

GnplS: an information system for plant breeding

21th october 2010

Thematic day on Integrative genomics - Nantes

Hadi Quesneville





URGI: Unité de Recherche en Génomique-Ingo

- **Research Unit**

- ◆ INRA unit (French National Institute for Agricultural Research)
- ◆ Plant breeding and Genetics Department
- ◆ Strong connexions with other plant INRA departments

- **Host a Bioinformatic platform**

- ◆ IBISA Grade
- ◆ Member of the French National Network of Bioinformatic Platforms (ReNaBi)

- **Research**

- ◆ Data integration
- ◆ Functional and evolutionary genome dynamics

Databases

- **Central role for data analysis**
 - ◆ Repository
 - ◆ Navigation
- **Link heterogeneous informations**
 - ◆ Experiments: Transcriptomes, EST, SNP, chIP
 - ◆ Genome annotations: genes, repeats
 - ◆ Genetic informations: markers, recombination rates
 - ◆ Lineages: lines, populations, species
 - ◆ ...
- **Performance issues**
 - ◆ Adapted schema for query

Genome data integration



Data integration

- Data integration involves combining data residing in different sources and providing users with a unified view of these data. *(from wikipedia)*

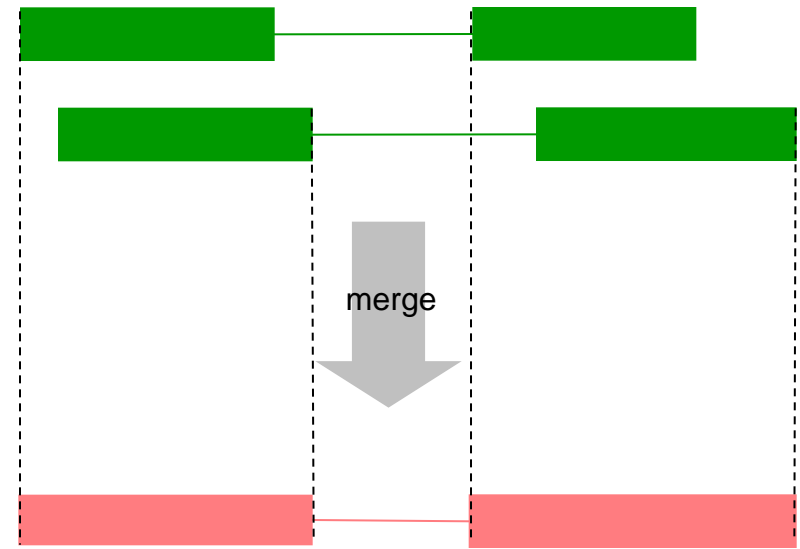
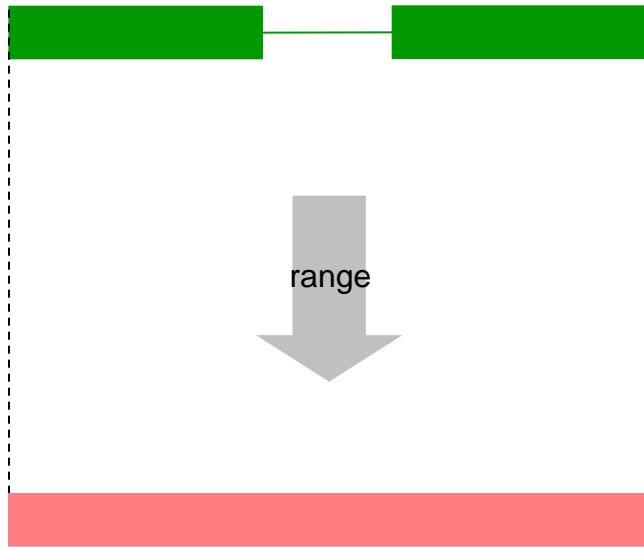


Genome base data integration

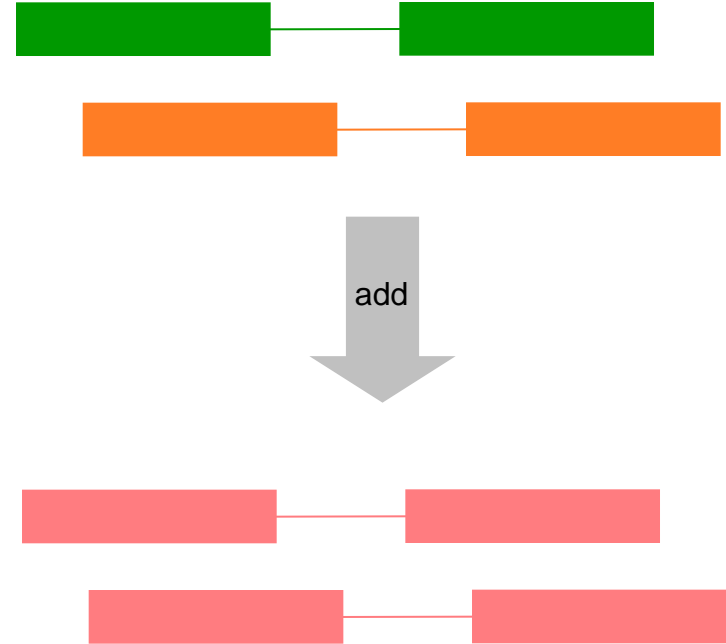
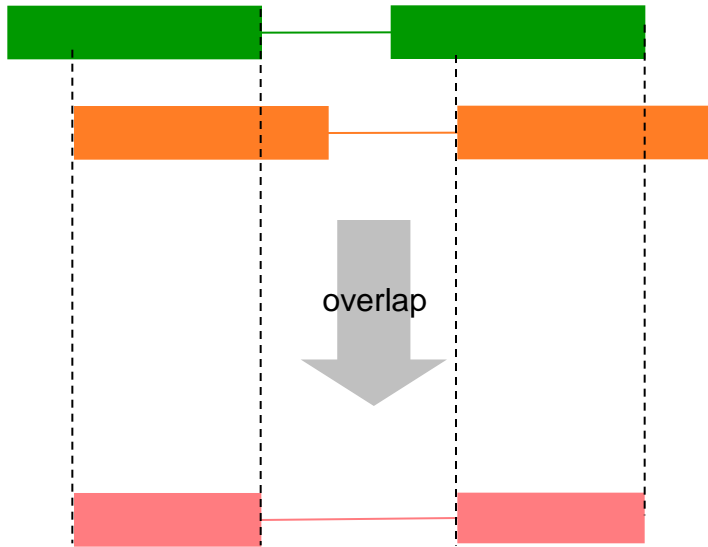
- **Natural way to integrate « omics » data**
 - ◆ Map data to genome sequence
 - ◆ Compare genome coordinates
 - ◆ Identify relationships
 - ◆ Compute correlations

- **Few generic operations needed**
 - ◆ Generic unary and binary operators
 - ◆ Density, counts
 - ◆ Statistics
 - ◆ Visualisations

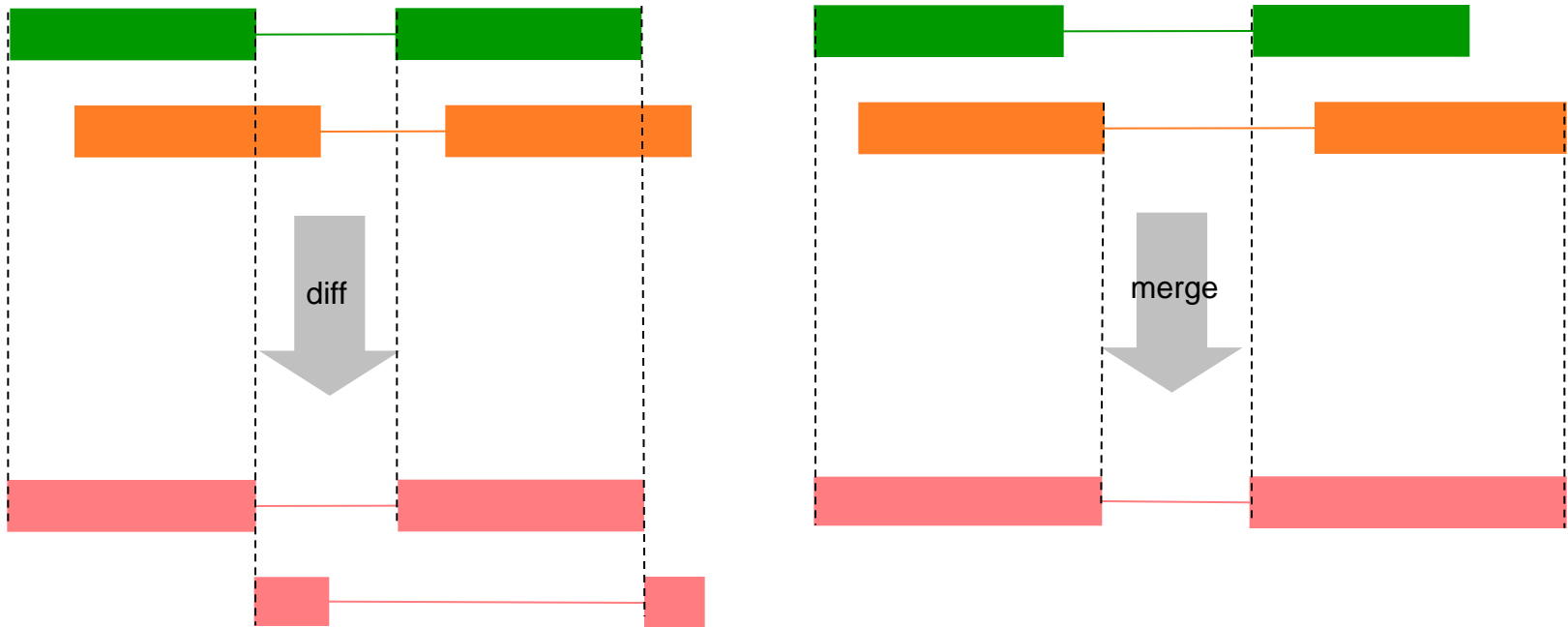
Unary operators (1 source)



