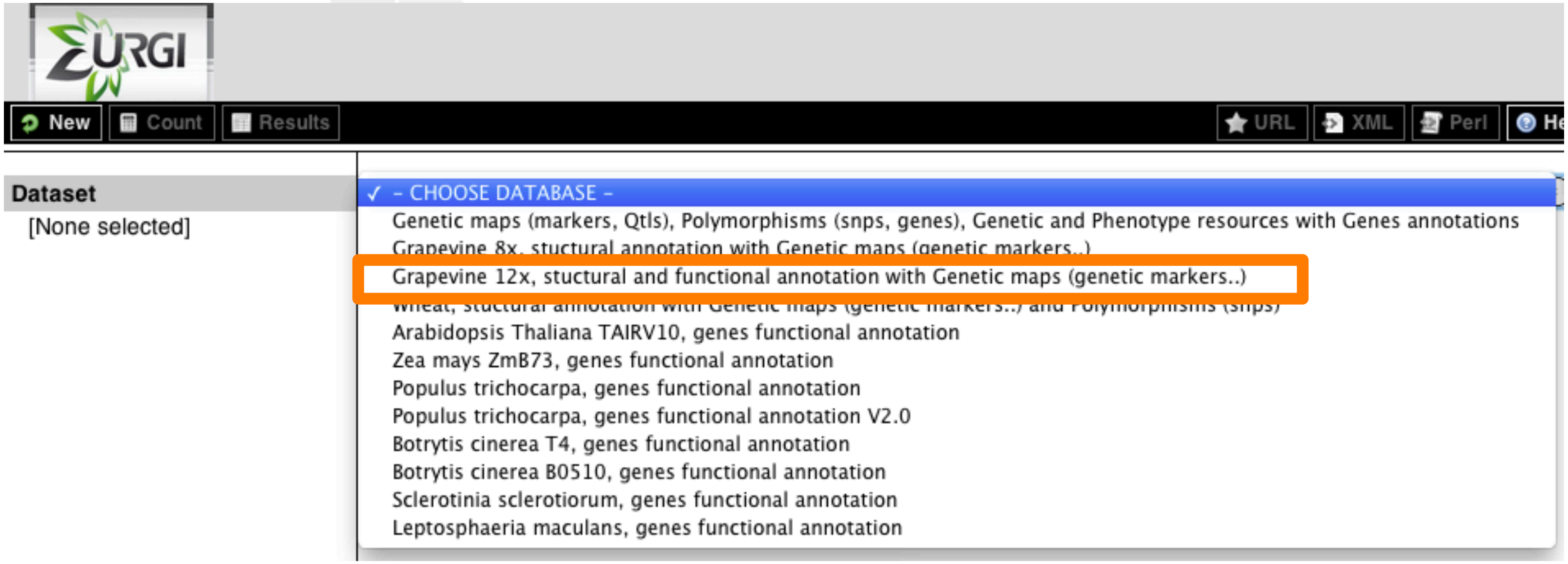
Functional annotation

BofuT4_P000210.1

Analysis	Start	End	Length	Location	Reliability class	Signal Peptide CutOff	Mitochondrion CutOff
targetp	1	19	19	Secretory pathway	3	0.820	0.301

Analysis	Start	End	Length
signalp	1	19	18
tmhmm	Not result		

biomart for GnpIS interoperability

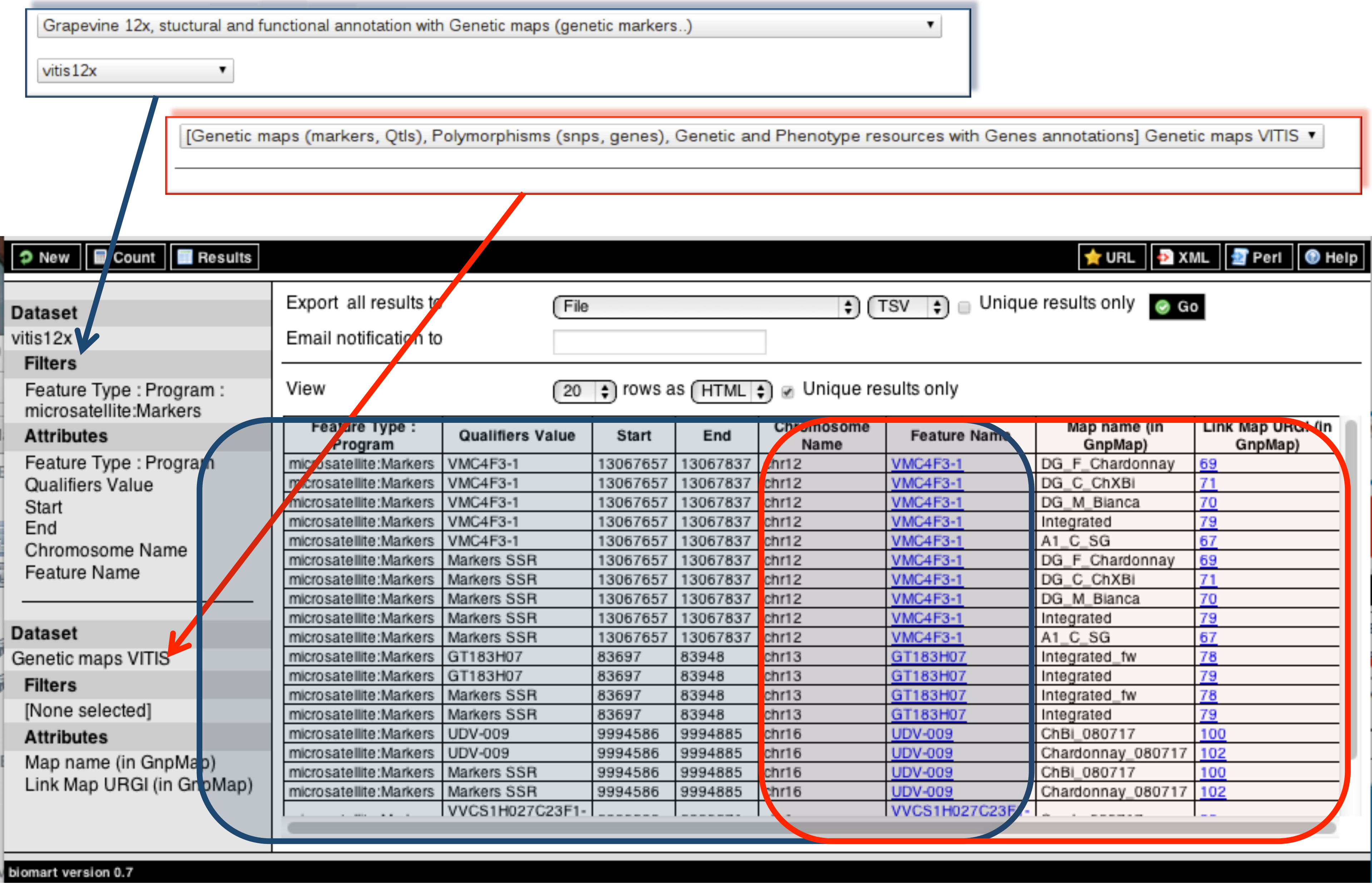


Dataset
[None selected]

✓ - CHOOSE DATABASE -
- Genetic maps (markers, Qtls), Polymorphisms (snps, genes), Genetic and Phenotype resources with Genes annotations
- Grapevine 8x, stuctural annotation with Genetic maps (genetic markers..)
- Grapevine 12x, stuctural and functional annotation with Genetic maps (genetic markers..)
- Wheat, stuctural annotation with Genetic maps (genetic markers..) and Polymorphisms (snps)
- Arabidopsis Thaliana TAIRV10, genes functional annotation
- Zea mays ZmB73, genes functional annotation
- Populus trichocarpa, genes functional annotation
- Populus trichocarpa, genes functional annotation V2.0
- Botrytis cinerea T4, genes functional annotation
- Botrytis cinerea B0510, genes functional annotation
- Sclerotinia sclerotiorum, genes functional annotation
- Leptosphaeria maculans, genes functional annotation

joelle.amselem@versailles.inra.fr

biomart for GnpIS interoperability



Grapevine 12x, stuctural and functional annotation with Genetic maps (genetic markers..)

vitis12x

[Genetic maps (markers, Qtls), Polymorphisms (snps, genes), Genetic and Phenotype resources with Genes annotations] Genetic maps VITIS

Dataset
vitis12x
Filters
Feature Type : Program : microsatellite:Markers
Attributes
Feature Type : Program
Qualifiers Value
Start
End
Chromosome Name
Feature Name

Dataset
Genetic maps VITIS
Filters
[None selected]
Attributes
Map name (in GnpMap)
Link Map URGI (in GnpMap)

Export all results to: File, TSV ☐ Unique results only

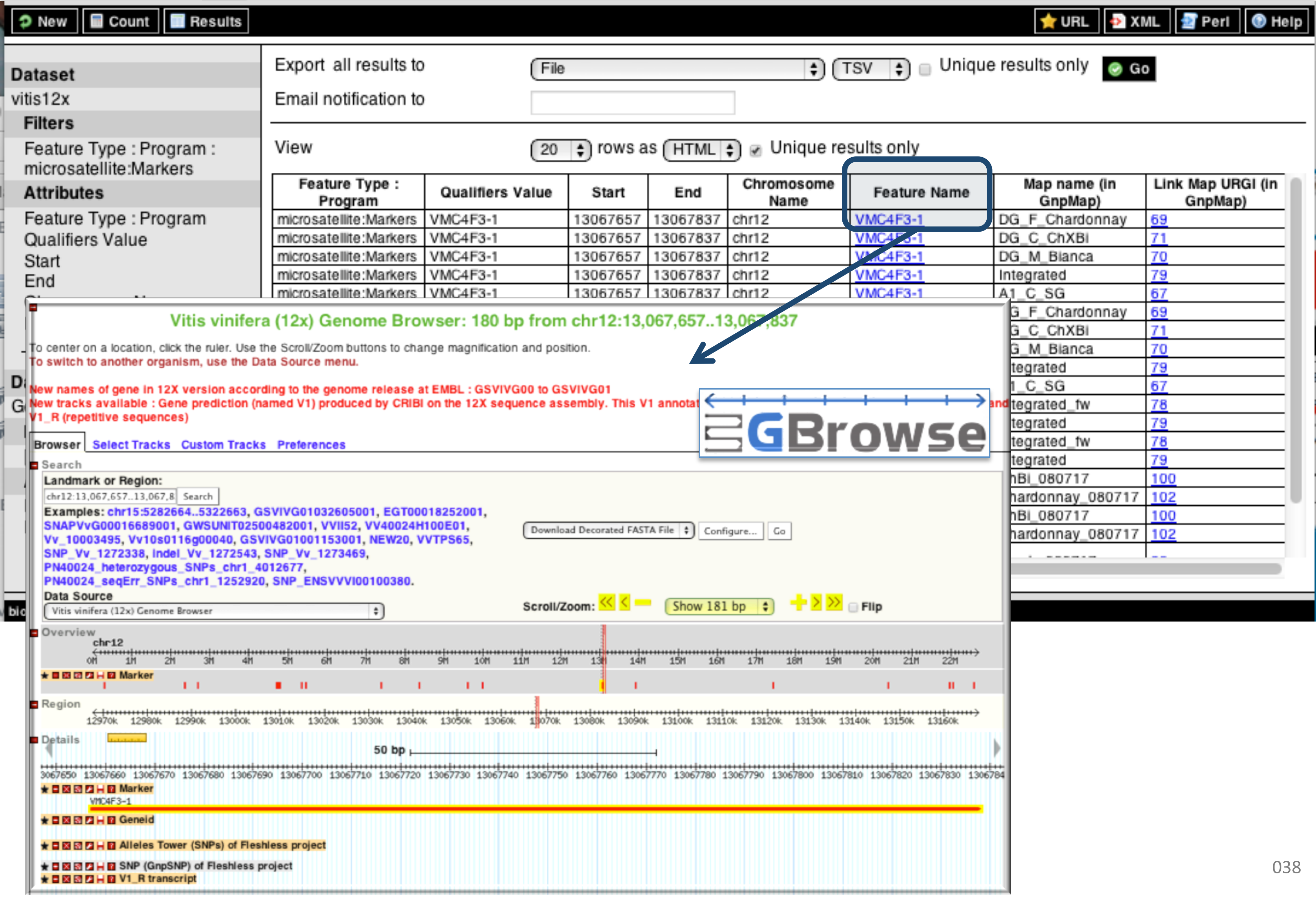
Email notification to

View 20 rows as HTML ☑ Unique results only

Feature Type : Program	Qualifiers Value	Start	End	Chromosome Name	Feature Name	Map name (in GnpMap)	Link Map URGI (in GnpMap)
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_F_Chardonnay	69
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_C_ChXBi	71
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_M_Bianca	70
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	Integrated	79
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	A1_C_SG	67
microsatellite:Markers	Markers SSR	13067657	13067837	chr12	VMC4F3-1	DG_F_Chardonnay	69
microsatellite:Markers	Markers SSR	13067657	13067837	chr12	VMC4F3-1	DG_C_ChXBi	71
microsatellite:Markers	Markers SSR	13067657	13067837	chr12	VMC4F3-1	DG_M_Bianca	70
microsatellite:Markers	Markers SSR	13067657	13067837	chr12	VMC4F3-1	Integrated	79
microsatellite:Markers	Markers SSR	13067657	13067837	chr12	VMC4F3-1	A1_C_SG	67
microsatellite:Markers	GT183H07	83697	83948	chr13	GT183H07	Integrated_fw	78
microsatellite:Markers	GT183H07	83697	83948	chr13	GT183H07	Integrated	79
microsatellite:Markers	Markers SSR	83697	83948	chr13	GT183H07	Integrated_fw	78
microsatellite:Markers	Markers SSR	83697	83948	chr13	GT183H07	Integrated	79
microsatellite:Markers	UDV-009	9994586	9994885	chr16	UDV-009	ChBi_080717	100
microsatellite:Markers	UDV-009	9994586	9994885	chr16	UDV-009	Chardonnay_080717	102
microsatellite:Markers	Markers SSR	9994586	9994885	chr16	UDV-009	ChBi_080717	100
microsatellite:Markers	Markers SSR	9994586	9994885	chr16	UDV-009	Chardonnay_080717	102
	VVCS1H027C23F1-				VVCS1H027C23F-		

biomart version 0.7

biomart ...to GBrowse



Dataset
vitis12x
Filters
Feature Type : Program : microsatellite:Markers
Attributes
Feature Type : Program
Qualifiers Value
Start
End

Export all results to File TSV ☐ Unique results only Go

Email notification to

View 20 rows as HTML ☑ Unique results only

Feature Type : Program	Qualifiers Value	Start	End	Chromosome Name	Feature Name	Map name (in GnpMap)	Link Map URGI (in GnpMap)
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_F_Chardonnay	69
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_C_ChXBi	71
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_M_Bianca	70
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	Integrated	79
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	A1_C_SG	67
						G_F_Chardonnay	69
						G_C_ChXBi	71
						G_M_Bianca	70
						tegrated	79
						1_C_SG	67
						tegrated_fw	78
						tegrated	79
						tegrated_fw	78
						tegrated	79
						hBi_080717	100
						hardonnay_080717	102
						hBi_080717	100
						hardonnay_080717	102

Vitis vinifera (12x) Genome Browser: 180 bp from chr12:13,067,657..13,067,837

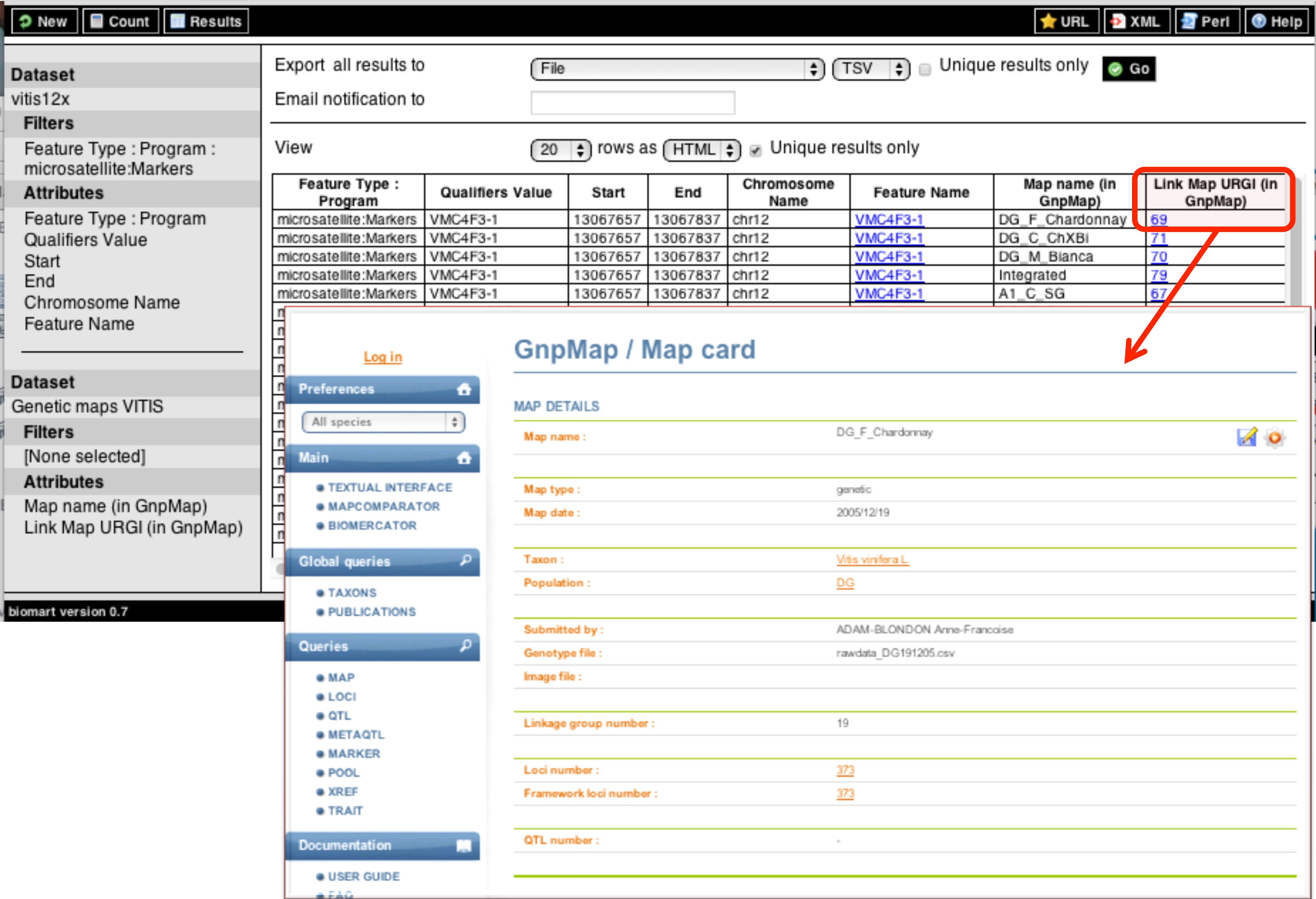
To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.
To switch to another organism, use the Data Source menu.

