An Integrated Information System Dedicated to Oak Genomics and Genetics

EPGV meeting
Thursday 4th, 2018

joelle.amselem@inra.fr

Quercus robur genome in numbers

- 12 pairs of chromosomes
- Genetic size: 933 cM
  - ✓ 12 linkage groups from 5.5k SNP (Bodénès et al. 2016)
- Physical size: 1.5 Gb (Kremer et al. 2007)
- Assembly and annotation
  - ✓ Diploïd version (2N): 8,827 scaffolds covering 1.45 Gb
  - ✓ Haploïd version (1N): 1,409 scaffolds covering 814 Mb
    - 871 scaffolds ordered in 12 pseudomolecules and assigned to the 12 linkage groups of the genetic map
    - 533 unassigned scaffold
    - 25,808 high-quality protein-coding gene models (4% manually curated/validated)
    - 52% of diverse Transposable Elements

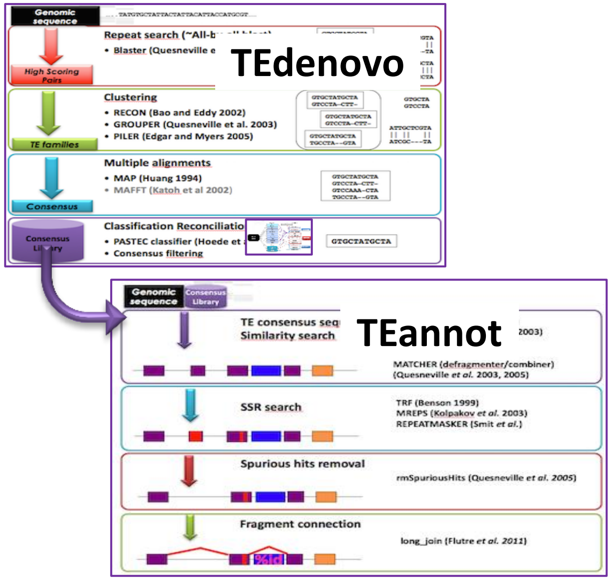
- Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, et al. Oak genome reveals facets of long lifespan. Nature Plants. 2018 Jun 18; doi:[10.1038/s41477-018-0172-3](https://doi.org/10.1038/s41477-018-0172-3)

Structural annotation pipelines

Transposable Elements

REPET package

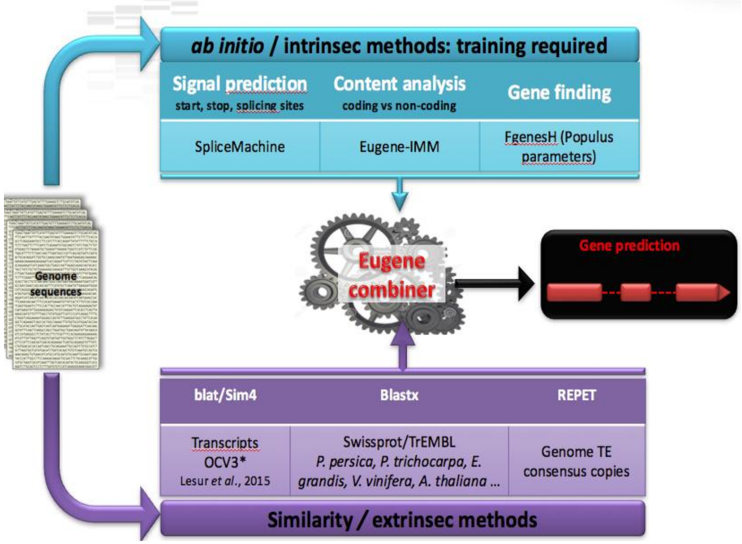
https://urgi.versailles.inra.fr/Tools/REPET