Binary operators (2 sources)

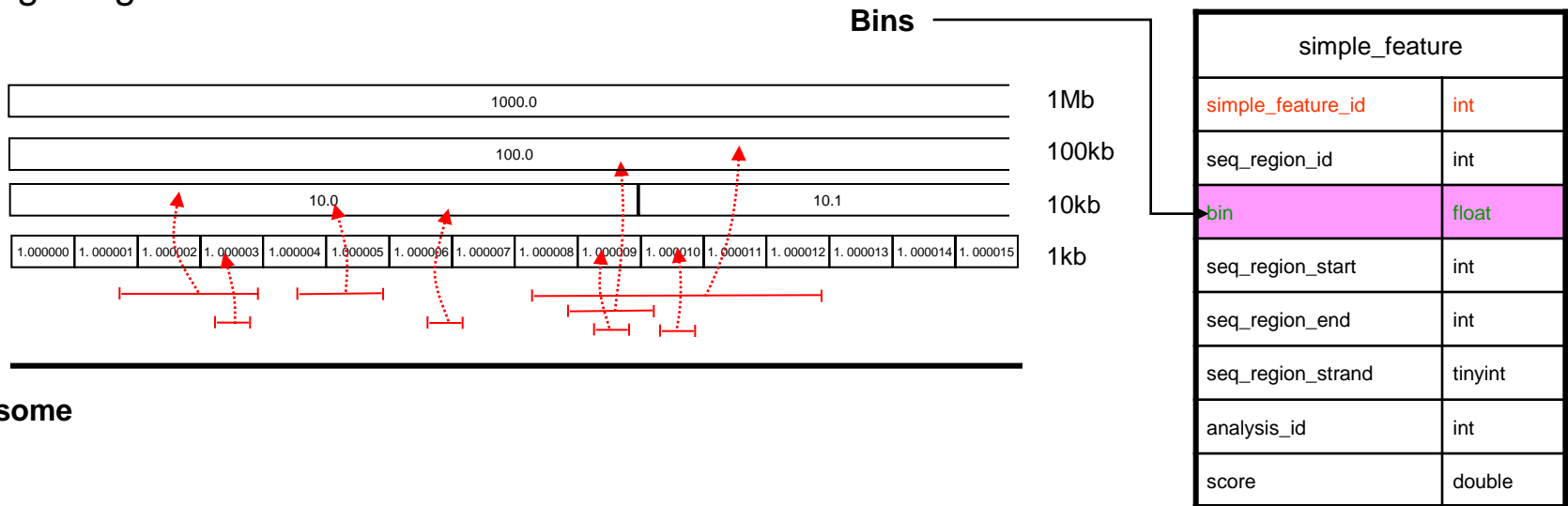


Binary operators (2 sources)

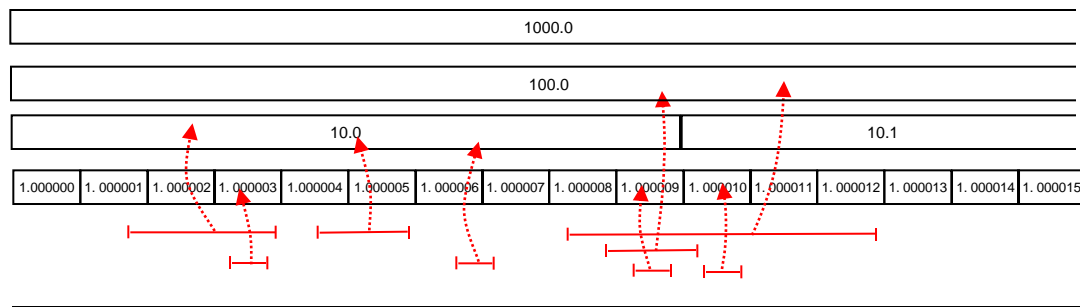


Coordinate indexes

- Inspired from R-tree (Guttman 1984)
- Hierarchical set of bin
- Adjacent bin at a level have adjacent ID
- Bin from different level do not overlap
- Assign segment to smallest bin that contain it



Coordinate indexes



Chromosome

12010000 12020000

1Mb

100kb

10kb

1kb

Bins

Select * from simple_feature
where seq_region_id=1
AND

(
bin = 1000.000012
OR bin = 100.000120
OR bin between 10.001201 and 10.001202
OR bin between 1.012010 and 1.012020
)

Bin level
(x 10')

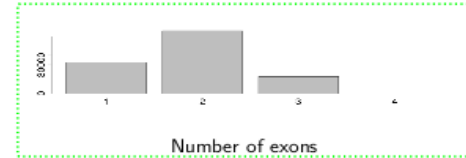
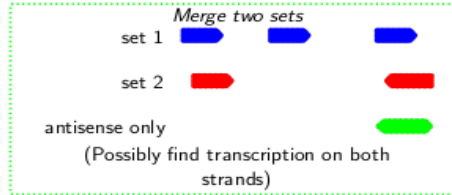
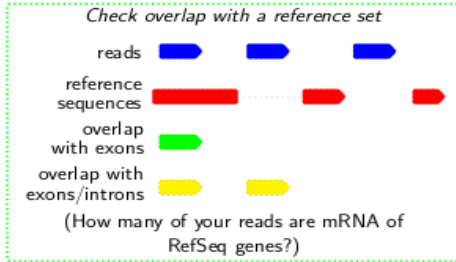
Bin number
(/ 10')

simple_feature	
simple_feature_id	int
seq_region_id	int
bin	float
seq_region_start	int
seq_region_end	int
seq_region_strand	tinyint
analysis_id	int
score	double

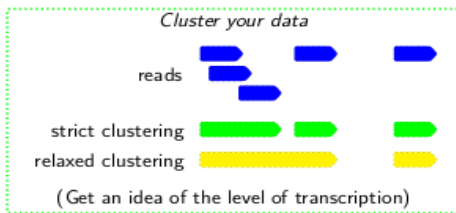
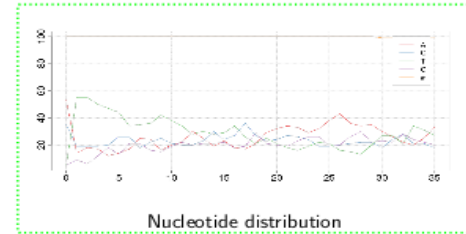
AND seq_region_start <= 12010000
AND seq_region_end >= 12020000

18/11/2010

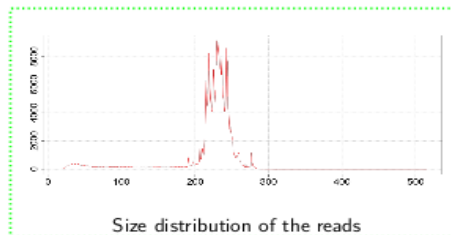
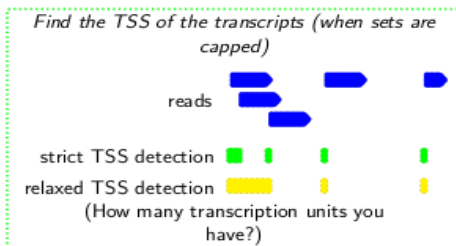
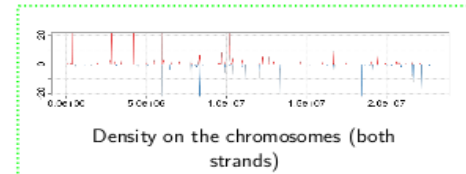
S-MART



← ↑ **Transformation tools**
 S-MART can select, remove, gather all or part of you data, depending on the task you want to perform



S-MART provides you with different graphs in order to control your data or publish them
Visualization tools ↓ →



