GMOD tools in the frame of GnpAnnot and GnpIS projects at URGI

GMOD Meeting Jan 13th, 2010

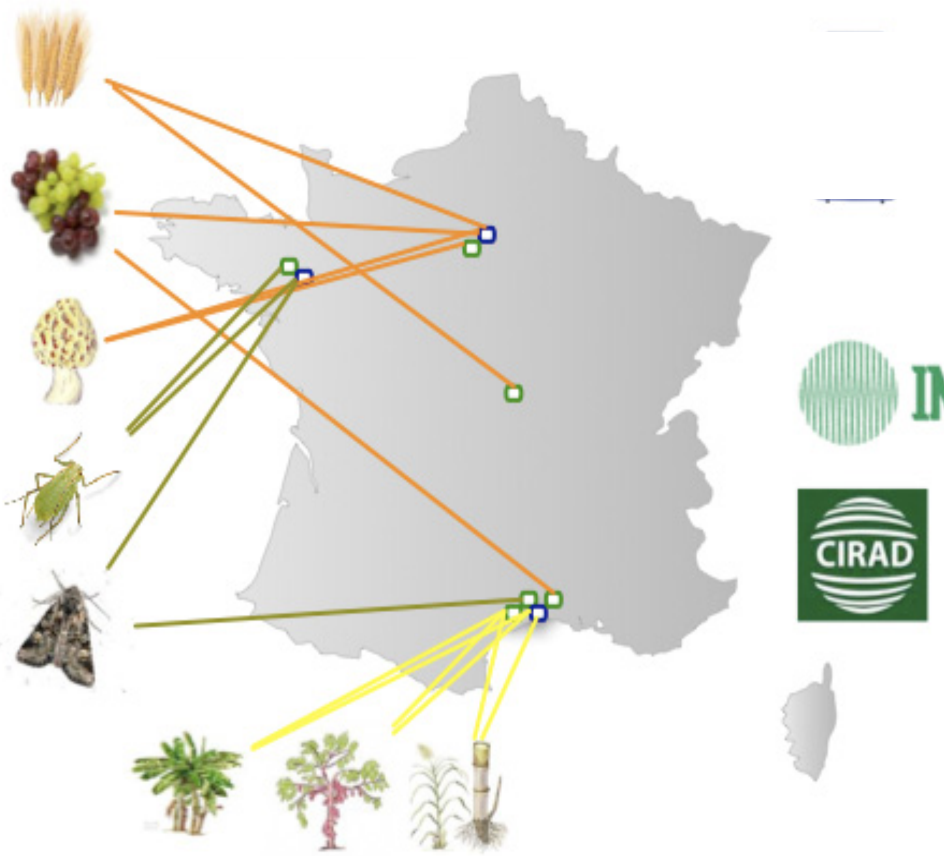
J. Amselem

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

GnpAnnot project

- Granted by ANR (French National Research Agency)
- A structural and functional annotation platform supported by comparative genomics and dedicated to plant and bio-aggressor genomes
- 5 work packages
    - database and flow management
    - annotator interfaces implementation
    - interoperability with other systems
    - sequence exploitation and platform release
    - manual annotation and platform validation

GnpAnnot species

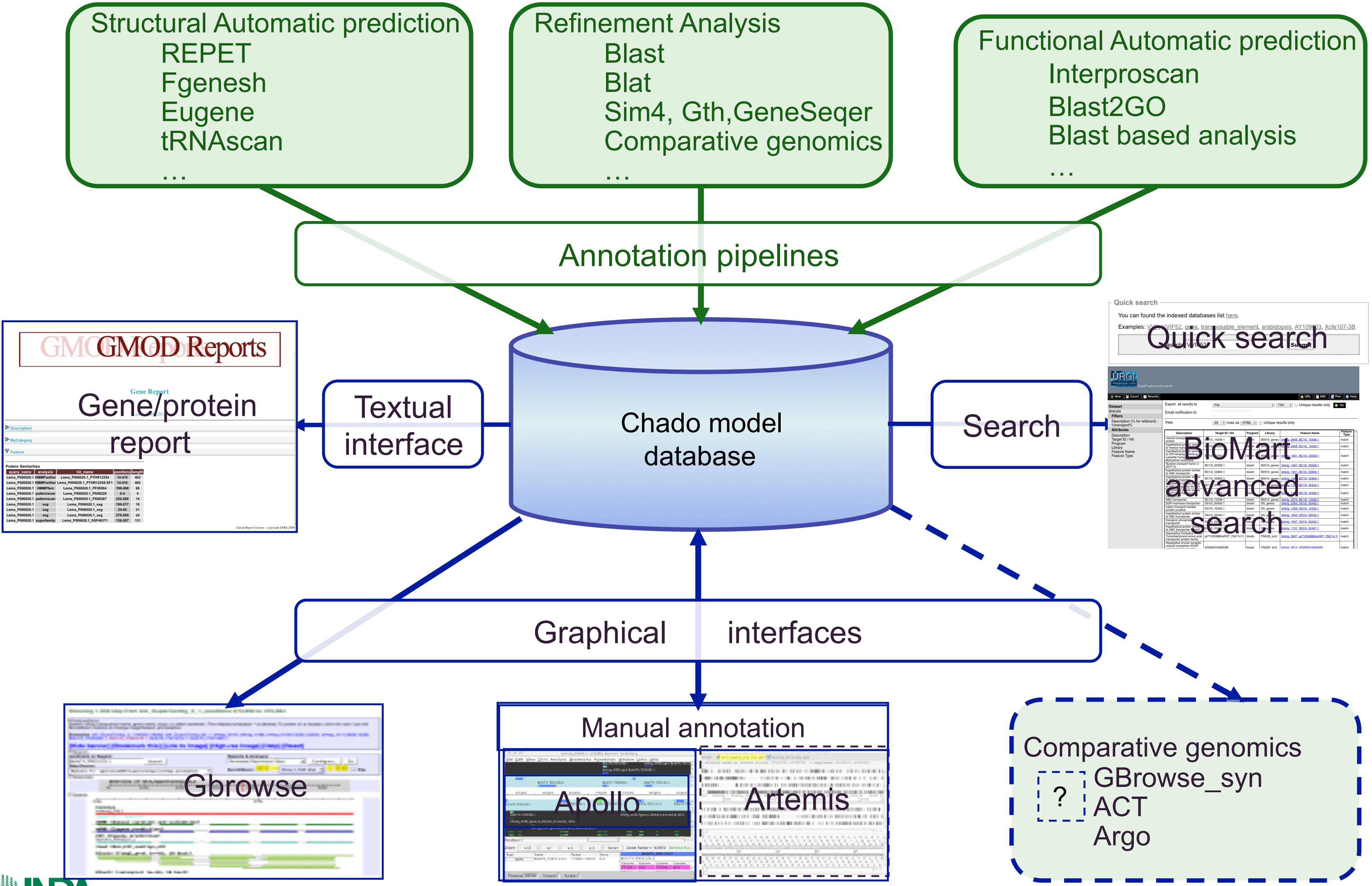


URGI
- Wheat
- Grapevine
- Fungi

BIO3P
- Insects

UMR-PIA
- Musa, Sugarcane…

Pipelines, DBs, Interfaces at URGI



Structural Automatic prediction
- REPET
- Fgenesh
- Eugene
- tRNAscan
- …

Refinement Analysis
- Blast
- Blat
- Sim4, Gth,GeneSeqer
- Comparative genomics
- …

Functional Automatic prediction
- Interproscan
- Blast2GO
- Blast based analysis
- …

Annotation pipelines

Chado model database

Textual interface

GMOD Reports

Gene/protein report

Search

Quick search

BioMart advanced search

Graphical interfaces

Gbrowse

Manual annotation

Apollo

Artemis

Comparative genomics
- GBrowse_syn
- ACT
- Argo

?

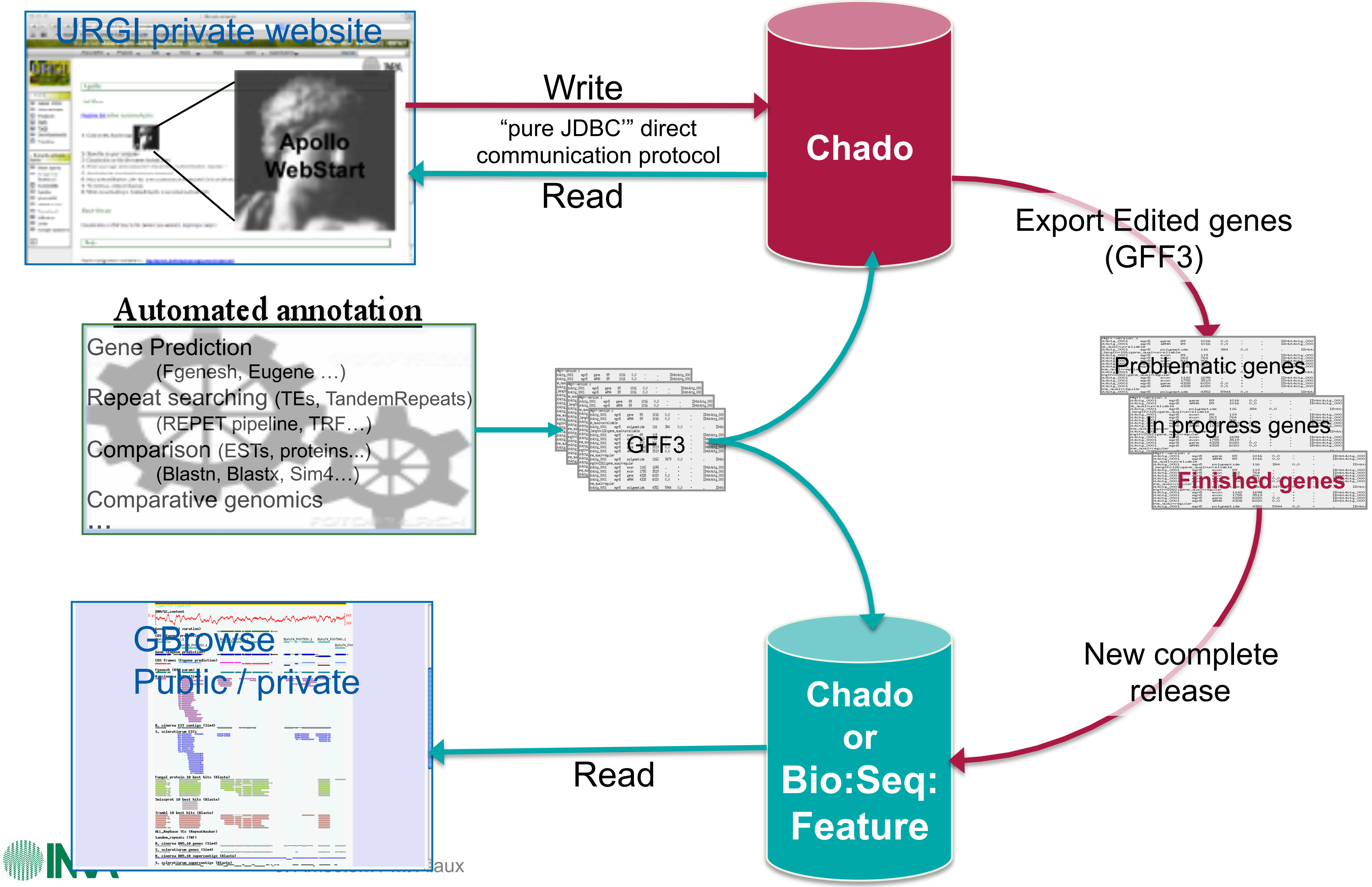
J. Amselem / M. Alaux

Distributed Manual Annotation System

- A DAS defines a communication protocol used to exchange/share sequences and annotations
  - Relies on sequence references to be annotated
  - A reference sequence server
  - One or more annotation servers

- Set up at URGI : a DAS for manual annotation
  - One annotation server
  - One can upload and display it’s own annotation tracks
  - Manual annotations shared by all partners in real time

Manual annotation roundtrip at URGI



URGI private website

Write

"pure JDBC" direct communication protocol

Read

Chado

Export Edited genes (GFF3)

Problematic genes

In progress genes

Finished genes

Automated annotation

Gene Prediction
(Fgenesh, Eugene …)
Repeat searching (TEs, TandemRepeats)
(REPET pipeline, TRF…)
Comparison (ESTs, proteins...)
(Blastn, Blastx, Sim4…)
Comparative genomics
…

GFF3

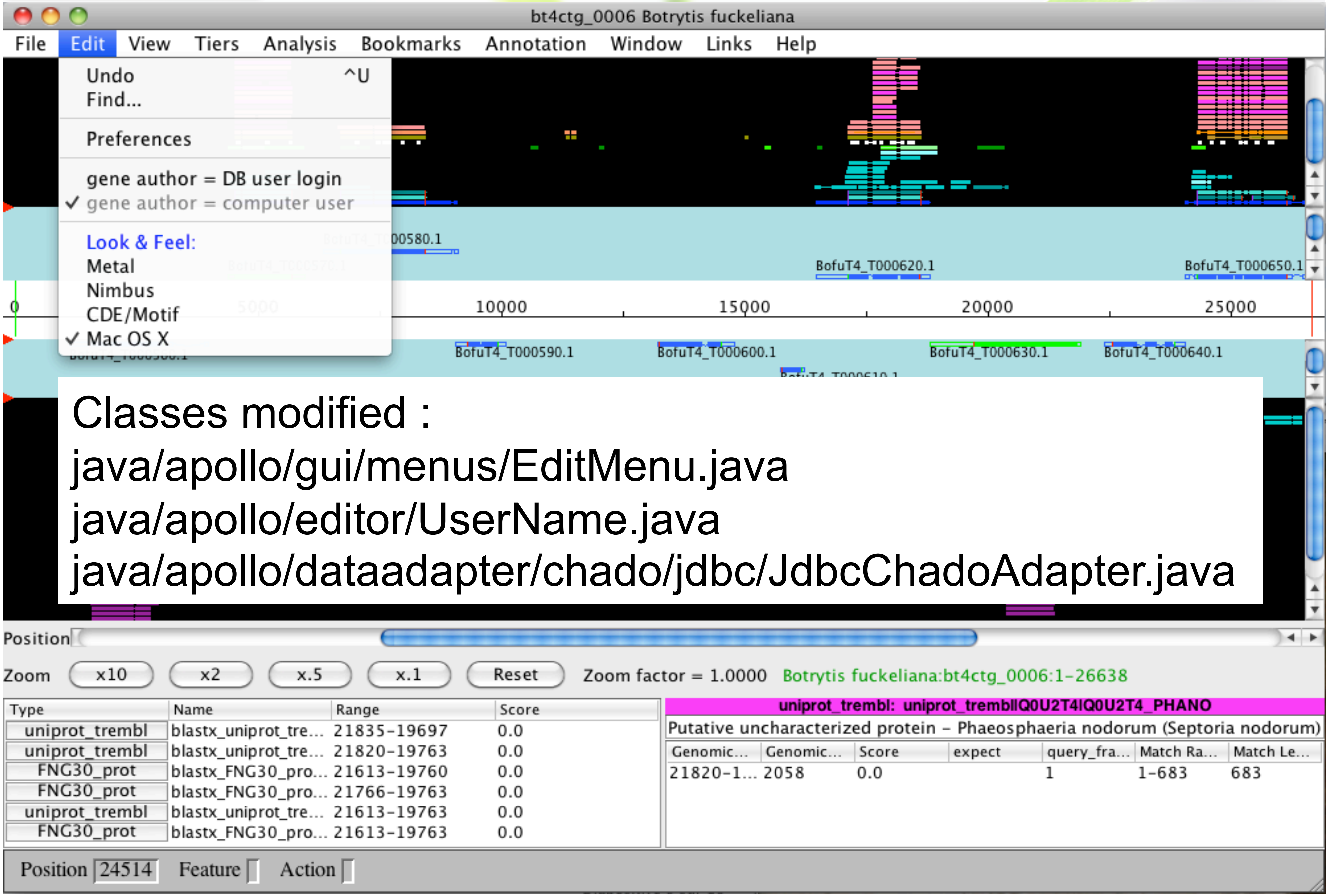
New complete release

Chado or Bio:Seq:Feature

Read

GBrowse
Public / private

Apollo add…



gene author = DB user login

✓ gene author = computer user

✓ Mac OS X

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

uniprot_trembl: uniprot_trembl|Q0U2T4|Q0U2T4_PHANO

Putative uncharacterized protein - Phaeosphaeria nodorum (Septoria nodorum)

Genomic...	Genomic...	Score	expect	query_fra...	Match Ra...	Match Le...
21820-1...	2058	0.0		1	1-683	683