Genes prediction

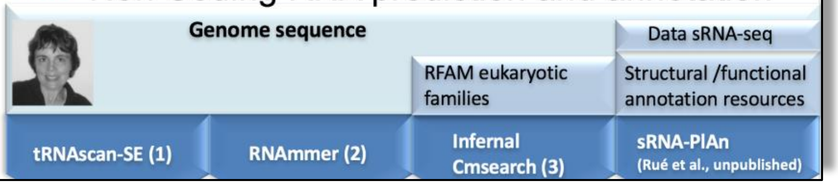
Eugene

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



Long Non Coding RNA

Non Coding RNA prediction and annotation



Functional annotation pipeline

Predicted polypeptides

Blast similarities

KEGG PDBe

swissprot

rpsBlast
Conserved domains

NCBI Conserved Domains

KOG

InterProScan

Protein domain identification

Pfam HAMAP PIRSF TIGRFAMs PANTHER
PRINTS Gene3D ProDom SMART Superfamily

Coils, Patternscan, Profilescan, Scanregexp, seg …

InterPro

Localization Targeting

TMHMM
SignalP
TargetP
Phobius

Protein definition* based on results of

EC Number

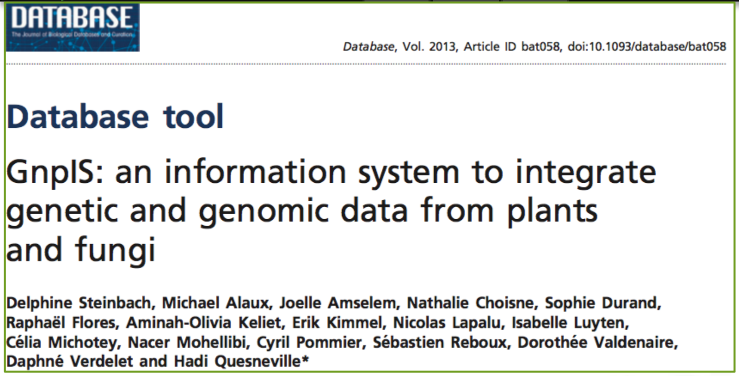
KEGG
Orthology

PANTHER Pfam KOG

*script from David Goodman, personal communication

GnpIS: A Genetic and genomic Information System

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



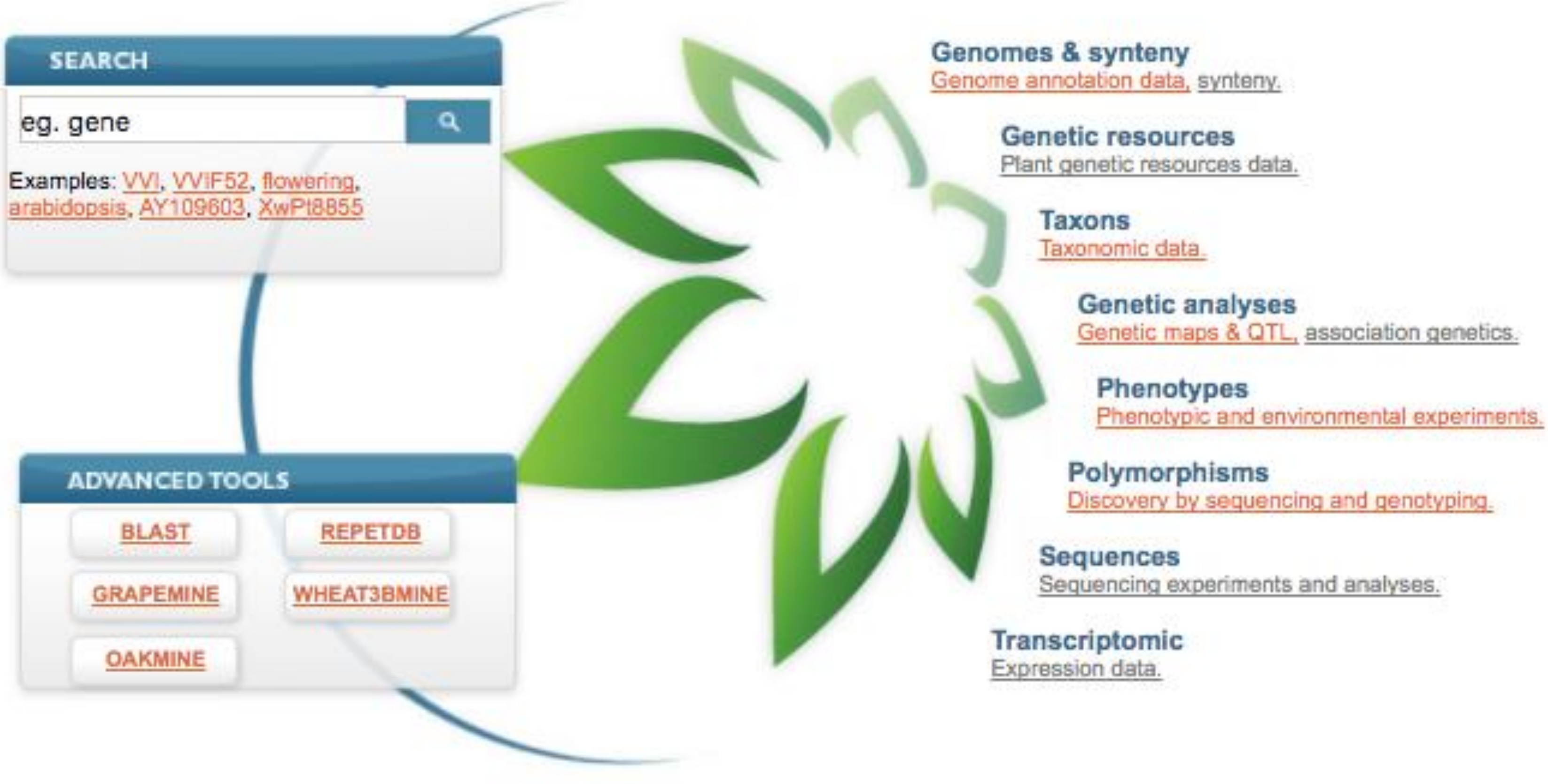
DATABASE

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*



Examples: VVI, VVIF52, flowering, arabidopsis, AY109603, XwPI8855

Genomes & synteny
Genome annotation data, synteny.

Genetic resources
Plant genetic resources data.

Taxons
Taxonomic data.

Genetic analyses
Genetic maps & QTL, association genetics.

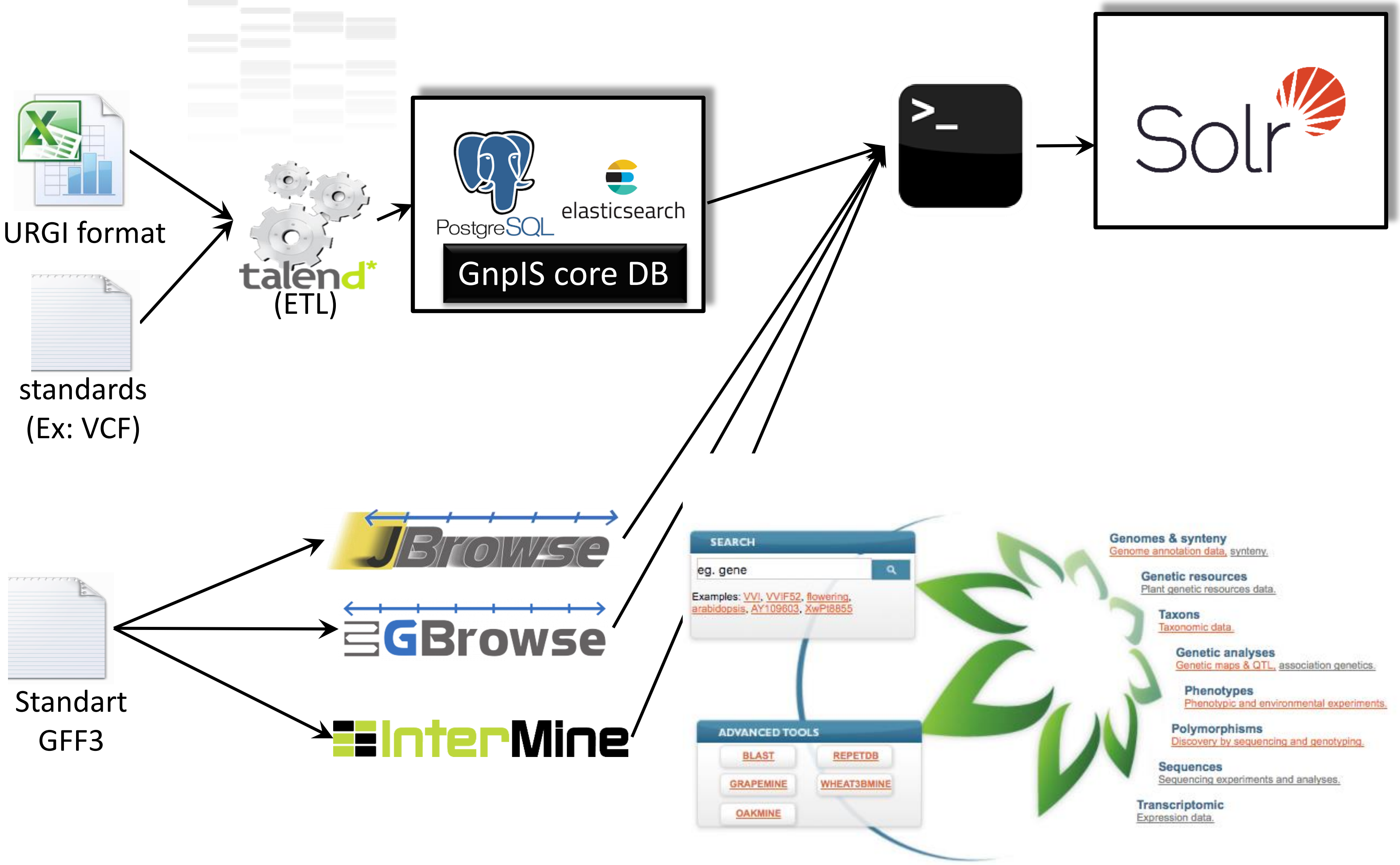
Phenotypes
Phenotypic and environmental experiments.

Polymorphisms
Discovery by sequencing and genotyping.

Sequences
Sequencing experiments and analyses.

Transcriptomic
Expression data.

GnpIS architecture



URGI format

standards
(Ex: VCF)

talend
(ETL)

PostgreSQL

elasticsearch

GnpIS core DB

Solr

Standart
GFF3

JBrowse

GBrowse

InterMine

SEARCH

eg. gene

Examples: VVI, VVIF52, flowering, arabidopsis, AY109603, XwPt8855

ADVANCED TOOLS

BLAST
REPETDB
GRAPEMINE
WHEAT3BMINE
OAKMINE

Genomes & synteny
Genome annotation data, synteny.

Genetic resources
Plant genetic resources data.

Taxons
Taxonomic data.

Genetic analyses
Genetic maps & QTL, association genetics.

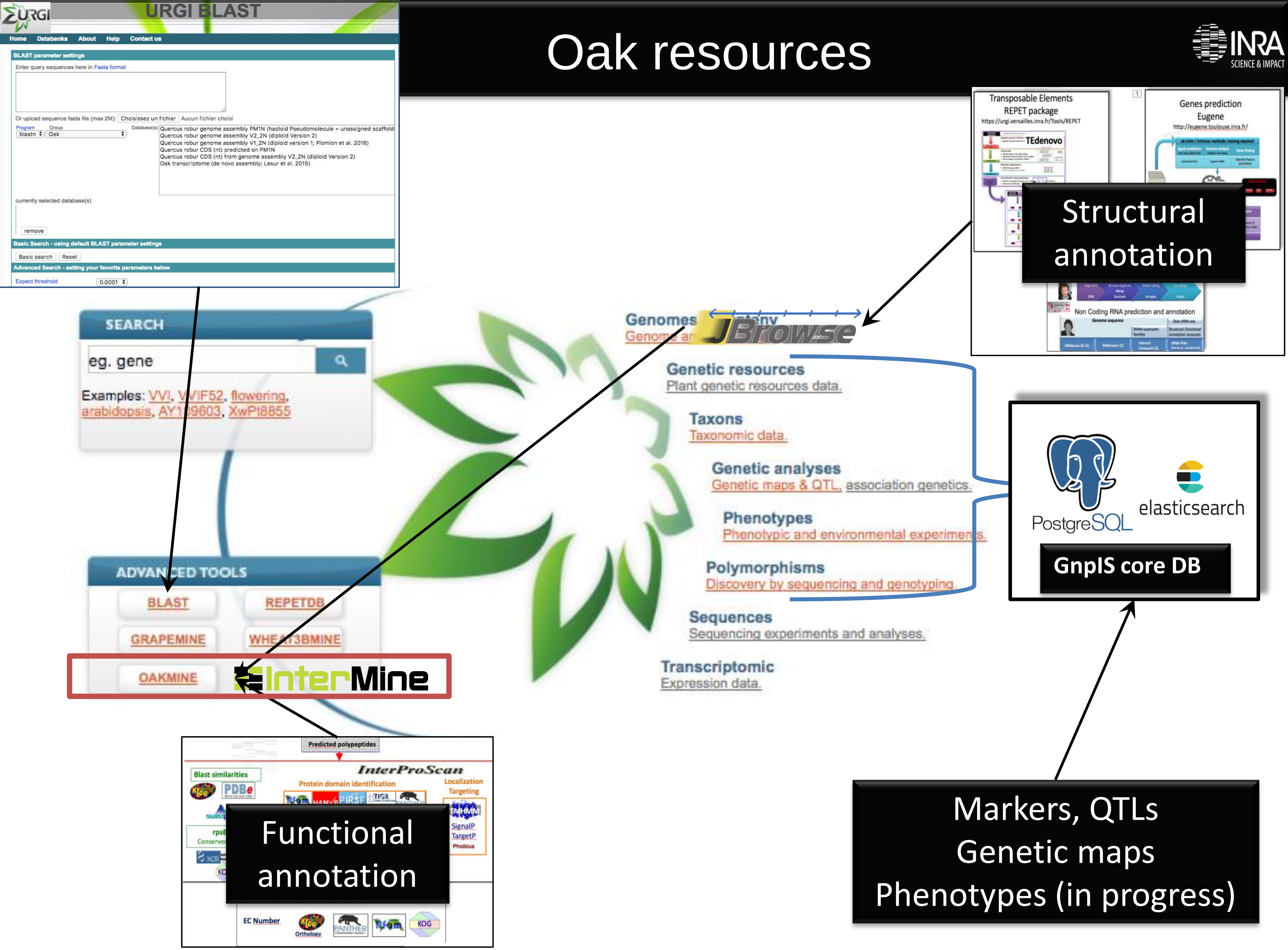
Phenotypes
Phenotypic and environmental experiments.