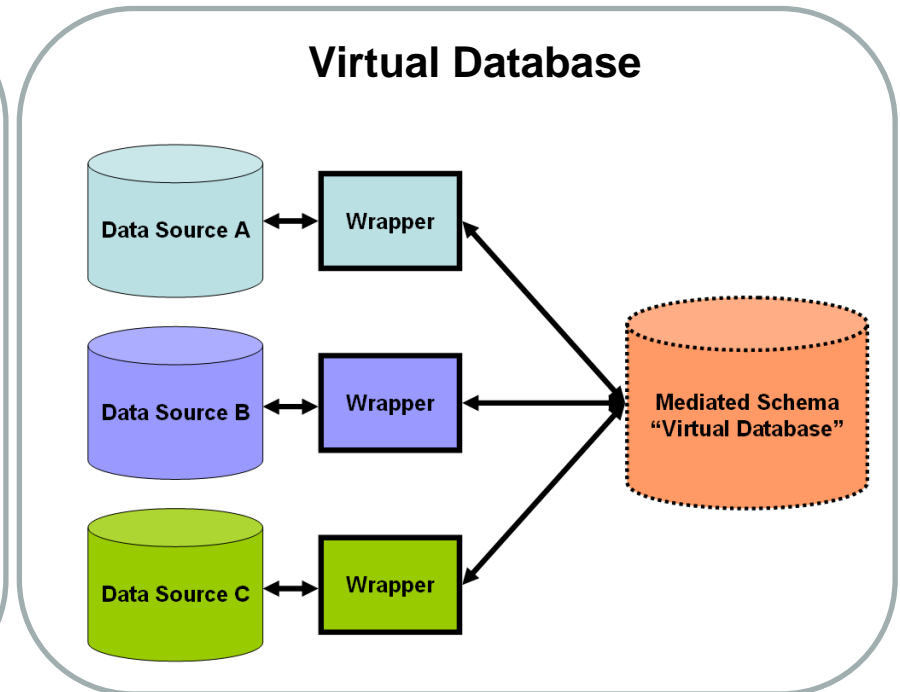
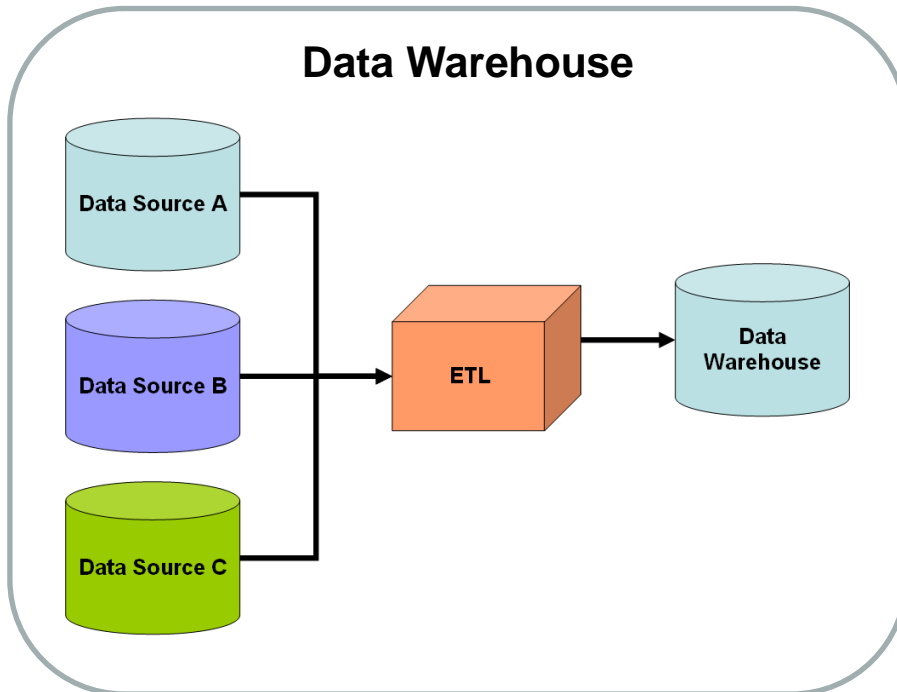
Database data integration



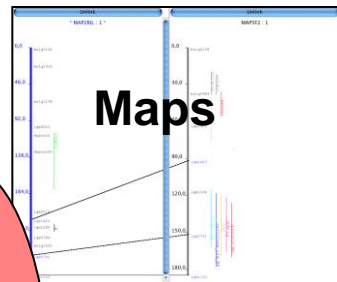
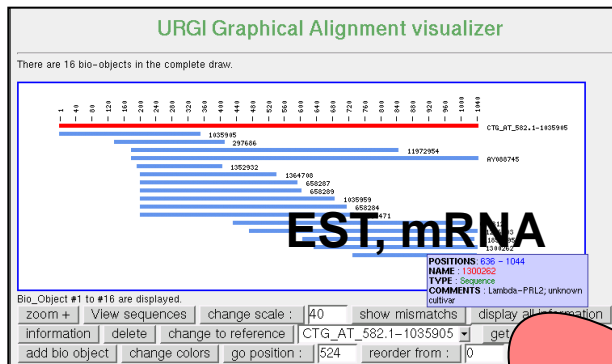
Architectures

- **Data integration is defined as a triple $\langle G, S, M \rangle$ where:**
 - ♦ G is the global (or mediated) schema,
 - ♦ S is the heterogeneous set of source schemas,
 - ♦ M is the mapping that maps queries between the source and the global schemas.

(from wikipedia)
- **Two architectures**

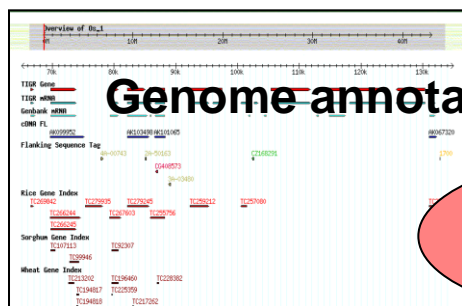


GnplS data warehouse



DNA Polymorphisms

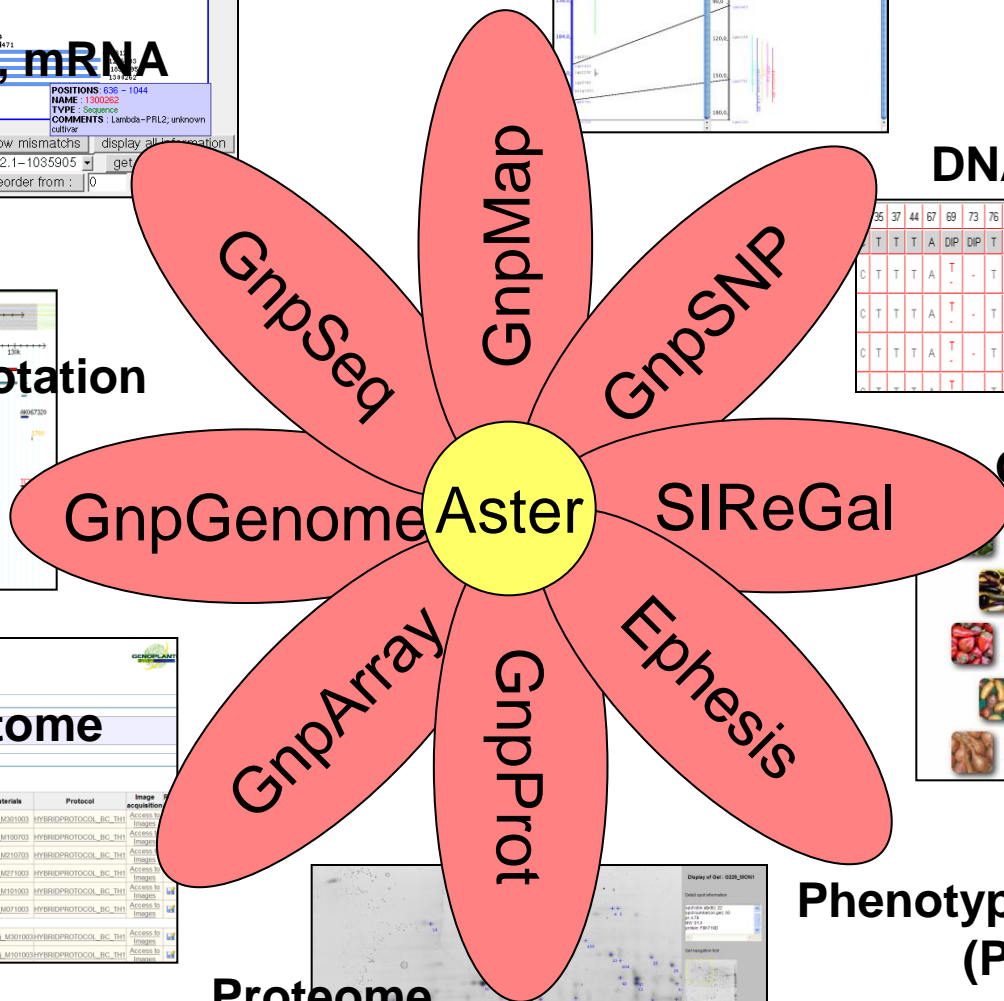
35	37	44	67	69	73	76	84	85	110	112	120	121	129	162	190	192	211	223	226	234	235	250	291	297	
T	T	T	A	DIP	DIP	T	C	T	G	G	T	C	G	C	G	T	C	C	T	A	C	C	T		
C	T	T	T	A	T	.	T	C	T	T	G	G	T	C	G	C	G	T	C	C	T	A	C	C	T
C	T	T	T	A	T	.	T	C	T	T	G	G	T	C	G	C	G	T	C	C	T	A	C	C	T
C	T	T	T	A	T	.	T	C	T	T	G	G	T	C	G	A	G	T	C	C	T	A	C	C	T