Position 24514 Feature Action

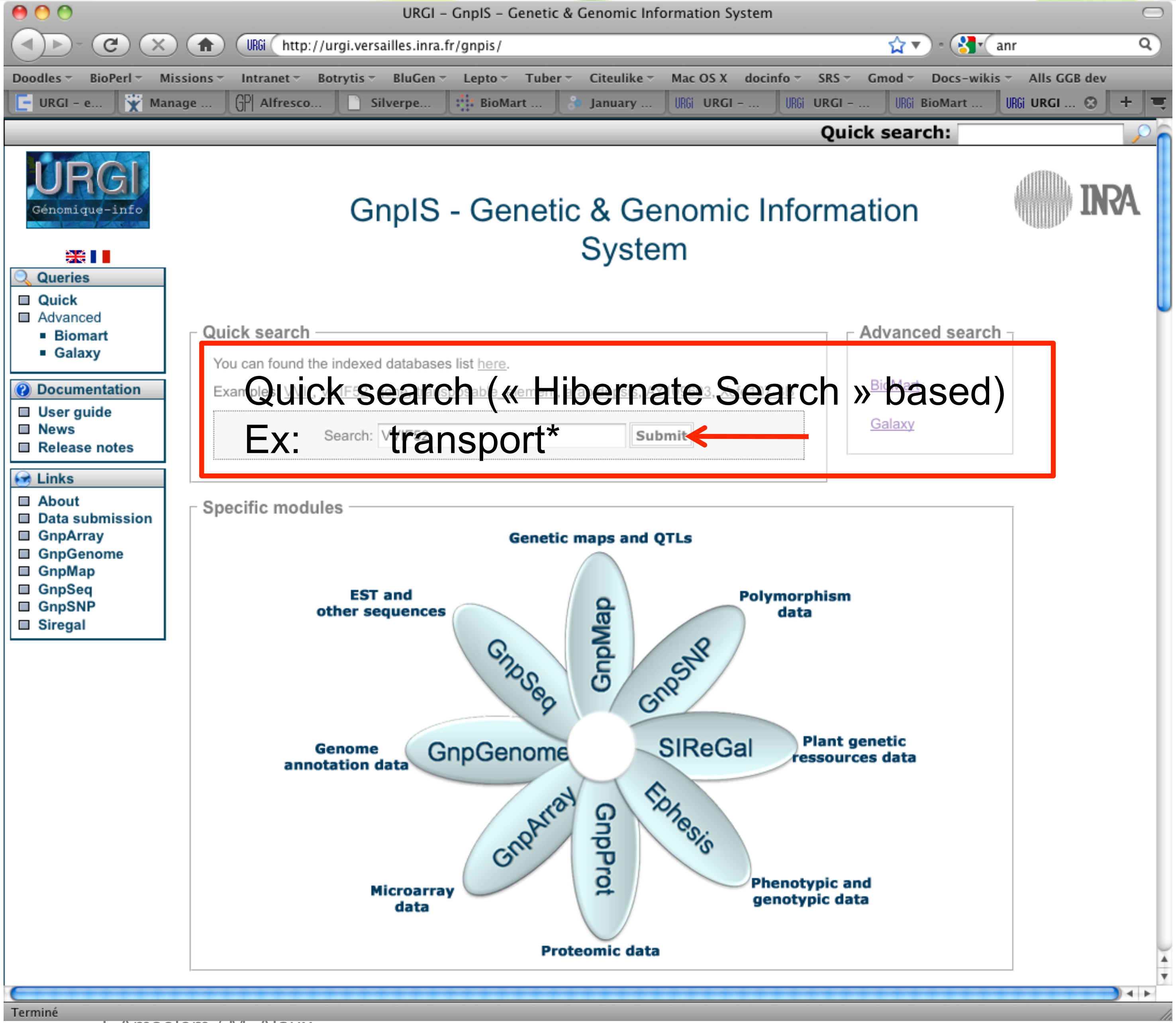
Classes modified :

java/apollo/gui/menus/EditMenu.java

java/apollo/editor/UserName.java

java/apollo/dataadapter/chado/jdbc/JdbcChadoAdapter.java

Quick and advanced search



GnpIS - Genetic & Genomic Information System

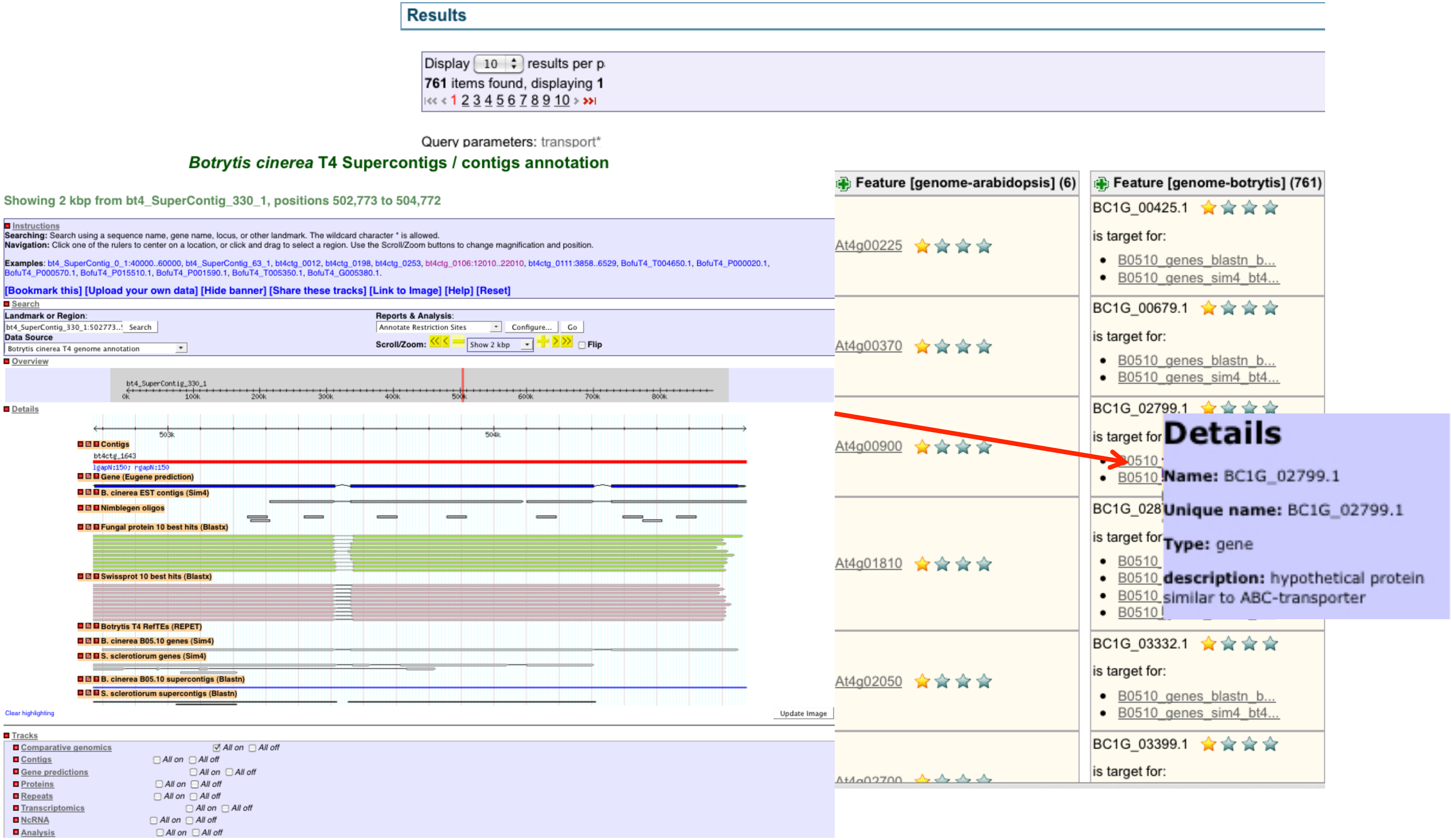
Quick search

You can found the indexed databases list here.

Quick search (« Hibernate Search » based)

Ex: transport*

Quick search : Hibernate Search



Results

Display 10 results per p

761 items found, displaying **1**

1 2 3 4 5 6 7 8 9 10

Query parameters: transport*

***Botrytis cinerea* T4 Supercontigs / contigs annotation**

Showing 2 kbp from bt4_SuperContig_330_1, positions 502,773 to 504,772

Instructions

Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed.

Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.

Examples: bt4_SuperContig_0_1:40000..60000, 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, BofuT4_P015510.1, BofuT4_P001590.1, BofuT4_T005350.1, BofuT4_G005380.1.

[Bookmark this] [Upload your own data] [Hide banner] [Share these tracks] [Link to Image] [Help] [Reset]

Search

Landmark or Region: bt4_SuperContig_330_1:502773..!

Data Source: Botrytis cinerea T4 genome annotation

Reports & Analysis: Annotate Restriction Sites

Scroll/Zoom: Show 2 kbp ☐ Flip

Tracks

- Comparative genomics ☑ All on ☐ All off
- Contigs ☐ All on ☐ All off
- Gene predictions ☐ All on ☐ All off
- Proteins ☐ All on ☐ All off
- Repeats ☐ All on ☐ All off
- Transcriptomics ☐ All on ☐ All off
- NcRNA ☐ All on ☐ All off
- Analysis ☐ All on ☐ All off

Feature [genome-arabidopsis] (6)	Feature [genome-botrytis] (761)
At4g00225 ★☆☆☆	BC1G_00425.1 ★☆☆☆ is target for: • B0510_genes_blastn_b... • B0510_genes_sim4_bt4...
At4g00370 ★☆☆☆	BC1G_00679.1 ★☆☆☆ is target for: • B0510_genes_blastn_b... • B0510_genes_sim4_bt4...
At4g00900 ★☆☆☆	BC1G_02799.1 ★☆☆☆ is target for • B0510 • B0510
At4g01810 ★☆☆☆	BC1G_028 is target for • B0510 • B0510 • B0510 • B0510
At4g02050 ★☆☆☆	BC1G_03332.1 ★☆☆☆ is target for: • B0510_genes_blastn_b... • B0510_genes_sim4_bt4...
At4g02700 ★☆☆☆	BC1G_03399.1 ★☆☆☆ is target for:

Details

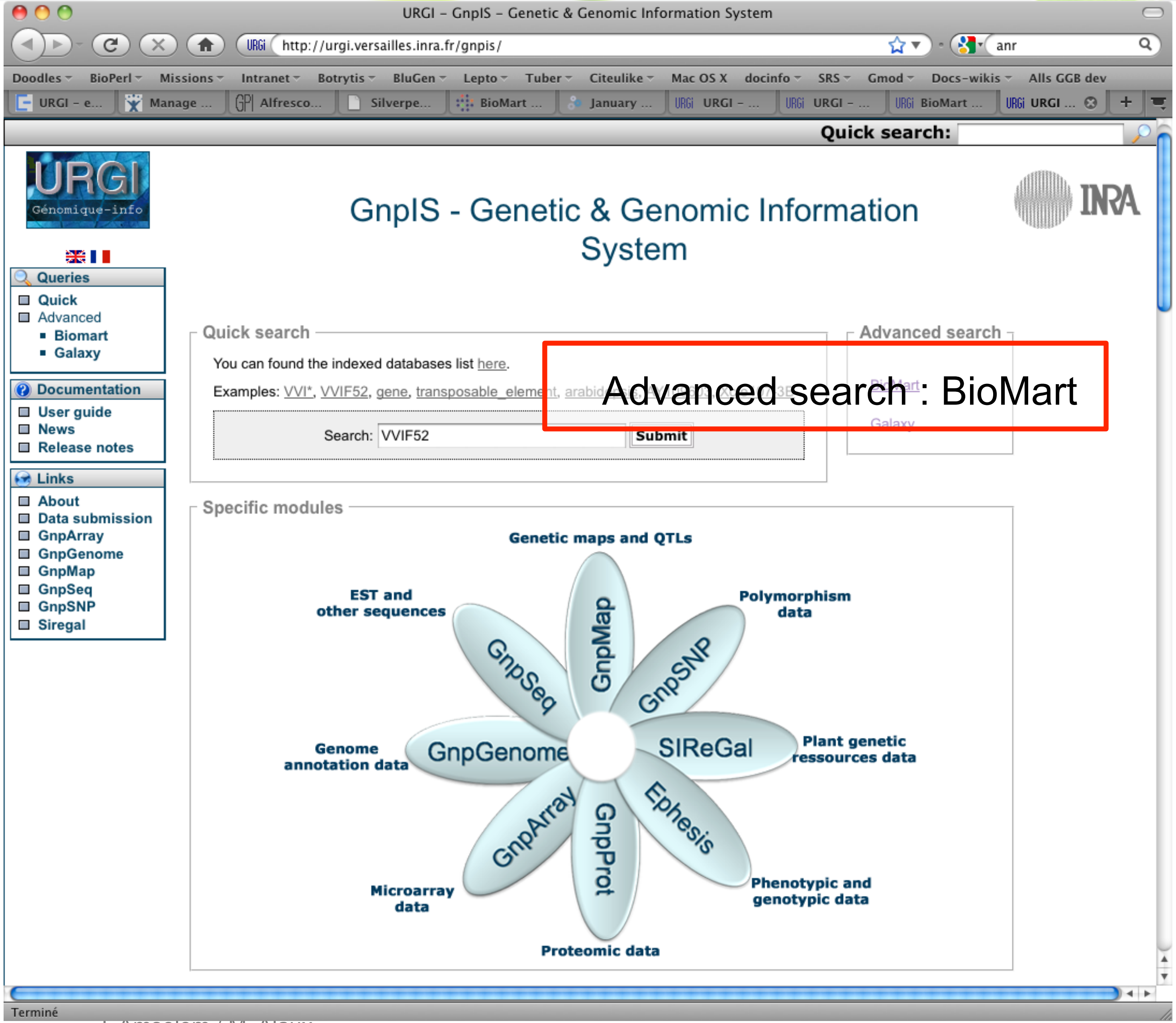
Name: BC1G_02799.1

Unique name: BC1G_02799.1

Type: gene

description: hypothetical protein similar to ABC-transporter

Quick and advanced search



Quick search:

GnpIS - Genetic & Genomic Information System

Quick search

You can found the indexed databases list here.

Examples: VVI*, VVIF52, gene, transposable_element, arabid...

Search: VVIF52

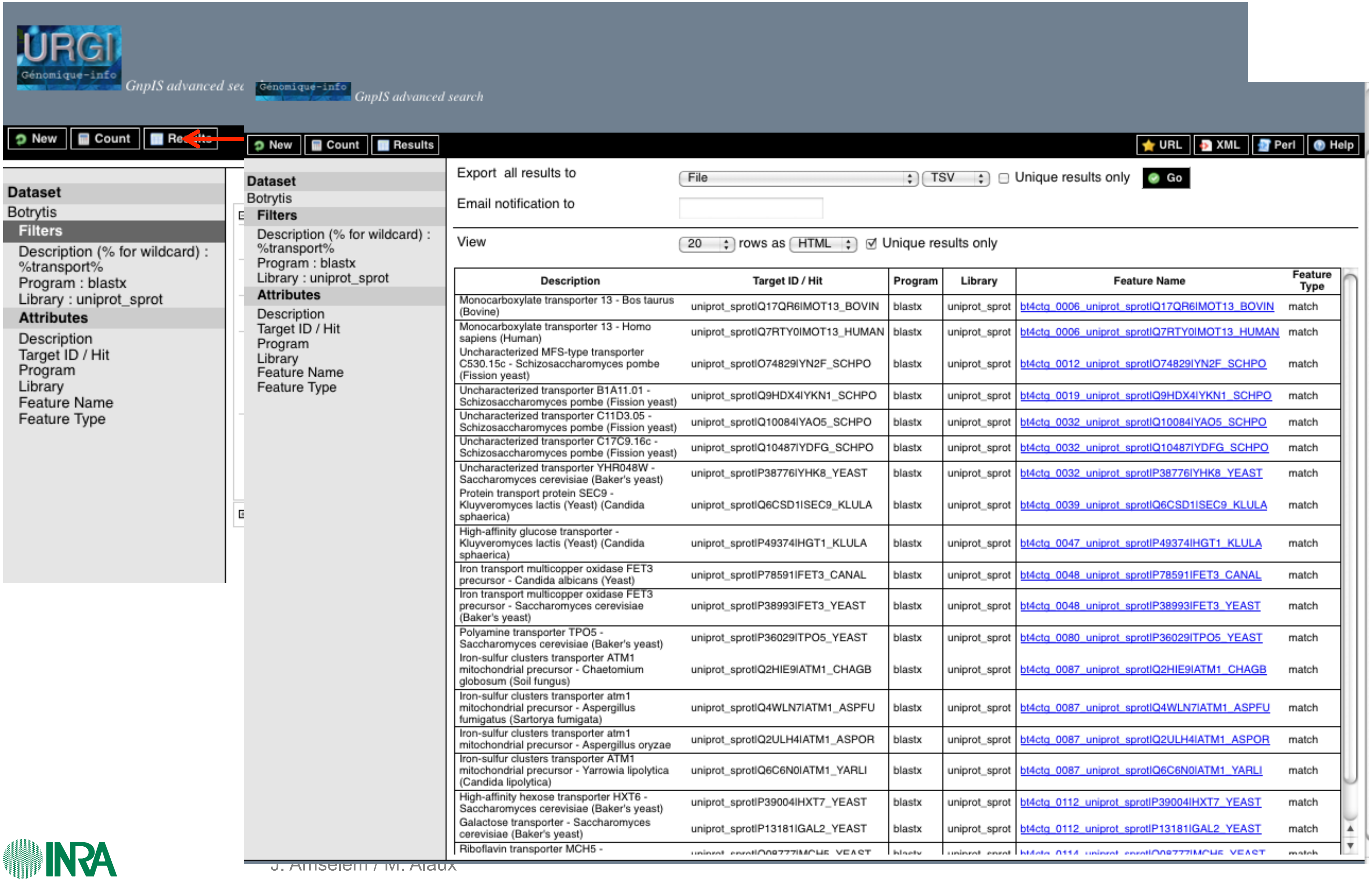
Advanced search

Specific modules

Advanced search : BioMart

J. Amselem / M. Alaux

Advanced search : BioMart



Dataset
Botrytis
Filters
Description (% for wildcard) : %transport%
Program : blastx
Library : uniprot_sprot
Attributes
Description
Target ID / Hit
Program
Library
Feature Name
Feature Type

Dataset
Botrytis
Filters
Description (% for wildcard) : %transport%
Program : blastx
Library : uniprot_sprot
Attributes
Description
Target ID / Hit
Program
Library
Feature Name
Feature Type

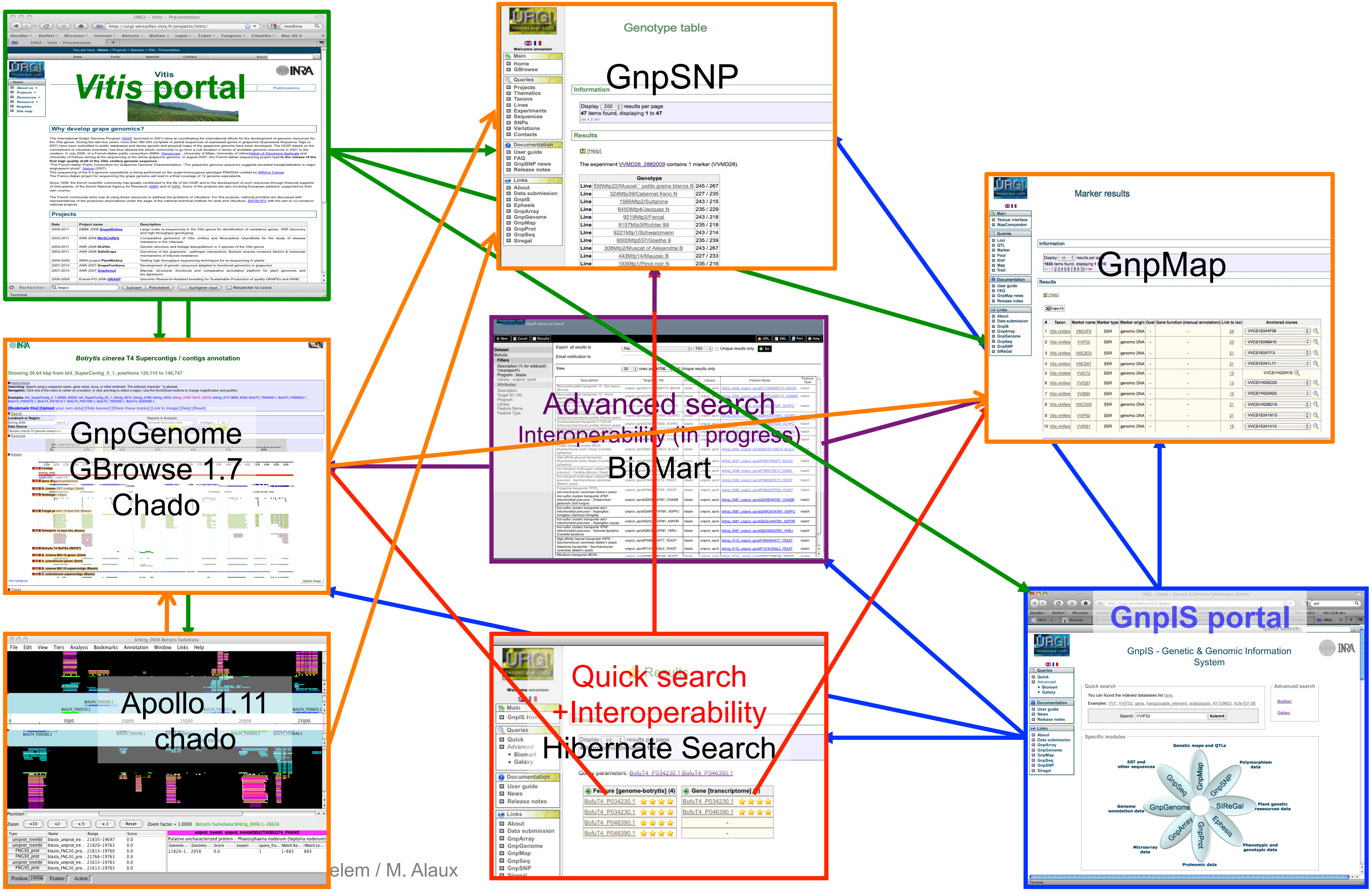
Export all results to File TSV ☐ Unique results only Go