Polymorphisms
Discovery by sequencing and genotyping.

Sequences
Sequencing experiments and analyses.

Transcriptomic
Expression data.

Oak resources



URGI BLAST

BLAST parameter settings

Enter query sequences here in Fasta format

Or upload sequence fasta file (max 2M): Choisissez un fichier Aucun fichier choisi

Program: blastn
Group: Oak
Database(s):
Quercus robur genome assembly PM1N (haploid Pseudomolecule + unassigned scaffold
Quercus robur genome assembly V2_2N (diploid Version 2)
Quercus robur genome assembly V1_2N (diploid version 1; Plomion et al. 2016)
Quercus robur CDS (nt) predicted on PM1N
Quercus robur CDS (nt) from genome assembly V2_2N (diploid Version 2)
Oak transcriptome (de novo assembly; Lesur et al. 2015)

currently selected database(s)

Basic Search - using default BLAST parameter settings

Advanced Search - setting your favorite parameters below

Expect threshold: 0.0001

Structural annotation

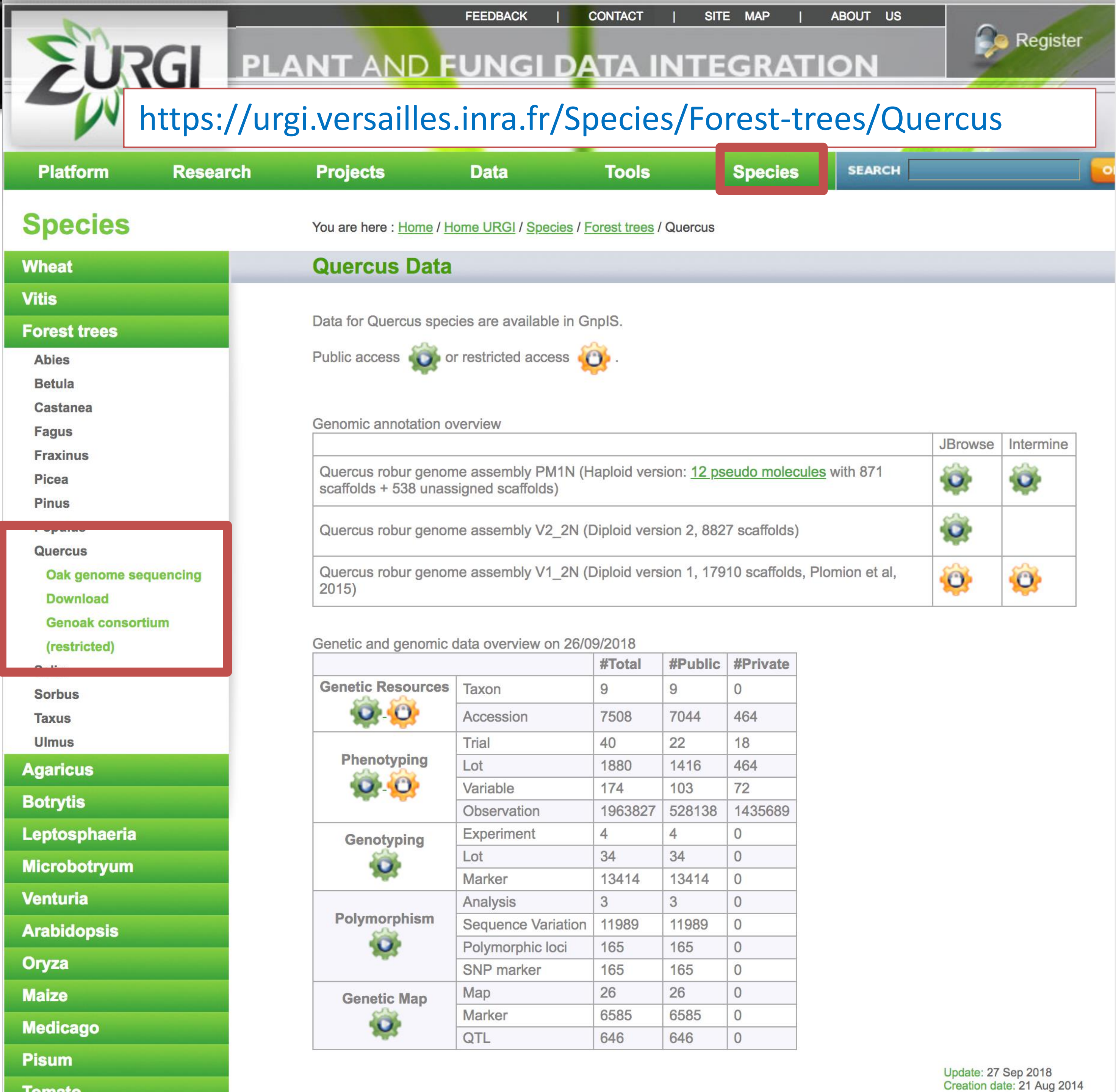
GnpIS core DB

Functional annotation

Markers, QTLs
Genetic maps
Phenotypes (in progress)

Navigating oak genomic and genetic data

https://urgi.versailles.inra.fr/Species/Forest-trees/Quercus



Species

You are here : Home / Home URGI / Species / Forest trees / Quercus

- **Wheat**
- **Vitis**
- **Forest trees**
  - Abies
  - Betula
  - Castanea
  - Fagus
  - Fraxinus
  - Picea
  - Pinus
  - Quercus
    - Oak genome sequencing
    - Download
    - Genoak consortium (restricted)
  - Sorbus
  - Taxus
  - Ulmus
- **Agaricus**
- **Botrytis**
- **Leptosphaeria**
- **Microbotryum**
- **Venturia**
- **Arabidopsis**
- **Oryza**
- **Maize**
- **Medicago**
- **Pisum**

Quercus Data

Data for Quercus species are available in GnpIS.

Public access ⚙ or restricted access 🔒 .