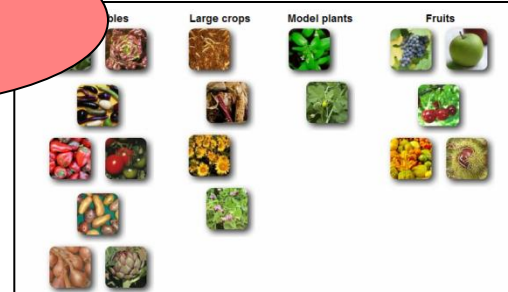
Hybridization results

Transcriptome

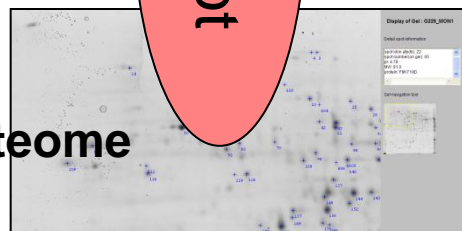
Result number	Hybridization date	Description	Experiments	Array name	Design name	Bio materials	Protocol	Image acquisition
1	30/10/2003	T4_MM3_1001003	STUDY OF BOTRYTIS	BC2002_288/BCARRAY_MM302	T4_MM3_M301003	HYBRIDPROTOCOL_BC_Th1	Access to Images	
2	10/07/2003	T4_MM3_1001003	STUDY OF BOTRYTIS	BC2002_48/BCARRAY_MM302	T4_MM3_M101003	HYBRIDPROTOCOL_BC_Th1	Access to Images	
3	21/07/2003	T4_MM3_H210703	STUDY OF BOTRYTIS	BC2002_30/BCARRAY_MM302	T4_MM3_M210703	HYBRIDPROTOCOL_BC_Th1	Access to Images	
4	27/10/2003	T4_MM3_H210703	STUDY OF BOTRYTIS	BC2002_240/BCARRAY_MM302	T4_MM3_M271003	HYBRIDPROTOCOL_BC_Th1	Access to Images	
5	10/10/2003	T4_MM3_H101003	STUDY OF BOTRYTIS	BC2002_258/BCARRAY_MM302	T4_MM3_M101003	HYBRIDPROTOCOL_BC_Th1	Access to Images	
6	07/10/2003	T4_MM3_H210703	STUDY OF BOTRYTIS	BC2002_3A/BCARRAY_MM302	T4_MM3_M271003	HYBRIDPROTOCOL_BC_Th1	Access to Images	
7	30/10/2003	T4_Sher3_H091003	STUDY OF BOTRYTIS	BC2002_288/BCARRAY_MM302	T4_Sher3_M301003	HYBRIDPROTOCOL_BC_Th1	Access to Images	
8	10/10/2003	T4_Sher3_H101003	STUDY OF BOTRYTIS	BC2002_258/BCARRAY_MM302	T4_Sher3_M101003	HYBRIDPROTOCOL_BC_Th1	Access to Images	



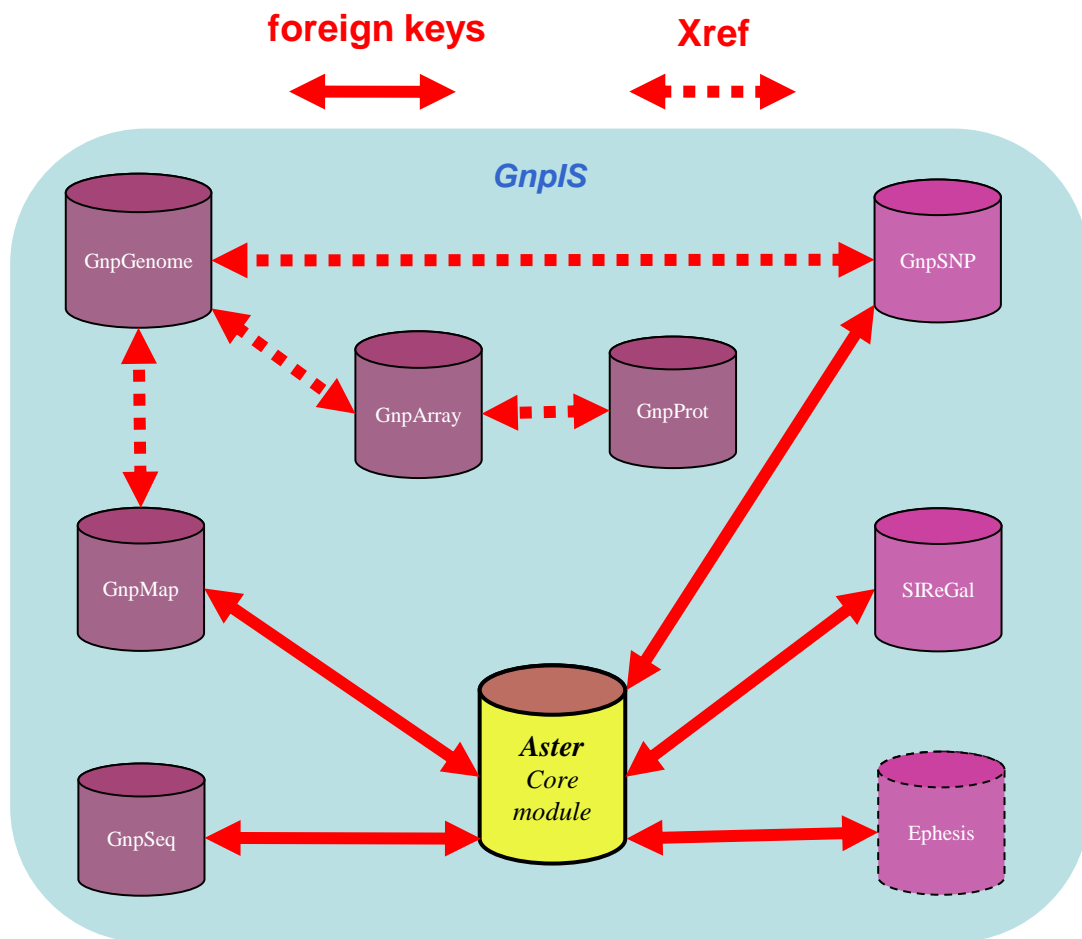
Genetic collections



Phenotype evaluations (P=GxE)



Data consistency



- A **foreign key** is a relationship or link between two tables which ensures that the data stored in a database is consistent.
- Deleting a record that contains a value referred to by a foreign key in another table would **break referential integrity**.
- Some relational database management systems (RDBMS) can **enforce referential integrity**
 - by deleting the foreign key rows as well to maintain integrity,
 - by returning an error and not performing the delete.

Architecturally, this offers a **tightly coupled** approach because the data reside together in a single repository at query-time

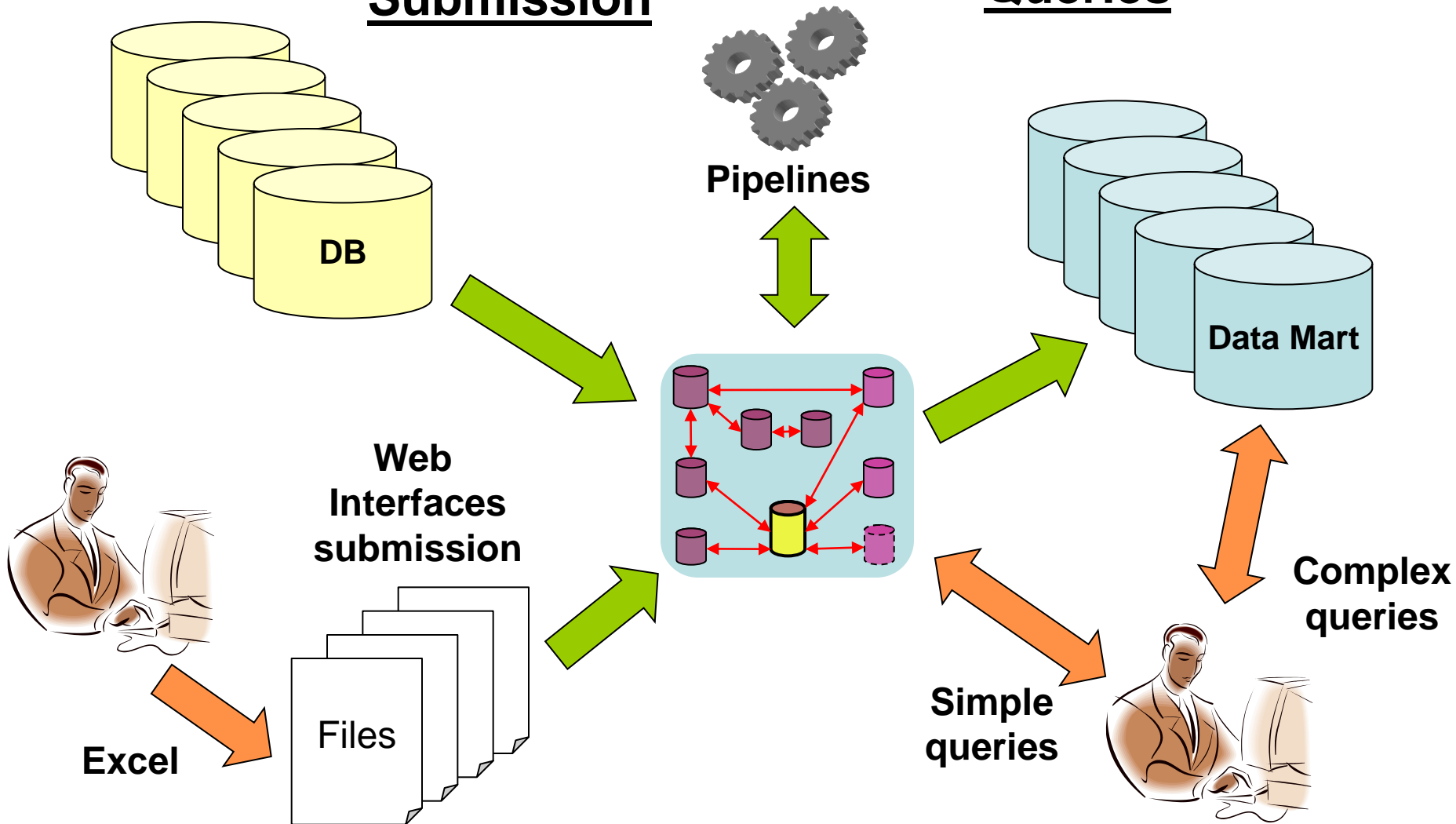
Interoperability

- **A property referring to the ability of diverse systems to work together (inter-operate)**
 - ◆ capability of different programs to exchange data via a common set of exchange formats, to read and write the same file formats, and to use the same protocols.
- (from wikipedia)*
- **(loose) links between RDBMS**
 - ◆ Xrefs
 - ◆ Web services
 - **Tools over the data warehouse**
 - ◆ Quick search
 - ◆ Biomart
 - ◆ Galaxy