Email notification to

View 20 rows as HTML ☑ Unique results only

Description	Target ID / Hit	Program	Library	Feature Name	Feature Type
Monocarboxylate transporter 13 - Bos taurus (Bovine)	uniprot_sprot|Q17QR6|MOT13_BOVIN	blastx	uniprot_sprot	bt4ctg_0006_uniprot_sprot|Q17QR6|MOT13_BOVIN	match
Monocarboxylate transporter 13 - Homo sapiens (Human)	uniprot_sprot|Q7RTY0|MOT13_HUMAN	blastx	uniprot_sprot	bt4ctg_0006_uniprot_sprot|Q7RTY0|MOT13_HUMAN	match
Uncharacterized MFS-type transporter C530.15c - Schizosaccharomyces pombe (Fission yeast)	uniprot_sprot|O74829|YN2F_SCHPO	blastx	uniprot_sprot	bt4ctg_0012_uniprot_sprot|O74829|YN2F_SCHPO	match
Uncharacterized transporter B1A11.01 - Schizosaccharomyces pombe (Fission yeast)	uniprot_sprot|Q9HDX4|YKN1_SCHPO	blastx	uniprot_sprot	bt4ctg_0019_uniprot_sprot|Q9HDX4|YKN1_SCHPO	match
Uncharacterized transporter C11D3.05 - Schizosaccharomyces pombe (Fission yeast)	uniprot_sprot|Q10084|YAO5_SCHPO	blastx	uniprot_sprot	bt4ctg_0032_uniprot_sprot|Q10084|YAO5_SCHPO	match
Uncharacterized transporter C17C9.16c - Schizosaccharomyces pombe (Fission yeast)	uniprot_sprot|Q10487|YDFG_SCHPO	blastx	uniprot_sprot	bt4ctg_0032_uniprot_sprot|Q10487|YDFG_SCHPO	match
Uncharacterized transporter YHR048W - Saccharomyces cerevisiae (Baker's yeast)	uniprot_sprot|P38776|YHK8_YEAST	blastx	uniprot_sprot	bt4ctg_0032_uniprot_sprot|P38776|YHK8_YEAST	match
Protein transport protein SEC9 - Kluyveromyces lactis (Yeast) (Candida sphaerica)	uniprot_sprot|Q6CSD1|SEC9_KLULA	blastx	uniprot_sprot	bt4ctg_0039_uniprot_sprot|Q6CSD1|SEC9_KLULA	match
High-affinity glucose transporter - Kluyveromyces lactis (Yeast) (Candida sphaerica)	uniprot_sprot|P49374|HGT1_KLULA	blastx	uniprot_sprot	bt4ctg_0047_uniprot_sprot|P49374|HGT1_KLULA	match
Iron transport multicopper oxidase FET3 precursor - Candida albicans (Yeast)	uniprot_sprot|P78591|FET3_CANAL	blastx	uniprot_sprot	bt4ctg_0048_uniprot_sprot|P78591|FET3_CANAL	match
Iron transport multicopper oxidase FET3 precursor - Saccharomyces cerevisiae (Baker's yeast)	uniprot_sprot|P38993|FET3_YEAST	blastx	uniprot_sprot	bt4ctg_0048_uniprot_sprot|P38993|FET3_YEAST	match
Polyamine transporter TPO5 - Saccharomyces cerevisiae (Baker's yeast)	uniprot_sprot|P36029|TPO5_YEAST	blastx	uniprot_sprot	bt4ctg_0080_uniprot_sprot|P36029|TPO5_YEAST	match
Iron-sulfur clusters transporter ATM1 mitochondrial precursor - Chaetomium globosum (Soil fungus)	uniprot_sprot|Q2HIE9|ATM1_CHAGB	blastx	uniprot_sprot	bt4ctg_0087_uniprot_sprot|Q2HIE9|ATM1_CHAGB	match
Iron-sulfur clusters transporter atm1 mitochondrial precursor - Aspergillus fumigatus (Sartorya fumigata)	uniprot_sprot|Q4WLN7|ATM1_ASPFU	blastx	uniprot_sprot	bt4ctg_0087_uniprot_sprot|Q4WLN7|ATM1_ASPFU	match
Iron-sulfur clusters transporter atm1 mitochondrial precursor - Aspergillus oryzae	uniprot_sprot|Q2ULH4|ATM1_ASPOR	blastx	uniprot_sprot	bt4ctg_0087_uniprot_sprot|Q2ULH4|ATM1_ASPOR	match
Iron-sulfur clusters transporter ATM1 mitochondrial precursor - Yarrowia lipolytica (Candida lipolytica)	uniprot_sprot|Q6C6N0|ATM1_YARLI	blastx	uniprot_sprot	bt4ctg_0087_uniprot_sprot|Q6C6N0|ATM1_YARLI	match
High-affinity hexose transporter HXT6 - Saccharomyces cerevisiae (Baker's yeast)	uniprot_sprot|P39004|HXT7_YEAST	blastx	uniprot_sprot	bt4ctg_0112_uniprot_sprot|P39004|HXT7_YEAST	match
Galactose transporter - Saccharomyces cerevisiae (Baker's yeast)	uniprot_sprot|P13181|GAL2_YEAST	blastx	uniprot_sprot	bt4ctg_0112_uniprot_sprot|P13181|GAL2_YEAST	match
Riboflavin transporter MCH5 -	uniprot_sprot|Q08777|MCH5_YEAST	blastx	uniprot_sprot	bt4ctg_0114_uniprot_sprot|Q08777|MCH5_YEAST	match

Usecase *Vitis vinifera*



Vitis portal

Genotype table

GnpSNP

Marker results

GnpMap

Advanced search
Interoperability (in progress)
BioMart

GnpGenome
GBrowse 1.7
Chado

Apollo 1.11
chado

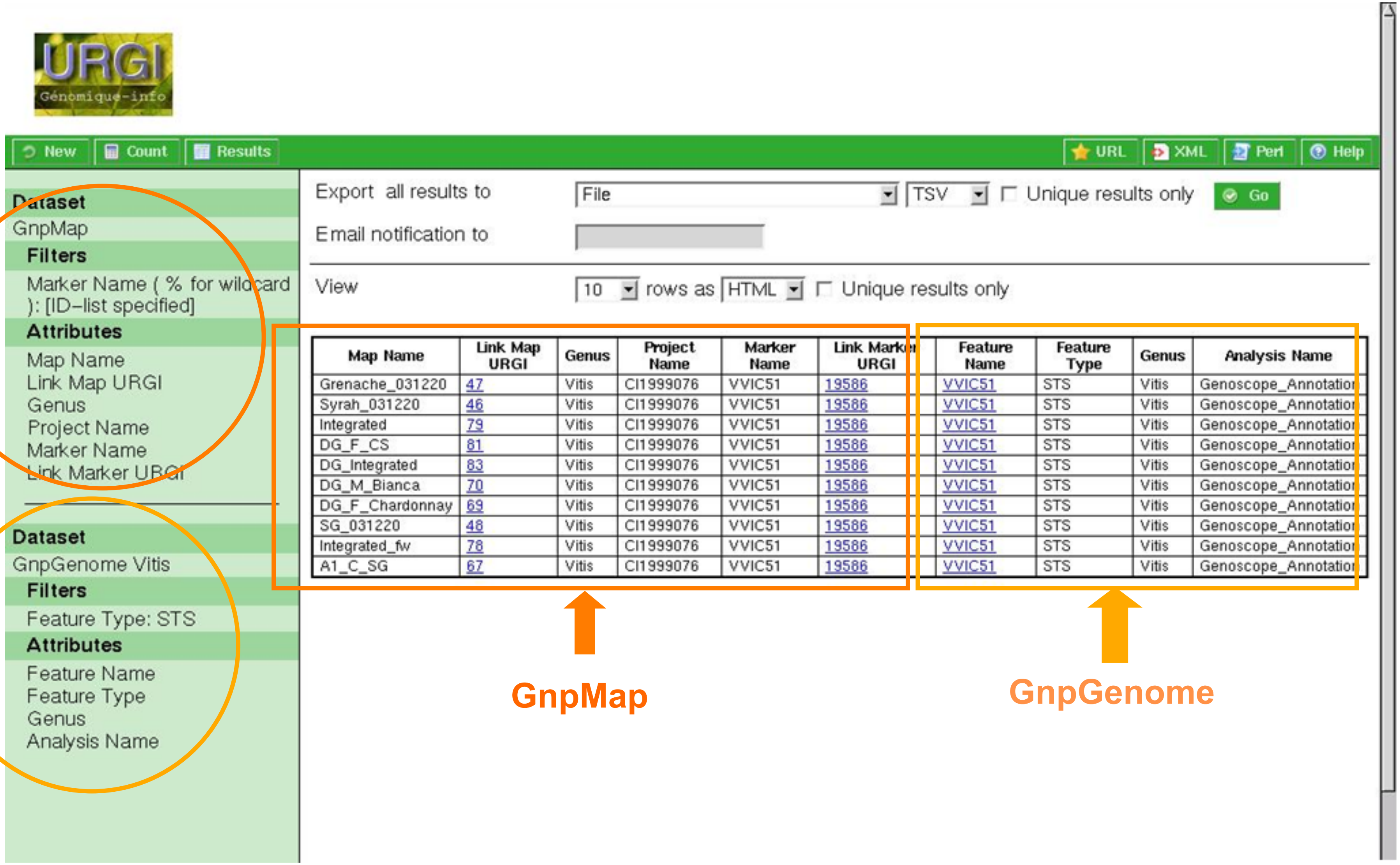
Quick search
+Interoperability
Hibernate Search

GnpIS portal

GnpIS - Genetic & Genomic Information System

elem / M. Alaux

BioMart based Interoperability
Usecase : *Vitis vinifera*



Dataset
GnpMap
Filters
Marker Name (% for wildcard): [ID–list specified]
Attributes
Map Name
Link Map URGI
Genus
Project Name
Marker Name
Link Marker URGI

Dataset
GnpGenome Vitis
Filters
Feature Type: STS
Attributes
Feature Name
Feature Type
Genus
Analysis Name

Export all results to File TSV ☐ Unique results only

Email notification to

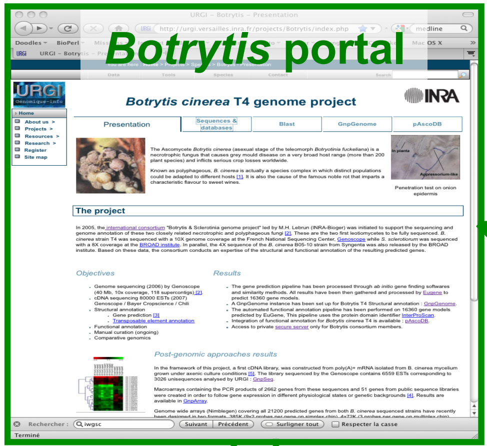
View 10 rows as HTML ☐ Unique results only

Map Name	Link Map URGI	Genus	Project Name	Marker Name	Link Marker URGI	Feature Name	Feature Type	Genus	Analysis Name
Grenache_031220	47	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
Syrah_031220	46	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
Integrated	79	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
DG_F_CS	81	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
DG_Integrated	83	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
DG_M_Bianca	70	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
DG_F_Chardonnay	69	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
SG_031220	48	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
Integrated_fw	78	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio
A1_C_SG	67	Vitis	CI1999076	VVIC51	19586	VVIC51	STS	Vitis	Genoscope_Annotatio

GnpMap

GnpGenome

Usecase *Botrytis cinerea*



Botrytis portal

Botrytis cinerea T4 genome project

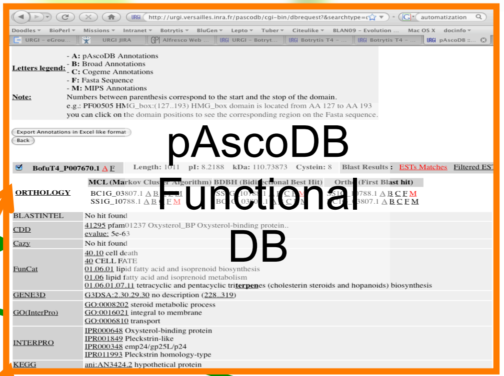
Presentation | Sequences & databases | Blast | GnpGenome | pAscoDB

The project

Objectives

Results

Post-genomic approaches results



pAscoDB

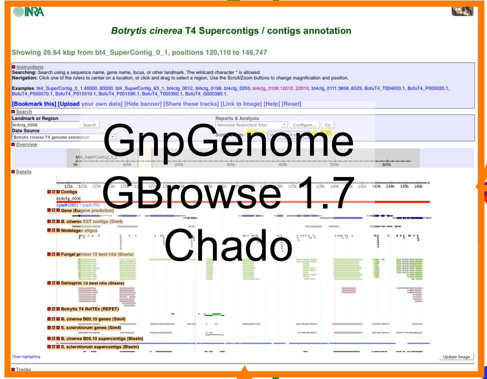
Functional DB

Letters legend:
- A: pAscoDB Annotations
- B: Broad Annotations
- C: Cogeme Annotations
- F: Fasta Sequence
- M: MIPS Annotations

Note: Numbers between parenthesis correspond to the start and the stop of the domain. e.g.: PF00505 HMG_box:(127..193) HMG_box domain is located from AA 127 to AA 193. you can click on the domain positions to see the corresponding region on the Fasta sequence.

BofuT4_P007670.1

ORTHOLOGY	
BLASTINTEL	No hit found
CDD	41295 pfam01237 Oxysterol_BP Oxysterol-binding protein. evalue: 5e-63
Cazy	No hit found
FunCat	40.10 cell death 40 CELL FATE 01.06.01 lipid fatty acid and isoprenoid biosynthesis 01.06 lipid fatty acid and isoprenoid metabolism 01.06.01.07.11 tetracyclic and pentacyclic triterpenes (cholesterin steroids and hopanoids) biosynthesis
GENE3D	G3DSA:2.30.29.30 no description (228..319)
GO(InterPro)	GO:0008202 steroid metabolic process GO:0016021 integral to membrane GO:0006810 transport
INTERPRO	IPR000648 Oxysterol-binding protein IPR001849 Pleckstrin-like IPR000348 emp24/gp25L/p24 IPR011993 Pleckstrin homology-type
KEGG	ani:AN3424.2 hypothetical protein



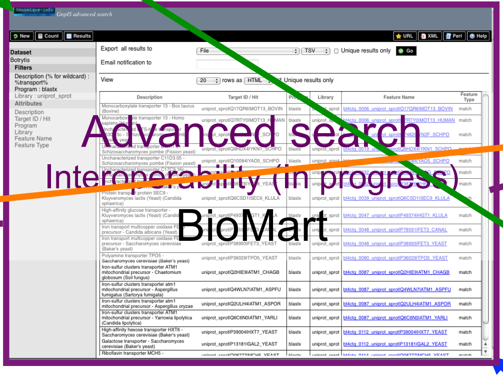
GnpGenome

GBrowse 1.7

Chado

Botrytis cinerea T4 Supercontigs / contigs annotation

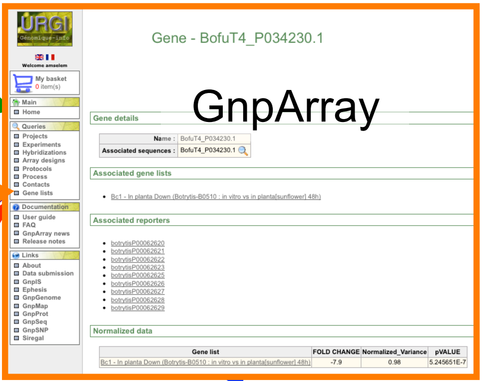
Showing 26.64 kbp from bt4_SuperContig_0_1, positions 120,110 to 146,747



Advanced search

Interoperability (in progress)

BioMart



Gene - BofuT4_P034230.1

GnpArray

Gene details

Name : BofuT4_P034230.1

Associated sequences : BofuT4_P034230.1

Associated gene lists

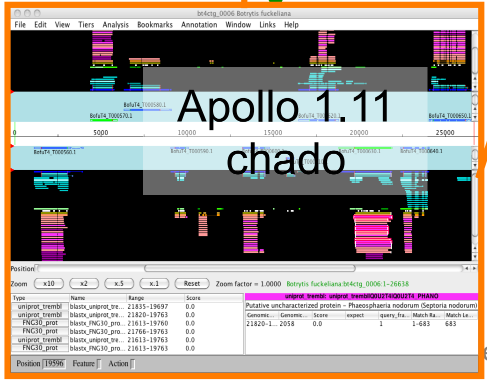
Bc1 - In planta Down (Botrytis-B0510 : in vitro vs in planta[sunflower] 48h)

Associated reporters

botrytisP00062620
botrytisP00062621
botrytisP00062622
botrytisP00062623
botrytisP00062625
botrytisP00062626
botrytisP00062627
botrytisP00062628
botrytisP00062629

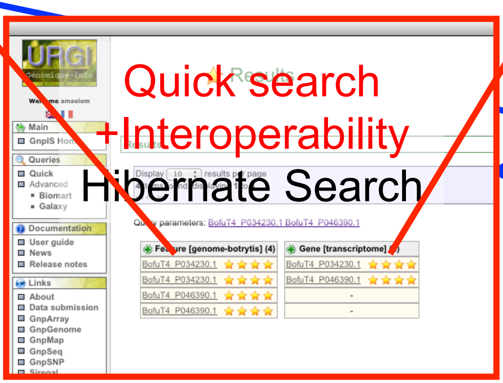
Normalized data

Gene list	FOLD CHANGE	Normalized_Variance	pVALUE
Bc1 - In planta Down (Botrytis-B0510 : in vitro vs in planta[sunflower] 48h)	-7.9	0.98	5.245651E-7



Apollo 1.11

chado



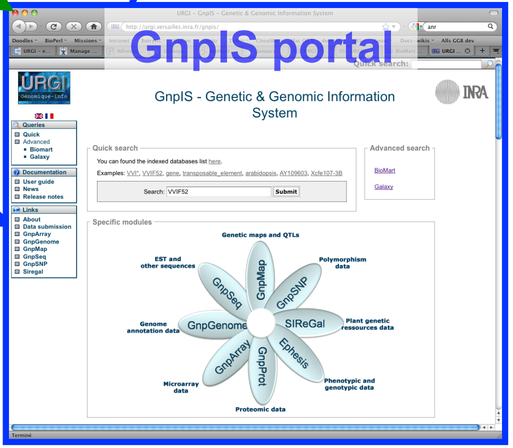
Quick search

+Interoperability

Hibernate Search

Query parameters: BofuT4_P034230.1 BofuT4_P046390.1

Feature [genome-botrytis] (4)	Gene [transcriptome] (2)
BofuT4_P034230.1 ★★★★	BofuT4_P034230.1 ★★★★
BofuT4_P034230.1 ★★★★	BofuT4_P046390.1 ★★★★
BofuT4_P046390.1 ★★★★	-
BofuT4_P046390.1 ★★★★	-



GnpIS portal

GnpIS - Genetic & Genomic Information System

Quick search

You can found the indexed databases list here.