New names of gene in 12X version according to the genome release at EMBL : GSVIVG00 to GSVIVG01
New tracks available : Gene prediction (named V1) produced by CRIBI on the 12X sequence assembly. This V1 annotat and
V1_R (repetitive sequences)

Landmark or Region:
chr12:13,067,657..13,067,8 Search
Examples: chr15:5282664..5322663, GSVIVG01032605001, EGT00018252001, SNAPVvG00016689001, GWSUNIT02500482001, VVII52, VV40024H100E01, Vv_10003495, Vv10s0116g00040, GSVIVG01001153001, NEW20, VVTPS65, SNP_Vv_1272338, indel_Vv_1272543, SNP_Vv_1273469, PN40024_heterozygous_SNPs_chr1_4012677, PN40024_seqErr_SNPs_chr1_1252920, SNP_ENSVVVI00100380.
Data Source
Vitis vinifera (12x) Genome Browser

biomart ...to Genetic maps

Genetic maps and QTLs. GnpMap



Dataset
vitis12x
Filters
Feature Type : Program : microsatellite:Markers
Attributes
Feature Type : Program
Qualifiers Value
Start
End
Chromosome Name
Feature Name

Dataset
Genetic maps VITIS
Filters
[None selected]
Attributes
Map name (in GnpMap)
Link Map URGI (in GnpMap)

biomart version 0.7

Export all results to: File, TSV ☐ Unique results only — Go

Email notification to

View 20 rows as HTML ☑ Unique results only

Feature Type : Program	Qualifiers Value	Start	End	Chromosome Name	Feature Name	Map name (in GnpMap)	Link Map URGI (in GnpMap)
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_F_Chardonnay	69
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_C_ChXBi	71
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	DG_M_Bianca	70
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	Integrated	79
microsatellite:Markers	VMC4F3-1	13067657	13067837	chr12	VMC4F3-1	A1_C_SG	67

GnpMap / Map card

MAP DETAILS

Map name : DG_F_Chardonnay

Map type : genetic
Map date : 2005/12/19

Taxon : Vitis vinifera L.
Population : DG

Submitted by : ADAM-BLONDON Anne-Francoise
Genotype file : rawdata_DG191205.csv
Image file :

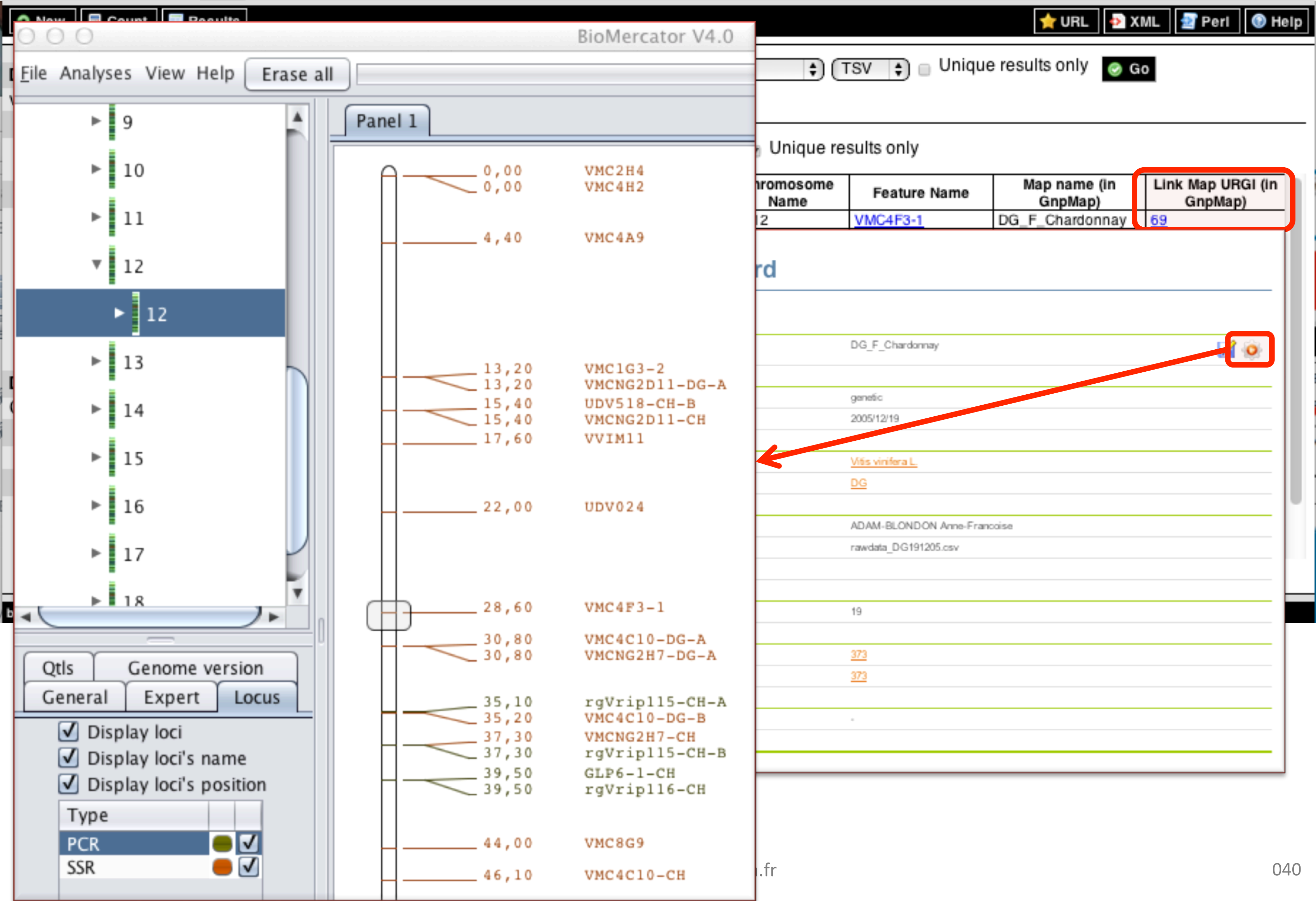
Linkage group number : 19

Loci number : 373
Framework loci number : 373

QTL number : -

Run Biomercator for

Genetic maps
Genetic maps and QTLs. GnpMap



☐ Unique results only

Unique results only

Chromosome Name	Feature Name	Map name (in GnpMap)	Link Map URGI (in GnpMap)
12	VMC4F3-1	DG_F_Chardonnay	69

DG_F_Chardonnay

genetic

2005/12/19

Vitis vinifera L.

DG

ADAM-BLONDON Anne-Francoise

rawdata_DG191205.csv

19

373

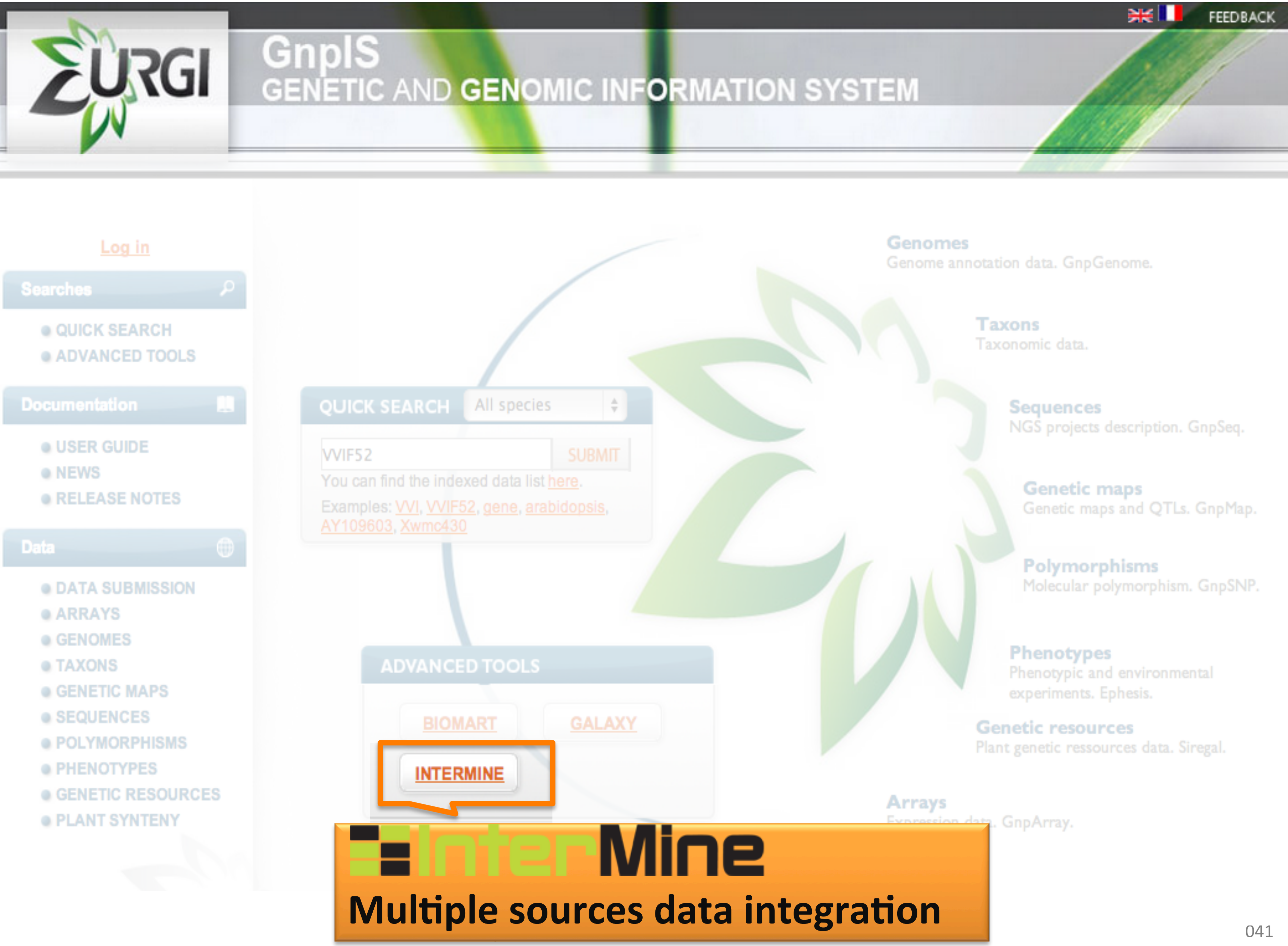
373

-

Position	Locus
0,00	VMC2H4
0,00	VMC4H2
4,40	VMC4A9
13,20	VMC1G3-2
13,20	VMCNG2D11-DG-A
15,40	UDV518-CH-B
15,40	VMCNG2D11-CH
17,60	VVIM11
22,00	UDV024
28,60	VMC4F3-1
30,80	VMC4C10-DG-A
30,80	VMCNG2H7-DG-A
35,10	rgVrip115-CH-A
35,20	VMC4C10-DG-B
37,30	VMCNG2H7-CH
37,30	rgVrip115-CH-B
39,50	GLP6-1-CH
39,50	rgVrip116-CH
44,00	VMC8G9
46,10	VMC4C10-CH

☑ Display loci
☑ Display loci's name
☑ Display loci's position

Type		
PCR	🟩	☑
SSR	🟧	☑



VVIF52

You can find the indexed data list here.

Examples: VVI, VVIF52, gene, arabidopsis, AY109603, Xwmc430

Genomes
Genome annotation data. GnpGenome.

Taxons
Taxonomic data.

Sequences
NGS projects description. GnpSeq.

Genetic maps
Genetic maps and QTLs. GnpMap.

Polymorphisms
Molecular polymorphism. GnpSNP.

Phenotypes
Phenotypic and environmental experiments. Ephesis.

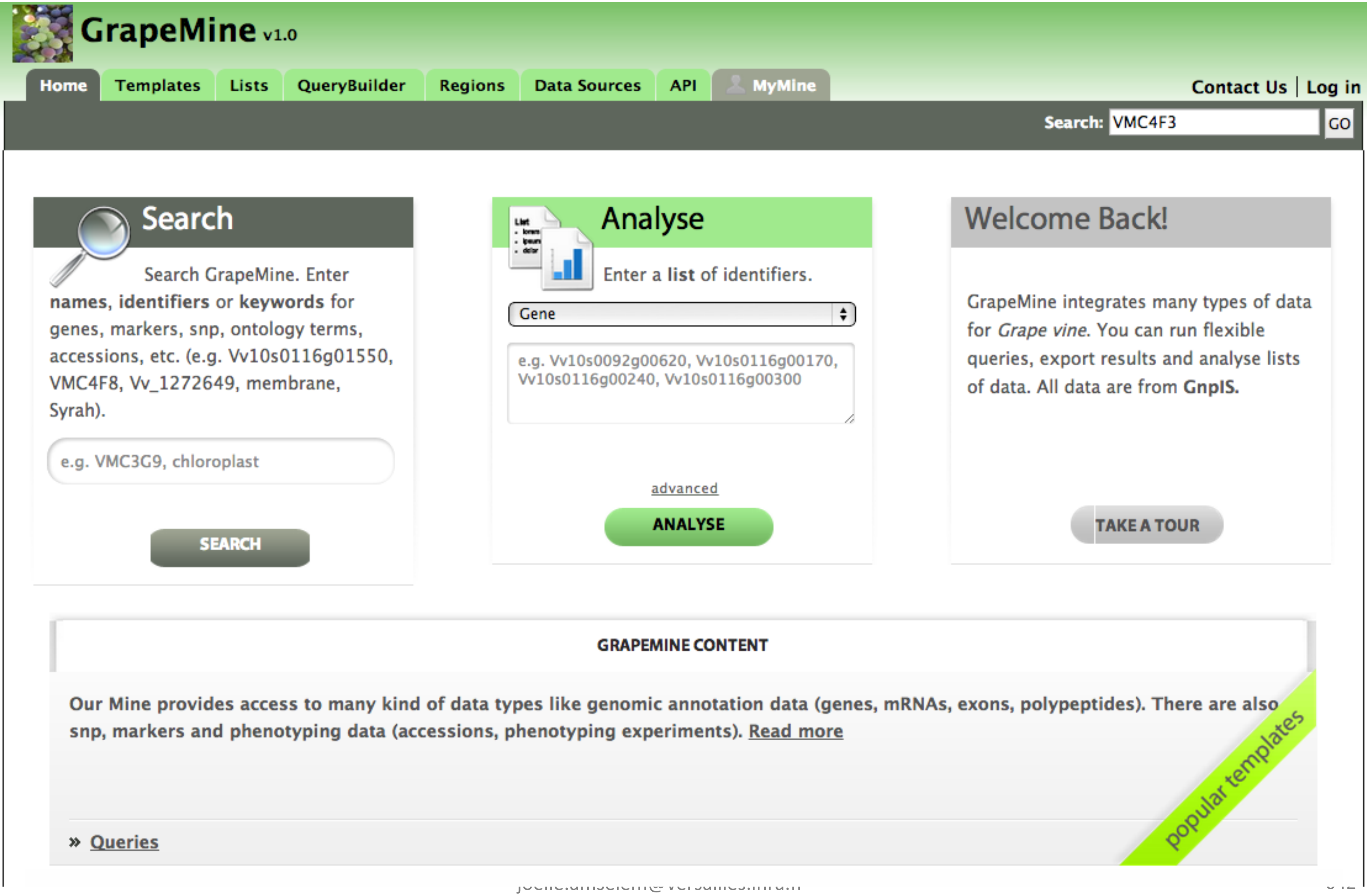
Genetic resources
Plant genetic ressources data. Siregal.

Arrays
Expression data. GnpArray.

InterMine

Multiple sources data integration

Grape data integration



Search

Search GrapeMine. Enter **names**, **identifiers** or **keywords** for genes, markers, snp, ontology terms, accessions, etc. (e.g. Vv10s0116g01550, VMC4F8, Vv_1272649, membrane, Syrah).

Analyse

Enter a **list** of identifiers.

Welcome Back!