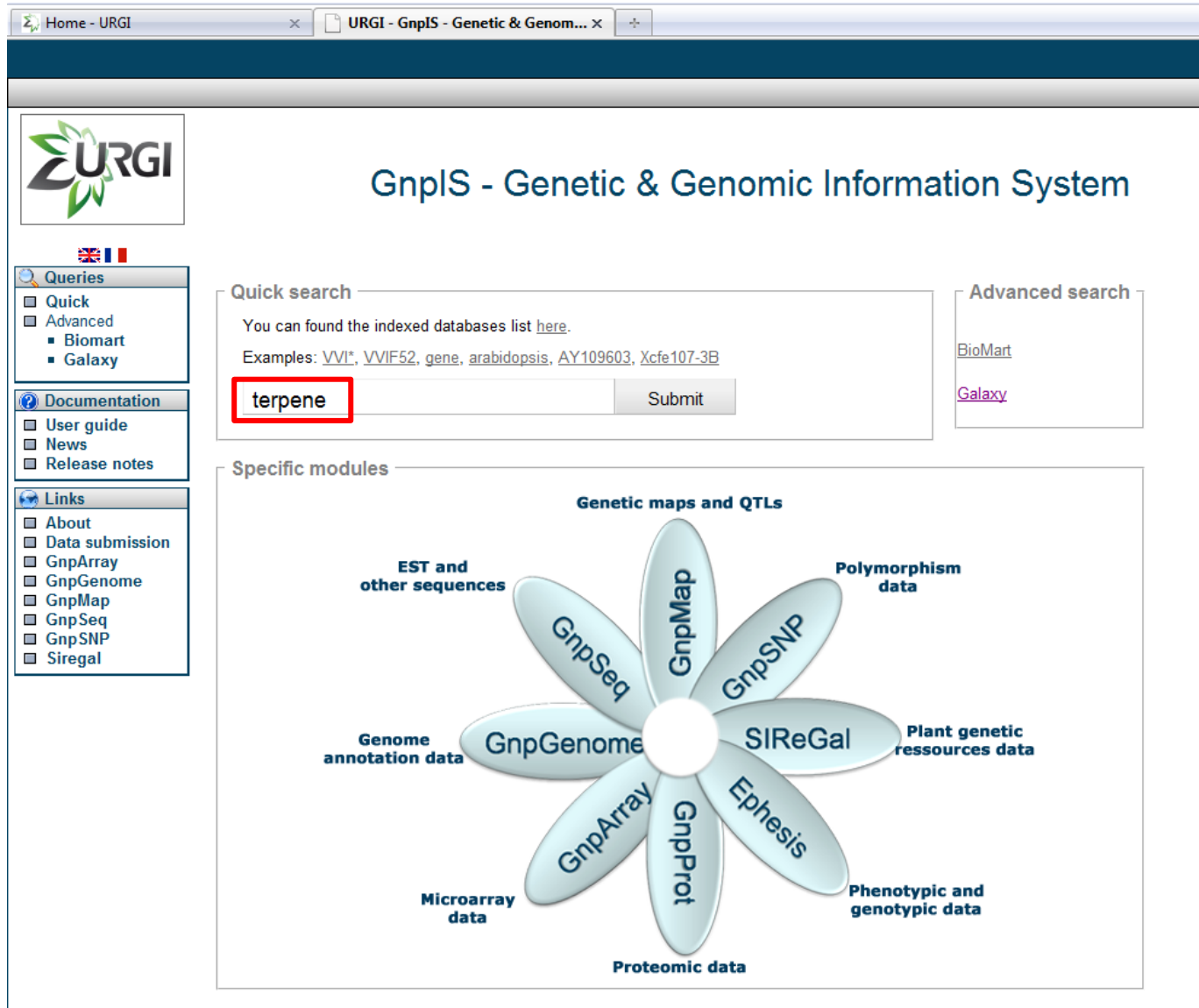
GnplS interoperability

Submission

Queries




Quick search



The screenshot shows a web browser window with two tabs: "Home - URGI" and "URGI - GnpIS - Genetic & Genom...". The main content area displays the "GnpIS - Genetic & Genomic Information System" title. On the left, there is a navigation menu with sections: "Queries" (Quick, Advanced, Biomart, Galaxy), "Documentation" (User guide, News, Release notes), and "Links" (About, Data submission, GnpArray, GnpGenome, GnpMap, GnpSeq, GnpSNP, Siregal). The main search area includes a "Quick search" box with a search input field containing the word "terpene" (highlighted with a red box) and a "Submit" button. Below the search box is a "Specific modules" section featuring a flower-shaped diagram with petals labeled with database names: GnpSeq, GnpMap, GnpSNP, SReGal, Ephesis, GnpProt, GnpArray, and GnpGenome. Surrounding the diagram are labels for data types: "Genetic maps and QTLs", "EST and other sequences", "Polymorphism data", "Plant genetic resources data", "Phenotypic and genotypic data", "Proteomic data", "Microarray data", and "Genome annotation data".

Quick search results

Home - URGI | URGI - GnpIS - Genetic & Genom... | +



common
genome-arabidopsis
genome-poplar

+ Library (18)	+ SeqSample (18)
Y11188 ★★★★★	Y11188 ★★★★★
AF497486 ★★★★★	AF497486 ★★★★★
AF497487 ★★★★★	AF497487 ★★★★★
AF497488 ★★★★★	AF497488 ★★★★★
AF497489 ★★★★★	AF497489 ★★★★★
AF497490 ★★★★★	AF497490 ★★★★★
AF497492 ★★★★★	AF497492 ★★★★★
AY062741 ★★★★★☆	AY062741 ★★★★★☆
AY081341 ★★★★★☆	AY081341 ★★★★★☆
AF529266 ★★★★★☆	AF529266 ★★★★★☆

« 1 2 3 4 5 6 7 8 9 10 » | 129 items

Name: AF529266

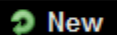
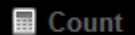
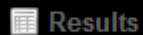
Description: Zea mays terpene synthase (tps1) mRNA, complete cds.

10 results per page

- Main
 - GnpIS Home
- Queries
 - Quick
 - Advanced
 - Biomart
 - Galaxy
- Documentation
 - User guide
 - News
 - Release notes
- Links
 - About
 - Data submission
 - GnpArray
 - GnpGenome
 - GnpMap
 - GnpSeq
 - GnpSNP
 - Siregal



GnpIS advanced search

Dataset

[None selected]

- CHOOSE DATABASE -
- CHOOSE DATABASE -
- Query local mart: GnpMap (markers, Qtls), GnpSNP (snps, genes) with GnpGenome (annotations)**
- Query local mart: GnpGenome (annotations) with GnpMap (genetic markers..) for Grapevine 8x
- Query local single mart: GnpGenome (annotations) for Poplar
- Query local single mart: GnpGenome (annotations) for Poplar V2.0
- Query local single mart: GnpGenome (annotations) for Botrytis
- Query local single mart: GnpGenome Arabidopsis, Rice (gene annotations) with GnpSNP (snps)
- Query distant single mart: MSD (EBI UK)
- Query distant single mart: Uniprot (EBI UK)

New Count Results

Dataset
GnpMap

Filters
Trait : %dry_matter_120

Attributes
Marker Name (in GnpMap)
Map Name (in GnpMap)
Link Map URGI (in GnpMap)

Dataset
[None Selected]

Please restrict your query using criteria below

Get All Assigned Mappable Element (Locus, Marker and QTL)

Get Marker by Distance, Position, Marker Name or from File

Get Locus by Locus Name or from file

Get QTL by Genus, Theme, Trait, QTL Name, QTL Detection, Prox Marker, Dist Mark or from File

Genus Brassica ▾

Theme name Biochemical
Chilling tolerance
Development
Digestibility
Drought tolerance ▾

Trait %N plant at Beginning of Seed Filling
%N seeds
%N straw
%dry_matter_120
%dry_matter_300
%dry_matter_600
%dry_matter_800
(5'OH)G_subunit_CW
AAL
ABS
ADF
ADL
ADM
ADMF
AGN%
CWD
CWD_Lg
DMC
DOM
DOM_Lg ▾

Qtl Name (% for wildcard)

Parcourir...