Genomic annotation overview

	JBrowse	Intermine
Quercus robur genome assembly PM1N (Haploid version: 12 pseudo molecules with 871 scaffolds + 538 unassigned scaffolds)	⚙	⚙
Quercus robur genome assembly V2_2N (Diploid version 2, 8827 scaffolds)	⚙	
Quercus robur genome assembly V1_2N (Diploid version 1, 17910 scaffolds, Plomion et al, 2015)	🔒	🔒

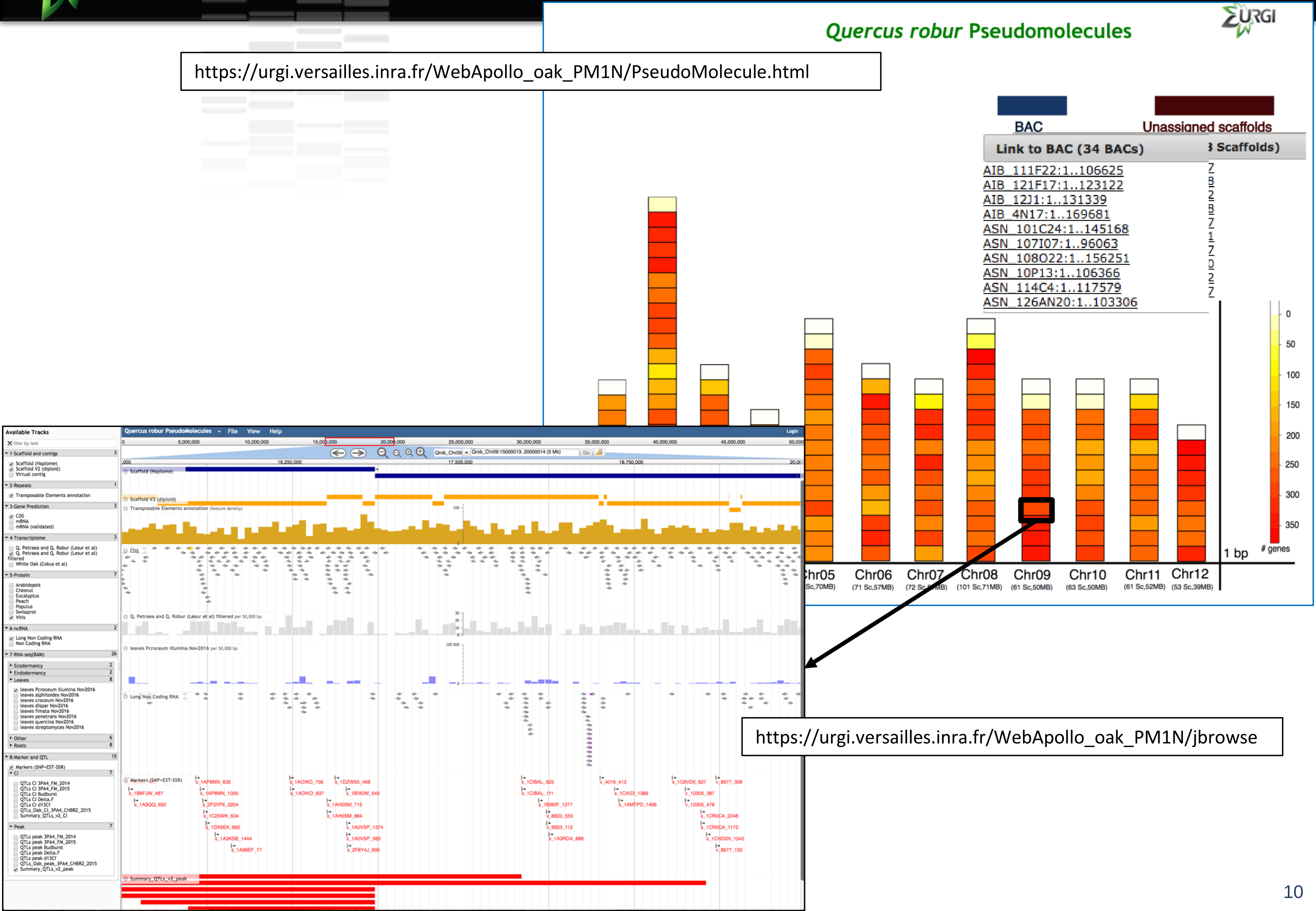
Genetic and genomic data overview on 26/09/2018

		#Total	#Public	#Private
Genetic Resources ⚙ - 🔒	Taxon	9	9	0
	Accession	7508	7044	464
Phenotyping ⚙ - 🔒	Trial	40	22	18
	Lot	1880	1416	464
	Variable	174	103	72
	Observation	1963827	528138	1435689
Genotyping ⚙	Experiment	4	4	0
	Lot	34	34	0
	Marker	13414	13414	0
Polymorphism ⚙	Analysis	3	3	0
	Sequence Variation	11989	11989	0
	Polymorphic loci	165	165	0
	SNP marker	165	165	0
Genetic Map ⚙	Map	26	26	0
	Marker	6585	6585	0
	QTL	646	646	0