Examples: VVI*, VVIF52, gene, transposable_element, arabidopsis, AY109603, Xcfe107-3B

Search: VVIF52 Submit

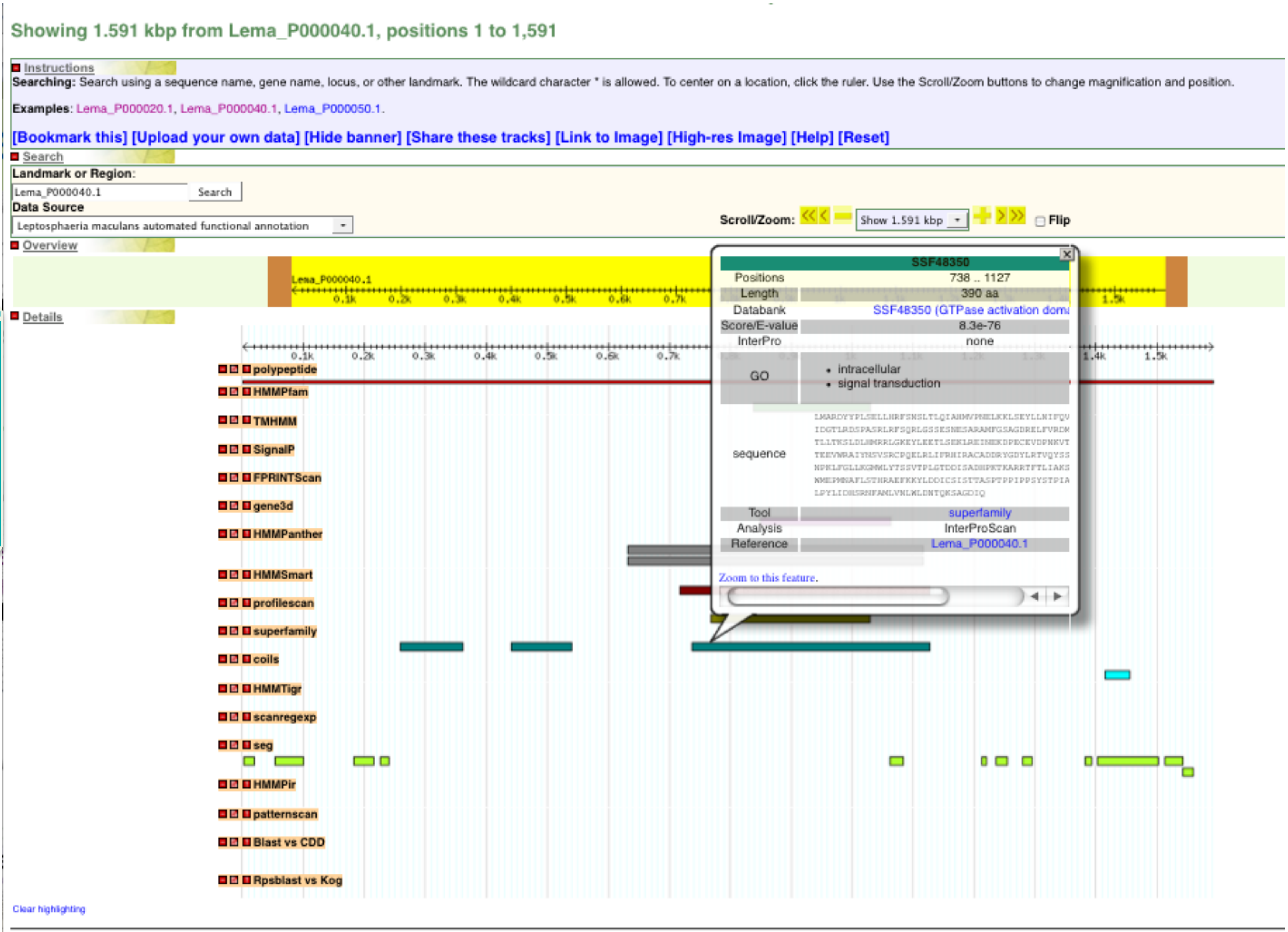
Advanced search

BioMart

Galaxy

Specific modules

Functional annotation
Usecase *Leptosphaeria maculans*



Showing 1.591 kbp from Lema_P000040.1, positions 1 to 1,591

Instructions

Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: Lema_P000020.1, Lema_P000040.1, Lema_P000050.1.

[Bookmark this] [Upload your own data] [Hide banner] [Share these tracks] [Link to Image] [High-res Image] [Help] [Reset]

Search

Landmark or Region: Lema_P000040.1 Search

Data Source: Leptosphaeria maculans automated functional annotation

Scroll/Zoom: Show 1.591 kbp ☐ Flip

Overview

Details

SSF48350	
Positions	738 .. 1127
Length	390 aa
Databank	SSF48350 (GTPase activation dom
Score/E-value	8.3e-76
InterPro	none
GO	• intracellular • signal transduction
Tool	superfamily
Analysis	InterProScan
Reference	Lema_P000040.1

Zoom to this feature.

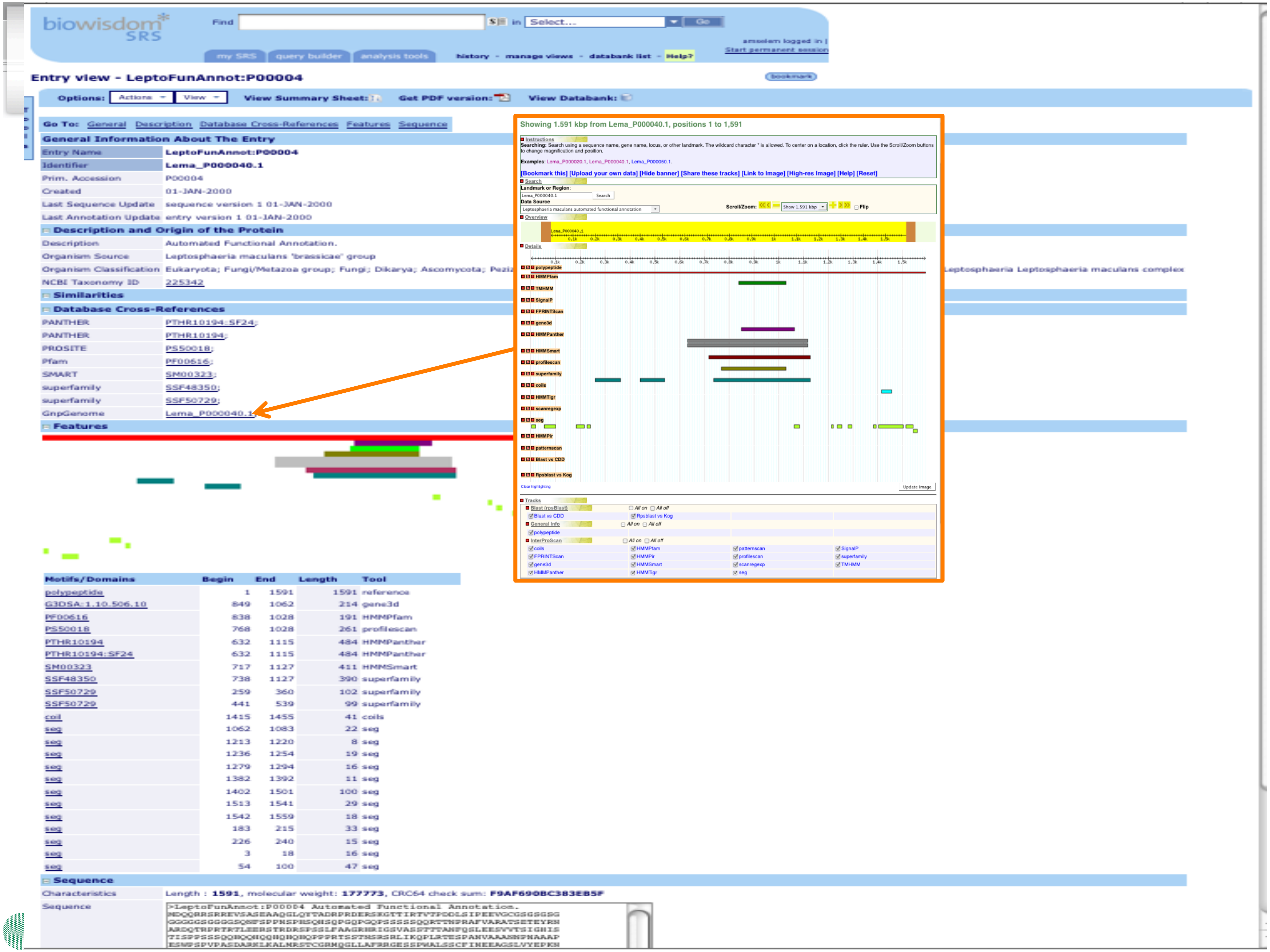
Chado Ref = polypeptide

Functional annotation
Usecase *Leptosphaeria maculans*

Lema_P000040.1

Length	1591 aa
CAZy	not identified as CAZy
Blast2GO	- Description: gtpase activating protein (bud2 cla2) - GO: GTPase activator activity - GO: intracellular - GO: regulation of small GTPase mediated signal transduction
sequence	

Link to URGI SRS environment



Entry view - LeptoFunAnnot:P00004

General Information About The Entry

Entry Name	LeptoFunAnnot:P00004
Identifier	Lema_P000040.1
Prim. Accession	P00004
Created	01-JAN-2000
Last Sequence Update	sequence version 1 01-JAN-2000
Last Annotation Update	entry version 1 01-JAN-2000

Description and Origin of the Protein

Description	Automated Functional Annotation.
Organism Source	Leptosphaeria maculans 'brassicae' group
Organism Classification	Eukaryota; Fungi/Metazoa group; Fungi; Dikarya; Ascomycota; Pezi Leptosphaeria Leptosphaeria maculans complex
NCBI Taxonomy ID	225342

Similarities

Database Cross-References

PANTHER	PTHR10194:SF24;
PANTHER	PTHR10194;
PROSITE	PS50018;
Pfam	PF00616;
SMART	SM00323;
superfamily	SSF48350;
superfamily	SSF50729;
GnpGenome	Lema_P000040.1

Features

Motifs/Domains	Begin	End	Length	Tool
polypeptide	1	1591	1591	reference
G3DSA:1.10.506.10	849	1062	214	gene3d
PF00616	838	1028	191	HMMPfam
PS50018	768	1028	261	profilescan
PTHR10194	632	1115	484	HMMPanther
PTHR10194:SF24	632	1115	484	HMMPanther
SM00323	717	1127	411	HMMSmart
SSF48350	738	1127	390	superfamily
SSF50729	259	360	102	superfamily
SSF50729	441	539	99	superfamily
coil	1415	1455	41	coils
seg	1062	1083	22	seg
seg	1213	1220	8	seg
seg	1236	1254	19	seg
seg	1279	1294	16	seg
seg	1382	1392	11	seg
seg	1402	1501	100	seg
seg	1513	1541	29	seg
seg	1542	1559	18	seg
seg	183	215	33	seg
seg	226	240	15	seg
seg	3	18	16	seg
seg	54	100	47	seg

Sequence

Characteristics	Length : **1591**, molecular weight: **177773**, CRC64 check sum: **F9AF690BC383EB5F**
Sequence	>LeptoFunAnnot:P00004 Automated Functional Annotation.

Showing 1.591 kbp from Lema_P000040.1, positions 1 to 1,591

Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.

Examples: Lema_P000020.1, Lema_P000040.1, Lema_P000050.1.

[Bookmark this] [Upload your own data] [Hide banner] [Share these tracks] [Link to Image] [High-res Image] [Help] [Reset]

Landmark or Region: Lema_P000040.1

Data Source: Leptosphaeria maculans automated functional annotation

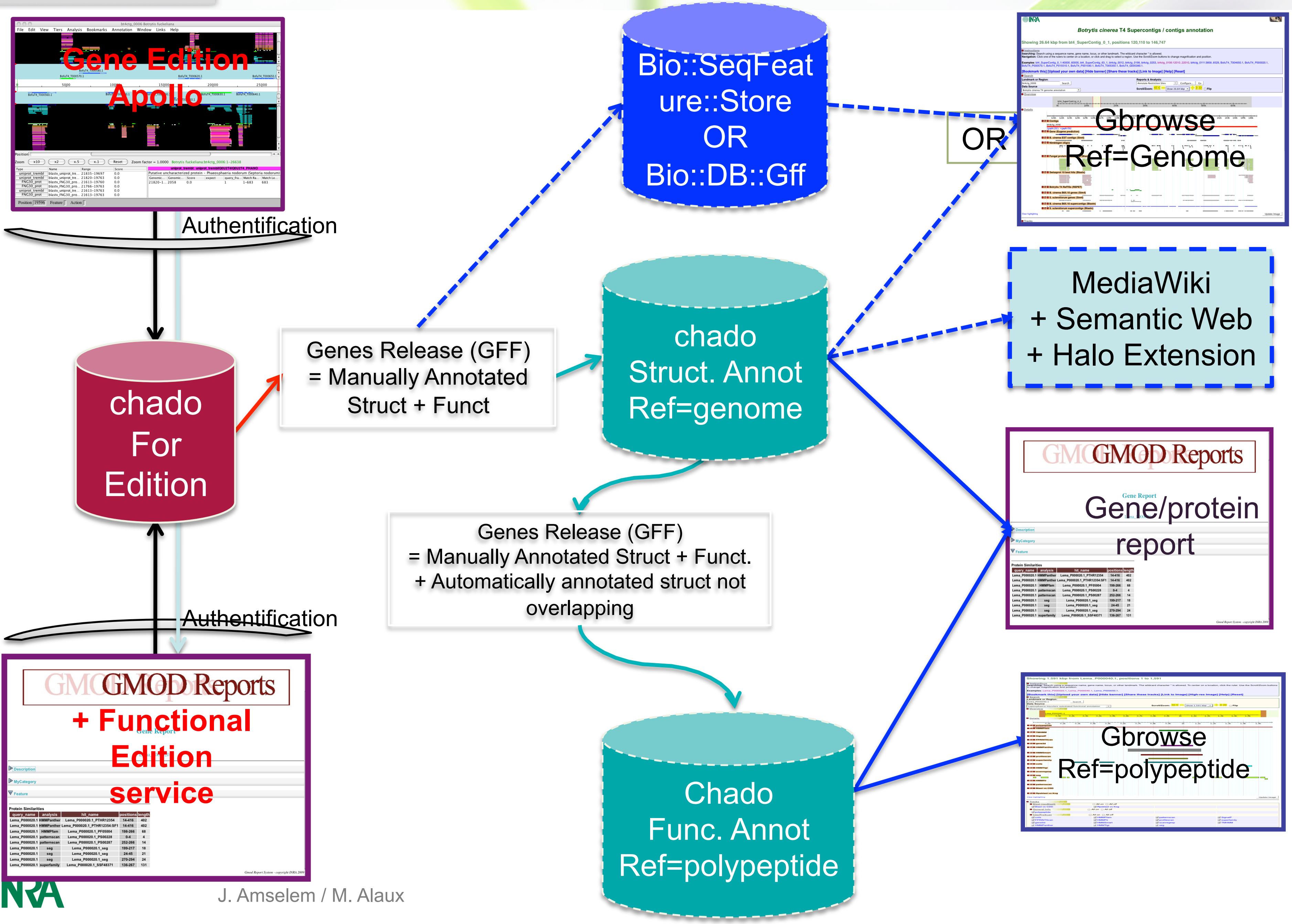
Scroll/Zoom: Show 1.591 kbp ☐ Flip

Blast (rpsBlast) ☐ All on ☐ All off
☑ Blast vs CDD ☑ Rpsblast vs Kog

General Info ☐ All on ☐ All off
☑ polypeptide

InterProScan ☐ All on ☐ All off
☑ coils ☑ HMMPfam ☑ patternscan ☑ SignalP
☑ FPRINTScan ☑ HMMPir ☑ profilescan ☑ superfamily
☑ gene3d ☑ HMMSmart ☑ scanregexp ☑ TMHMM
☑ HMMPanther ☑ HMMTigr ☑ seg

Perspectives…



Gene Edition
Apollo

Authentification

chado
For
Edition

Authentification

GMOD Reports
+ Functional
Edition
service

Genes Release (GFF)
= Manually Annotated
Struct + Funct

Bio::SeqFeature::Store
OR
Bio::DB::Gff

chado
Struct. Annot
Ref=genome

Genes Release (GFF)
= Manually Annotated Struct + Funct.
+ Automatically annotated struct not overlapping

Chado
Func. Annot
Ref=polypeptide

OR

Gbrowse
Ref=Genome

MediaWiki
+ Semantic Web
+ Halo Extension

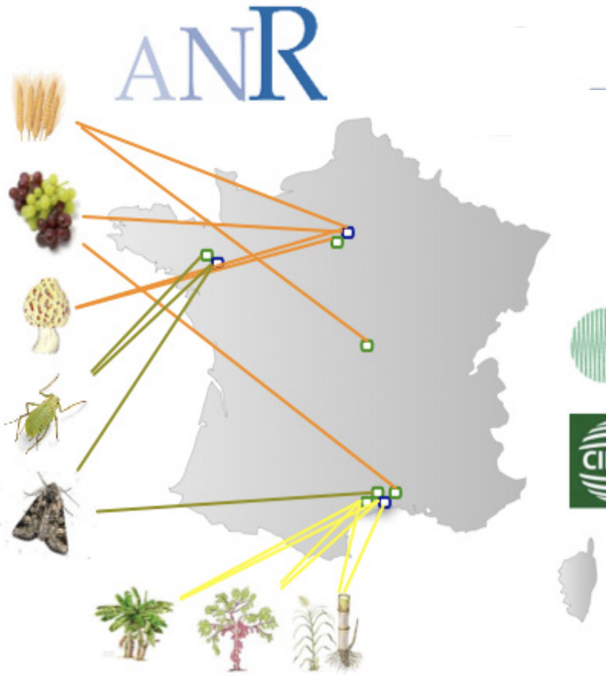
GMOD Reports
Gene/protein
report

Gbrowse
Ref=polypeptide

Aknowledgements

- INRA – URGI/BIOGER
  - Michaël Alaux
  - Joelle Amselem
  - Baptiste Brault
  - Nathalie Choisne
  - Erik Kimmel
  - Veronique Jamilloux
  - Aminah Keliet
  - Nicolas Lapalu
  - Isabelle Luyten
  - Nacer Mohelebi
  - Sebastien Reboux
  - Delphine Steinbach
  - Daphné Verdelet
  - Hadi Quesneville
- INRA – BIO3P
  - Fabrice Legeai
  - Denis Tagu
- KORILOG
  - Patrick durand
  - Goulven Kerbellec