Qtl detection CIM
CIM QTL Cartographer (Basten et al, 1997)
Composite interval mapping
MIM
SIM ▾

Prox Marker (% for wildcard)

Dist Marker (% for wildcard)

Attributes for results

Please select columns to be included in the output and hit 'Results' when ready

Dataset
GnpMap

Filters
Genus : Zea
Trait : %dry_matter_120

Attributes
 Marker Name (in GnpMap)
 Map Name (in GnpMap)
 Link Map URGI (in GnpMap)
 Link Marker URGI (in GnpMap)
 Distance (in GnpMap)
 Link Population URGI (in GnpMap)
 Population Name (in GnpMap)
 Prox Marker
 Dist Marker
 Trait Name (in GnpMap)
 Link Trait URGI (in GnpMap)
 Species

Dataset
[None Selected]

Marker

Marker

<input checked="" type="checkbox"/> Distance (in GnpMap)	<input type="checkbox"/> Gene Function (in GnpMap)
<input type="checkbox"/> Position 2 (in GnpMap)	<input type="checkbox"/> Contig Name (in GnpMap)
<input type="checkbox"/> Position 1 (in GnpMap)	<input checked="" type="checkbox"/> Link Marker URGI (in GnpMap)
<input type="checkbox"/> Laboratory Origine (in GnpMap)	<input checked="" type="checkbox"/> Marker Name (in GnpMap)
<input type="checkbox"/> Insert Length (in GnpMap)	

Map

Measure

Population

Population

<input checked="" type="checkbox"/> Link Population URGI (in GnpMap)	<input type="checkbox"/> Population Author (in GnpMap)
<input type="checkbox"/> Population Comment (in GnpMap)	<input checked="" type="checkbox"/> Population Name (in GnpMap)
<input type="checkbox"/> Population Size (in GnpMap)	

Qtl

Markers

<input checked="" type="checkbox"/> Prox Marker	<input checked="" type="checkbox"/> Dist Marker
---	---

Qtl detection

<input type="checkbox"/> Detection Threshold (in GnpMap)	<input type="checkbox"/> Parameter (in GnpMap)
<input type="checkbox"/> R2 Corrected (in GnpMap)	<input type="checkbox"/> Qtl detection date (in GnpMap)
<input type="checkbox"/> R2 Global (in GnpMap)	<input type="checkbox"/> Method (in GnpMap)

Theme

<input type="checkbox"/> Theme Id (in GnpMap)	<input type="checkbox"/> Theme Name (in GnpMap)
---	---

Trait

<input type="checkbox"/> Description (in GnpMap)	<input type="checkbox"/> Trait Unit (in GnpMap)
<input checked="" type="checkbox"/> Trait Name (in GnpMap)	<input checked="" type="checkbox"/> Link Trait URGI (in GnpMap)

Qtl

<input type="checkbox"/> Genus (in GnpMap)	<input type="checkbox"/> Qtl Name (in GnpMap)
<input type="checkbox"/> Link Taxon URGI (in GnpMap)	

Taxon

Taxon

<input checked="" type="checkbox"/> Species	<input type="checkbox"/> Kingdom
<input type="checkbox"/> Family	<input type="checkbox"/> Domain
<input type="checkbox"/> Division	<input type="checkbox"/> Common Name
<input type="checkbox"/> Phylum	<input type="checkbox"/> Scientific Name

Results and links for details

Dataset
GnpMap

Filters
Genus : Zea
Trait : %dry_matter_120

Attributes
 Marker Name (in GnpMap)
 Map Name (in GnpMap)
 Link Map URGI (in GnpMap)
 Link Marker URGI (in GnpMap)
 Distance (in GnpMap)
 Link Population URGI (in GnpMap)
 Population Name (in GnpMap)
 Prox Marker
 Dist Marker
 Trait Name (in GnpMap)
 Link Trait URGI (in GnpMap)
 Species

Export all results to Unique results only

Email notification to

View rows as Unique results only

Marker Name (in GnpMap)	Map Name (in GnpMap)	Link Map URGI (in GnpMap)	Link Marker URGI (in GnpMap)	Distance (in GnpMap)	Link Population URGI (in GnpMap)	Population Name (in GnpMap)	Prox Marker	Dist Marker	Trait Name (in GnpMap)	Link Trait URGI (in GnpMap)	Species
	REFMAP_031127	3		158.23	1	LHRE-F3			%dry_matter_120	213	Zea mays
	LHRE-F3-1201	35		56	38	LHRE-F3	csu11	bnlg1094	%dry_matter_120	213	Zea mays
	REFMAP_050110	54		158.23	1	LHRE-F3			%dry_matter_120	213	Zea mays
	REFMAP_050110	54		142.61	1	LHRE-F3			%dry_matter_120	213	Zea mays
	REFMAP_031127	3		144.72	1	LHRE-F3			%dry_matter_120	213	Zea mays
	LHRE-F3-1201	35		168	38	LHRE-F3	bnlg1627	bnlg1083	%dry_matter_120	213	Zea mays
	LHRE-F3-1201	35		96	38	LHRE-F3	umc1484	umc1222	%dry_matter_120	213	Zea mays
	REFMAP_050110	54		68.95	1	LHRE-F3			%dry_matter_120	213	Zea mays
	REFMAP_031127	3		215.99	1	LHRE-F3			%dry_matter_120	213	Zea mays
	LHRE-F3-1201	35		372	38	LHRE-F3	umc1221	mmc0081	%dry_matter_120	213	Zea mays

Dataset
[None Selected]

Interoperability (QTL) with GnpMap

★

Dataset 84 / 428702 Entries
 poplar
Filters
 Feature Type : QTL
Attributes
 Common Name
 Feature Name
 Feature Type

Export all results to TSV Unique results only

Email notification to

View rows as HTML Unique results only

Common Name	Feature Name	Feature Type	Map Name (in GnpMap)	Link Map URGI (in GnpMap)
poplar	qtl_tolCir3_pop2_Populustrichocarpa_16042009_05a_0	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_budburst_pop2_Populustrichocarpa_16042009_12b_52.1_UK	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_syllep1_pop2_Populustrichocarpa_16042009_05a_67.5_F	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_circum2_pop2_Populustrichocarpa_16042009_14_27.2_UK	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_budburst_pop2_Populustrichocarpa_16042009_08a_68.3_UK	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_circum2_pop2_Populustrichocarpa_16042009_04_62.4_F	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_budburst_pop2_Populustrichocarpa_16042009_02a_0_I	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_S93ID6_pop2_Populustrichocarpa_16042009_19_48.6	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_Canker_LL_pop2_Populustrichocarpa_16042009_02a_0_I	QTL	pop2_Populustrichocarpa_16042009	103
poplar	qtl_budburst_pop2_Populustrichocarpa_16042009_02b_89.8_I	QTL	pop2_Populustrichocarpa_16042009	103

GnpGenome Poplar
GnpMap

Dataset 5674 / 106662 Entries
 GnpMap
Filters
 Qtl Name (% for wildcard) : [ID-list specified]
Attributes
 Map Name (in GnpMap)
 Link Map URGI (in GnpMap)

QTL mapped on GnpGenome (via markers)

Applications Raccourcis Système

Populus trichocarpa Genome Browser: PTR02:6951373..7451372 - Icweasel

https://gpi.versailles.inra.fr/cgi-bin/gbrowse/populus_PTR/#search

ven 21 mai, 14:18

Chercher PTR02:6951373..7451372

Source de données: Populus trichocarpa Genome Browser

Sorties, recherches et autres opérations:
 Sortie Decorated FASTA File Configurer... Lancer

Défil./Zoom: Voir 500 kbp Inversion

qt1_budburst_pop2_Populstrichocarpa_16042009_02b_89.8

Feature Infos :

Positions	6233777 .. 8168969
Length	1935193
Source	pop2_Populstrichocarpa_16042009
Reference	PTR02
Mapping interval used	PMGC684 PMGC422
Link to GnpMap	

Zoom to this feature

qt1_S93CV1_pop2_Populstrichocarpa_16042009_02b_50.1
 position on genetic map:50.1..126.1
 qt1_budburst_pop2_Populstrichocarpa_16042009_02b_89.8_F
 position on genetic map:89.8..111
 qt1_budburst_pop2_Populstrichocarpa_16042009_02b_89.8_I
 position on genetic map:89.8..111
 qt1_budburst_pop2_Populstrichocarpa_16042009_02b_89.8_UK
 position on genetic map:89.8..111

SSR

PMGC684	http://www.ornl.gov/sci/ipgc/ssr_resource.htm
ORPM207	Tuskan et al. 2004. Can J Forest Res. 34(1):85-93.
GCPM3732-1	http://www.ornl.gov/sci/ipgc/ssr_resource.htm
GCPM3732-2	http://www.ornl.gov/sci/ipgc/ssr_resource.htm
GCPM3540-1	http://www.ornl.gov/sci/ipgc/ssr_resource.htm
GCPM2805-1	http://www.ornl.gov/sci/ipgc/ssr_resource.htm

Supprimer le surlignage

Mise à jour de l'image

Terminé

gpi.versailles.inra.fr

QTLs found in GnpMap

Quick search:



QTL results



Welcome a kelet

- Main
- Textual interface
- MapComparator

- Queries
- Loci
- QTL
- Marker
- Pool
- Xref
- Map
- Trait
- Blast
- Keywords for submission

- Documentation
- User guide
- FAQ
- GnpMap news
- Release notes

- Links
- About
- Data submission
- GnpIS
- GnpArray
- GnpGenome
- GnpSeq
- GnpSNP
- SIReGal

Information

Display results per page
 158 items found, displaying 1 to 10
 < 1 2 3 4 5 >

Results

[Help]