Update: 27 Sep 2018
Creation date: 21 Aug 2014

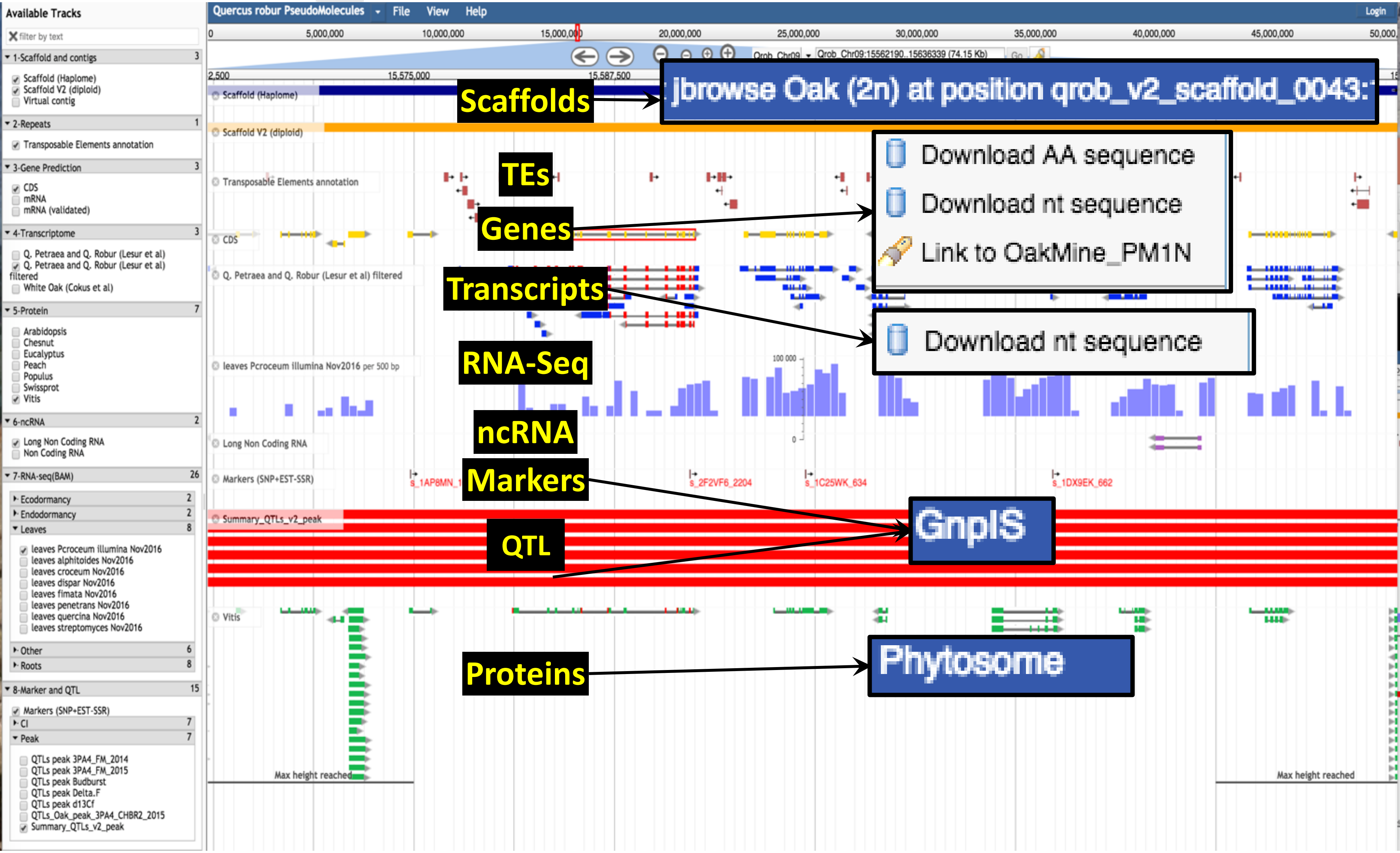
Pseudomolecules … to JBrowse

https://urgi.versailles.inra.fr/WebApollo_oak_PM1N/PseudoMolecule.html



https://urgi.versailles.inra.fr/WebApollo_oak_PM1N/jbrowse

Jbrowse overview and cross references



Available Tracks

filter by text

- 1-Scaffold and contigs 3
  - ☑ Scaffold (Haplome)
  - ☑ Scaffold V2 (diploid)
  - ☐ Virtual contig
- 2-Repeats 1
  - ☑ Transposable Elements annotation
- 3-Gene Prediction 3
  - ☑ CDS
  - ☐ mRNA
  - ☐ mRNA (validated)
- 4-Transcriptome 3
  - ☐ Q. Petraea and Q. Robur (Lesur et al)
  - ☑ Q. Petraea and Q. Robur (Lesur et al) filtered
  - ☐ White Oak (Cokus et al)
- 5-Protein 7
  - ☐ Arabidopsis
  - ☐ Chesnut
  - ☐ Eucalyptus
  - ☐ Peach
  - ☐ Populus
  - ☐ Swissprot
  - ☑ Vitis
- 6-ncRNA 2
  - ☑ Long Non Coding RNA
  - ☐ Non Coding RNA
- 7-RNA-seq(BAM) 26
  - Ecodormancy 2
  - Endodormancy 2
  - Leaves 8
    - ☑ leaves Pcroceum illumina Nov2016
    - ☐ leaves alphitoides Nov2016
    - ☐ leaves croceum Nov2016
    - ☐ leaves dispar Nov2016
    - ☐ leaves fimata Nov2016
    - ☐ leaves penetrans Nov2016
    - ☐ leaves quercina Nov2016
    - ☐ leaves streptomyces Nov2016
  - Other 6
  - Roots 8
- 8-Marker and QTL 15
  - ☑ Markers (SNP+EST-SSR)
  - CI 7
  - Peak 7
    - ☐ QTLs peak 3PA4_FM_2014
    - ☐ QTLs peak 3PA4_FM_2015
    - ☐ QTLs peak Budburst
    - ☐ QTLs peak Delta.F
    - ☐ QTLs peak d13Cf
    - ☐ QTLs_Oak_peak_3PA4_CHBR2_2015
    - ☑ Summary_QTLs_v2_peak

Scaffolds → jbrowse Oak (2n) at position qrob_v2_scaffold_0043:

TEs

Genes →
- Download AA sequence
- Download nt sequence
- Link to OakMine_PM1N

Transcripts → Download nt sequence

RNA-Seq

ncRNA

Markers → GnpIS

QTL → GnpIS

Proteins → Phytosome

OakMine portal

Search OakMine_PM1N. Enter names, identifiers or keywords for genes, proteins, pathways, ontology terms, etc. (e.g. LRR, glycolysis,).

e.g. X, Y, Z

Enter a list of identifiers.

e.g. Qrob_P0000010.2, Qrob_P0000020.2

Welcome Back on OakMine_PM1N !

OakMine_PM1N integrates many types of data for *QuercusRobur genes*. based on domain, blast, interproscan analysis. Information about the genoak project can be found at the quercus portal on the evoltree

OakMine_PM1N load many description (based on InterPro and blast); domain data from InterPro; data from Reactome and KEGG. Read more

Query for description:

- Blast description ➡ Quercusrobur Protein.
- DomainMotif Description ➡ Quercusrobur Protein
- Gene ontology Description ➡ Quercusrobur Protein

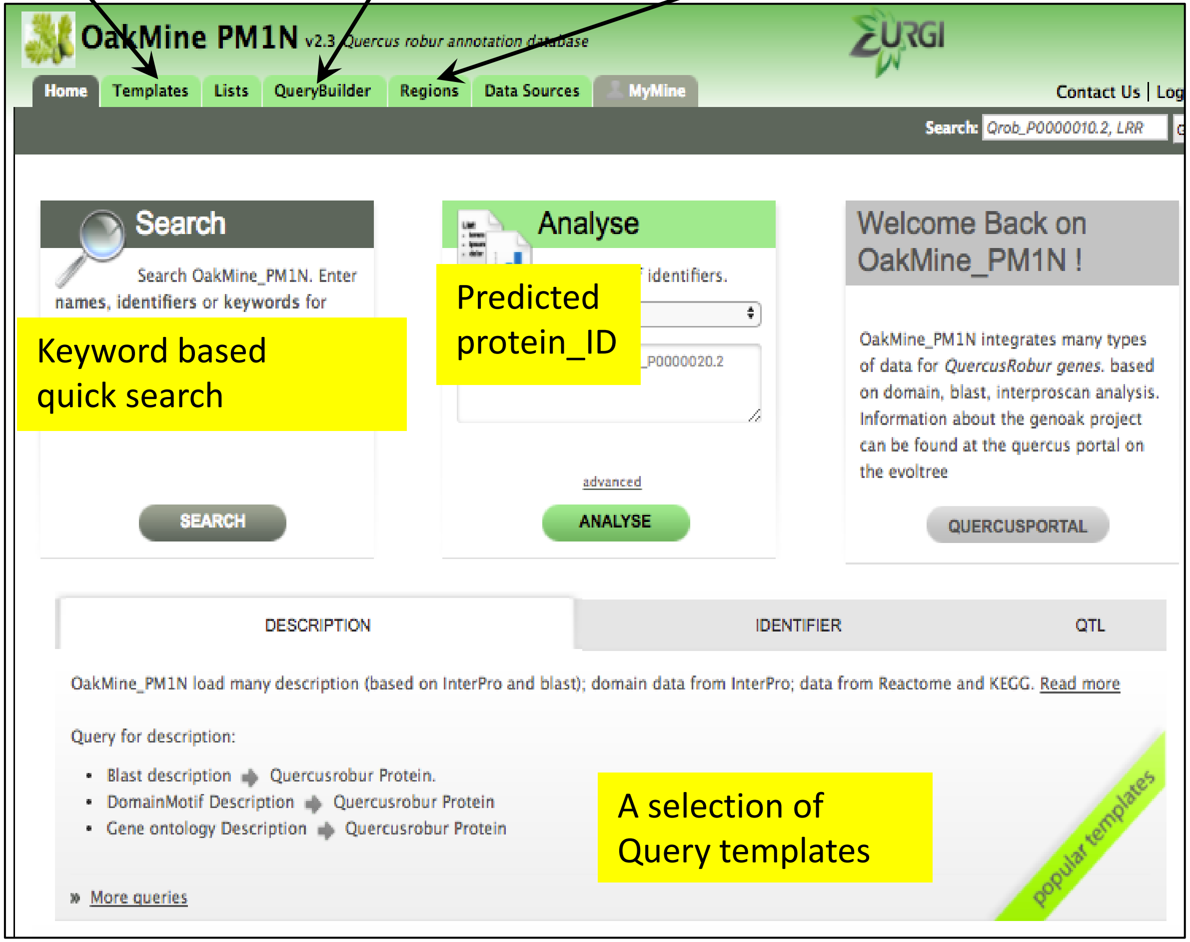
» More queries

Several ways to query data in OakMine

Query templates

Advanced search using Query the builder

Search for genes in genome regions



Search OakMine_PM1N. Enter names, identifiers or keywords for

Welcome Back on OakMine_PM1N !