GnpAnnot Funding



- CIRAD
  - Stéphanie Sidibe Bocs
  - Gaëtan Drocs
  - Valentin Guignon
  - Manuel Ruiz
- Bioversity
  - Mathieu Rouard

GMOD Report

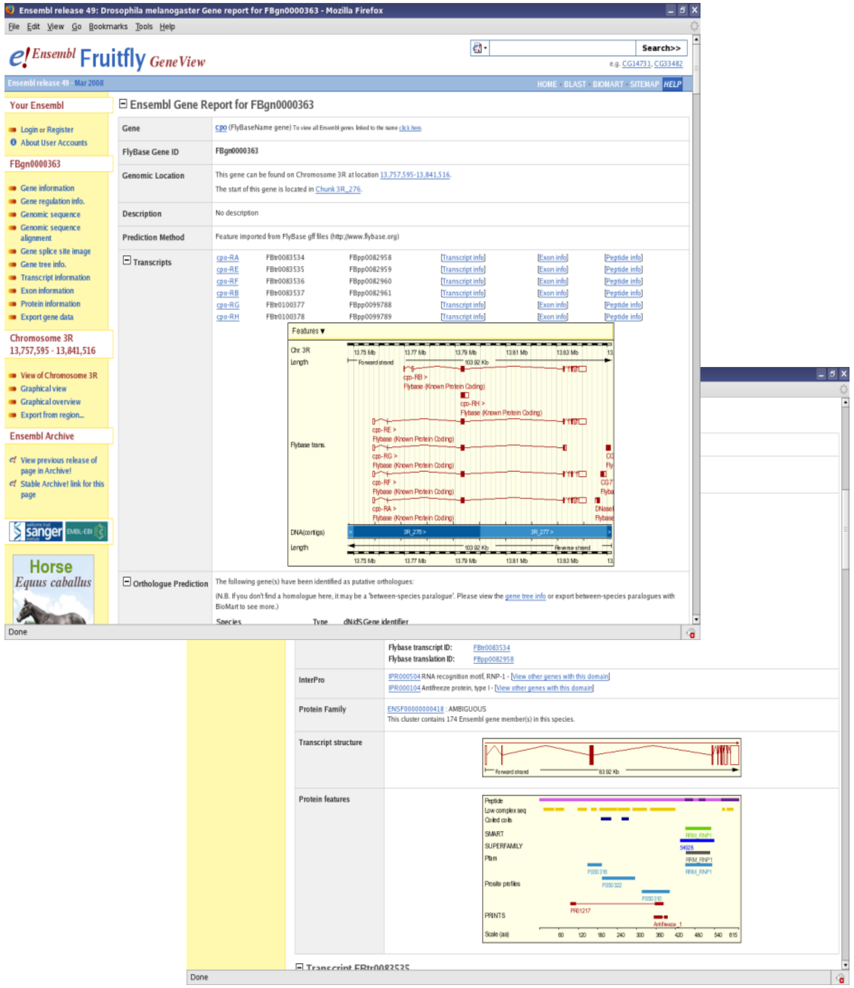
GNPAnnot project
`www.gnpannot.org`

GMOD Community

Fabrice Legeai (INRA, Bio3P)

Presentation by Michael Alaux (INRA, URGI)

GMOD Report



Ensembl release 49: Drosophila melanogaster Gene report for FBgn0000363 - Mozilla Firefox

e! Ensembl Fruitfly GeneView

Search>>

e.g. CG14731, CG33482

Ensembl release 49 - Mar 2008

HOME · BLAST · BIOMART · SITEMAP HELP

Your Ensembl

- Login or Register
- About User Accounts

FBgn0000363

- Gene information
- Gene regulation info.
- Genomic sequence
- Genomic sequence alignment
- Gene splice site image
- Gene tree info.
- Transcript information
- Exon information
- Protein information
- Export gene data

Chromosome 3R
13,757,595 - 13,841,516

- View of Chromosome 3R
- Graphical view
- Graphical overview
- Export from region...

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Horse
Equus caballus

Ensembl Gene Report for FBgn0000363

Gene	cpo (FlyBaseName gene) To view all Ensembl genes linked to the name click here.
FlyBase Gene ID	FBgn0000363
Genomic Location	This gene can be found on Chromosome 3R at location 13,757,595-13,841,516. The start of this gene is located in Chunk 3R_276.
Description	No description
Prediction Method	Feature imported from FlyBase gff files (http://www.flybase.org)

Transcripts

cpo-RA	FBtr0083534	FBpp0082958	[Transcript info]	[Exon info]	[Peptide info]
cpo-RE	FBtr0083535	FBpp0082959	[Transcript info]	[Exon info]	[Peptide info]
cpo-RF	FBtr0083536	FBpp0082960	[Transcript info]	[Exon info]	[Peptide info]
cpo-RB	FBtr0083537	FBpp0082961	[Transcript info]	[Exon info]	[Peptide info]
cpo-RG	FBtr0100377	FBpp0099788	[Transcript info]	[Exon info]	[Peptide info]
cpo-RH	FBtr0100378	FBpp0099789	[Transcript info]	[Exon info]	[Peptide info]

Orthologue Prediction

The following gene(s) have been identified as putative orthologues:

(N.B. If you don't find a homologue here, it may be a 'between-species paralogue'. Please view the gene tree info or export between-species paralogues with BioMart to see more.)

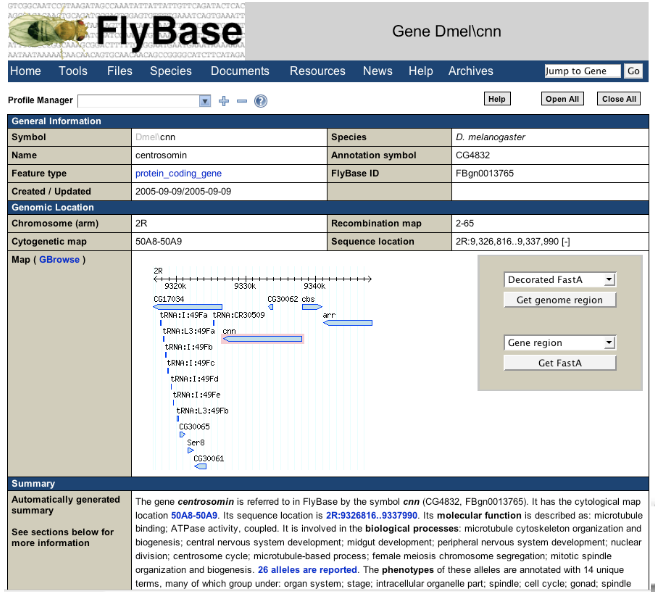
Species Type dN/dS Gene identifier

Done

	Flybase transcript ID: FBtr0083534 Flybase translation ID: FBpp0082958
InterPro	IPR000504 RNA recognition motif, RNP-1 - [View other genes with this domain] IPR000104 Antifreeze protein, type I - [View other genes with this domain]
Protein Family	ENSF00000000418 : AMBIGUOUS This cluster contains 174 Ensembl gene member(s) in this species.
Transcript structure	
Protein features	

Transcript FBtr0083535

Done



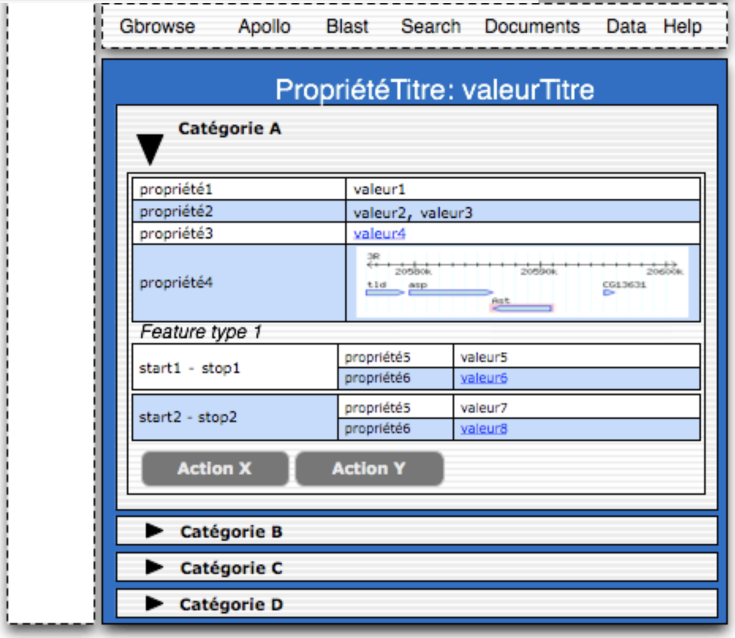
FlyBase

Gene Dmel\cnn

General Information			
Symbol	Dmel\cnn	Species	*D. melanogaster*
Name	centrosomin	Annotation symbol	CG4832
Feature type	protein_coding_gene	FlyBase ID	FBgn0013765
Created / Updated	2005-09-09/2005-09-09		
Genomic Location			
Chromosome (arm)	2R	Recombination map	2-65
Cytogenetic map	50A8-50A9	Sequence location	2R:9,326,816..9,337,990 [-]
Map (GBrowse)			

Summary

Automatically generated summary See sections below for more information	The gene ***centrosomin*** is referred to in FlyBase by the symbol ***cnn*** (CG4832, FBgn0013765). It has the cytological map location **50A8-50A9**. Its sequence location is **2R:9326816..9337990**. Its **molecular function** is described as: microtubule binding; ATPase activity, coupled. It is involved in the **biological processes**: microtubule cytoskeleton organization and biogenesis; central nervous system development; midgut development; peripheral nervous system development; nuclear division; centrosome cycle; microtubule-based process; female meiosis chromosome segregation; mitotic spindle organization and biogenesis. **26 alleles are reported**. The **phenotypes** of these alleles are annotated with 14 unique terms, many of which group under: organ system; stage; intracellular organelle part; spindle; cell cycle; gonad; spindle



PropriétéTitre: valeurTitre

▼ Catégorie A

propriété1	valeur1
propriété2	valeur2, valeur3
propriété3	valeur4
propriété4	

Feature type 1

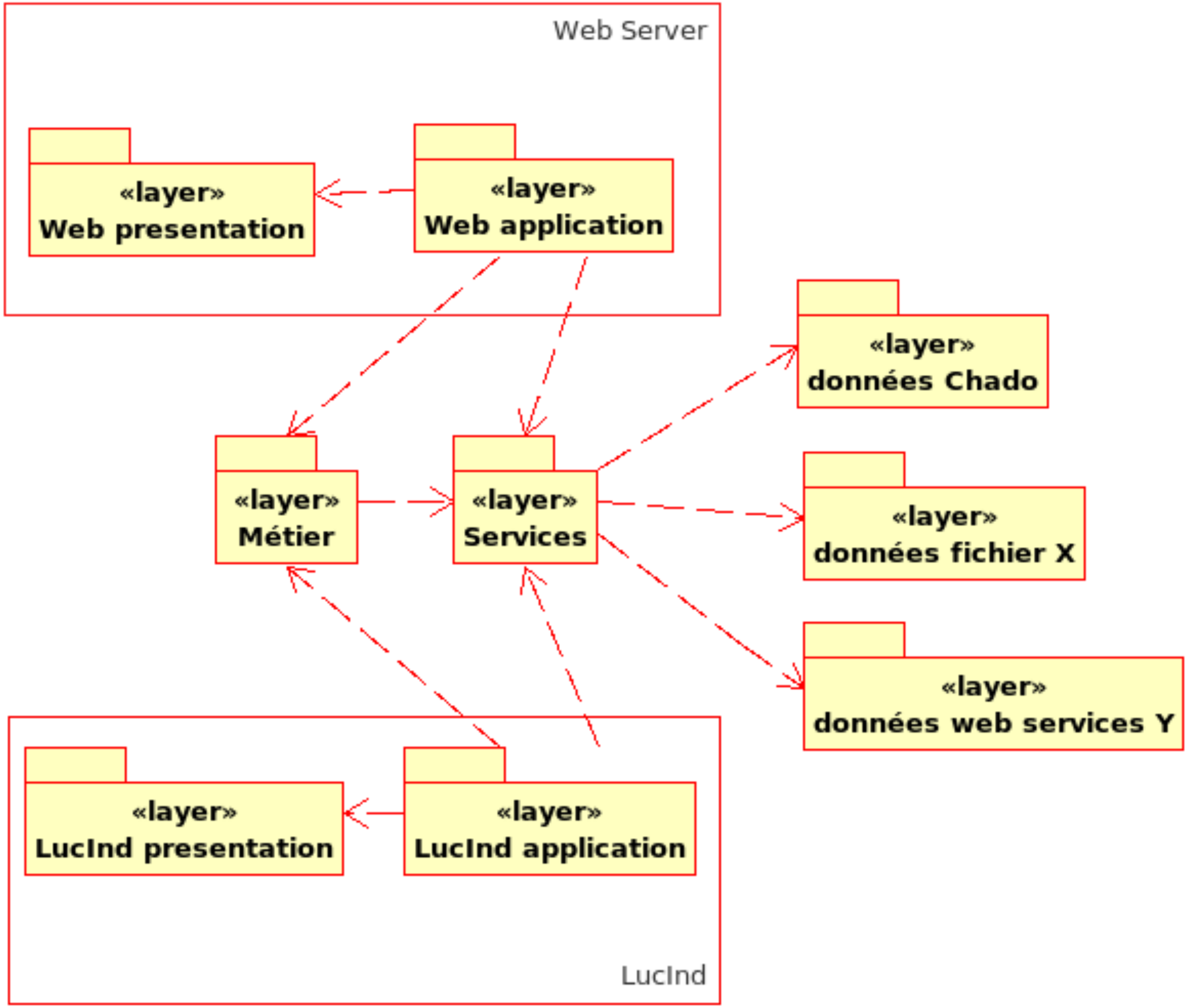
start1 - stop1	propriété5	valeur5
	propriété6	valeur6
start2 - stop2	propriété5	valeur7
	propriété6	valeur8