#	QTL name	Theme	Trait	Trait description	Map name	Linkage group	LOD	R2	Distance	From	to	Reliability / (source map name)	Link to loci	Link to loci large
1	qtl_budburst_pop2_Populustrichocarpa_16042009_03_0_F	phenology	phenology	-	pop2_Populustrichocarpa_16042009	3	-	-	-	0	153.8	unprojected	7	7
2	qtl_budburst_pop2_Populustrichocarpa_16042009_03_0_I	phenology	phenology	-	pop2_Populustrichocarpa_16042009	3	-	-	-	0	153.8	unprojected	7	7
3	qtl_budburst_pop2_Populustrichocarpa_16042009_03_0_UK	phenology	phenology	-	pop2_Populustrichocarpa_16042009	3	-	-	-	0	153.8	unprojected	7	7
4	qtl_S98AR1_pop2_Populustrichocarpa_16042009_03_7.7	Resistance/Tolerance	pest resistance/tolerance	-	pop2_Populustrichocarpa_16042009	3	-	-	7.7	50.4		unprojected	1	1
5	qtl_tolH5_pop2_Populustrichocarpa_16042009_03_108	Resistance/Tolerance	pest resistance/tolerance	-	pop2_Populustrichocarpa_16042009	3	-	-	108	248.2		unprojected	6	6
6	qtl_tolCir3_pop2_Populustrichocarpa_16042009_03_90.9	Resistance/Tolerance	pest resistance/tolerance	-	pop2_Populustrichocarpa_16042009	3	-	-	90.9	274.9		unprojected	9	9
7	qtl_syllep1_pop2_Populustrichocarpa_16042009_04_0_F	yield	yield	-	pop2_Populustrichocarpa_16042009	4	-	-	-	0	206.3	unprojected	11	11
8	qtl_syllep1_pop2_Populustrichocarpa_16042009_04_0_I	yield	yield	-	pop2_Populustrichocarpa_16042009	4	-	-	-	0	206.3	unprojected	11	11
9	qtl_syllep1_pop2_Populustrichocarpa_16042009_04_0_UK	yield	yield	-	pop2_Populustrichocarpa_16042009	4	-	-	-	0	206.3	unprojected	11	11
10	qtl_budburst_pop2_Populustrichocarpa_16042009_04_0_F	phenology	phenology	-	pop2_Populustrichocarpa_16042009	4	-	-	-	0	128.6	unprojected	7	7

Rechercher : Précédent Suivant Surligner tout Respecter la casse

A data integration workbench



- Tools
- Get Data
 - Text Manipulation
 - Filter and Sort
 - Join, Subtract and Group
 - Convert Formats
 - Extract Features
 - Fetch Sequences
 - Fetch Alignments
 - Get Genomic Scores
 - Operate on Genomic Intervals
 - Statistics
 - Graph/Display Data
 - Regional Variation
 - Solexa tools
 - SOLiD Data Analysis
 - FASTA manipulation
 - Short Read QC and Manipulation
 - SAM Tools

i We are hiring!
 Thanks to your support and an unprecedented level of usage, we are looking for an experienced software developer to join our team. For more information about this position, please, click [here](#).

- ### Get Data
- Upload File from your computer
 - UCSC Main table browser
 - UCSC Test table browser
 - UCSC Archaea table browser
 - Get Microbial Data
 - BioMart Central server
 - BioMart INRA URGI GnpIs**
 - GrameneMart Central server
 - Flymine server
 - Flymine test server
 - modMine server
 - Ratmine server
 - Wormbase server
 - Wormbase test server
 - EuPathDB server
 - EncodeDB at NHGRI
 - EpiGRAPH server
 - EpiGRAPH test server
 - HbVar Human Hemoglobin Variants and Thalassemias

Galactic Quickies

short screencasts that are *always* under 5 minutes. We thought it may be a good idea to provide Galaxy's functionality while keeping the "annoyance factor" to the minimum. The sky.

Joining Features

Galactic quickie # 5

at Penn State.
 part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.
 e3752e\$

History Options ▾

refresh | collapse all

Unnamed history 0

71: Merge on data 56 👁 🗑

70: BioMart 👁 🗑

68: sam_file.sam 👁 🗑
 14 lines, format: sam, database: ?
 Info: uploaded txt file
 save | rerun

```

1. QNAME      2
oHD
  VN: 1.0
oSQ
  SN: chr20 LN: 62435964
oRG
  ID: L1 PU: SC_1_10 LB: SC_1 SN: NA12891
  < ... >
  
```

64: UCSC Main on D. viridis: xenoRefGene (genome) 👁 🗑
 158,792 regions, format: bed, database: droVir2
 Info: UCSC Main on D. viridis: xenoRefGene (genome)
 save | rerun | display at UCSC main test

1. Chrom	2. Start	3. End	4. Name
scaffold_34	371	489	NM_057619
scaffold_55	1232	1350	NM_057619
scaffold_152	381	583	NM_001110885
scaffold_152	858	1593	NM_001015102
scaffold_152	858	1782	NM_001110576
scaffold_179	2	391	NM_005044

56: MAF file.maf 👁 🗑
 10 lines, format: maf, database: mm8
 Info: uploaded maf file
 save | rerun

```

Species: hg18 panTro2 rhesMac2 mm8 canFam
##maf version=1
  
```


Get Data

Text Manipulation

- Add column to an existing query
- Compute an expression on every row
- Concatenate queries tail-to-head
- **Cut** columns from a table
- Merge Columns together
- Convert delimiters to TAB
- Create single interval as a new query
- Change Case of selected columns
- Paste two files side by side
- Remove beginning of a file
- Select first lines from a Query
- Select last lines from a Query

Filter and Sort

Cut

Cut columns:

Delimited by:

From:

Exemple:

Input dataset (six columns: c1, c2, c3, c4, c5, and c6):

```
chr1 10 1000 gene1 0 +
chr2 100 1500 gene2 0 +
```

cut on columns "c1,c4,c6" will return:

```
chr1 gene1 +
chr2 gene2 +
```

cut on columns "c6,c5,c4,c1" will return:

```
+ 0 gene1 chr1
+ 0 gene2 chr2
```

cut on columns "c8,c7,c4" will return:

```
. . gene1
. . gene2
```

Other manipulations

Galaxy

Analyze Data Workflow Data Libraries Help User

Tools

- Get Data
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Solexa tools
- SOLID Data Analysis
- FASTA manipulation
- Short Read QC and Manipulation
- SAM Tools

scaffold_34	371	489	NM_057619	0	-	489	489	0	2	78,15,	0,103,	
scaffold_55	1232	1350	NM_057619	0	+	1350	1350	0	2	15,78,	0,40,	
scaffold_152	381	583	NM_00110885	0	-	381	583	0	1	202,	0,	
s						858	1503	0	2	102,240,	0,405,	
s						858	1782	0	3	102,162,144,	0,405,780,	
s						2	391	0	2	85,245,	0,144,	
s						2	3816	0	5	89,243,3,109,171,	0,148,422,448,3643,	
s						2	3816	0	5	89,243,3,109,171,	0,148,422,448,3643,	
s						391	391	0	2	85,245,	0,144,	
s						11	397	0	2	74,253,	0,133,	
s						11	3816	0	5	76,251,75,117,30,	0,135,446,3634,3775,	
s						11	3816	0	5	76,245,75,117,30,	0,135,446,3634,3775,	
s						142	391	0	2	84,123,	0,126,	
s						142	391	0	2	84,123,	0,126,	
s						142	397	0	1	255,	0,	
s						142	403	0	1	261,	0,	
s						142	397	0	2	84,129,	0,126,	
s						142	391	0	2	84,123,	0,126,	
s						142	397	0	2	84,129,	0,126,	
s						142	391	0	2	84,123,	0,126,	
s						142	391	0	2	84,123,	0,126,	
s						142	403	0	1	261,	0,	
s						142	403	0	1	261,	0,	
s						142	400	0	2	84,132,	0,126,	
s						148	397	0	2	78,129,	0,120,	
s						157	397	0	2	69,129,	0,111,	
s						157	403	0	2	69,117,	0,129,	
s						157	397	0	2	69,129,	0,111,	
s						157	397	0	2	69,129,	0,111,	
s						157	397	0	2	69,129,	0,111,	
s						157	397	0	2	69,129,	0,111,	
s						199	358	0	2	36,54,	0,105,	
s						250	391	0	1	141,	0,	
s						265	397	0	1	132,	0,	
s						265	397	0	1	132,	0,	
s						265	388	0	1	123,	0,	
s						265	397	0	1	132,	0,	
s						265	388	0	1	123,	0,	
s						265	388	0	1	123,	0,	
s						265	388	0	1	123,	0,	
s						265	388	0	1	123,	0,	
s						265	388	0	1	123,	0,	
s						2436	4210	0	3	18,36,1723,	0,191,242,	
s						2436	4210	0	3	18,36,1723,	0,191,242,	
s						2436	4210	0	3	18,36,1723,	0,191,242,	
s						2436	4210	0	2	36,1722,	0,51,	
s						2436	4209	0	2	36,1722,	0,51,	
s						2546	4211	0	2	75,1547,	0,118,	
s						2670	4209	0	1	1539,	0,	
s						2688	4206	0	5	471,291,132,234,117,	0,492,930,1083,1401,	