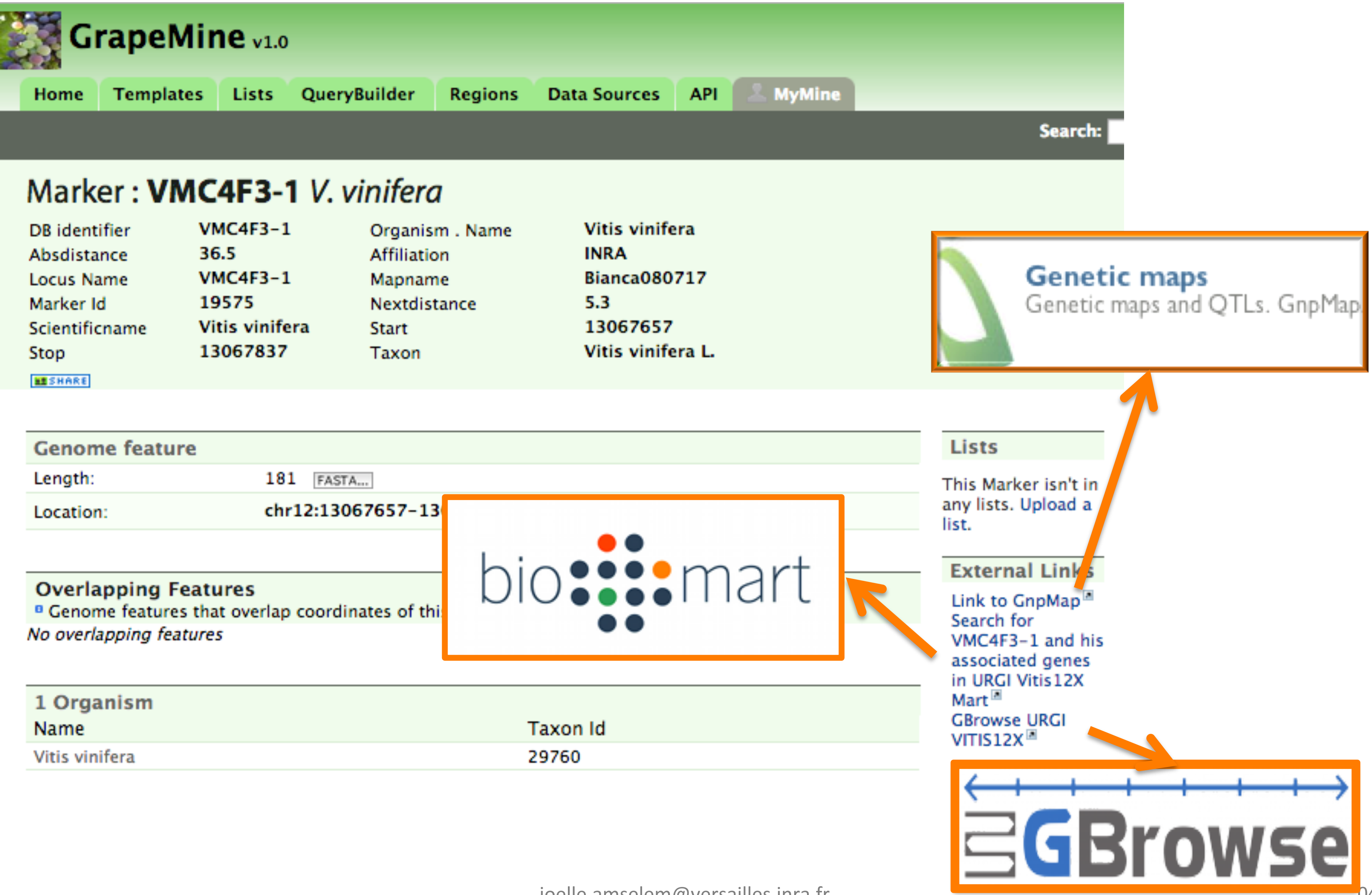
GrapeMine integrates many types of data for *Grape vine*. You can run flexible queries, export results and analyse lists of data. All data are from **GnpIS**.

GRAPEMINE CONTENT

Our Mine provides access to many kind of data types like genomic annotation data (genes, mRNAs, exons, polypeptides). There are also snp, markers and phenotyping data (accessions, phenotyping experiments). Read more

» Queries

Marker search



Marker : **VMC4F3-1** *V. vinifera*

DB identifier	**VMC4F3-1**	Organism . Name	**Vitis vinifera**
Absdistance	**36.5**	Affiliation	**INRA**
Locus Name	**VMC4F3-1**	Mapname	**Bianca080717**
Marker Id	**19575**	Nextdistance	**5.3**
Scientificname	**Vitis vinifera**	Start	**13067657**
Stop	**13067837**	Taxon	**Vitis vinifera L.**

Genome feature

Length: 181

Location: chr12:13067657-13

Overlapping Features

Genome features that overlap coordinates of thi

No overlapping features

1 Organism

Name	Taxon Id
Vitis vinifera	29760

Lists

This Marker isn't in any lists. Upload a list.

External Links

Link to GnpMap

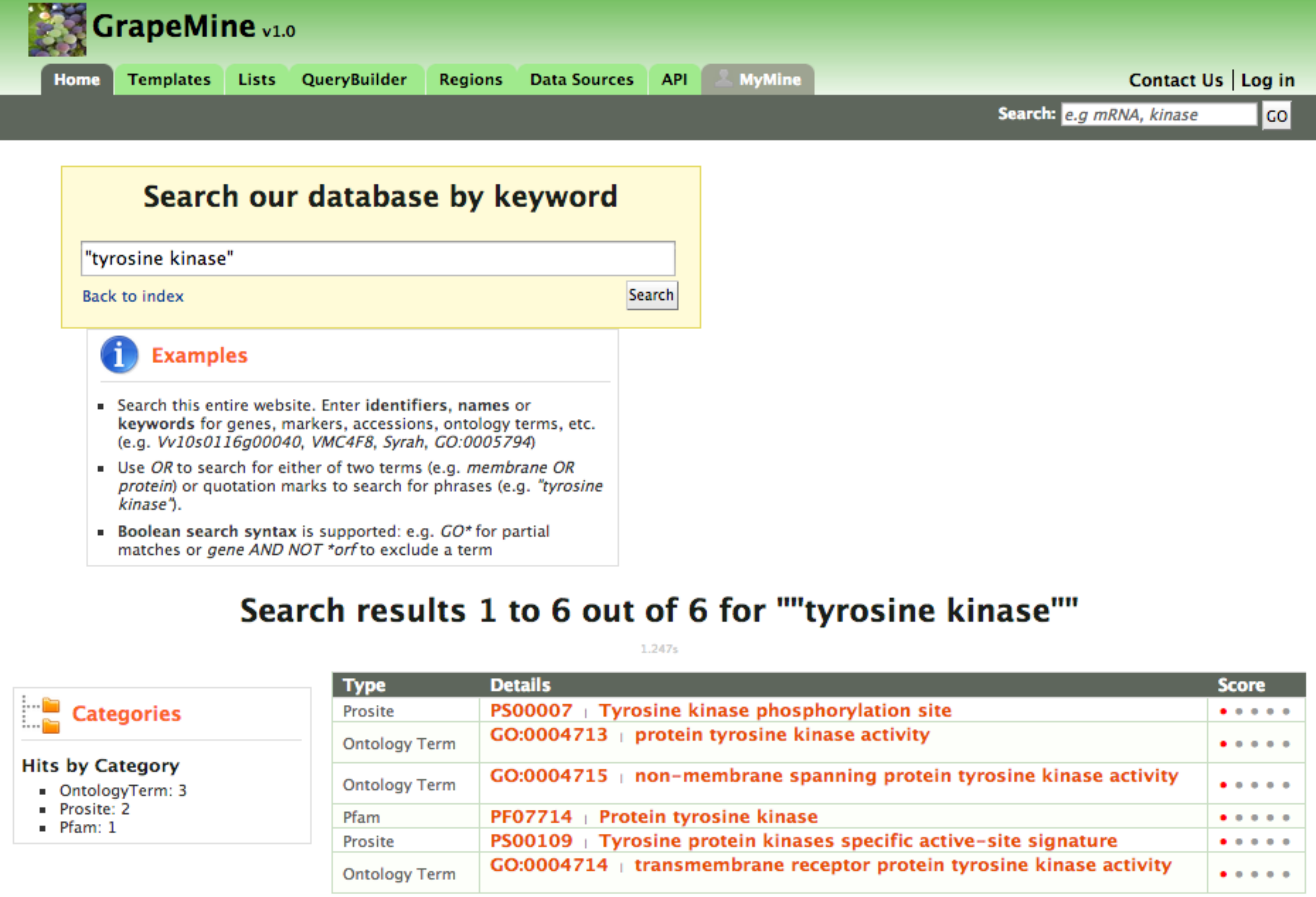
Search for VMC4F3-1 and his associated genes in URGI Vitis12X Mart

GBrowse URGI VITIS12X

Genetic maps

Genetic maps and QTLs. GnpMap

joelle.amselem@versailles.inra.fr



Search our database by keyword

"tyrosine kinase"

Examples

- Search this entire website. Enter **identifiers, names** or **keywords** for genes, markers, accessions, ontology terms, etc. (e.g. *Vv10s0116g00040, VMC4F8, Syrah, GO:0005794*)
- Use *OR* to search for either of two terms (e.g. *membrane OR protein*) or quotation marks to search for phrases (e.g. *"tyrosine kinase"*).
- **Boolean search syntax** is supported: e.g. *GO** for partial matches or *gene AND NOT *orf* to exclude a term

Search results 1 to 6 out of 6 for ""tyrosine kinase""

1.247s

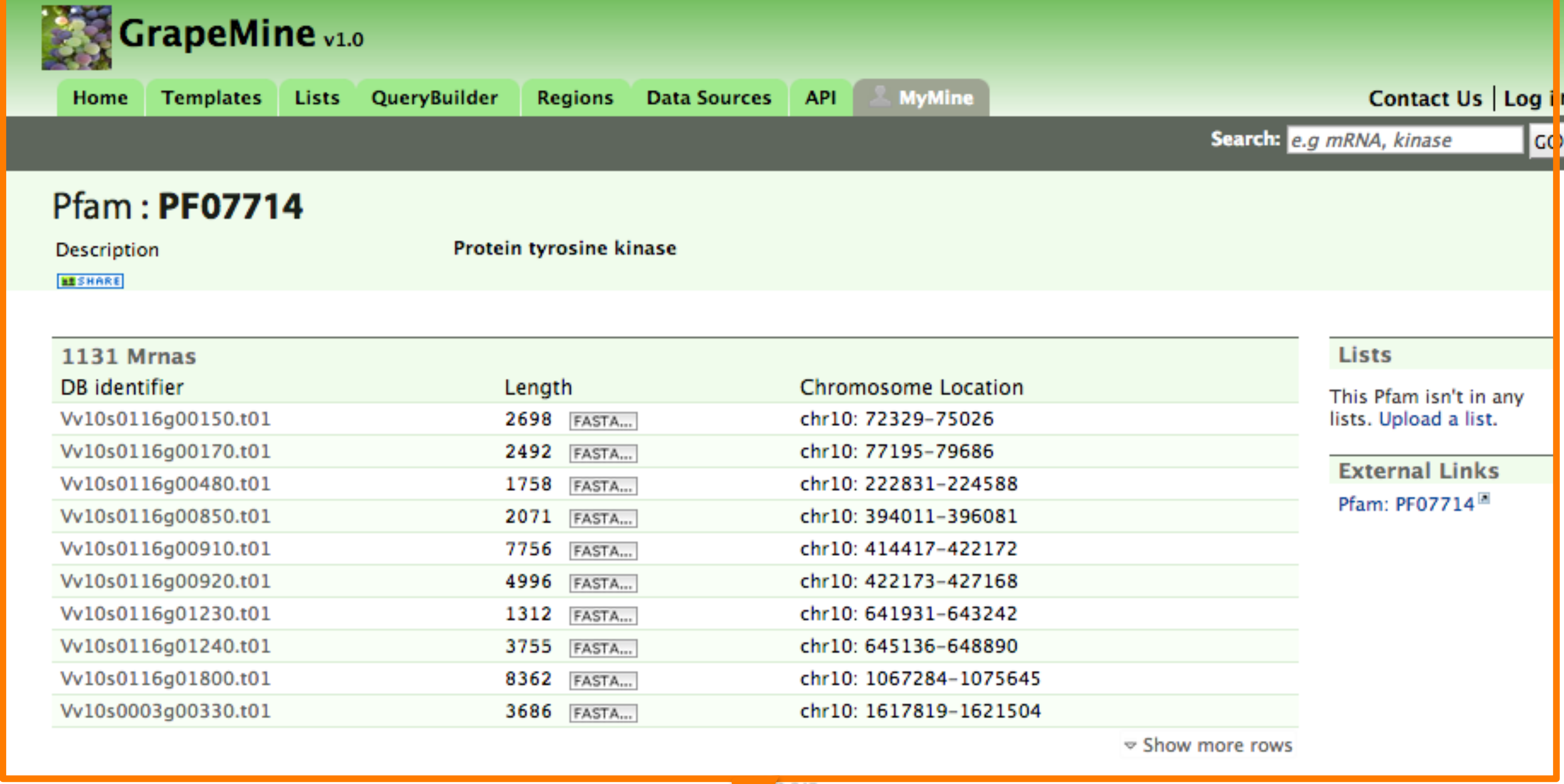
Categories

Hits by Category

- OntologyTerm: 3
- Prosite: 2
- Pfam: 1

Type	Details	Score
Prosite	PS00007 | Tyrosine kinase phosphorylation site	●○○○○
Ontology Term	GO:0004713 | protein tyrosine kinase activity	●○○○○
Ontology Term	GO:0004715 | non-membrane spanning protein tyrosine kinase activity	●○○○○
Pfam	PF07714 | Protein tyrosine kinase	●○○○○
Prosite	PS00109 | Tyrosine protein kinases specific active-site signature	●○○○○
Ontology Term	GO:0004714 | transmembrane receptor protein tyrosine kinase activity	●○○○○

joelle.amselem@versailles.inra.fr



Pfam : **PF07714**

Description **Protein tyrosine kinase**

1131 Mrnas

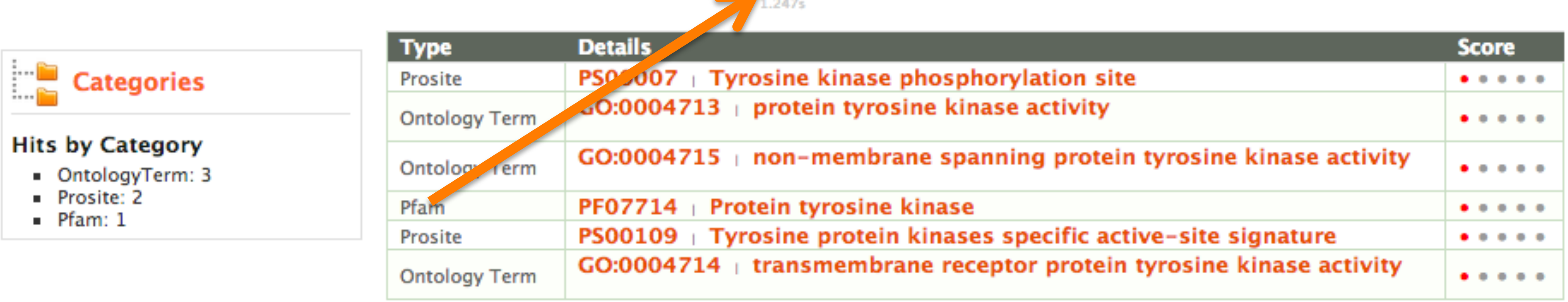
DB identifier	Length		Chromosome Location
Vv10s0116g00150.t01	2698	FASTA...	chr10: 72329-75026
Vv10s0116g00170.t01	2492	FASTA...	chr10: 77195-79686
Vv10s0116g00480.t01	1758	FASTA...	chr10: 222831-224588
Vv10s0116g00850.t01	2071	FASTA...	chr10: 394011-396081
Vv10s0116g00910.t01	7756	FASTA...	chr10: 414417-422172
Vv10s0116g00920.t01	4996	FASTA...	chr10: 422173-427168
Vv10s0116g01230.t01	1312	FASTA...	chr10: 641931-643242
Vv10s0116g01240.t01	3755	FASTA...	chr10: 645136-648890
Vv10s0116g01800.t01	8362	FASTA...	chr10: 1067284-1075645
Vv10s0003g00330.t01	3686	FASTA...	chr10: 1617819-1621504

Lists

This Pfam isn't in any lists. Upload a list.