OakMine_PM1N integrates many types of data for *QuercusRobur genes.* based on domain, blast, interproscan analysis. Information about the genoak project can be found at the quercus portal on the evoltree

OakMine_PM1N load many description (based on InterPro and blast); domain data from InterPro; data from Reactome and KEGG. Read more

Query for description:

- Blast description ➔ Quercusrobur Protein.
- DomainMotif Description ➔ Quercusrobur Protein
- Gene ontology Description ➔ Quercusrobur Protein

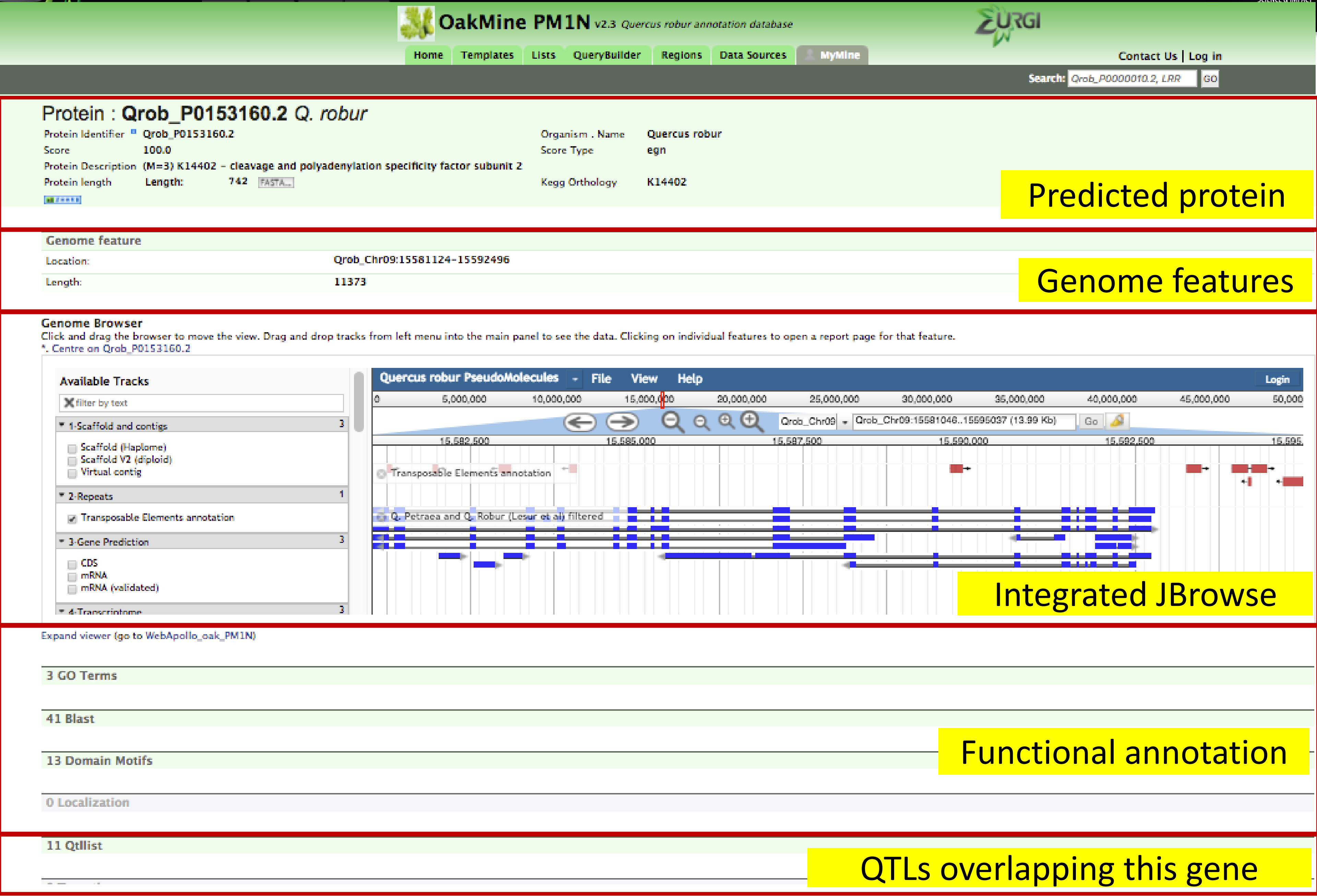
» More queries

Keyword based quick search

Predicted protein_ID

A selection of Query templates

OakMine_PM1N central object: the gene card



Protein : Qrob_P0153160.2 *Q. robur*

Protein Identifier	Qrob_P0153160.2	Organism . Name	Quercus robur
Score	100.0	Score Type	egn
Protein Description	(M=3) K14402 - cleavage and polyadenylation specificity factor subunit 2		
Protein length	Length: 742 FASTA...	Kegg Orthology	K14402

Genome feature

Location:	Qrob_Chr09:15581124-15592496
Length:	11373

Genome Browser

Click and drag the browser to move the view. Drag and drop tracks from left menu into the main panel to see the data. Clicking on individual features to open a report page for that feature.

*. Centre on Qrob_P0153160.2

Expand viewer (go to WebApollo_oak_PM1N)

3 GO Terms

41 Blast

13 Domain Motifs

0 Localization

11 Qtllist

Predicted protein

Genome features

Integrated JBrowse

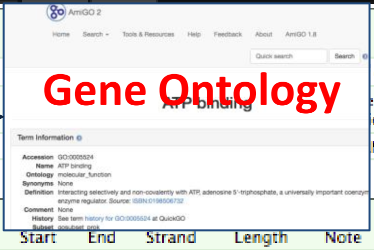
Functional annotation

QTLs overlapping this gene

Functional annotation : cross references to external DBs

3 GO Terms

Identifier	Name	
GO:0006378	mRNA polyadenylation	40-200 adenylyl residues at the 3' end of a eukaryotic mRNA primary transcript.
GO:0005847	mRNA cleavage and polyadenylation specificity factor complex	canonical AAUAAA hexamer and to U-rich upstream sequence elements on the pre-mRNA, thereby stimulating the otherwise weakly active and ently RNAs containing a poly(A) signal.
GO:0006379	mRNA cleavage	NA molecule is cleaved at specific sites or in a regulated manner.



Gene Ontology

ATP binding

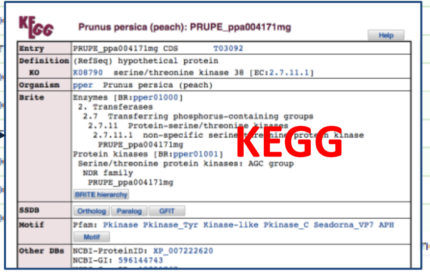
Term Information

Accession GO:0005524
Name ATP binding
Ontology molecular_function
Synonyms None
Definition Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzym enzyme regulator. Source: ISBN:0198506732
Comment None
History See term history for GO:0005524 at QuickGO
Subset gosubset_prok

41 Blast

Analysis	Hit	Start	End	Strand	Length	Note	Hit Coverage	Hit Length	Hit Pident	E Val	Hit Description
blastp_kegg	cmo:103484088	1	741	+	741	Caps:3	100.00	738	85.09	0.0	cleavage and polyadenylation specificity factor subunit 2
blastp_kegg	csv:101215673	1	741	+	741	Caps:3	100.00	738	84.96	0.0	cleavage and polyadenylation specificity factor subunit 2-like
blastp_kegg	pmum:103333205	1	741	+	741	Caps:1	100.00	740	83.51	0.0	cleavage and p
blastp_kegg	pvu:PHAVU_004G114000g	1	741	+	741	Caps:2	100.00	739	84.57	0.0	hypothetical pr
blastp_kegg	gmx:100777002	1	741	+	741	Caps:2	100.00	739	84.71	0.0	cleavage and p
blastp_kegg	mdm:103439774	1	741	+	741	Caps:1	100.00	740	83.11	0.0	cleavage and p
blastp_kegg	pper:PRUPE_ppa001928mg	1	741	+	741	Caps:1	100.00	740	82.97	0.0	hypothetical
blastp_kegg	pxb:103934917	1	741	+	741	Caps:1	100.00	740	82.70	0.0	cleavage and p
blastp_kegg	gmx:100796817	1	741	+	741	Caps:3	100.00	738	84.42	0.0	cleavage and p
blastp_kegg	cam:101489524	1	741	+	741	Caps:1	100.00	740	83.65	0.0	cleavage and p

more rows



Prunus persica (peach): PRUPE_ppa004171mg

KEGG

Entry	PRUPE_ppa004171mg CDS T03092
Definition	(RefSeq) hypothetical protein
KO	K08790 serine/threonine kinase 38 [EC:2.7.11.1]
Organism	pper Prunus persica (peach)
Brite	Enzymes [BR:pper01000] 2. Transferases 2.7 Transferring phosphorus-containing groups 2.7.11 Protein-serine/threonine kinases 2.7.11.1 non-specific serine/threonine protein kinase PRUPE_ppa004171mg Protein kinases [BR:pper01001] Serine/threonine protein kinases: AGC group NDR family PRUPE_ppa004171mg
SSDB	
Motif	Pfam: Pkinase Pkinase_Tyr Kinase-like Pkinase_C Seadorna_VP7 APH
Other DBs	NCBI-ProteinID: XP_007222620 NCBI-GI: 596144743

13 Domain Motifs

Analysis	Begin	End	Length	Domain Identifier	Cross Ref	Descriptio	Inter Pro
Gene3D	9	254	246	G3DSA:3.60.15.10	none	none	IPR001279
Pfam	655	738	84	PF13299	none	Cleavage a	IPR025069
SMART	243	366	124	SM01027	none	Beta-Cas	IPR022712
PANTHER	514	741	228	PTHR11203:SF5			IPR027075
PANTHER	1	366	366	PTHR11203:SF5			IPR027075
SUPERFAMILY	5	376	372	SSF56281			IPR001279
SUPERFAMILY	466	604	139	SSF56281			IPR001279
Pfam	20	157	138	PF00753		llo-beta-lactamase superfamily	IPR001279
Pfam	243	359	117	PF10996		-Casp domain	IPR022712
Pfam	536	575	40	PF07521		-metabolising metallo-beta-lactamase	IPR011108

Show more rows

Pfam

Family: *Pkinase* (PF00069)

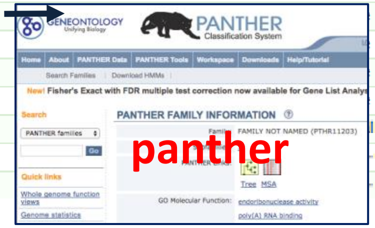
Summary: Protein kinase domain

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

This is the Wikipedia entry entitled "Protein kinase domain". More...