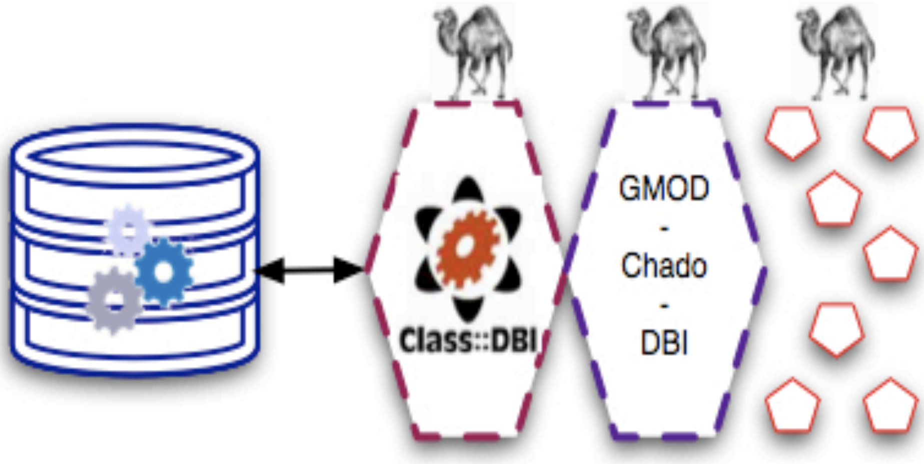
▶ Catégorie B

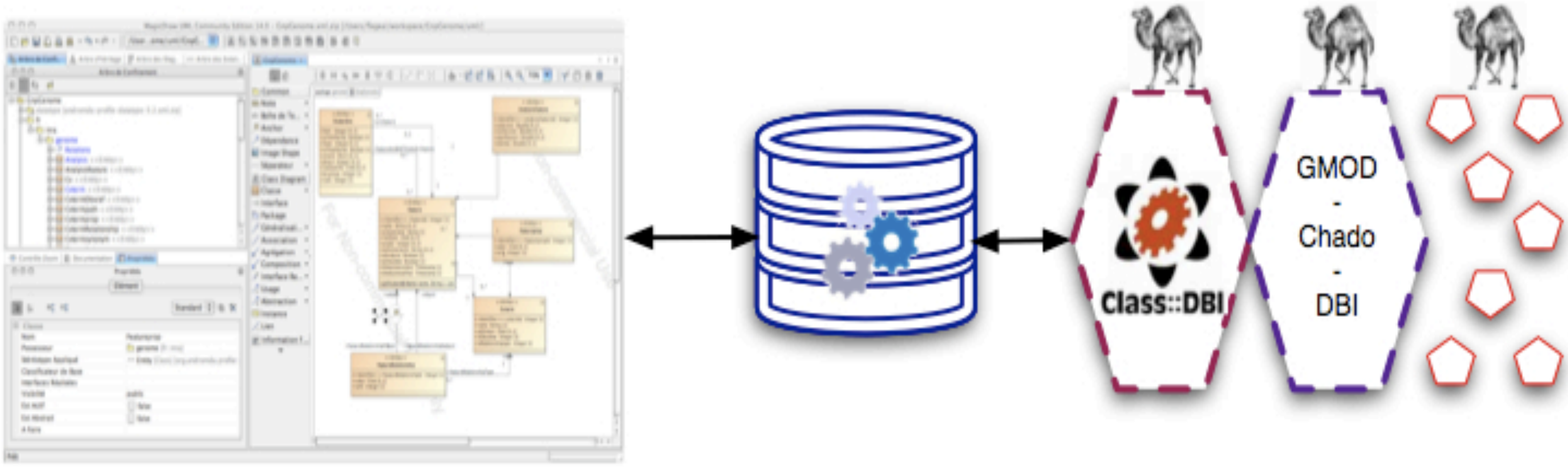
▶ Catégorie C

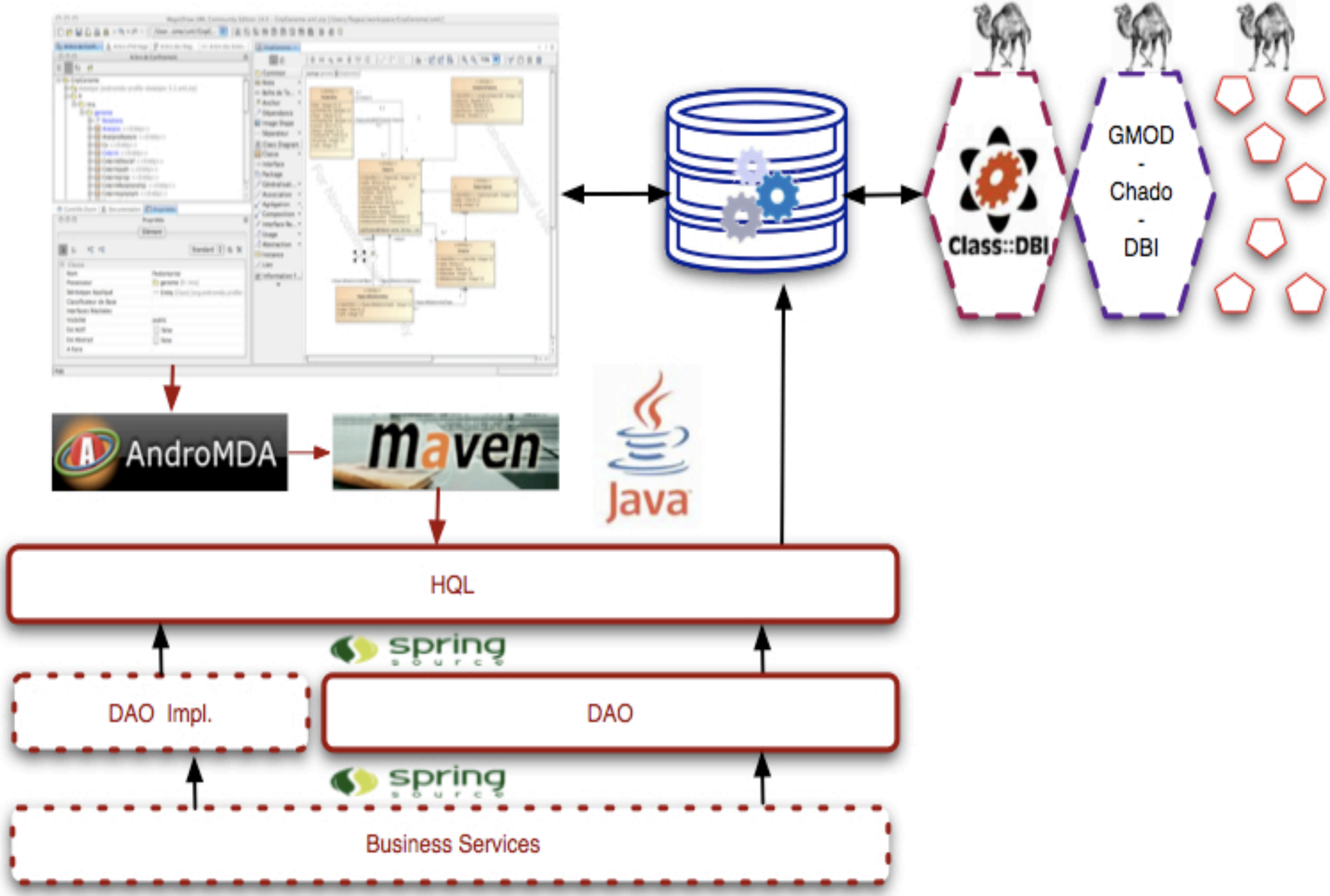
▶ Catégorie D

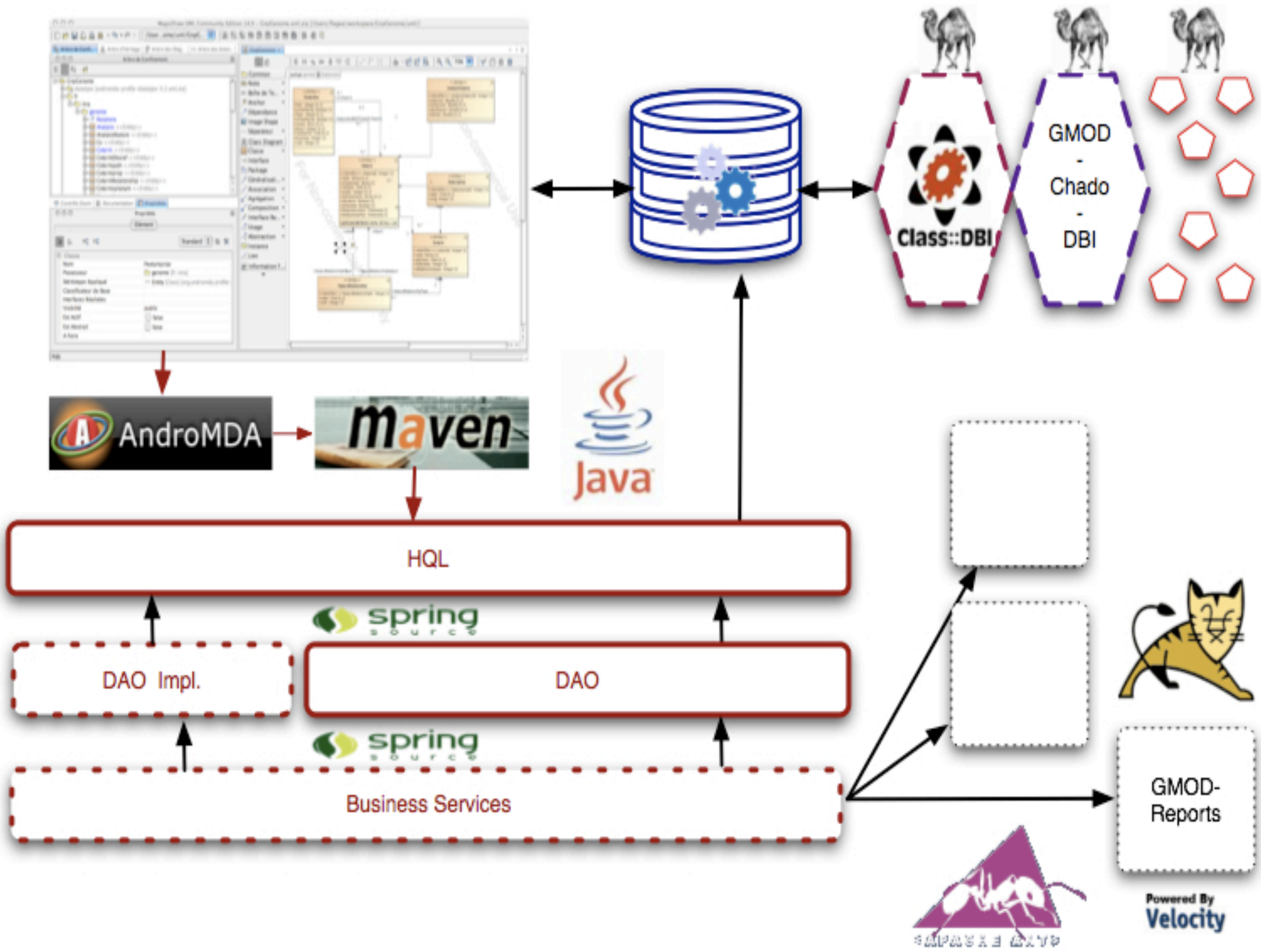
GMOD Report


Web Server
«layer» Web presentation
«layer» Web application
«layer» données Chado
«layer» Métier
«layer» Services
«layer» données fichier X
«layer» données web services Y
«layer» LucInd presentation
«layer» LucInd application
LucInd


Class::DBI
GMOD
-
Chado
-
DBI


Class::DBI
GMOD
-
Chado
-
DBI


AndroMDA
maven
Java
Class::DBI
GMOD
-
Chado
-
DBI
HQL
spring source
DAO Impl.
DAO
spring source
Business Services



Class::DBI

GMOD
-
Chado
-
DBI

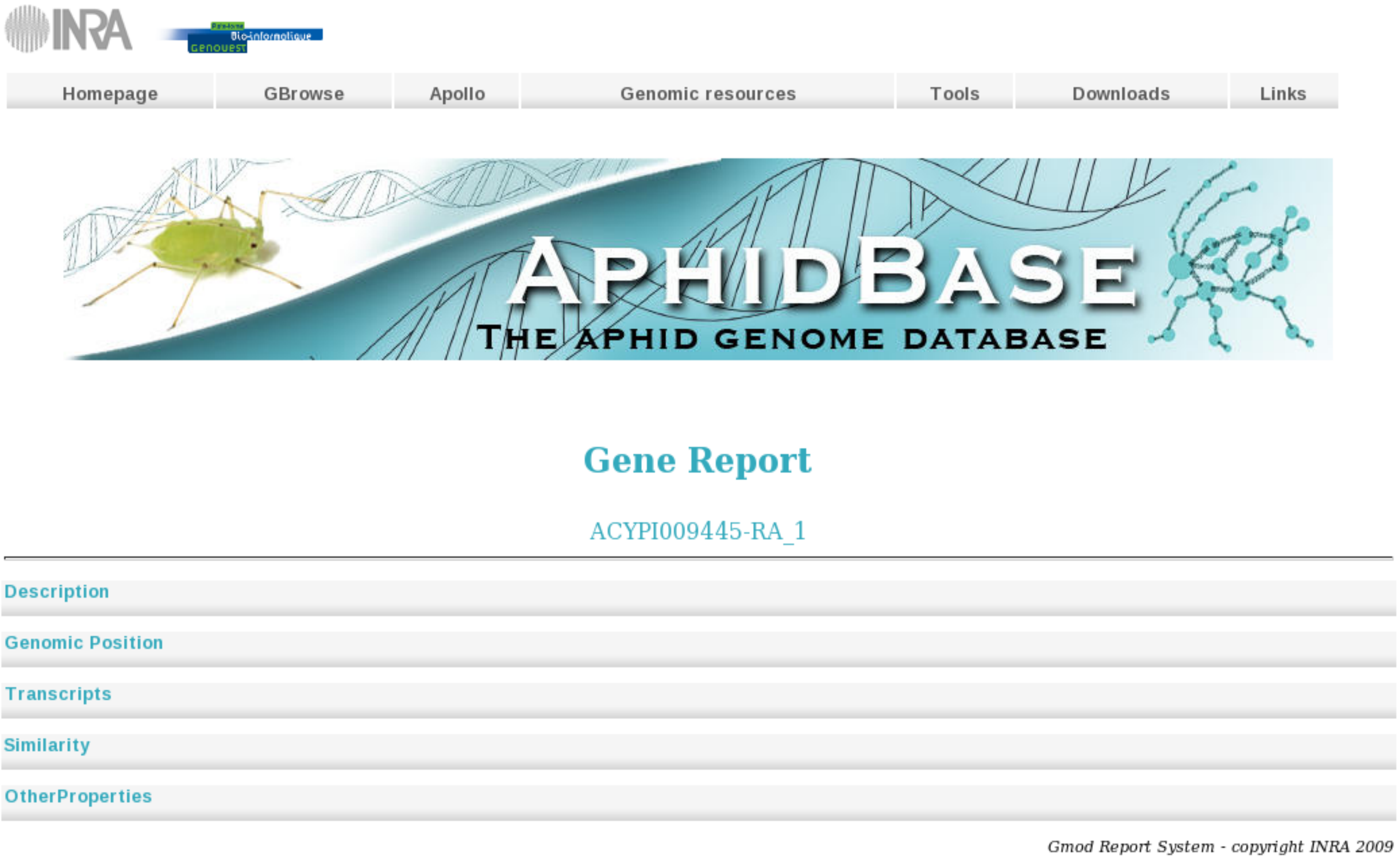
HQL

DAO Impl.

DAO

Business Services

GMOD-
Reports



APHIDBASE

THE APHID GENOME DATABASE

Gene Report

ACYPI009445-RA_1

Description

Genomic Position

Transcripts

Similarity

OtherProperties

Gmod Report System - copyright INRA 2009

http://webapps1.genouest.org/grs-1.6/grs?reportID=aphidbase_genome_report&objectID=ACYPI006415-RA#.

Gene Report

ACYPI006147-RA_1

Description

ID	ApisOBP2-transcript
Name	ApisOBP2-transcript
Synonym	ACYPI006147-RA_1
Synonym	ApisOBP2-transcript
Synonym	ACYPIG179180-jjzhou-RA
owner	Name : Jing-Jiang Zhou Institute : Rothamsted Research
comments	putative odorant-binding protein 2.

Genomic Position

Transcripts

Similarity

OtherProperties

Gene Report

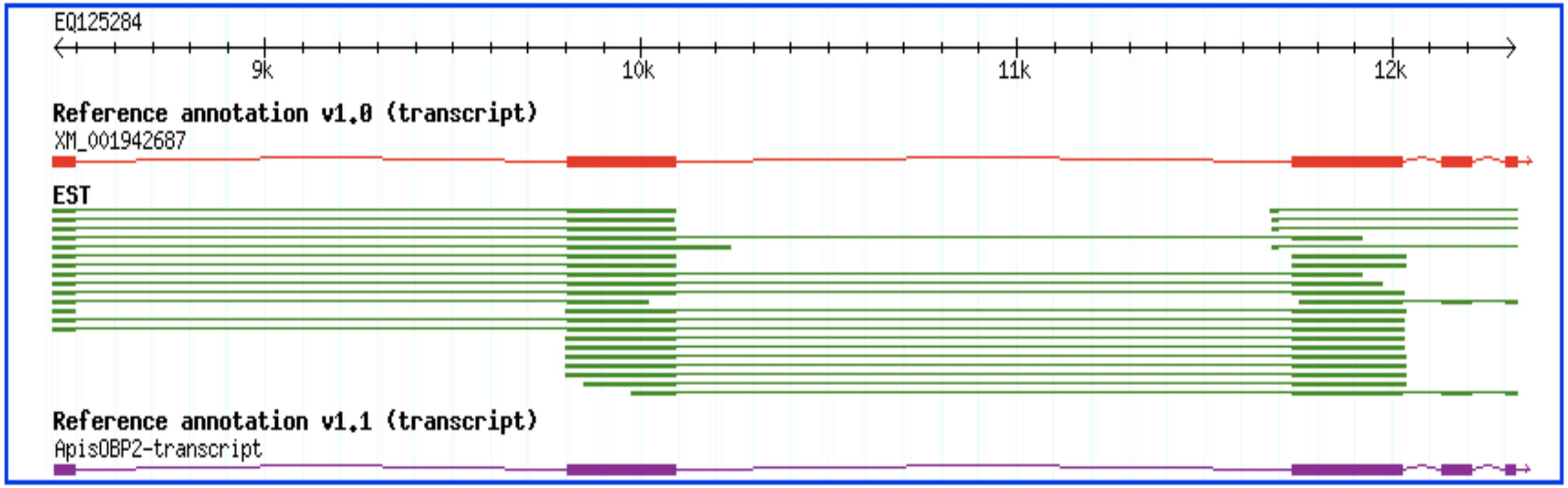
ACYPI006147-RA_1

▬ Description

ID	ApisOBP2-transcript
Name	ApisOBP2-transcript
Synonym	ACYPIG179180-jjzhou-RA
Synonym	ApisOBP2-transcript
Synonym	ACYPI006147-RA_1

+ Genomic Position

▬ Gbrowse



+ Transcripts

+ Similarity

Gene Report

ACYPI006147-RA_1

Description

ID	ApisOBP2-transcript
Name	ApisOBP2-transcript
Synonym	ACYPI006147-RA_1
Synonym	ApisOBP2-transcript
Synonym	ACYPIG179180-jjzhou-RA
owner	Name : Jing-Jiang Zhou Institute : Rothamsted Research
comments	putative odorant-binding protein 2.

Genomic Position

Start	Stop	Strand
8440	12325	PLUS

Scaffold
EQ125284

Transcripts

mRNA Name	Scaffold	Start	Stop	Strand
ApisOBP2-transcript	EQ125284	8440	12325	PLUS

```
>ApisOBP2-transcript_mRNA
ATGAAGGTATCTGCAGCGACCGCCGTCCTGGTCGCTCTGGTCGCCACCGTGCAGAGCTCGGACCCATGTAACATATCTACTTGCTACAAGAGCGGCACGA
CGAAGCCGCCAATGGCCGTTACGCCCACTCACCTGCCGGTACAGTCATCGTCCACCCAGACCAGCCACCCACAGACCACGTACGCCAAGGATCACGTGCA
CGGTTCAACCACCACCAAGTCCGGTGTCAATGCTACGGTCACGACAGCCAGCGGAGCGTCCGTCAACGGCACTGAACCGCCGGCCGTCGTCAAGTCTTCC
GCCGGAGTAACCGGAAACTCCACCACGCCCAAGCCTACAATGACCGAAGGACATGTGGCTCTGAAACAGAAGTTGAACACAATTGCGGTGAAGTGCAAGG
ACGAGCTGCACGCGCCCCAAGAGATCATGGCACTGGTCAGTAACACGGTAGTACCGCAGAACGAGCAACAAAGGTGCTACTTAGAGTGCGTGTACAAAAA
TCTTAATTTGATCAAAAATAACAAGTTCAGTGTGGAAGACGGCAAGGCGATGGCTAGAATACGCTTCGCAAATCAACCAGAAGAGCACAAGAAGGCAGTA
ACCATAATAGAAACTTGCGAAAAAGAAGCTGTTATCGATCCGAAAACCACTGAAAAATGCGCAGCTGGACGAGTAATCAGAAACTGCTTCGTCAAAAACG
GAGAAAAAATAAATTTCTTCCCTAAAGCATAA

>ApisOBP2-transcript_prot
MKVSAATAVLVALVATVQSSDPCNISTCYKSGTTKPPMAVTPTHLPVQSSSTQTSHPQTTYAKDHVHGSTTTKSGVNATVTTASGASVNGTEPPAVVKSS
AGVTGNSTTPKPTMTEGHVALKQKLNTIAVKCKDELHAPQEIMALVSNTVVPQNEQQRCYLECVYKNLNLIKNNKFSVEDGKAMARIRFANQPEEHKKAV
TIIETCEKEAVIDPKTTEKCAAGRVIRNCFVKNGEKINFFPKA
```

Exon	Start	Stop	Strand
ACYPIE9001-jjzhou	8440	8494	PLUS
ACYPIE9002-jjzhou	9804	10089	PLUS
ACYPIE9003-jjzhou	11734	12023	PLUS
ACYPIE9004-jjzhou	12130	12210	PLUS
ACYPIE9005-jjzhou	12301	12325	PLUS

Acyrthosiphon pisum	Whole body, nymphs and adults, winged and wingless	Pea aphid whole body normalized full length cDNA library	12456	13152	+	FF308174
Acyrthosiphon pisum	Whole body, nymphs and adults, winged and wingless	Pea aphid whole body normalized full length cDNA library	9976	12465	-	EX631171
Acyrthosiphon pisum	Whole body, nymphs and adults, winged and wingless	Pea aphid whole body normalized full length cDNA library	8221	11917	+	EX645090
Acyrthosiphon pisum	Whole body, nymphs and adults, winged and wingless	Pea aphid whole body normalized full length cDNA library	11754	12677	-	EX645235
Acyrthosiphon pisum	Whole body, nymphs and adults, winged and wingless	Pea aphid whole body normalized full length cDNA library	8246	10019	+	EX646992
Acyrthosiphon pisum	Whole body, nymphs and adults, winged and wingless	Pea aphid whole body normalized full length cDNA library	12335	12465	-	EX648195
Acyrthosiphon pisum	head	ApHL3SD	12696	13007	-	DY228241
Acyrthosiphon pisum	whole body, nymphs and adults	Acyrthosiphon pisum, Pea Aphid	9801	12032	+	CN582915
Acyrthosiphon pisum	whole body, nymphs and adults	Acyrthosiphon pisum, Pea Aphid	8283	8494	+	CN583930
Acyrthosiphon pisum	whole insect	ApMS	9801	12030	+	CN754376
Acyrthosiphon pisum	whole insect	ApMS	9801	12030	+	CN758381
Acyrthosiphon pisum	whole insect	ApMS	9801	12032	+	CN760196
Acyrthosiphon pisum	whole insect	ApMS	9801	12032	+	CN764432
Acyrthosiphon pisum	whole insect	ApMS	8231	12030	+	CN764449
Acyrthosiphon pisum	whole insect	ApMS	8283	12030	+	CN764329
Acyrthosiphon pisum	whole insect	ApMS	8283	12030	+	CN754121
Acyrthosiphon pisum	whole insect	ApMS	8231	11969	+	CN758303
Acyrthosiphon pisum	whole insect	ApMS	9801	12032	+	CN759740

Summary of organism

value	Number	frequency
Acyrthosiphon pisum	24	1.0

Summary of tissue_type

value	Number	frequency
whole insect	9	0.375
Whole body, nymphs and adults, winged and wingless	12	0.5
head	1	0.041666668
whole body, nymphs and adults	2	0.083333336

Summary of clone_lib

value	Number	frequency
Pea aphid whole body normalized full length cDNA library	12	0.5
ApHL3SD	1	0.041666668
Acyrthosiphon pisum, Pea Aphid	2	0.083333336
ApMS	9	0.375

Protein Similarities

query_name	analysis	hit_name	description	interpro_id	go	positions
XP_001942722	Gene3D	G3DSA:1.10.238.20	no description	IPR006170	Molecular Function: odorant binding (GO:0005549), Biological Process: transport (GO:0006810)	120-241
XP_001942722	HMMPfam	PF01395	PBP_GOBP	IPR006170	Molecular Function: odorant binding (GO:0005549), Biological Process: transport (GO:0006810)	113-235
XP_001942722	HMMSmart	SM00708	no description	IPR006625		127-234
XP_001942722	Seg	seg	seg			2-17,47-60
XP_001942722	superfamily	SSF47565	Insect pheromone/odorant-binding proteins	IPR006170	Molecular Function: odorant binding (GO:0005549), Biological Process: transport (GO:0006810)	113-241

OtherProperties

New services and perspectives

- More graphics and more statistics
- Cross reference towards paralogous and orthologous genes
- Gene ontology trees
- Publication References
- Add annotation by authorised curators
- Versionning

http://webapps1.genouest.org/grs-1.6/grs?reportID=aphidbase_genome_report&objectID=ACYPI006415-RA#.