External Links

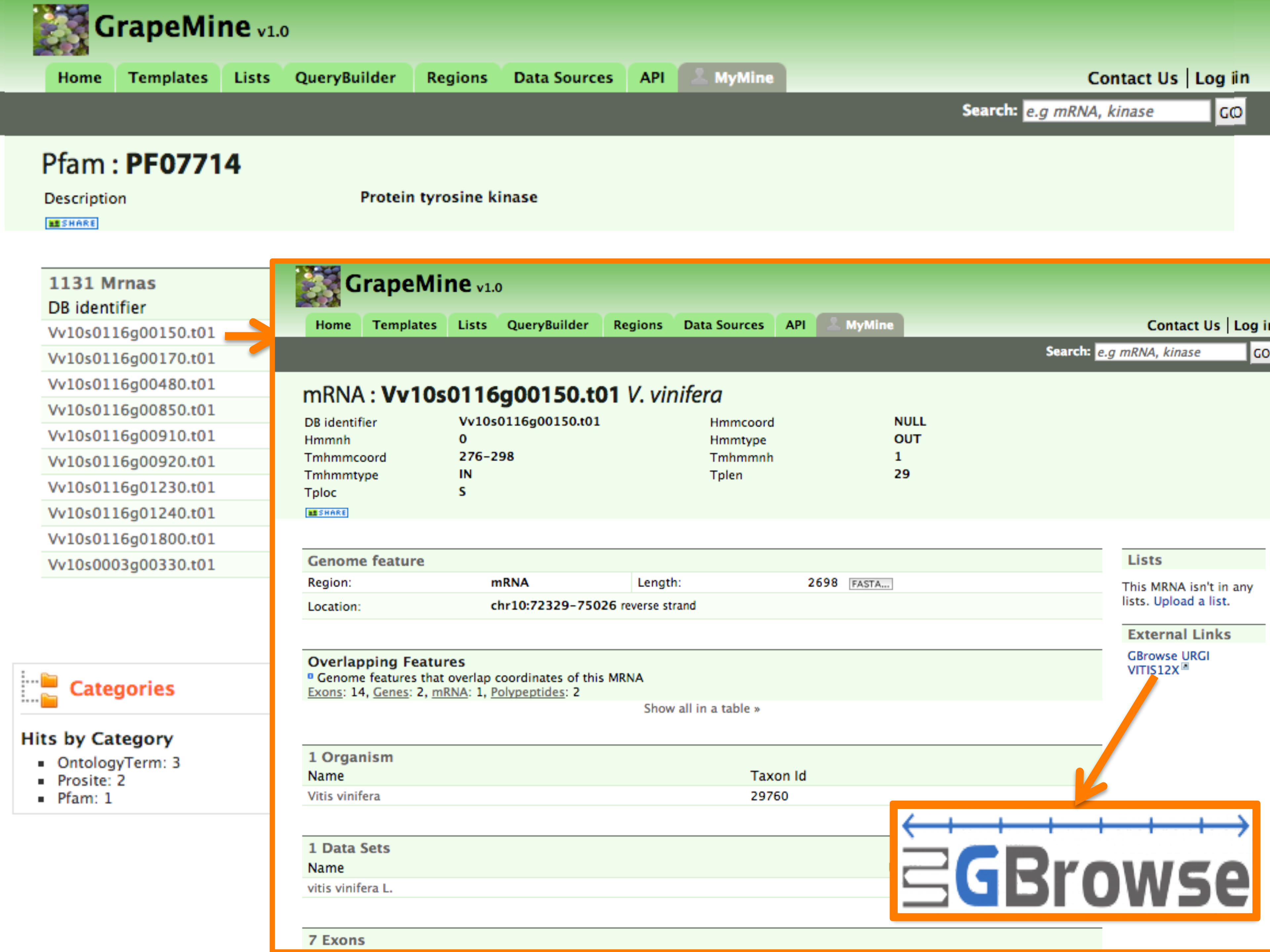
Pfam: PF07714



Hits by Category

- OntologyTerm: 3
- Prosite: 2
- Pfam: 1

Type	Details	Score
Prosite	PS00007 | Tyrosine kinase phosphorylation site	● ● ● ● ●
Ontology Term	GO:0004713 | protein tyrosine kinase activity	● ● ● ● ●
Ontology Term	GO:0004715 | non-membrane spanning protein tyrosine kinase activity	● ● ● ● ●
Pfam	PF07714 | Protein tyrosine kinase	● ● ● ● ●
Prosite	PS00109 | Tyrosine protein kinases specific active-site signature	● ● ● ● ●
Ontology Term	GO:0004714 | transmembrane receptor protein tyrosine kinase activity	● ● ● ● ●



Pfam : **PF07714**

Description	**Protein tyrosine kinase**

1131 Mrnas

DB identifier
Vv10s0116g00150.t01
Vv10s0116g00170.t01
Vv10s0116g00480.t01
Vv10s0116g00850.t01
Vv10s0116g00910.t01
Vv10s0116g00920.t01
Vv10s0116g01230.t01
Vv10s0116g01240.t01
Vv10s0116g01800.t01
Vv10s0003g00330.t01

Categories

Hits by Category

- OntologyTerm: 3
- Prosite: 2
- Pfam: 1

mRNA : **Vv10s0116g00150.t01** *V. vinifera*

DB identifier	**Vv10s0116g00150.t01**	Hmmcoord	**NULL**
Hmmnh	**0**	Hmmtype	**OUT**
Tmhmmcoord	**276-298**	Tmhmmnh	**1**
Tmhmmtype	**IN**	Tplen	**29**
Tploc	**S**		

Genome feature

Region:	**mRNA**	Length:	**2698** FASTA...
Location:	**chr10:72329-75026** reverse strand		

Overlapping Features

Genome features that overlap coordinates of this MRNA
Exons: 14, Genes: 2, mRNA: 1, Polypeptides: 2

Show all in a table »

1 Organism

Name	Taxon Id
Vitis vinifera	29760

1 Data Sets

Name
vitis vinifera L.

7 Exons

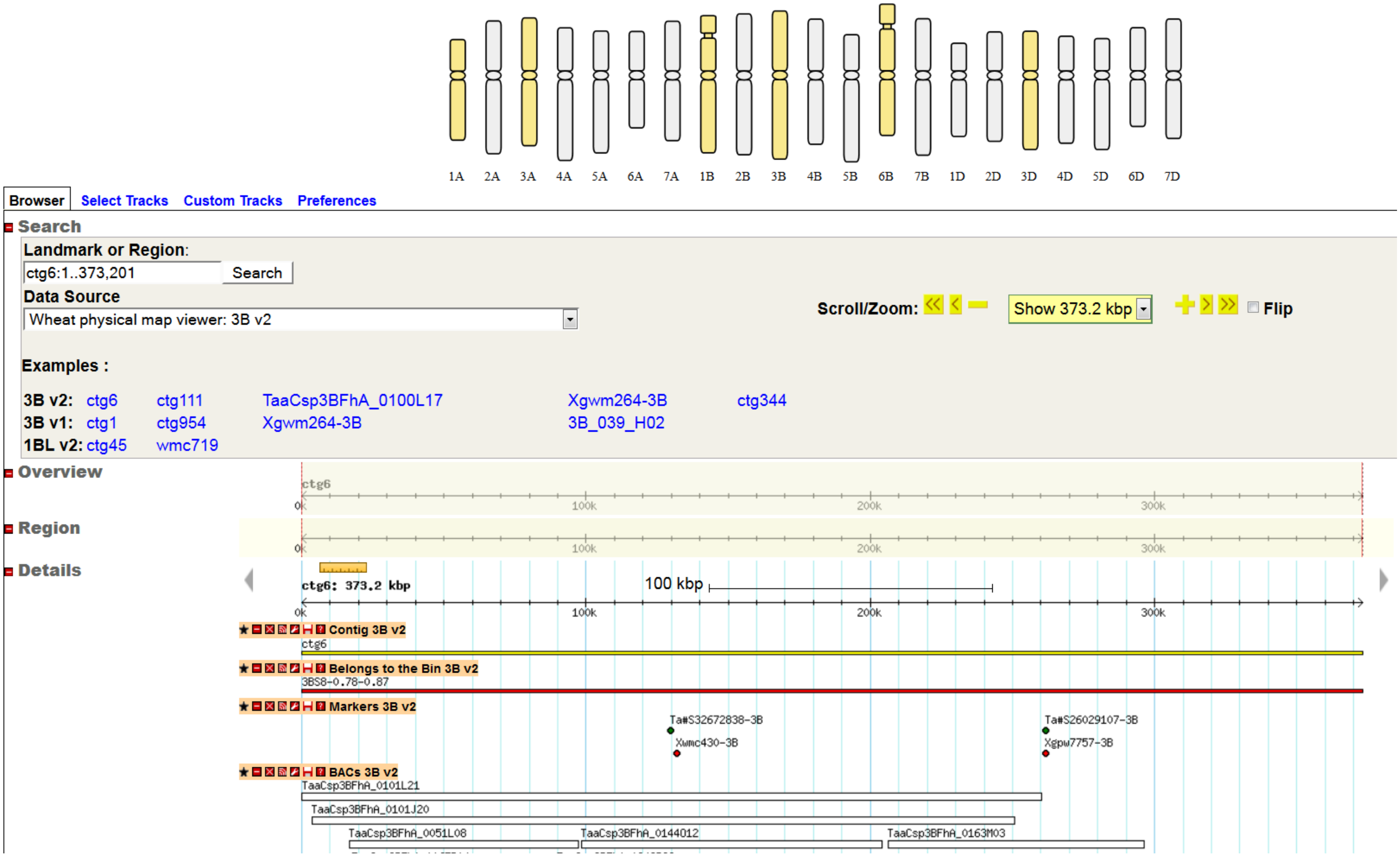
Lists

This MRNA isn't in any lists. Upload a list.

External Links

GBrowse URGI VITIS12X

GBrowse wheat physical maps



Browser Select Tracks Custom Tracks Preferences

Search

Landmark or Region: ctg6:1..373,201 Search

Data Source: Wheat physical map viewer: 3B v2

Scroll/Zoom: Show 373.2 kbp ☐ Flip

Examples :

3B v2:	ctg6	ctg111	TaaCsp3BFhA_0100L17	Xgwm264-3B	ctg344
3B v1:	ctg1	ctg954	Xgwm264-3B	3B_039_H02	
1BL v2:	ctg45	wmc719			

Overview

Region

Details

Contig 3B v2: ctg6

Belongs to the Bin 3B v2: 3BS8-0.78-0.87

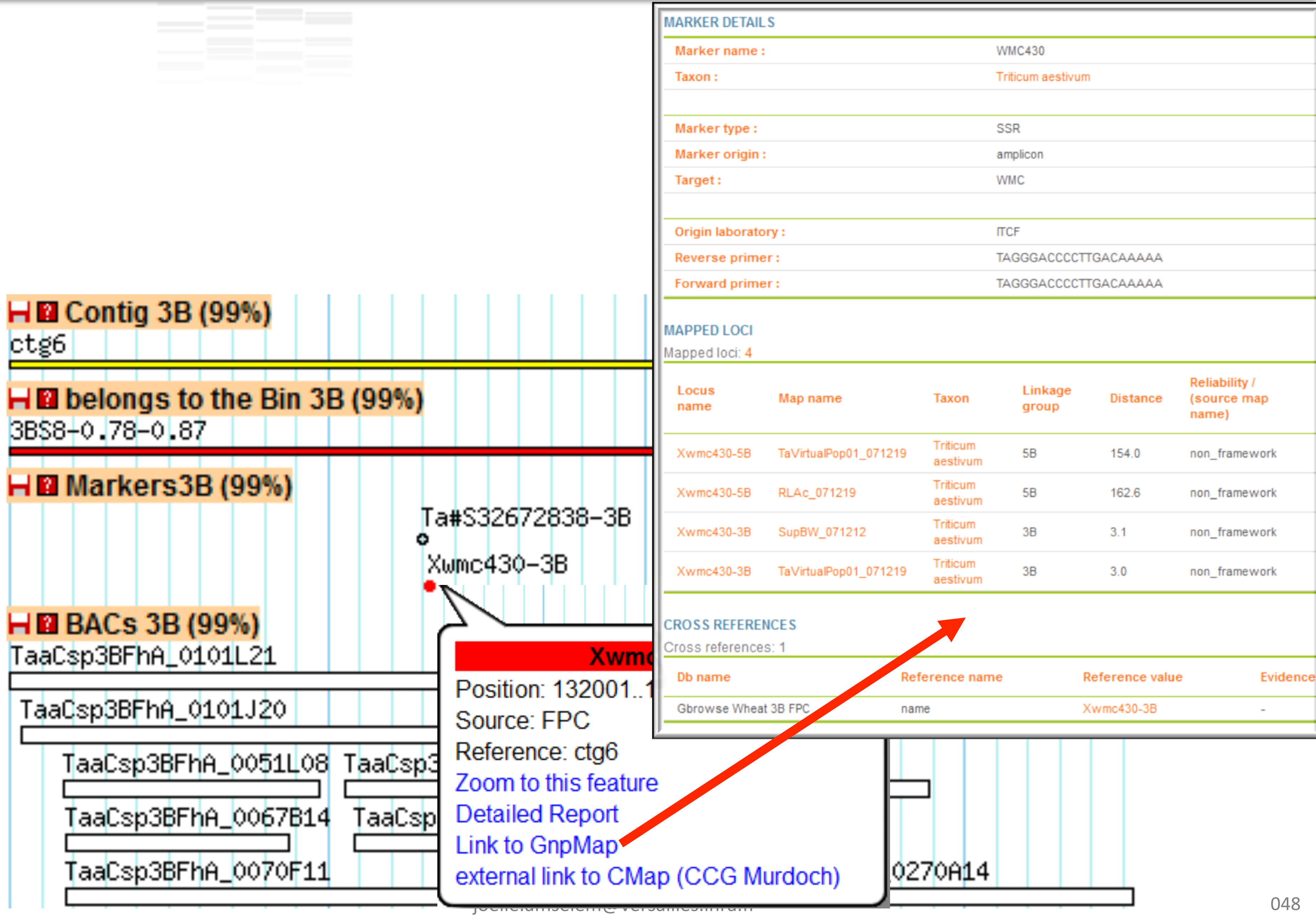
Markers 3B v2: Ta#S32672838-3B, Xwmc430-3B, Ta#S26029107-3B, Xgpw7757-3B

BACs 3B v2: TaaCsp3BFhA_0101L21, TaaCsp3BFhA_0101J20, TaaCsp3BFhA_0051L08, TaaCsp3BFhA_0144O12, TaaCsp3BFhA_0163M03

joelle.amselem@versailles.inra.fr

GBrowse wheat physical maps to Genetic maps

Genetic maps and QTLs. GnpMap



Contig 3B (99%)
ctg6

belongs to the Bin 3B (99%)
3BS8-0.78-0.87

Markers3B (99%)
Ta#S32672838-3B
Xwmc430-3B

BACs 3B (99%)
TaaCsp3BFhA_0101L21
TaaCsp3BFhA_0101J20
TaaCsp3BFhA_0051L08 TaaCsp3
TaaCsp3BFhA_0067B14 TaaCsp
TaaCsp3BFhA_0070F11
0270A14

Xwmc
Position: 132001..1
Source: FPC
Reference: ctg6
Zoom to this feature
Detailed Report
Link to GnpMap
external link to CMap (CCG Murdoch)

MARKER DETAILS

Marker name :	WMC430
Taxon :	Triticum aestivum
Marker type :	SSR
Marker origin :	amplicon
Target :	WMC
Origin laboratory :	ITCF
Reverse primer :	TAGGGACCCCTTGACAAAAA
Forward primer :	TAGGGACCCCTTGACAAAAA

MAPPED LOCI

Mapped loci: 4

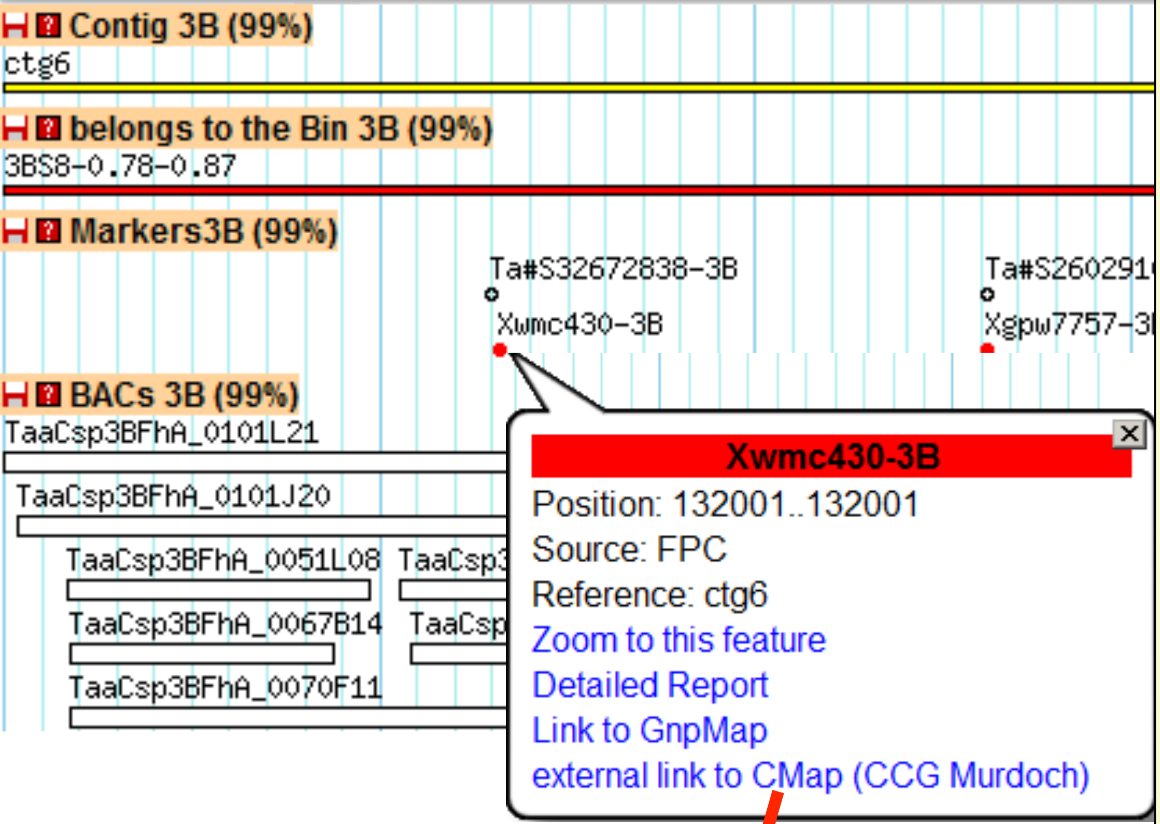
Locus name	Map name	Taxon	Linkage group	Distance	Reliability / (source map name)
Xwmc430-5B	TaVirtualPop01_071219	Triticum aestivum	5B	154.0	non_framework
Xwmc430-5B	RLAc_071219	Triticum aestivum	5B	162.6	non_framework
Xwmc430-3B	SupBW_071212	Triticum aestivum	3B	3.1	non_framework
Xwmc430-3B	TaVirtualPop01_071219	Triticum aestivum	3B	3.0	non_framework