Protein kinase domain

The **protein kinase domain** is a structurally conserved protein domain containing the catalytic function of protein kinases.[2][3][4] Protein kinases are a group of enzymes that move a phosphate group onto proteins, in a process called phosphorylation. This functions as an on/off switch for many cellular processes, including metabolism, transcription, cell cycle progression, cytoskeletal rearrangement and cell movement, apoptosis, and differentiation. They also function in embryonic development, physiological responses, and in the nervous and immune system. Abnormal phosphorylation causes many human diseases, including cancer,



New! Fisher's Exact with FDR multiple test correction now available for Gene List Analys

PANTHER FAMILY INFORMATION

panther

Family: FAMILY NOT NAMED (PTHR11203)
PANTHER Links:
GO Molecular Function: endoribonuclease activity
poly(A) RNA binding

Interpro

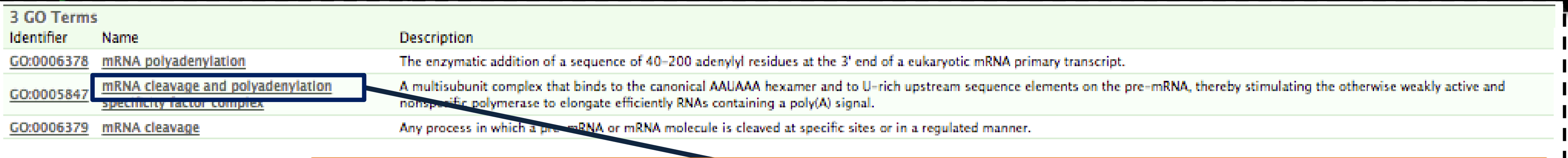
Serine/threonine-protein kinase, active site (IPR008271)

0 Localization

11 Qtllist

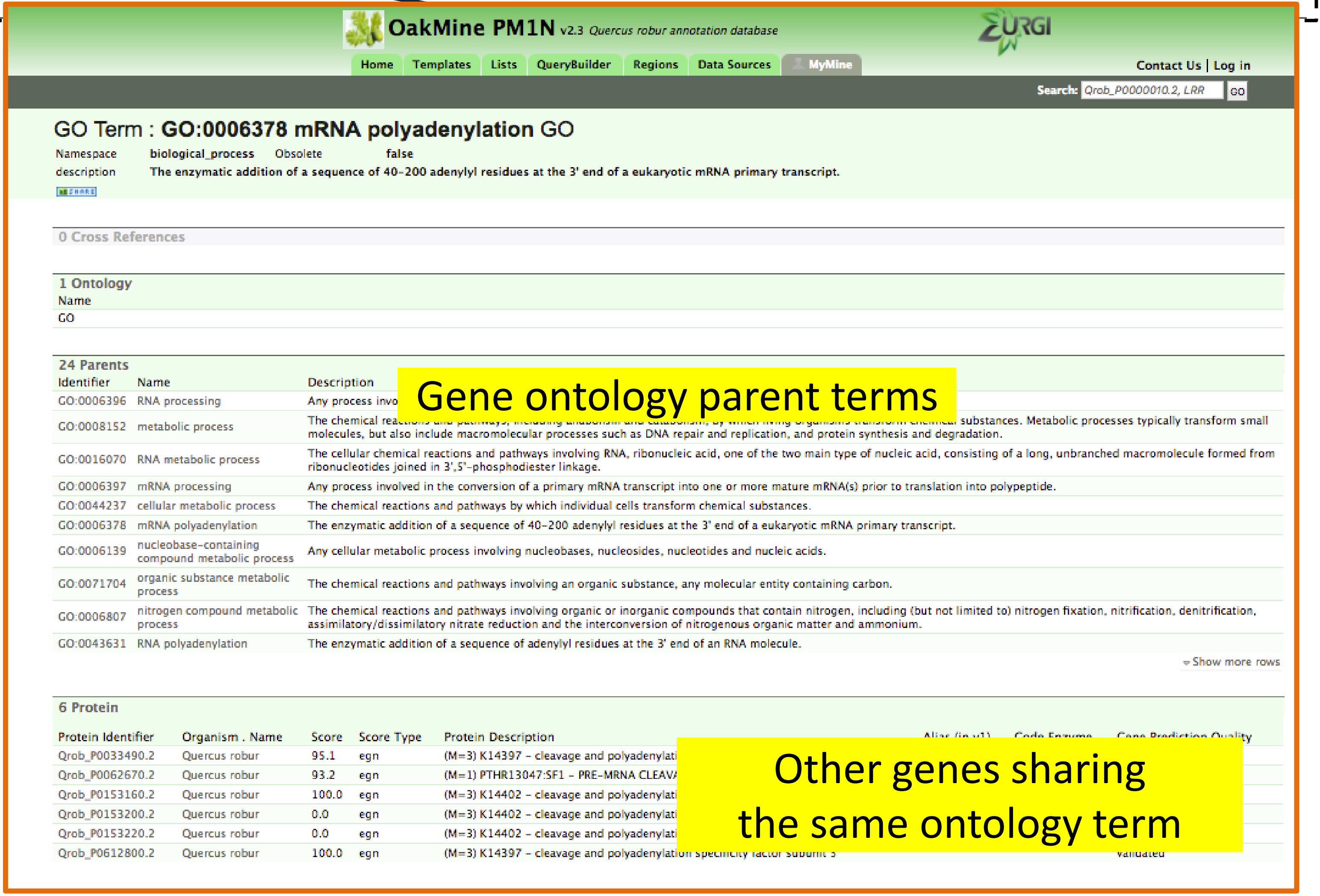
Qtl Name	Chromosome Name	Linkage Group	Prox Marker	Dist Marker	Position QTL	Pos One	Pos Two	Test Type	Test Value	R 2
Bourran1_2004_QTL4_peak_Bud_burst_A4	Qrob_Chr09	9	s_1BY6BQ_440	s_1AOIKO_756	16,45	0	27,45	lod	4,5	10,6
Bourran_2000_2002_QTL7_Delta.F	Qrob_Chr09	9	v_5944_442	s_1BA1PC_866	23.51	10,96	35,74	lod	4.1466	0.041
Bourran2_2003_QTL11_peak_Bud_burst_3P	Qrob_Chr09	9	s_1CGP2H_273	v_15801_330	27,16	4,16	48,16	lod	2,3	5,1
Bourran2_2003_QTL13_peak_Bud_burst_A4	Qrob_Chr09	9	s_1AP8MN_635	s_1A3QQ_692	18,18	10,88	25,88	lod	3,4	7,2
Bourran2_2004_QTL12_peak_Bud_burst_3P	Qrob_Chr09	9	s_1BDO6G_250	s_1A83AM_496	34,31	9,31	44,31	lod	3,6	7,6
Bourran2_2004_QTL14_peak_Bud_burst_A4	Qrob_Chr09	9	s_1BY6BQ_440	s_1AOIKO_756	16,83	10,33	22,33	lod	3,8	9

Interoperability with other features in OakMine



3 GO Terms

Identifier	Name	Description
GO:0006378	mRNA polyadenylation	The enzymatic addition of a sequence of 40-200 adenylyl residues at the 3' end of a eukaryotic mRNA primary transcript.
GO:0005847	mRNA cleavage and polyadenylation specificity factor complex	A multisubunit complex that binds to the canonical AAUAAA hexamer and to U-rich upstream sequence elements on the pre-mRNA, thereby stimulating the otherwise weakly active and nonspecific polymerase to elongate efficiently RNAs containing a poly(A) signal.
GO:0006379	mRNA cleavage	Any process in which a pre-mRNA or mRNA molecule is cleaved at specific sites or in a regulated manner.