Chado Controller

GNPAnnot project
`www.gnpannot.org`

GMOD Community

LA RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPEMENT

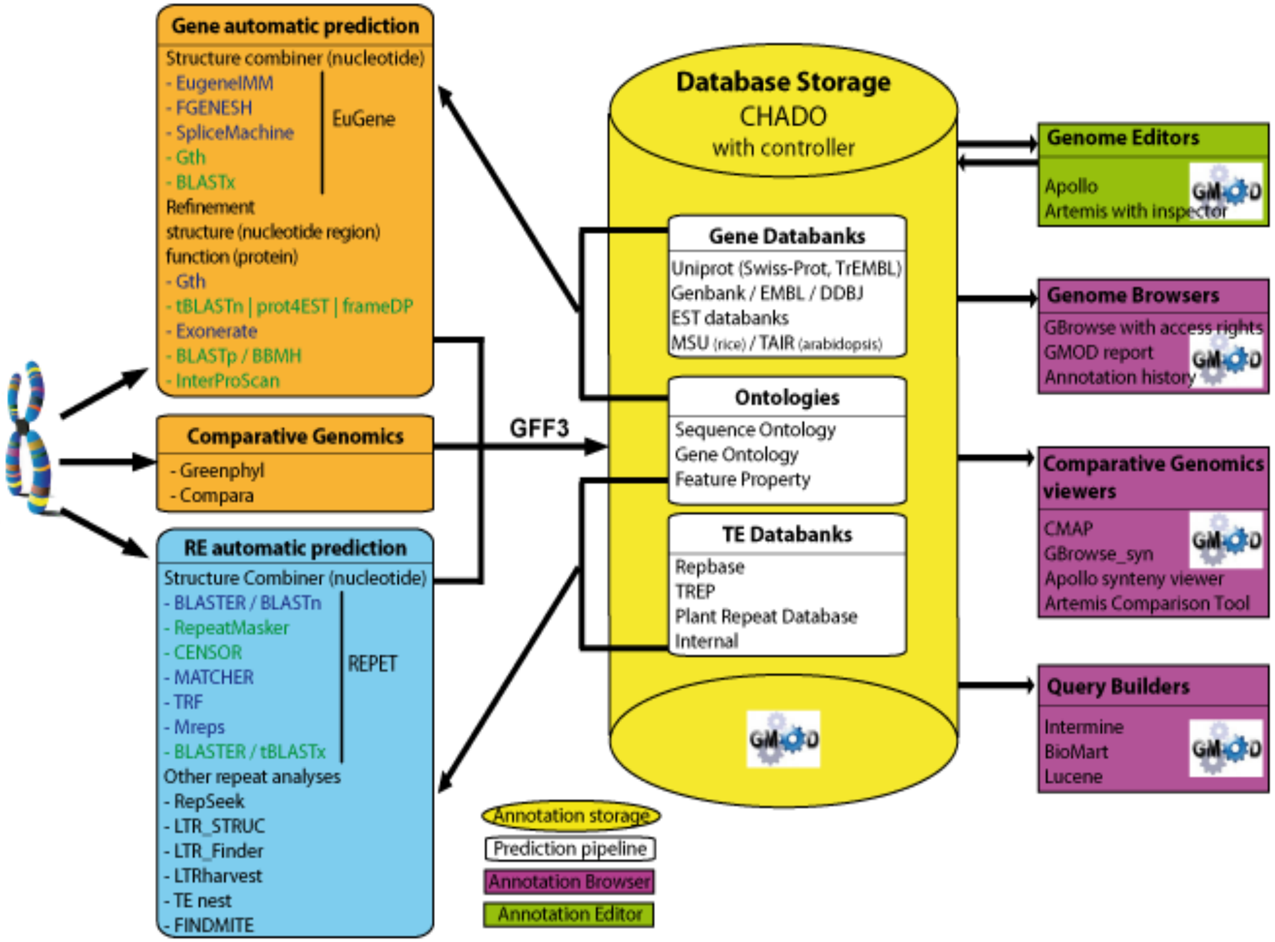
Valentin Guignon (CIRAD, UMR DAP)

Presentation by Michael Alaux (INRA, URGI)

Summary

1. Context
2. Annotator Access Restriction on Biological Features
    1. Architecture
    2. PostgreSQL Embedded Part
    3. Genome Browser Integration
3. History of Manual Annotations
    1. Architecture
    2. PostgreSQL Embedded Part
    3. Genome Browser Integration
4. Inspector of Manual Annotations
    1. Architecture
    2. PostgreSQL Embedded Part
    3. Genome Browser Integration
5. Conclusion

1. Context


Gene automatic prediction
Structure combiner (nucleotide)
- EugeneIMM
- FGENESH
- SpliceMachine
- Gth
- BLASTx
EuGene
Refinement
structure (nucleotide region)
function (protein)
- Gth
- tBLASTn | prot4EST | frameDP
- Exonerate
- BLASTp / BBMH
- InterProScan
Comparative Genomics
- Greenphyl
- Compara
RE automatic prediction
Structure Combiner (nucleotide)
- BLASTER / BLASTn
- RepeatMasker
- CENSOR
- MATCHER
- TRF
- Mreps
- BLASTER / tBLASTx
REPET
Other repeat analyses
- RepSeek
- LTR_STRUC
- LTR_Finder
- LTRharvest
- TE nest
- FINDMITE
GFF3
Database Storage
CHADO
with controller
Gene Databanks
Uniprot (Swiss-Prot, TrEMBL)
Genbank / EMBL / DDBJ
EST databanks
MSU (rice) / TAIR (arabidopsis)
Ontologies
Sequence Ontology
Gene Ontology
Feature Property
TE Databanks
Repbase
TREP
Plant Repeat Database
Internal
GMOD
Genome Editors
Apollo
Artemis with inspector
GMOD
Genome Browsers
GBrowse with access rights
GMOD report
Annotation history
GMOD
Comparative Genomics viewers
CMAP
GBrowse_syn
Apollo synteny viewer
Artemis Comparison Tool
GMOD
Query Builders
Intermine
BioMart
Lucene
GMOD
Annotation storage
Prediction pipeline
Annotation Browser
Annotation Editor

1. Context

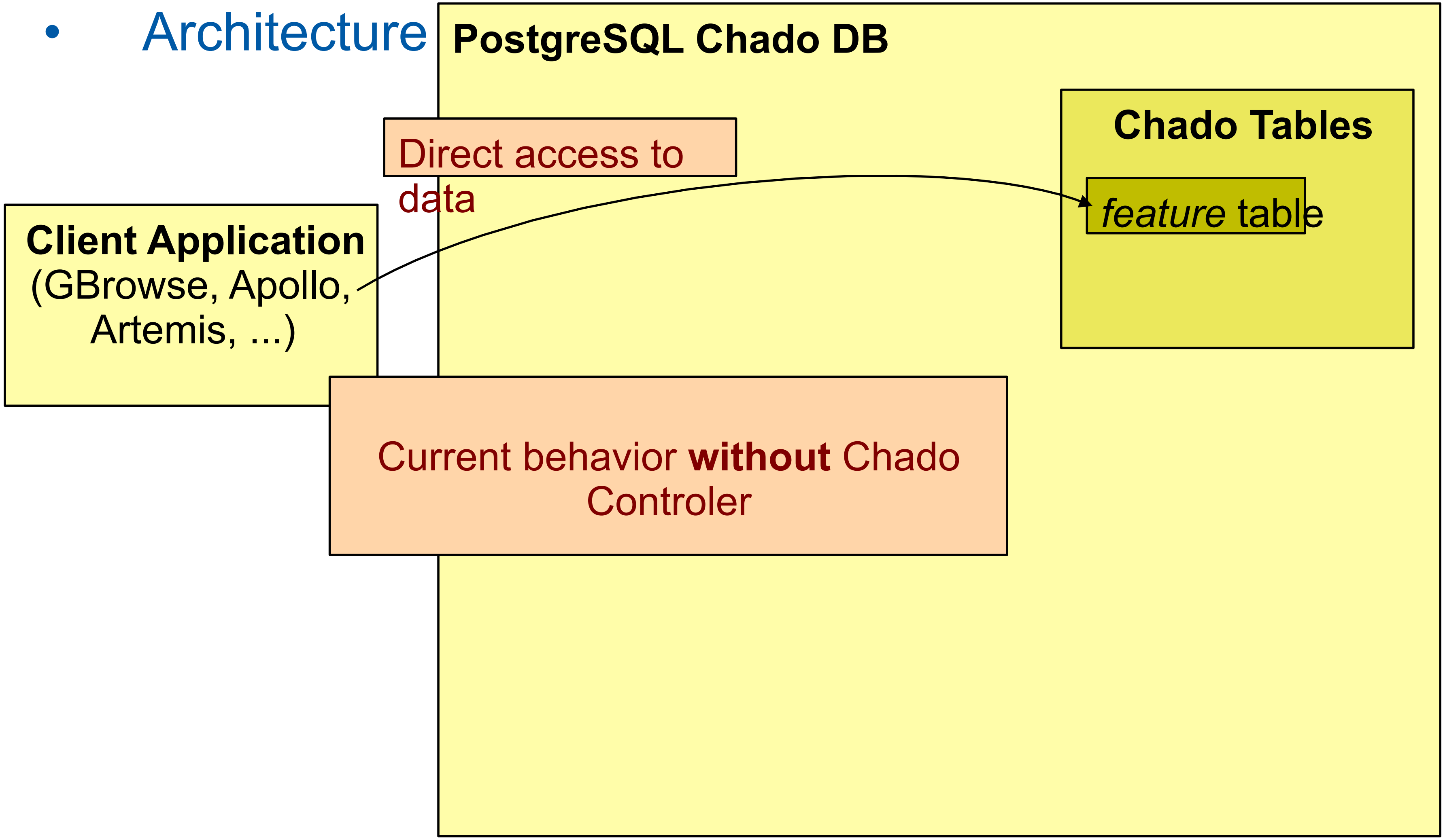
- Needs
- Restrict access to specific contigs
  - contig visibility and write access
  - manage access by groups
- Help manual annotation process
  - Auto-fill properties
  - Improve quality
- Keep track of annotations
  - old records with date and author

Goals

- Manage access right to Chado features
- Inspect manual annotations
- Manage history of manual annotations

1. Context

- Architecture



PostgreSQL Chado DB

Chado Tables

feature table

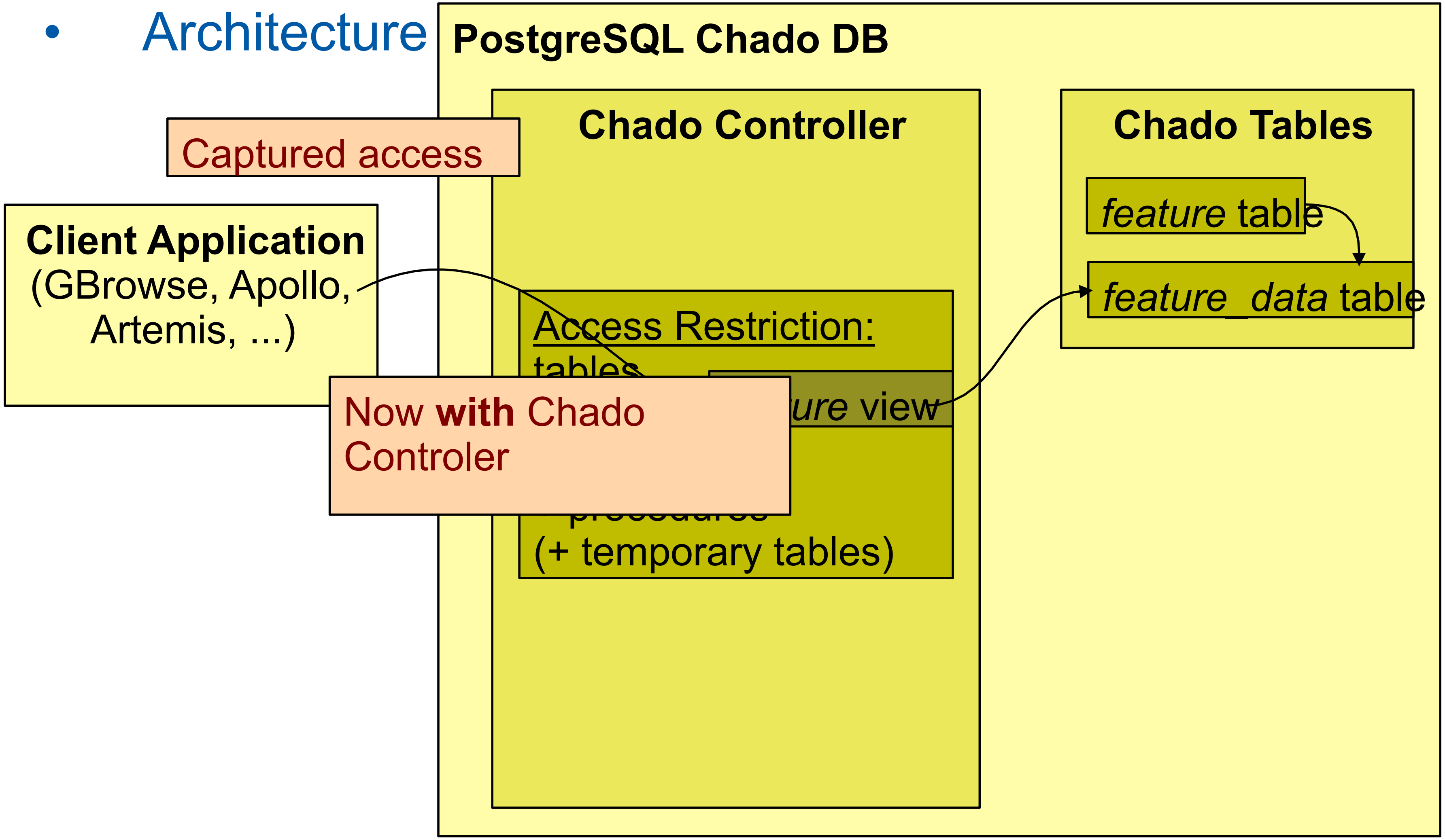
Direct access to data

Client Application
(GBrowse, Apollo, Artemis, ...)

Current behavior **without** Chado Controler

2. Annotator Access Restriction

- Architecture



PostgreSQL Chado DB

Chado Controller

Chado Tables

feature table

feature_data table

Access Restriction:
tables
ure view
procedures
(+ temporary tables)

Captured access

Client Application
(GBrowse, Apollo,
Artemis, ...)

Now **with** Chado
Controler

2. Annotator Access Restriction

- PostgreSQL Embedded Part
  - Access to feature table through a view with rules (for *insert*/*update*/*delete*)
  - Restrictions using a table of users/groups and a table to associate an access level for each feature to a user or a group
  - To improve queries speed, a temporary access table is created for each PostgreSQL session (for softwares with persistent connexion like Artemis or Apollo)

2. Annotator Access Restriction

•Genome Browser Integration

- Web access → each page = a new PostgreSQL session
- Temporary access tables init. slows down GBrowse too much → solution: integrate access restriction to GBrowse queries using a special account (with read access on everything)

Login:
Password:

The landmark named *scaffold_0009:1..20000* is not recognized. See the help pages for suggestions.

Instructions
Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed.
Navigation: Click one of the rulers to center on a location, or click and drag to select a region. Use the Scroll/Zoom buttons to change magnification and position.