CROSS REFERENCES

Cross references: 1

Db name	Reference name	Reference value	Evidence
Gbrowse Wheat 3B FPC	name	Xwmc430-3B	-

GBrowse wheat physical maps to CMap



Xwmc430-3B

Position: 132001..132001

Source: FPC

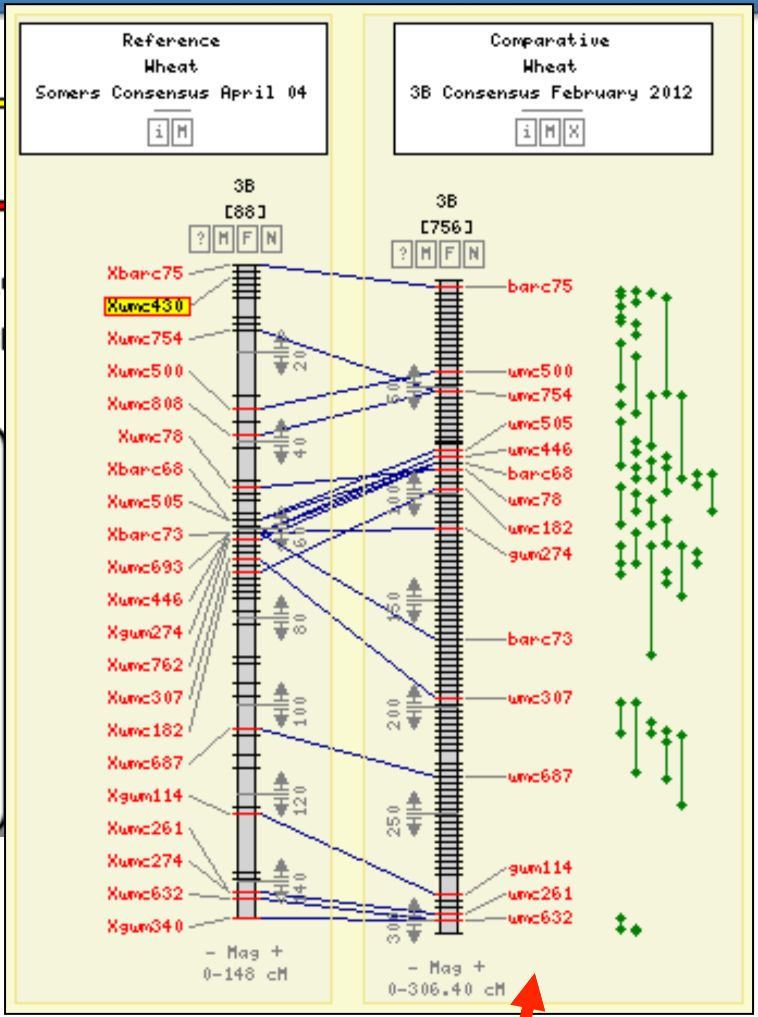
Reference: ctg6

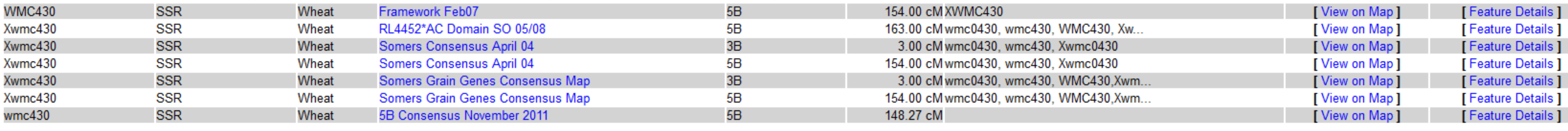
Zoom to this feature

Detailed Report

Link to GnpMap

external link to CMap (CCG Murdoch)





WMC430	SSR	Wheat	Framework Feb07	5B	154.00 cM	XWMC430	[View on Map]	[Feature Details]
Xwmc430	SSR	Wheat	RL4452*AC Domain SO 05/08	5B	163.00 cM	wmc0430, wmc430, WMC430, Xw...	[View on Map]	[Feature Details]
Xwmc430	SSR	Wheat	Somers Consensus April 04	3B	3.00 cM	wmc0430, wmc430, Xwmc0430	[View on Map]	[Feature Details]
Xwmc430	SSR	Wheat	Somers Consensus April 04	5B	154.00 cM	wmc0430, wmc430, Xwmc0430	[View on Map]	[Feature Details]
Xwmc430	SSR	Wheat	Somers Grain Genes Consensus Map	3B	3.00 cM	wmc0430, wmc430, WMC430,Xwm...	[View on Map]	[Feature Details]
Xwmc430	SSR	Wheat	Somers Grain Genes Consensus Map	5B	154.00 cM	wmc0430, wmc430, WMC430,Xwm...	[View on Map]	[Feature Details]
wmc430	SSR	Wheat	5B Consensus November 2011	5B	148.27 cM		[View on Map]	[Feature Details]

Galaxy@URGI

joelle.amselem@versailles.inra.fr

Galaxy @URGI

https://urgi.versailles.inra.fr/galaxy2

Welcome to URGI server

- This Galaxy server has 3 home-made toolbox installed: the URGI Maphits toolbox for SNP polymorphisms detection, the URGI Aplibio DEA toolbox for RNASeq differential analysis and S-Mart toolbox for RNASeq and ChipSeq analysis. It has also 2 workflows installed one for MapHits and one for DEA.
- To launch theses workflows respectively follow the tutorials:
- MAPHiTS: Mapping Analysis Pipeline for High-Throughput Sequences.
- Go to Shared Data -> Published Pages -> MAPHITS Tutorial.
- DEA: Differential Expression Analysis.
- Got to Shared Data -> Published Pages -> DEA tutorial.
- SMART Documentation is here
- You can also download data from GnpIS URGI information system marts by selecting Get Data -> Biomart INRA URGI GnpIS

- Before running workflows and Ad-Hoc tutorials, you need to be registered as URGI user.

- URGI accounts information is here
- Tools in the left panel can be used without registration (a storage quota up to 2Go of data is allowed)
- For any questions please contact: urgi-support at versailles.inra.fr

intensive biomedical research. The Galaxy team is a part of BX
at Emory University. The Galaxy Project is supported in part b
enn State, and Emory University.

getData from URGI BioMart

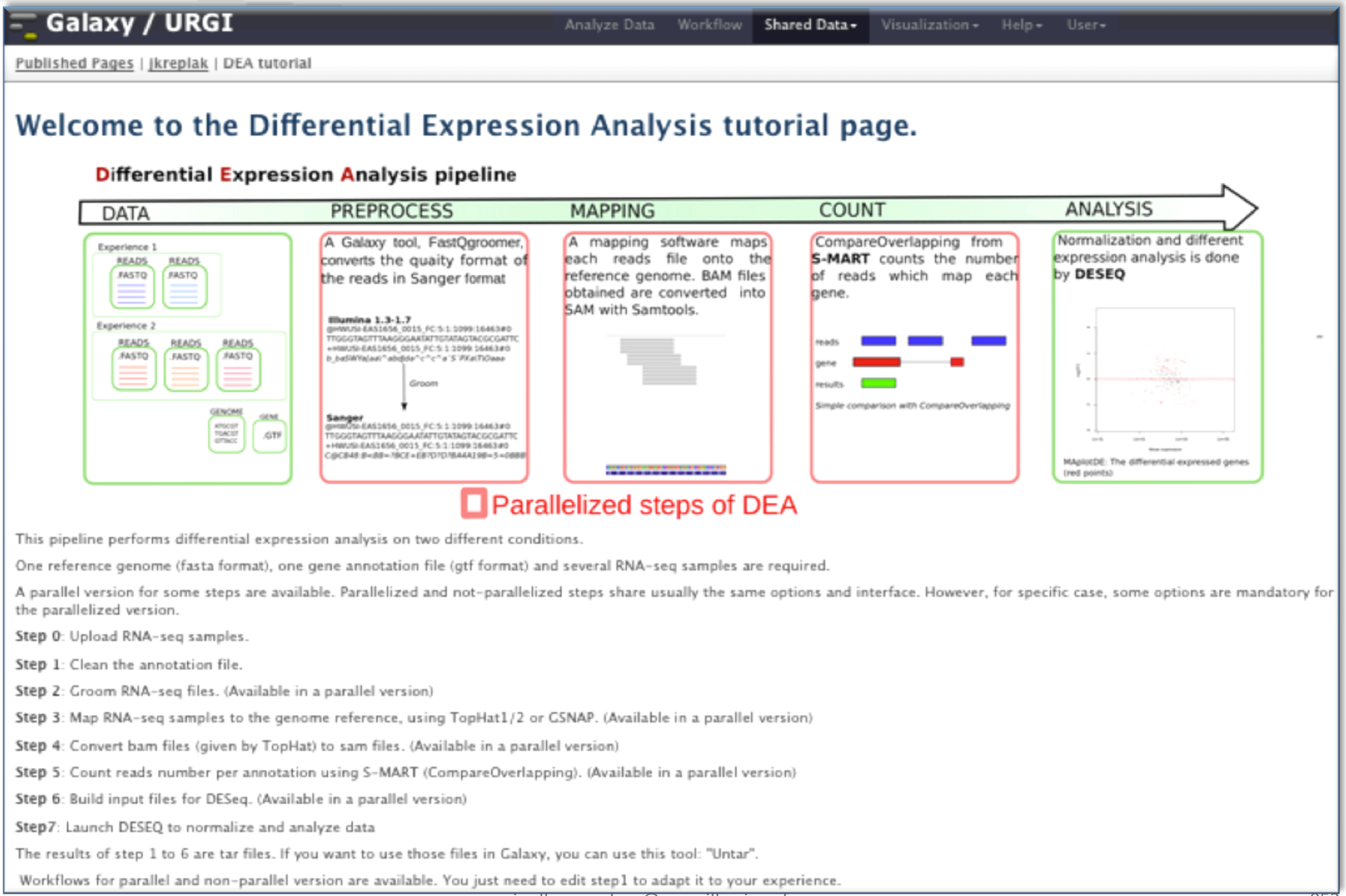
Pipelines: MAPHiTS, DEA

S-MART toolbox

joelle.amselem@versailles.inra.fr

Galaxy workflows@URGI

DEA



Welcome to the Differential Expression Analysis tutorial page.

Differential Expression Analysis pipeline

A Galaxy tool, FastQgroomer, converts the quaity format of the reads in Sanger format

A mapping software maps each reads file onto the reference genome. BAM files obtained are converted into SAM with Samtools.

CompareOverlapping from **S-MART** counts the number of reads which map each gene.

Normalization and different expression analysis is done by **DESEQ**

🟥 Parallelized steps of DEA

This pipeline performs differential expression analysis on two different conditions.

One reference genome (fasta format), one gene annotation file (gtf format) and several RNA-seq samples are required.

A parallel version for some steps are available. Parallelized and not-parallelized steps share usually the same options and interface. However, for specific case, some options are mandatory for the parallelized version.

Step 0: Upload RNA-seq samples.

Step 1: Clean the annotation file.

Step 2: Groom RNA-seq files. (Available in a parallel version)

Step 3: Map RNA-seq samples to the genome reference, using TopHat1/2 or GSNAP. (Available in a parallel version)

Step 4: Convert bam files (given by TopHat) to sam files. (Available in a parallel version)

Step 5: Count reads number per annotation using S-MART (CompareOverlapping). (Available in a parallel version)

Step 6: Build input files for DESeq. (Available in a parallel version)

Step7: Launch DESEQ to normalize and analyze data

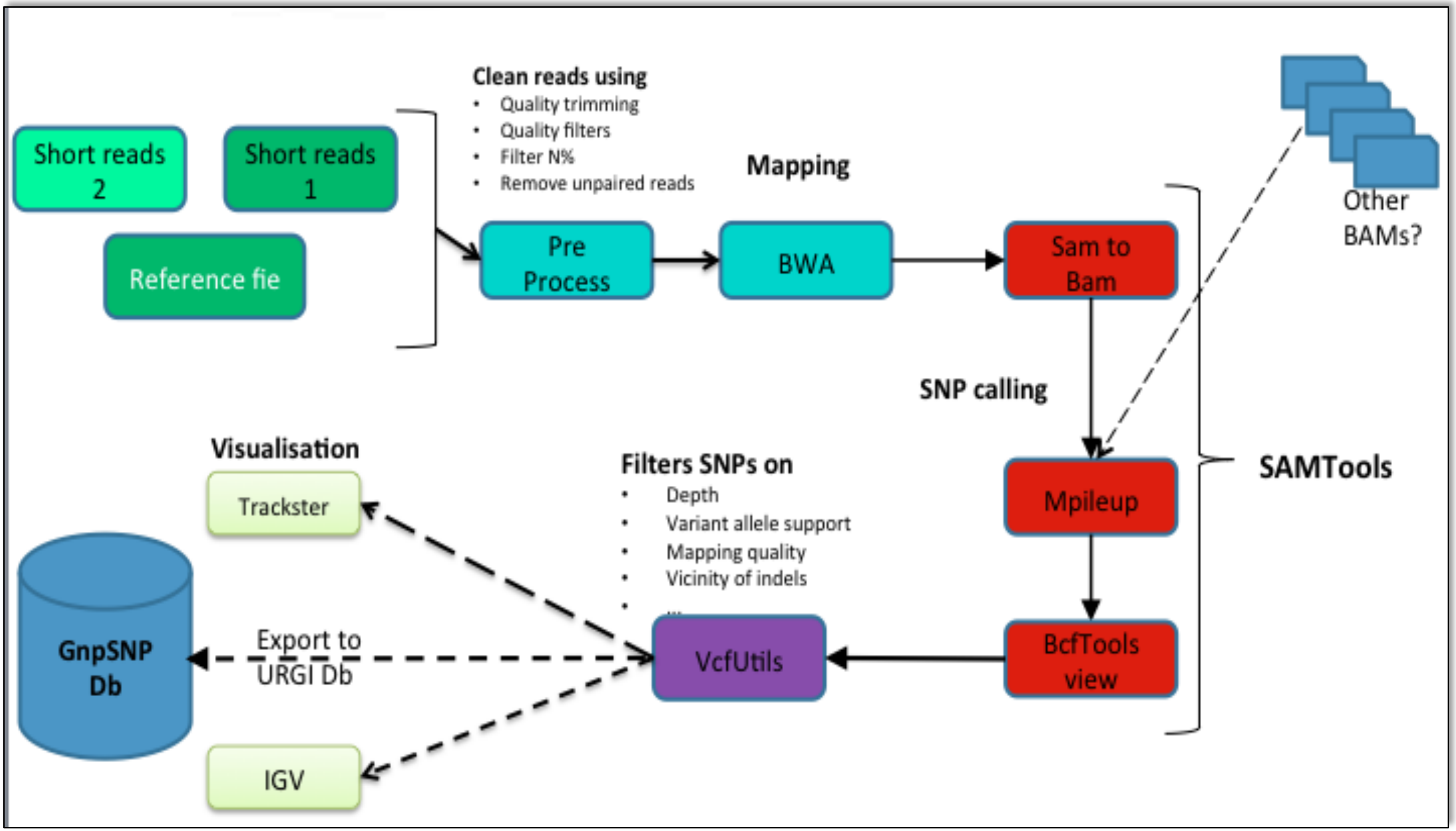
The results of step 1 to 6 are tar files. If you want to use those files in Galaxy, you can use this tool: "Untar".