Join, Subtract and Group

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

Multiple regression

- Perform Linear Regression
- Perform Best-subsets Regression
- Compute RCVE

History

Options

14 lines, format: sam, database: 2

Info: uploaded bt file

save | rerun

1. ONAME 2

#HD

VN: 1.0

ES0

SN: chr20 LN: 62435964

QR6

ID: L1 PU: SC_1_10 LB: SC_1 SM: NA12891

64: UCSC Main on D. viridis: xenoRefGene (genome)

158,792 regions, format: bed, database: droVir2

Info: UCSC Main on D. viridis: xenoRefGene (genome)

save | rerun | display at UCSC main test

1. Chrom	2. Start	3. End	4. Name
scaffold_34	371	489	NM_057619
scaffold_55	1232	1350	NM_057619
scaffold_152	381	583	NM_00110885
scaffold_152	858	1503	NM_00110885
scaffold_152	858	1782	NM_00110876
scaffold_179	2	391	NM_005044

56: MAF file.maf

10 lines, format: maf, database: mm8

Info: uploaded maf file

save | rerun

Species: hg18 panTro2 rhesMac2 mm8 canFam

##maf version=1

a scores=69686 000000

s hg18.chr20 56827368 75 + 62435964

s panTro2.chr20 56528685 75 + 62293572

s rhesMac2.chr10 89144112 69 - 94855758

s mm8.chr2 173910832 61 + 181976762

55: Split on data 33

104.0 bytes, format: fastqsolexa, database: 2

Galaxy

Analyze Data | Workflow | Data Libraries | Help | User

Tools

- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- FASTA manipulation
- NGS: QC and manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: Peak Calling
- Rg Data
- Rg Simulate
- Rg Visualise
- Rg Model Data
- URGI NGS
 - Map with Bwa 0.5.7 Map Short Reads to Reference sequence with BWA 0.5.7.
 - Map with Bwa 0.5.7 (cmd lines) Map Short Reads to Reference sequence with BWA 0.5.7 (and command lines).
 - Header Fasta Filter Modify all of this header file that contains one, or multiple, tabulation and informations after the name of the sequence.
 - Sam Filter All alignments that are outside the subject sequence (partially or not), are removed from the input file and copy to an output file.
 - VarScan VarScan: convert bam file to pileup file with different filters.
 - Tablet (v.1.10.03.04) Alignment Viewer (can use GFF3, ACE, AFG, MAQ, SOAP, SAM and BAM Files).
 - GenomeView (v.922) Alignment Viewer (can use BAM, GFF, FASTA and Annotation Files).
- URGI Mapping BES

<http://localhost:8080/workflow/editor?id=f06>

Convert Formats

- **AXT to concatenated FASTA**
Converts an AXT formatted file to a concatenated FASTA alignment
- **AXT to FASTA** Converts an AXT formatted file to FASTA format
- **BED-to-GFF** converter
- **FASTA-to-Tabular** converts FASTA file to tabular format
- **FASTQSOLEXA-to-FASTA-QUAL** extracts sequences and quality scores from FASTQSOLEXA data
- **GFF-to-BED** converter
- **MAF to FASTA** Converts a MAF formatted file to FASTA format
- **Tabular-to-FASTA** converts tabular file to FASTA format
- **SOLID-to-FASTQ** converts SOLiD data to FASTQ data
- **FASTQ Conversions** converts between FASTQ data and other data formats

```

graph LR
    A[Convert SAM file] --> B[Generate pileup]
    B --> C[Tablet v.1.10.03.04]
  
```

Details

Tool: Tablet (v.1.10.03.04)

Fichier Bam / Sam / Map / autres...
Data input Input_FICHER (bam)

RÃefÃerence Input.fasta
Data input Input_REF (fasta)

Edit Step Attributes

Annotation / Notes:

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

Tablet:

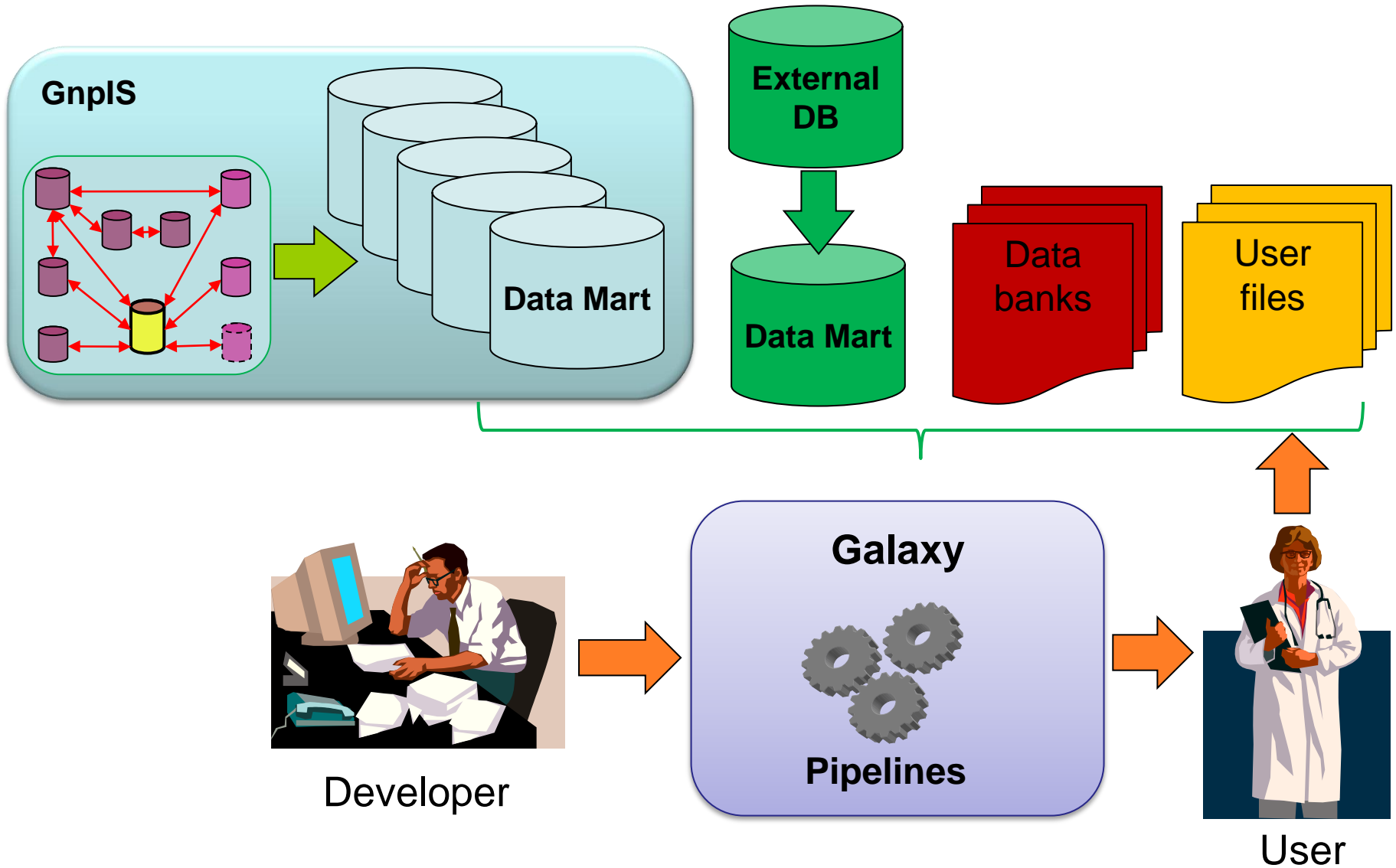
Tablet is a lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments. This software is able to view sequences alignments. It can use GFF3, ACE, AFG, MAQ, SOAP SAM and BAM Files.

Tablet features:

- High-performance visualization and data navigation.
- Display of reads in both packed and stacked formats.
- File format support for ACE, ACE, MAQ, SOAP, SAM and BAM.
- Import GFF features and quickly find/hightlight them.
- Search and locate reads by name across entire data sets.
- Entire contig overview, always data layout or coverage.
- Simple install routine via auto-updating graphical tool.
- Support for Windows, Apple Mac OS X, Linux and Solaris.

Website: <http://bioinf.scri.ac.uk/tablet/>

URGI - Data Integration Workbench



URGI

M. Alaux, N. Mohellibi, S. Arnoux (CDD),
F. Alfama, C. Pommier, M. Bras (CDD),
J. Amselem, H. Quesneville, B. Brault (CDD),
N. Choisne, S. Reboux, L. Brigitte (CDD),
S. Durand, D. Steinbach, T. Flutre (PhD),
O. Inizan, M. Zytnicki, C. Hoede (CDD),
V. Jamilloux, H. Mors (Master2)
A. Keliet, E. Permal (Post-doc),
E. Kimmel, D. Verdelet (CDD),
N. Lapalu, D. Valdenaire (CDD),
I. Luyten,

left URGI

B. Hilseberger (CDD)



Wheat genomics:

P. Leroy, C. Ravel, E. Paux,
C. Feuillet

Grape genomics

A.F. Adam-Blondon,
M. Moroldo



SNP thematic

D. Brunel, F. Granier, H. McKhann
C. De Poittevin

Genetic resources

J.M. Prosperi, P. Roumet



Fungi genomics

M.H. Lebrun, T. Rouxel

Maize genomics

M. Falque, J. Joets,
A. Charcosset



Colot's Lab

V. Colot, I. Ahmed, A. Sarazin

Tree genomics

C.Plomion, C. Poittevin