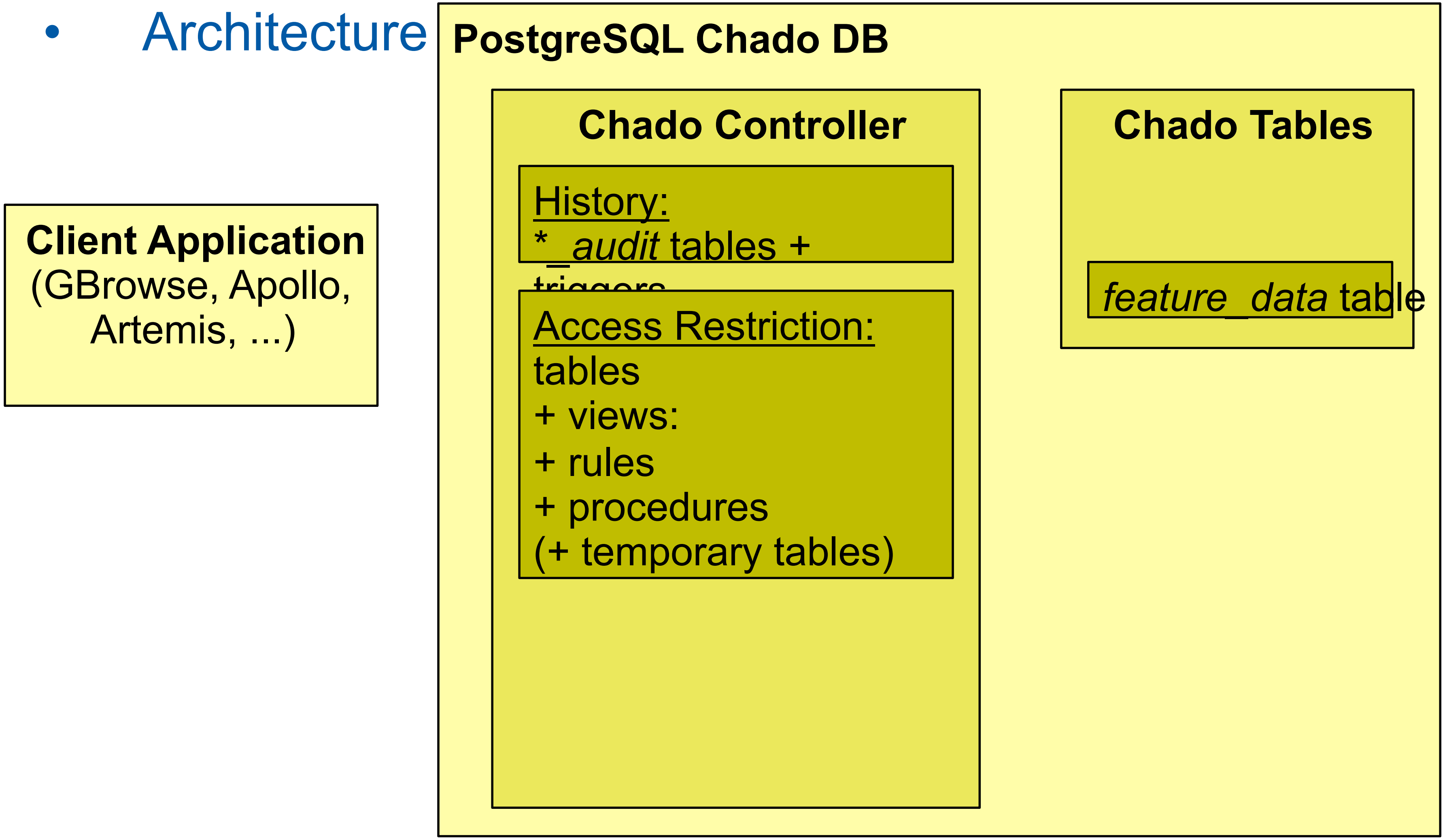
Change password...
New password: ●●●●●●●●●●
Confirmation: ●●●●●●●●●●

Showing 20 kbp from scaffold_0009, positions 1 to 20,000

Instructions
Searching: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed.

3. History of Manual Annotations

- Architecture



Client Application
(GBrowse, Apollo, Artemis, ...)

PostgreSQL Chado DB

Chado Controller

History:
*_*audit* tables + triggers

Access Restriction:
tables
+ views:
+ rules
+ procedures
(+ temporary tables)

Chado Tables

feature_data table

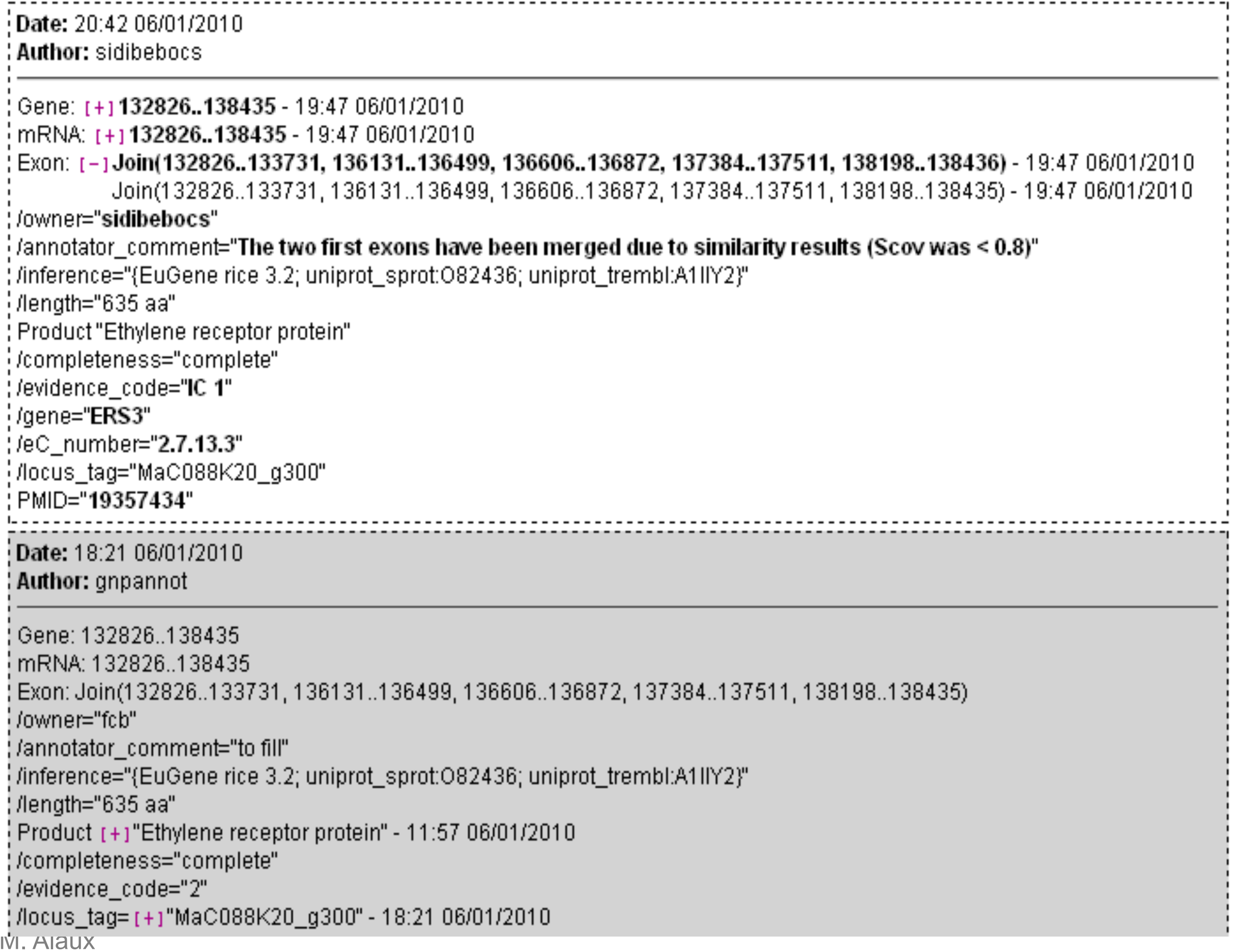
3. History of Manual Annotations

- PostgreSQL Embedded Part
- based on Chado audit module with some adds:
    - Author of the modification
    - Group of transaction identifier

3. History of Manual Annotations

- Genome Browser Integration: a *gbrowse_details*-like script

Curated annotation (modifications in bold)



Date: 20:42 06/01/2010
Author: sidibebocs

Gene: [+] **132826..138435** - 19:47 06/01/2010
mRNA: [+] **132826..138435** - 19:47 06/01/2010
Exon: [-] **Join(132826..133731, 136131..136499, 136606..136872, 137384..137511, 138198..138436)** - 19:47 06/01/2010
Join(132826..133731, 136131..136499, 136606..136872, 137384..137511, 138198..138435) - 19:47 06/01/2010
/owner="**sidibebocs**"
/annotator_comment="**The two first exons have been merged due to similarity results (Scov was < 0.8)**"
/inference="{EuGene rice 3.2; uniprot_sprot:O82436; uniprot_trembl:A1IIY2}"
/length="635 aa"
Product "Ethylene receptor protein"
/completeness="complete"
/evidence_code="**IC 1**"
/gene="**ERS3**"
/eC_number="**2.7.13.3**"
/locus_tag="MaC088K20_g300"
PMID="**19357434**"

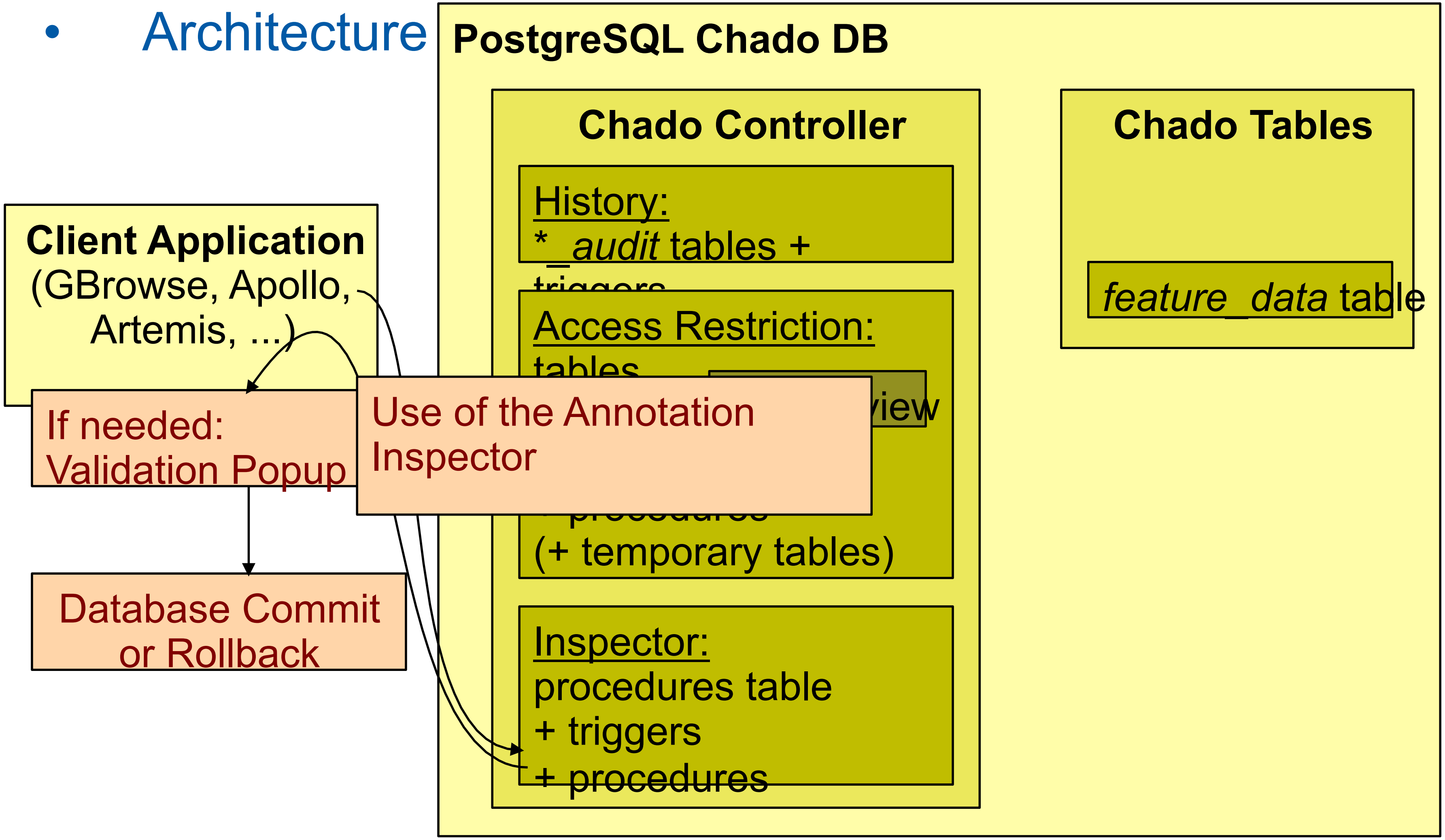
Previous annotation of MaC088K20_g300

Date: 18:21 06/01/2010
Author: gnpannot

Gene: 132826..138435
mRNA: 132826..138435
Exon: Join(132826..133731, 136131..136499, 136606..136872, 137384..137511, 138198..138435)
/owner="fcb"
/annotator_comment="to fill"
/inference="{EuGene rice 3.2; uniprot_sprot:O82436; uniprot_trembl:A1IIY2}"
/length="635 aa"
Product [+] "Ethylene receptor protein" - 11:57 06/01/2010
/completeness="complete"
/evidence_code="2"
/locus_tag= [+] "MaC088K20_g300" - 18:21 06/01/2010

4. Inspector of Manual Annotations

- Architecture



PostgreSQL Chado DB

Chado Controller

History:
* _*audit* tables +
triggers

Access Restriction:
tables
view
+ procedures
(+ temporary tables)

Inspector:
procedures table
+ triggers
+ procedures

Chado Tables

feature_data table

Client Application
(GBrowse, Apollo, Artemis, ...)

If needed:
Validation Popup

Use of the Annotation Inspector

Database Commit or Rollback

4. Inspector of Manual Annotations

•PostgreSQL Embedded Part

- Based on triggers to auto-fill some feature properties (*e.g. "/owner*", "*/color*", TE structure,...)
- Additional procedures for inspection of the structure and the function of curated genes to be called by the genome editor (*e.g.* start/stop codon, CDS length, splicing sites,...)

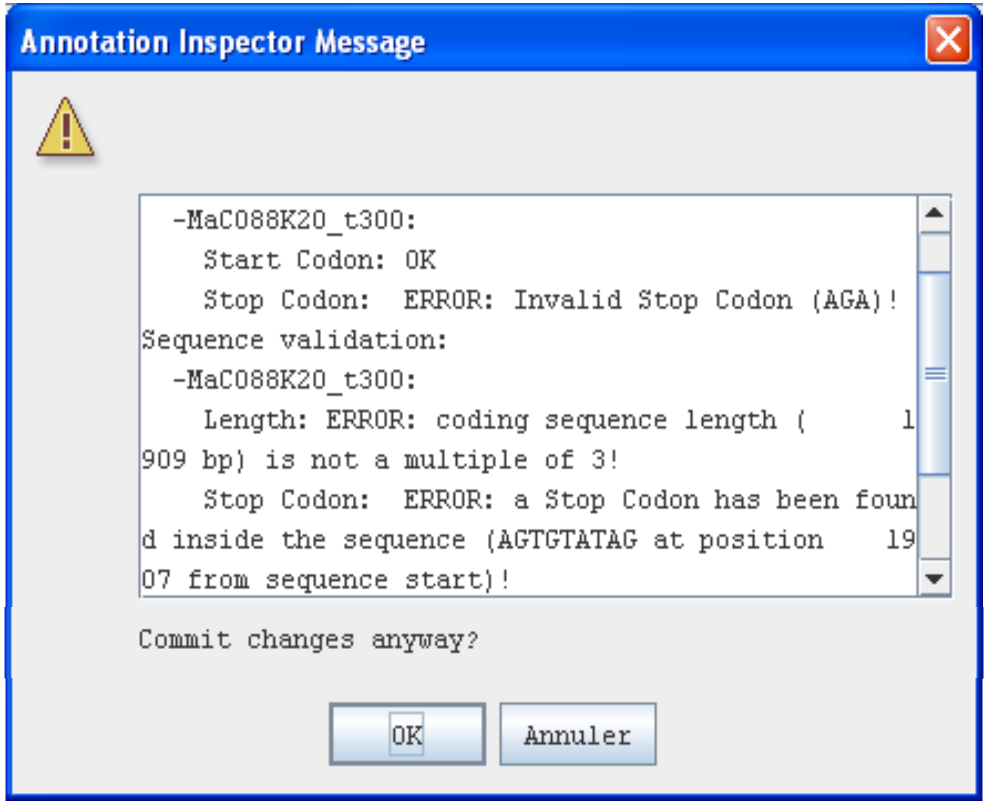
4. Inspector of Manual Annotations

- Genome Editor Integration
- Artemis calls the **initialization** procedure once connecting to Chado.
- Then, before each set of transactions to commit, it calls a procedure to get a new **group of transaction identifier**.
- Then, all the queries are executed and just before the commit, an **inspection procedure** is called with the identifier obtained earlier.
- Finally, the procedure returns the **inspection status** and Artemis can either **commit or rollback**.

4. Inspector of Manual Annotations

- Genome Editor Integration

Example of error message of invalid structural curation of MaC088K20_g300



```
  -MaC088K20_t300:
    Start Codon: OK
    Stop Codon:  ERROR: Invalid Stop Codon (AGA)!
Sequence validation:
  -MaC088K20_t300:
    Length: ERROR: coding sequence length (      1
909 bp) is not a multiple of 3!
    Stop Codon:  ERROR: a Stop Codon has been foun
d inside the sequence (AGTGTATAG at position    19
07 from sequence start)!
```

Commit changes anyway?

5. Conclusion

•Features

•Annotator Access Restriction

  - Feature-level access control
  - User and group management
  - Administration interface
  - Backward compatibility

•History of Manual Annotations

  - Keeps track of user actions

•Inspector of Manual Annotations

  - Many checks and annotation automation

5. Conclusion

•Extra-features

- install/update/uninstall script
- Chado Controller versions tracking
- compatibility mode script (for *COPY FROM* queries)
- annotation inspector is extensible
- some parts of the Chado Controller can be enabled/disabled "live"

5. Conclusion

- Perspectives
  - Include Annotation History into GMOD Report
  - “chado_undo.pl” script
  - Improve administration interface
  - Access restriction for Bio::Seq::Feature and JBrowse
  - Annotation inspector for Apollo

Contacts

- GMOD report:

fabrice.legeai@rennes.inra.fr

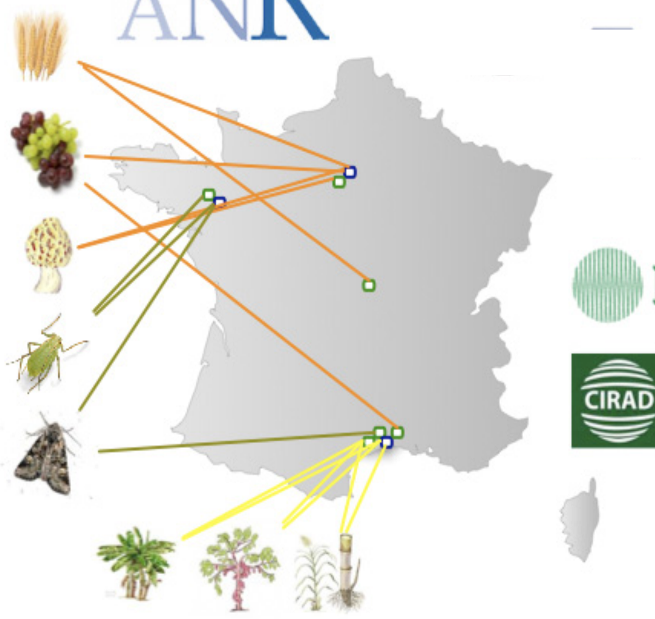
- Chado controller:

valentin.guignon@cirad.fr

Aknowledgements

- INRA – URGI/BIOGER
  - Michaël Alaux
  - Joelle Amselem
  - Baptiste Brault
  - Nathalie Choisne
  - Erik Kimmel
  - Veronique Jamilloux
  - Aminah Keliet
  - Nicolas Lapalu
  - Isabelle Luyten
  - Nacer Mohelebi
  - Sebastien Reboux
  - Delphine Steinbach
  - Daphné Verdelet
  - Hadi Quesneville
- INRA – BIO3P
  - Fabrice Legeai
  - Denis Tagu
- KORILOG
  - Patrick durand
  - Goulven Kerbellec

GnpAnnot Funding



- CIRAD
  - Stéphanie Sidibe Bocs
  - Gaëtan Drocs
  - Valentin Guignon
  - Manuel Ruiz
- Bioversity
  - Mathieu Rouard