Workflows for parallel and non-parallel version are available. You just need to edit step1 to adapt it to your experience.

joelle.amselem@versailles.inra.fr

Galaxy workflows@URGI

MAPHiTS: Mapping Analysis Pipeline for High-Throughput Sequences



Clean reads using

- Quality trimming
- Quality filters
- Filter N%
- Remove unpaired reads

Mapping

SNP calling

SAMTools

Visualisation

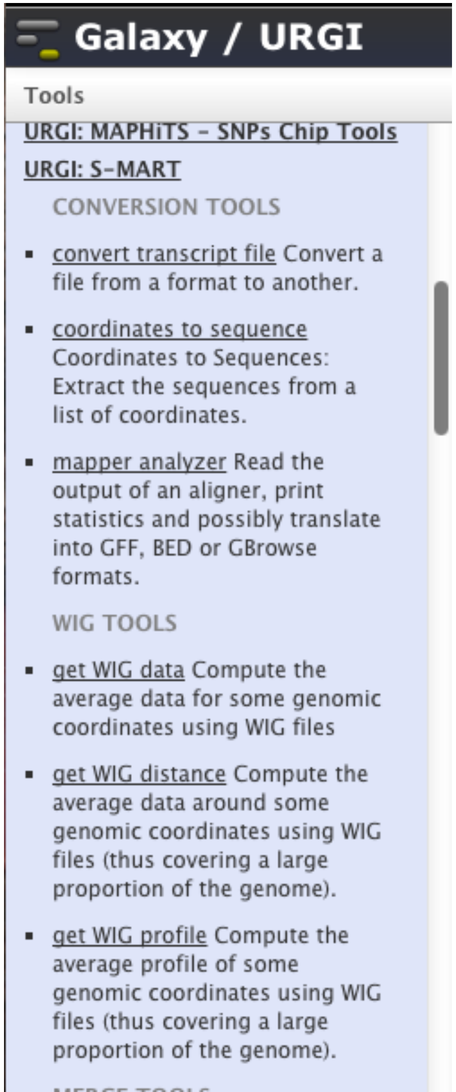
Filters SNPs on

- Depth
- Variant allele support
- Mapping quality
- Vicinity of indels
- …

joelle.amselem@versailles.inra.fr

S-MART a toolbox to convert, merge, select, compare and visualize data

Zytnicki and Quesneville, PLoS One, 2011



Galaxy / URGI

Tools

URGI: MAPHiTS - SNPs Chip Tools

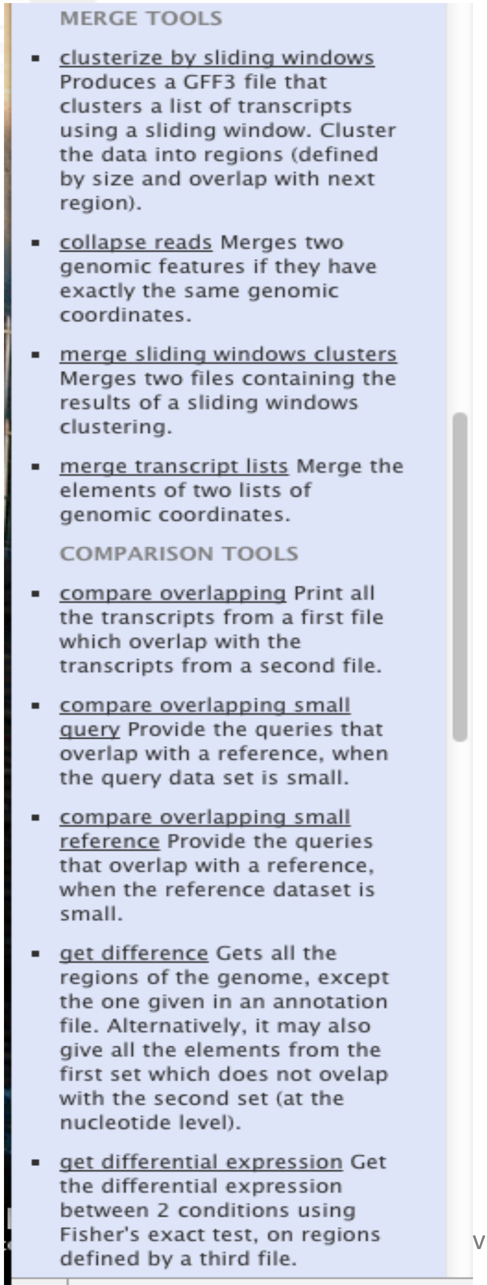
URGI: S-MART

CONVERSION TOOLS

- convert transcript file Convert a file from a format to another.
- coordinates to sequence Coordinates to Sequences: Extract the sequences from a list of coordinates.
- mapper analyzer Read the output of an aligner, print statistics and possibly translate into GFF, BED or GBrowse formats.

WIG TOOLS

- get WIG data Compute the average data for some genomic coordinates using WIG files
- get WIG distance Compute the average data around some genomic coordinates using WIG files (thus covering a large proportion of the genome).
- get WIG profile Compute the average profile of some genomic coordinates using WIG files (thus covering a large proportion of the genome).

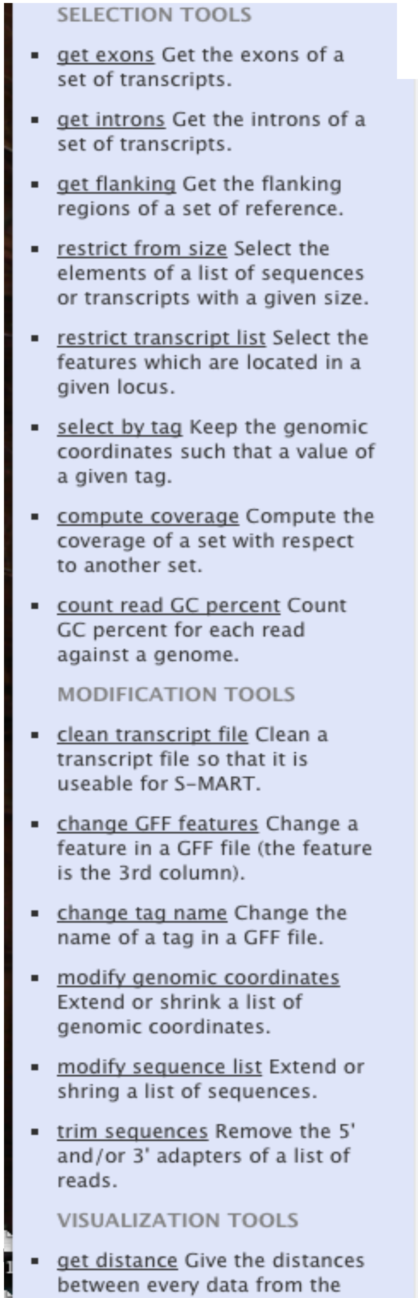


MERGE TOOLS

- clusterize by sliding windows Produces a GFF3 file that clusters a list of transcripts using a sliding window. Cluster the data into regions (defined by size and overlap with next region).
- collapse reads Merges two genomic features if they have exactly the same genomic coordinates.
- merge sliding windows clusters Merges two files containing the results of a sliding windows clustering.
- merge transcript lists Merge the elements of two lists of genomic coordinates.

COMPARISON TOOLS

- compare overlapping Print all the transcripts from a first file which overlap with the transcripts from a second file.
- compare overlapping small query Provide the queries that overlap with a reference, when the query data set is small.
- compare overlapping small reference Provide the queries that overlap with a reference, when the reference dataset is small.
- get difference Gets all the regions of the genome, except the one given in an annotation file. Alternatively, it may also give all the elements from the first set which does not ovelap with the second set (at the nucleotide level).
- get differential expression Get the differential expression between 2 conditions using Fisher's exact test, on regions defined by a third file.



SELECTION TOOLS

- get exons Get the exons of a set of transcripts.
- get introns Get the introns of a set of transcripts.
- get flanking Get the flanking regions of a set of reference.
- restrict from size Select the elements of a list of sequences or transcripts with a given size.
- restrict transcript list Select the features which are located in a given locus.
- select by tag Keep the genomic coordinates such that a value of a given tag.
- compute coverage Compute the coverage of a set with respect to another set.
- count read GC percent Count GC percent for each read against a genome.

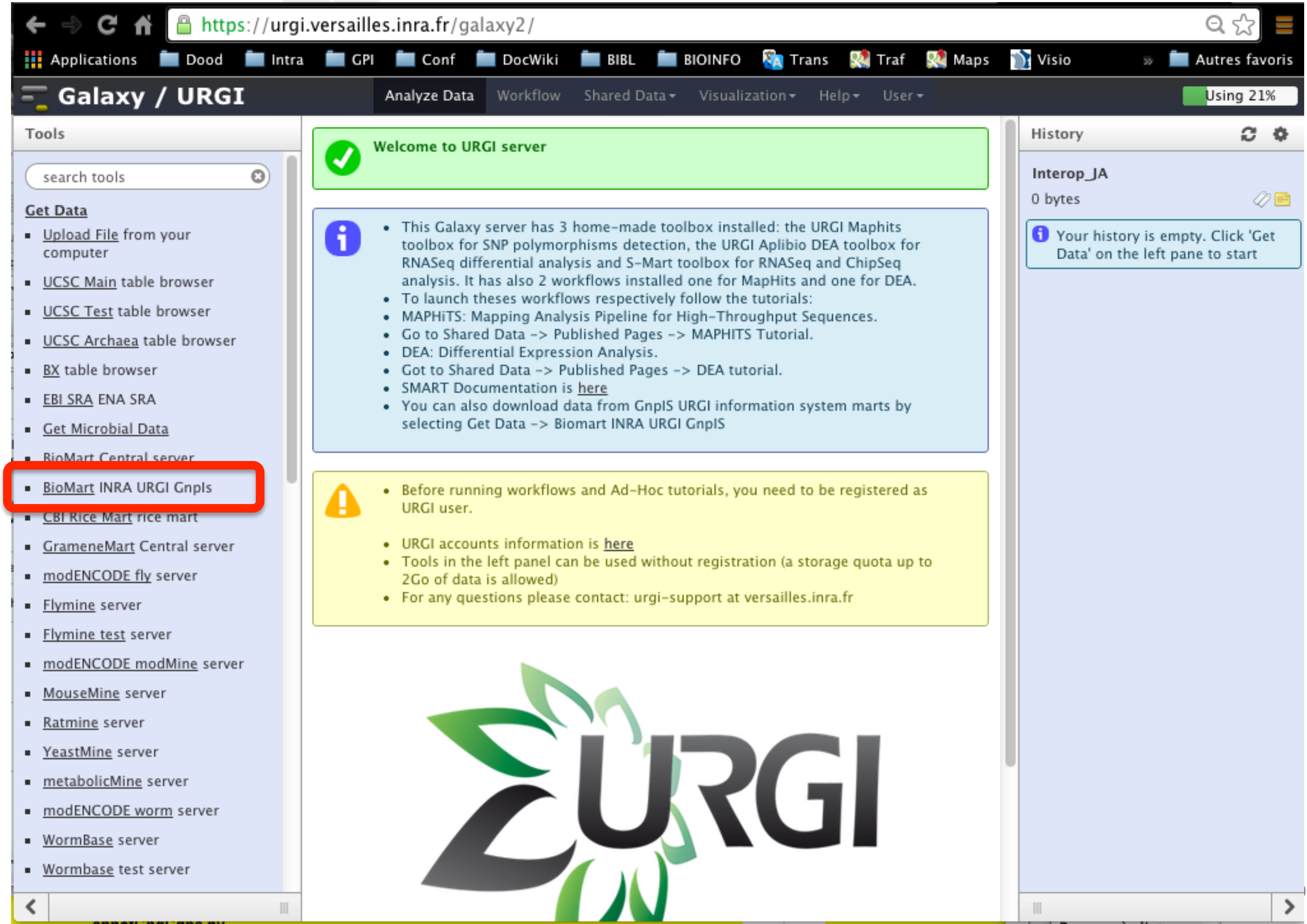
MODIFICATION TOOLS

- clean transcript file Clean a transcript file so that it is useable for S-MART.
- change GFF features Change a feature in a GFF file (the feature is the 3rd column).
- change tag name Change the name of a tag in a GFF file.
- modify genomic coordinates Extend or shrink a list of genomic coordinates.
- modify sequence list Extend or shring a list of sequences.
- trim sequences Remove the 5' and/or 3' adapters of a list of reads.

VISUALIZATION TOOLS

- get distance Give the distances between every data from the

Galaxy communication with BioMart



Welcome to URGI server

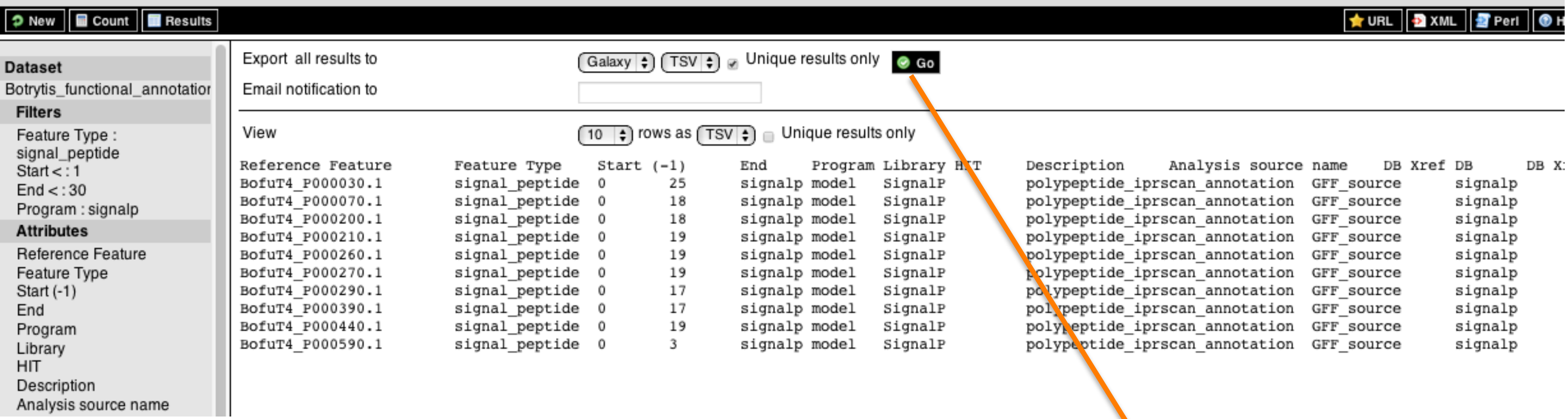
- This Galaxy server has 3 home-made toolbox installed: the URGI Maphits toolbox for SNP polymorphisms detection, the URGI Aplibio DEA toolbox for RNASeq differential analysis and S-Mart toolbox for RNASeq and ChipSeq analysis. It has also 2 workflows installed one for MapHits and one for DEA.
- To launch theses workflows respectively follow the tutorials:
- MAPHiTS: Mapping Analysis Pipeline for High-Throughput Sequences.
- Go to Shared Data -> Published Pages -> MAPHITS Tutorial.
- DEA: Differential Expression Analysis.
- Got to Shared Data -> Published Pages -> DEA tutorial.
- SMART Documentation is here
- You can also download data from GnpIS URGI information system marts by selecting Get Data -> Biomart INRA URGI GnpIS

- Before running workflows and Ad-Hoc tutorials, you need to be registered as URGI user.

- URGI accounts information is here
- Tools in the left panel can be used without registration (a storage quota up to 2Go of data is allowed)
- For any questions please contact: urgi-support at versailles.inra.fr

Your history is empty. Click 'Get Data' on the left pane to start

Biomart result: gene with signal peptide (putative secreted proteins)



Dataset
Botrytis_functional_annotatior

Filters
Feature Type : signal_peptide
Start < : 1
End < : 30
Program : signalp

Attributes
Reference Feature
Feature Type
Start (-1)
End
Program
Library
HIT
Description
Analysis source name

Export all results to Galaxy TSV ☑ Unique results only Go
Email notification to

View 10 rows as TSV ☐ Unique results only

```
Reference Feature     Feature Type   Start (-1)     End      Program Library HIT     Description      Analysis source name     DB Xref DB      DB X:
BofuT4_P000030.1      signal_peptide  0        25     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000070.1      signal_peptide  0        18     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000200.1      signal_peptide  0        18     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000210.1      signal_peptide  0        19     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000260.1      signal_peptide  0        19     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000270.1      signal_peptide  0        19     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000290.1      signal_peptide  0        17     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000390.1      signal_peptide  0        17     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000440.1      signal_peptide  0        19     signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
BofuT4_P000590.1      signal_peptide  0        3      signalp model  SignalP         polypeptide_iprscan_annotation  GFF_source       signalp
```

Welcome to URGI server

- This Galaxy server has 3 home-made toolbox installed: the URGI Maphits toolbox for SNP polymorphisms detection, the URGI Aplibio DEA toolbox for RNASeq differential analysis and S-Mart toolbox for RNASeq and ChipSeq analysis. It has also 2 workflows installed one for MapHits and one for DEA.
- To launch theses workflows respectively follow the tutorials:
- MAPHiTS: Mapping Analysis Pipeline for High-Throughput Sequences.
- Go to Shared Data -> Published Pages -> MAPHITS Tutorial.
- DEA: Differential Expression Analysis.
- Got to Shared Data -> Published Pages -> DEA tutorial.
- SMART Documentation is here
- You can also download data from GnpIS URGI information system marts by selecting Get Data -> Biomart INRA URGI GnpIS

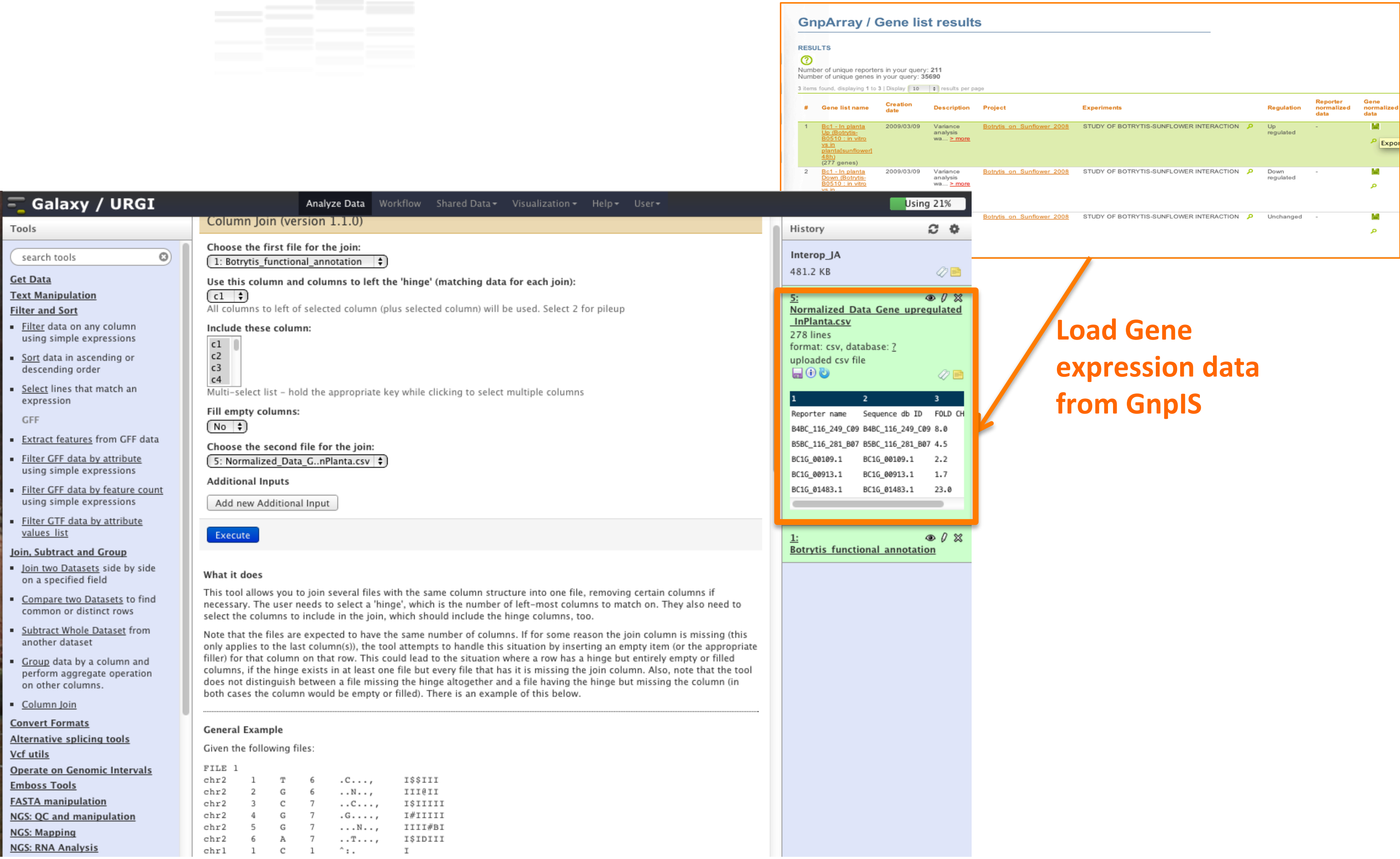
- Before running workflows and Ad-Hoc tutorials, you need to be registered as URGI user.

- URGI accounts information is here
- Tools in the left panel can be used without registration (a storage quota up to 2Go of data is allowed)
- For any questions please contact: urgi-support at versailles.inra.fr

Interop_JA
0 bytes

1: Botrytis_functional_annotation%20

Load Gene expression data from GnpIS transcriptome



GnpArray / Gene list results

RESULTS

Number of unique reporters in your query: **211**
Number of unique genes in your query: **35690**

3 items found, displaying 1 to 3 | Display 10 results per page

#	Gene list name	Creation date	Description	Project	Experiments	Regulation	Reporter normalized data	Gene normalized data
1	Bc1 - In planta Up (Botrytis-B0510 : in vitro vs in planta[sunflower] 48h) (277 genes)	2009/03/09	Variance analysis wa... > more	Botrytis_on_Sunflower_2008	STUDY OF BOTRYTIS-SUNFLOWER INTERACTION	Up regulated	-	
2	Bc1 - In planta Down (Botrytis-B0510 : in vitro vs in	2009/03/09	Variance analysis wa... > more	Botrytis_on_Sunflower_2008	STUDY OF BOTRYTIS-SUNFLOWER INTERACTION	Down regulated	-	
				Botrytis_on_Sunflower_2008	STUDY OF BOTRYTIS-SUNFLOWER INTERACTION	Unchanged	-	

Column Join (version 1.1.0)

Choose the first file for the join:
1: Botrytis_functional_annotation

Use this column and columns to left the 'hinge' (matching data for each join):
c1
All columns to left of selected column (plus selected column) will be used. Select 2 for pileup

Include these column:
c1
c2
c3
c4
Multi-select list - hold the appropriate key while clicking to select multiple columns

Fill empty columns:
No

Choose the second file for the join:
5: Normalized_Data_G..nPlanta.csv

Additional Inputs

Add new Additional Input

Execute

What it does

This tool allows you to join several files with the same column structure into one file, removing certain columns if necessary. The user needs to select a 'hinge', which is the number of left-most columns to match on. They also need to select the columns to include in the join, which should include the hinge columns, too.

Note that the files are expected to have the same number of columns. If for some reason the join column is missing (this only applies to the last column(s)), the tool attempts to handle this situation by inserting an empty item (or the appropriate filler) for that column on that row. This could lead to the situation where a row has a hinge but entirely empty or filled columns, if the hinge exists in at least one file but every file that has it is missing the join column. Also, note that the tool does not distinguish between a file missing the hinge altogether and a file having the hinge but missing the column (in both cases the column would be empty or filled). There is an example of this below.

General Example

Given the following files:

```
FILE 1
chr2    1    T    6    .C...,     I$$III
chr2    2    G    6    ..N..,     III@II
chr2    3    C    7    ..C...,    I$IIIII
chr2    4    G    7    .G....,    I#IIIII
chr2    5    G    7    ...N..,    IIII#BI
chr2    6    A    7    ..T...,    I$IDIII
chr1    1    C    1    ^:.        I
```

History

Interop_JA
481.2 KB

5: Normalized_Data_Gene_upregulated_InPlanta.csv
278 lines
format: csv, database: ?
uploaded csv file

1	2	3
Reporter name	Sequence db ID	FOLD CH
B4BC_116_249_C09	B4BC_116_249_C09	8.0
B5BC_116_281_B07	B5BC_116_281_B07	4.5
BC1G_00109.1	BC1G_00109.1	2.2
BC1G_00913.1	BC1G_00913.1	1.7
BC1G_01483.1	BC1G_01483.1	23.0

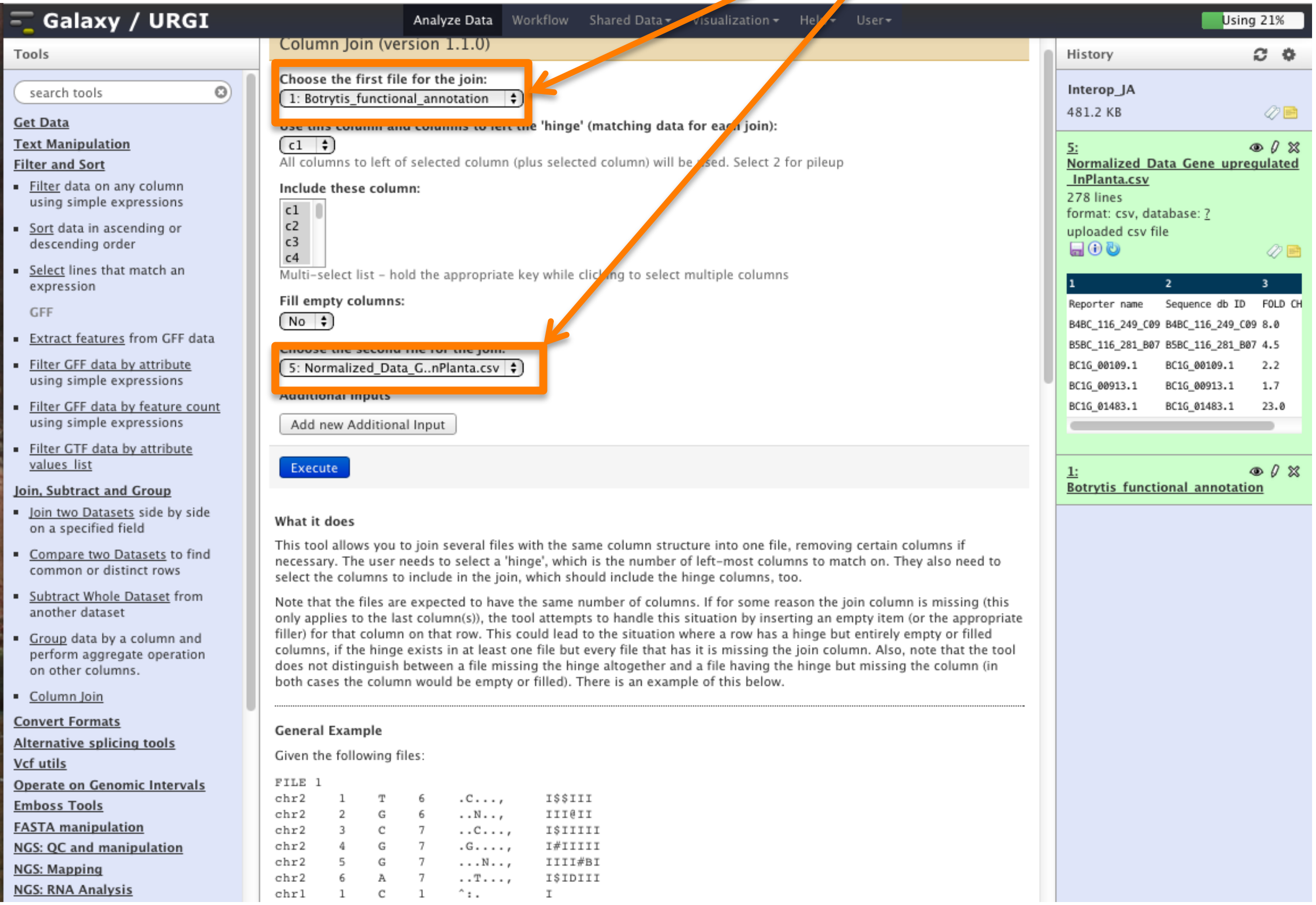
1: Botrytis_functional_annotation

Load Gene expression data from GnpIS

joelle.amselem@versailles.inra.fr

Biomart result: functional annotation with gene expression data

Join on Gene name column



Get Data

Text Manipulation

Filter and Sort

- Filter data on any column using simple expressions
- Sort data in ascending or descending order
- Select lines that match an expression

GFF

- Extract features from GFF data
- Filter GFF data by attribute using simple expressions
- Filter GFF data by feature count using simple expressions
- Filter GTF data by attribute values_list

Join, Subtract and Group

- Join two Datasets side by side on a specified field
- Compare two Datasets to find common or distinct rows
- Subtract Whole Dataset from another dataset
- Group data by a column and perform aggregate operation on other columns.
- Column Join

Convert Formats

Alternative splicing tools

Vcf utils

Operate on Genomic Intervals

Emboss Tools

FASTA manipulation

NGS: QC and manipulation

NGS: Mapping

NGS: RNA Analysis

Column Join (version 1.1.0)

Choose the first file for the join: 1: Botrytis_functional_annotation

Use this column and columns to left the 'hinge' (matching data for each join): c1

All columns to left of selected column (plus selected column) will be used. Select 2 for pileup

Include these column: c1 c2 c3 c4

Multi-select list - hold the appropriate key while clicking to select multiple columns

Fill empty columns: No

Choose the second file for the join: 5: Normalized_Data_G..nPlanta.csv

Additional Inputs

Add new Additional Input

Execute

What it does

This tool allows you to join several files with the same column structure into one file, removing certain columns if necessary. The user needs to select a 'hinge', which is the number of left-most columns to match on. They also need to select the columns to include in the join, which should include the hinge columns, too.

Note that the files are expected to have the same number of columns. If for some reason the join column is missing (this only applies to the last column(s)), the tool attempts to handle this situation by inserting an empty item (or the appropriate filler) for that column on that row. This could lead to the situation where a row has a hinge but entirely empty or filled columns, if the hinge exists in at least one file but every file that has it is missing the join column. Also, note that the tool does not distinguish between a file missing the hinge altogether and a file having the hinge but missing the column (in both cases the column would be empty or filled). There is an example of this below.

General Example

Given the following files:

```
FILE 1
chr2    1    T    6    .C...,      I$$III
chr2    2    G    6    ..N..,      III@II
chr2    3    C    7    ..C...,     I$IIIII
chr2    4    G    7    .G....,     I#IIIII
chr2    5    G    7    ...N..,     IIII#BI
chr2    6    A    7    ..T...,     I$IDIII
chr1    1    C    1    ^:.         I
```

Interop_JA

481.2 KB

5: Normalized_Data_Gene_upregulated_InPlanta.csv

278 lines

format: csv, database: ?

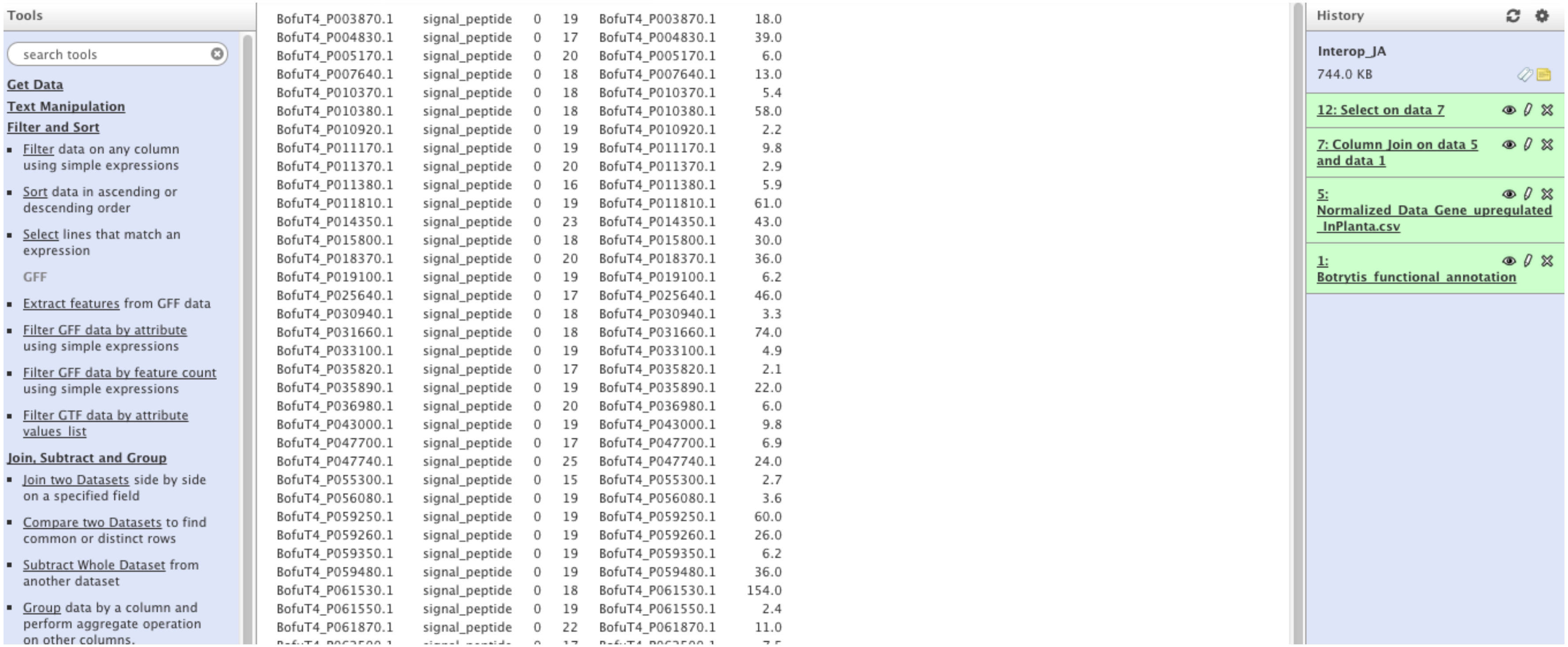
uploaded csv file

1	2	3
Reporter name	Sequence db ID	FOLD CH
B4BC_116_249_C09	B4BC_116_249_C09	8.0
B5BC_116_281_B07	B5BC_116_281_B07	4.5
BC1G_00109.1	BC1G_00109.1	2.2
BC1G_00913.1	BC1G_00913.1	1.7
BC1G_01483.1	BC1G_01483.1	23.0

1: Botrytis_functional_annotation

joelle.amselem@versailles.inra.fr

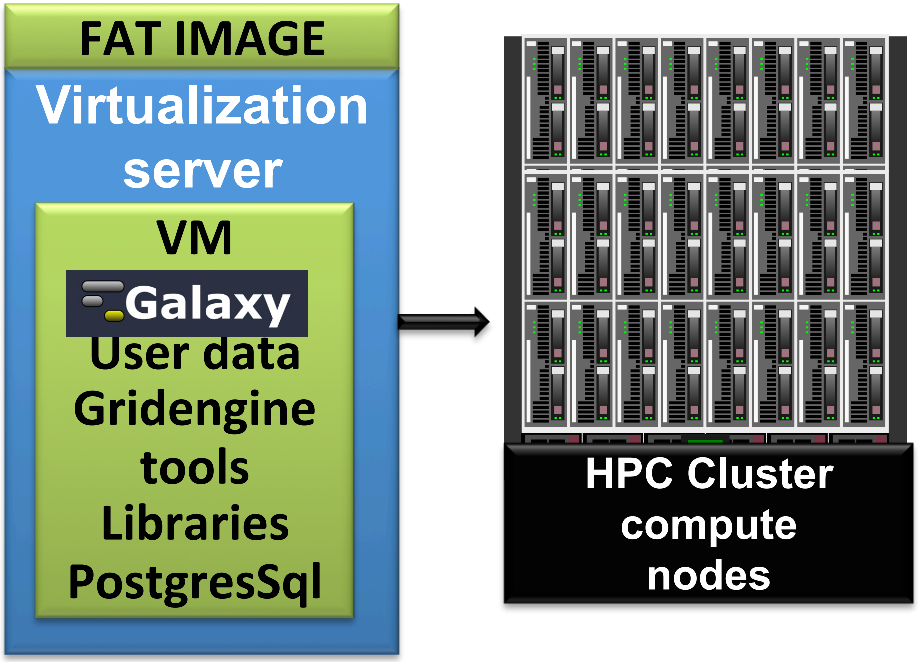
Result: expression level for genes with a signal peptide



```
BofuT4_P003870.1   signal_peptide   0   19   BofuT4_P003870.1    18.0
BofuT4_P004830.1   signal_peptide   0   17   BofuT4_P004830.1    39.0
BofuT4_P005170.1   signal_peptide   0   20   BofuT4_P005170.1     6.0
BofuT4_P007640.1   signal_peptide   0   18   BofuT4_P007640.1    13.0
BofuT4_P010370.1   signal_peptide   0   18   BofuT4_P010370.1     5.4
BofuT4_P010380.1   signal_peptide   0   18   BofuT4_P010380.1    58.0
BofuT4_P010920.1   signal_peptide   0   19   BofuT4_P010920.1     2.2
BofuT4_P011170.1   signal_peptide   0   19   BofuT4_P011170.1     9.8
BofuT4_P011370.1   signal_peptide   0   20   BofuT4_P011370.1     2.9
BofuT4_P011380.1   signal_peptide   0   16   BofuT4_P011380.1     5.9
BofuT4_P011810.1   signal_peptide   0   19   BofuT4_P011810.1    61.0
BofuT4_P014350.1   signal_peptide   0   23   BofuT4_P014350.1    43.0
BofuT4_P015800.1   signal_peptide   0   18   BofuT4_P015800.1    30.0
BofuT4_P018370.1   signal_peptide   0   20   BofuT4_P018370.1    36.0
BofuT4_P019100.1   signal_peptide   0   19   BofuT4_P019100.1     6.2
BofuT4_P025640.1   signal_peptide   0   17   BofuT4_P025640.1    46.0
BofuT4_P030940.1   signal_peptide   0   18   BofuT4_P030940.1     3.3
BofuT4_P031660.1   signal_peptide   0   18   BofuT4_P031660.1    74.0
BofuT4_P033100.1   signal_peptide   0   19   BofuT4_P033100.1     4.9
BofuT4_P035820.1   signal_peptide   0   17   BofuT4_P035820.1     2.1
BofuT4_P035890.1   signal_peptide   0   19   BofuT4_P035890.1    22.0
BofuT4_P036980.1   signal_peptide   0   20   BofuT4_P036980.1     6.0
BofuT4_P043000.1   signal_peptide   0   19   BofuT4_P043000.1     9.8
BofuT4_P047700.1   signal_peptide   0   17   BofuT4_P047700.1     6.9
BofuT4_P047740.1   signal_peptide   0   25   BofuT4_P047740.1    24.0
BofuT4_P055300.1   signal_peptide   0   15   BofuT4_P055300.1     2.7
BofuT4_P056080.1   signal_peptide   0   19   BofuT4_P056080.1     3.6
BofuT4_P059250.1   signal_peptide   0   19   BofuT4_P059250.1    60.0
BofuT4_P059260.1   signal_peptide   0   19   BofuT4_P059260.1    26.0
BofuT4_P059350.1   signal_peptide   0   19   BofuT4_P059350.1     6.2
BofuT4_P059480.1   signal_peptide   0   19   BofuT4_P059480.1    36.0
BofuT4_P061530.1   signal_peptide   0   18   BofuT4_P061530.1   154.0
BofuT4_P061550.1   signal_peptide   0   19   BofuT4_P061550.1     2.4
BofuT4_P061870.1   signal_peptide   0   22   BofuT4_P061870.1    11.0
```

joelle.amselem@versailles.inra.fr

The past



FAT IMAGE

Virtualization server

VM

Galaxy

User data

Gridengine

tools

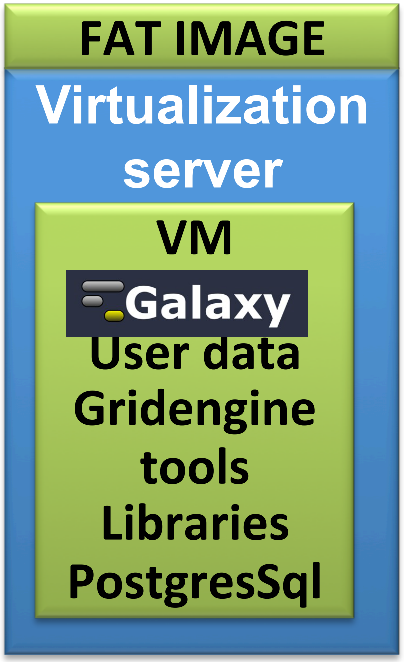
Libraries

PostgresSql

HPC Cluster compute nodes

- A Virtual vs physical instance
  - For easier monitoring
  - To better adapt memory, cpu, storage according to the use

joelle.amselem@versailles.inra.fr


FAT IMAGE

Virtualization server

VM

Galaxy

User data

Gridengine

tools

Libraries

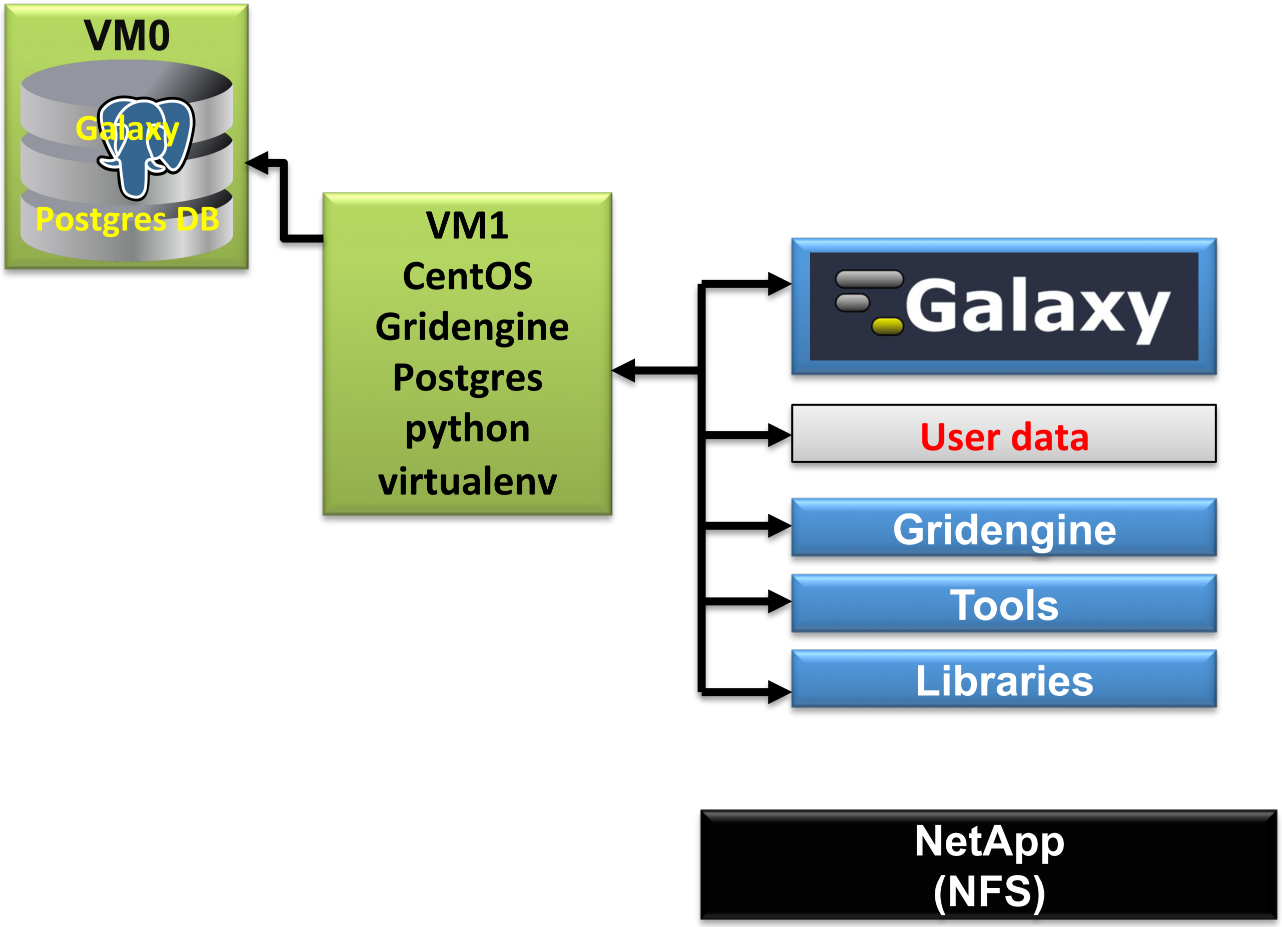
PostgresSql

CONS of FAT image

- At each distribution : an entire VM creation
- Duplicated application, dependencies, ...
- Maintenance issues (backup, size ...)

joelle.amselem@versailles.inra.fr

“SLIM” Image



VM0

Galaxy

Postgres DB

VM1
CentOS
Gridengine
Postgres
python
virtualenv

Galaxy

User data

Gridengine

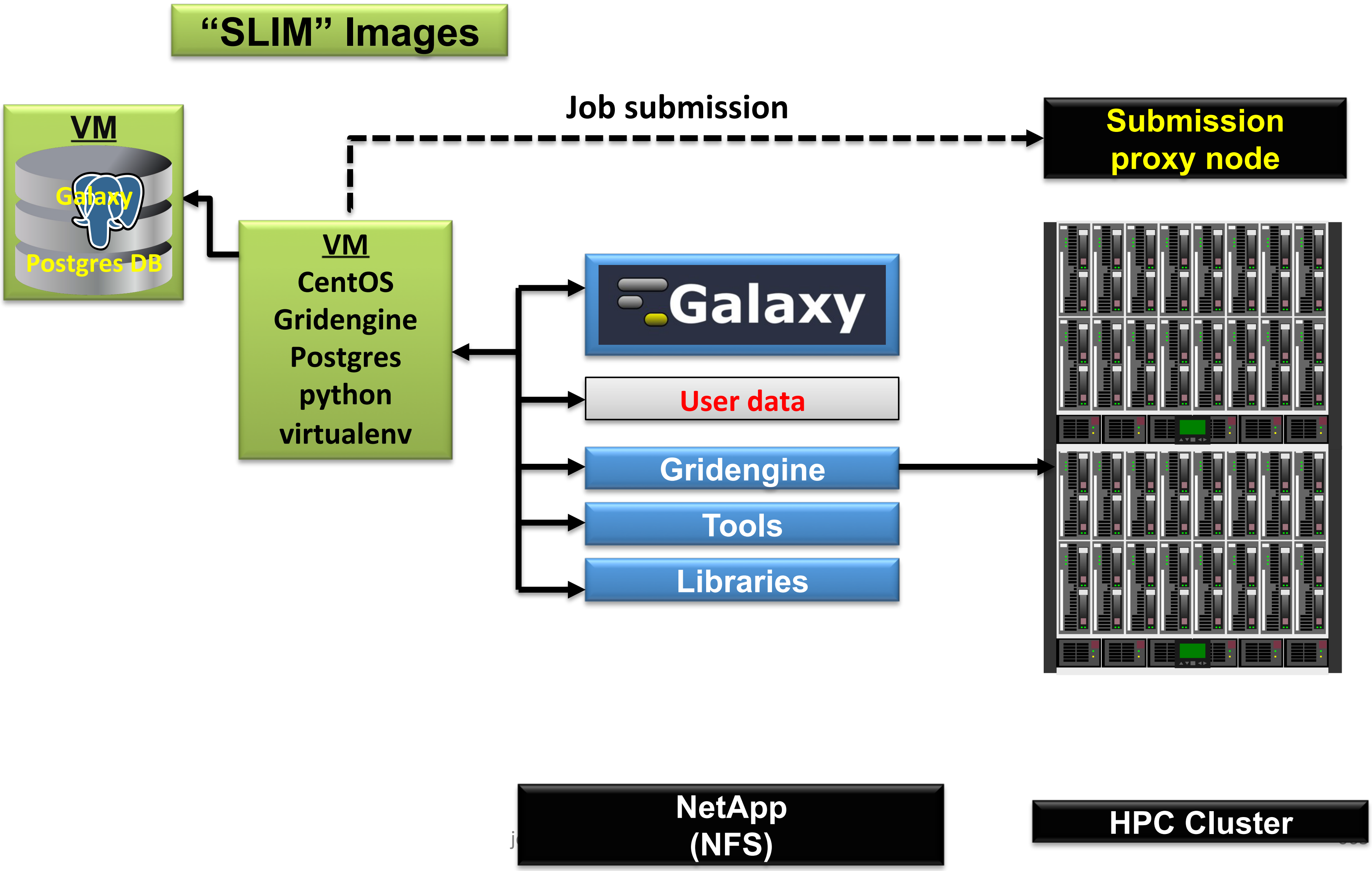
Tools

Libraries

NetApp
(NFS)

joelle.amselem@versailles.inra.fr

Galaxy @URGI



“SLIM” Images

VM

Galaxy

Postgres DB

VM
CentOS
Gridengine
Postgres
python
virtualenv

Job submission

Submission proxy node

User data

Gridengine

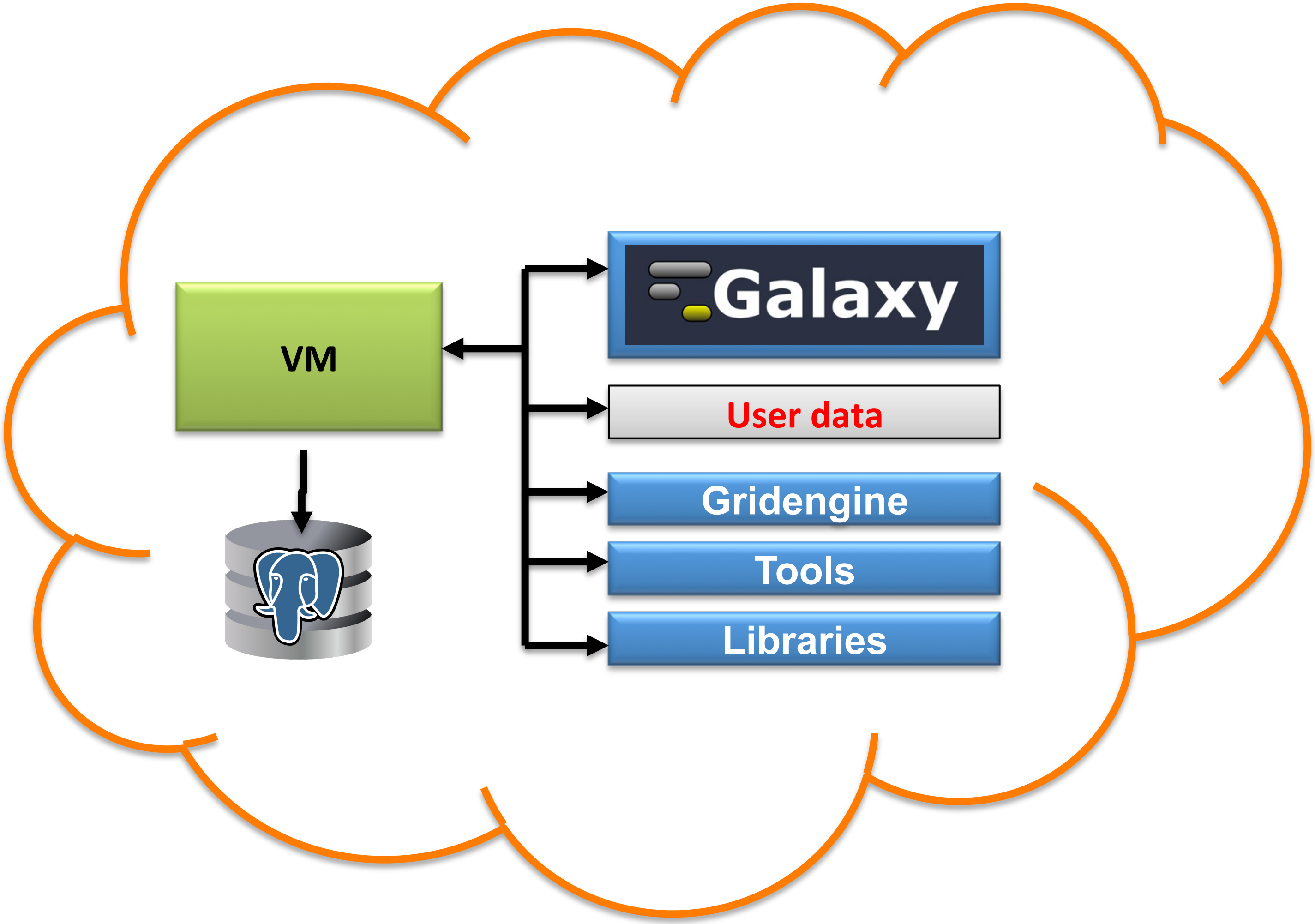
Tools

Libraries

NetApp (NFS)

HPC Cluster

Next step: cloud



VM

Galaxy

User data

Gridengine

Tools

Libraries

joelle.amselem@versailles.inra.fr

Joelle Amselem, Michael Alaux, Nathalie Choisne, Olivier Inizan, Nicolas Lapalu, Nacer Mohellibi, Françoise Alfama, Aminah Keliet, Erik Kimmel, Jonathan Kreplak, Claire Viseux-Guerche, Mikaël Loaec, Thomas Letellier, Yufei Luo, Cyril Pommier, Matthias Zytnicki, Marc-Henri Lebrun, Anne-Françoise Adam Blondon, Delphine Steinbach and Hadi Quesneville


URGI-team

joelle.amselem@versailles.inra.fr