GO Term : **GO:0006378 mRNA polyadenylation** GO

Namespace **biological_process** Obsolete **false**

description **The enzymatic addition of a sequence of 40-200 adenylyl residues at the 3' end of a eukaryotic mRNA primary transcript.**

0 Cross References

1 Ontology

Name
GO

24 Parents

Gene ontology parent terms

Identifier	Name	Description
GO:0006396	RNA processing	Any process invo
GO:0008152	metabolic process	The chemical rea substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation.
GO:0016070	RNA metabolic process	The cellular chemical reactions and pathways involving RNA, ribonucleic acid, one of the two main type of nucleic acid, consisting of a long, unbranched macromolecule formed from ribonucleotides joined in 3',5'-phosphodiester linkage.
GO:0006397	mRNA processing	Any process involved in the conversion of a primary mRNA transcript into one or more mature mRNA(s) prior to translation into polypeptide.
GO:0044237	cellular metabolic process	The chemical reactions and pathways by which individual cells transform chemical substances.
GO:0006378	mRNA polyadenylation	The enzymatic addition of a sequence of 40-200 adenylyl residues at the 3' end of a eukaryotic mRNA primary transcript.
GO:0006139	nucleobase-containing compound metabolic process	Any cellular metabolic process involving nucleobases, nucleosides, nucleotides and nucleic acids.
GO:0071704	organic substance metabolic process	The chemical reactions and pathways involving an organic substance, any molecular entity containing carbon.
GO:0006807	nitrogen compound metabolic process	The chemical reactions and pathways involving organic or inorganic compounds that contain nitrogen, including (but not limited to) nitrogen fixation, nitrification, denitrification, assimilatory/dissimilatory nitrate reduction and the interconversion of nitrogenous organic matter and ammonium.
GO:0043631	RNA polyadenylation	The enzymatic addition of a sequence of adenylyl residues at the 3' end of an RNA molecule.

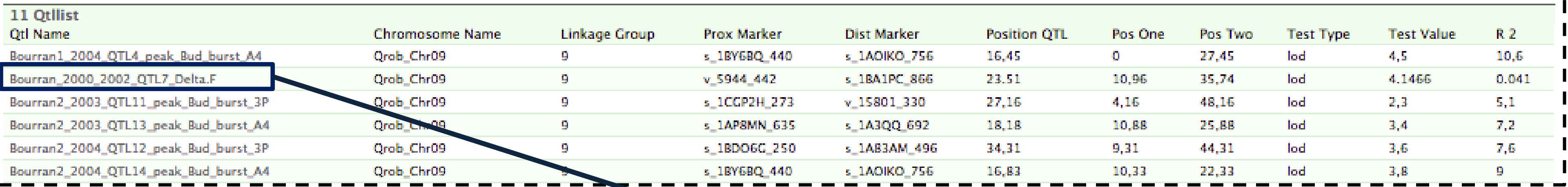
Show more rows

6 Protein

Protein Identifier	Organism . Name	Score	Score Type	Protein Description	Alias (in v1)	Code Enzyme	Gene Prediction Quality
Qrob_P0033490.2	Quercus robur	95.1	egn	(M=3) K14397 - cleavage and polyadenylati			
Qrob_P0062670.2	Quercus robur	93.2	egn	(M=1) PTHR13047:SF1 - PRE-MRNA CLEAVA			
Qrob_P0153160.2	Quercus robur	100.0	egn	(M=3) K14402 - cleavage and polyadenylati			
Qrob_P0153200.2	Quercus robur	0.0	egn	(M=3) K14402 - cleavage and polyadenylati			
Qrob_P0153220.2	Quercus robur	0.0	egn	(M=3) K14402 - cleavage and polyadenylati			
Qrob_P0612800.2	Quercus robur	100.0	egn	(M=3) K14397 - cleavage and polyadenylation specificity factor subunit 3			Validated

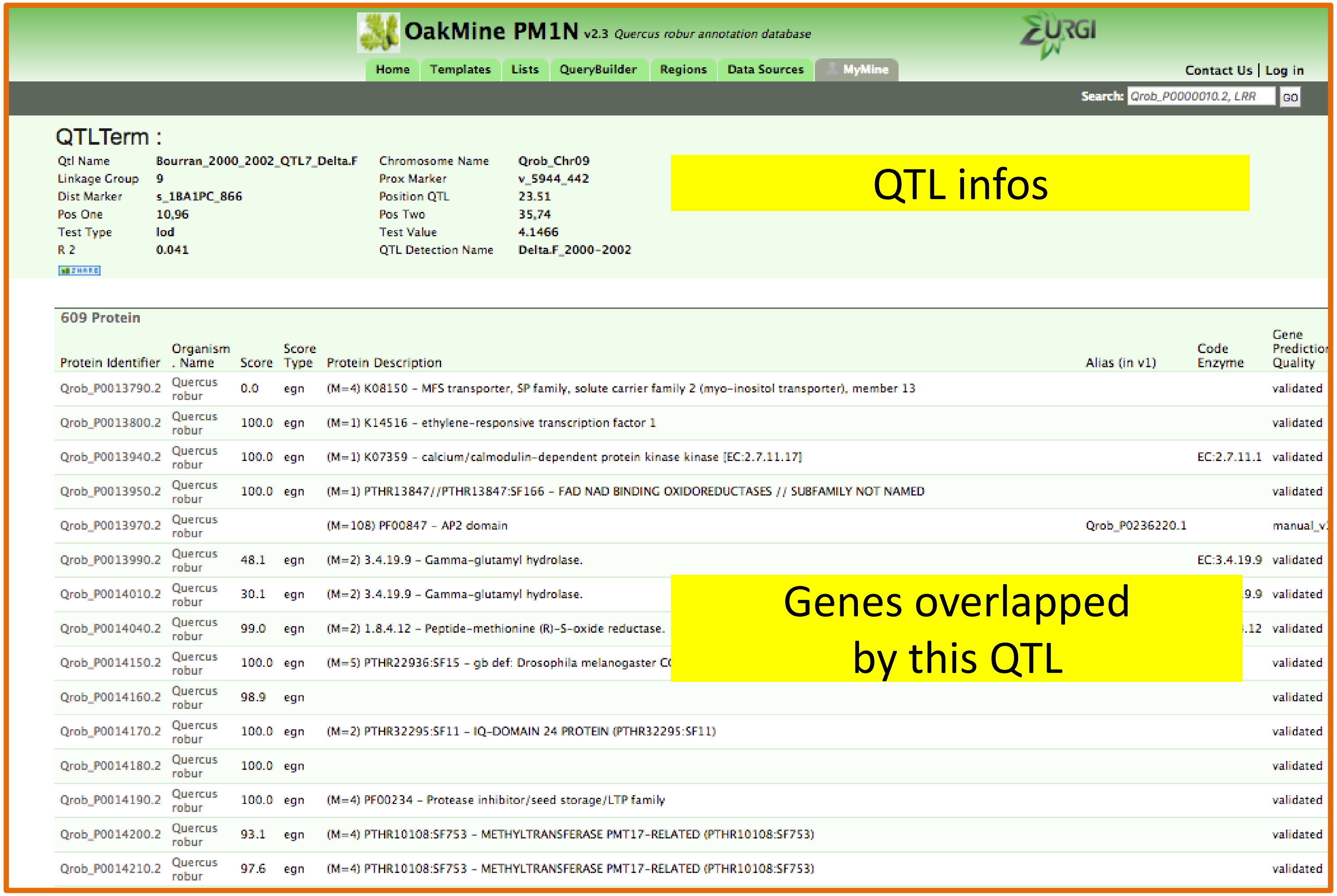
Other genes sharing the same ontology term

Interoperability with other features in the database



11 Qtllist

Qtl Name	Chromosome Name	Linkage Group	Prox Marker	Dist Marker	Position QTL	Pos One	Pos Two	Test Type	Test Value	R 2
Bourran1_2004_QTL4_peak_Bud_burst_A4	Qrob_Chr09	9	s_1BY6BQ_440	s_1AOIKO_756	16,45	0	27,45	lod	4,5	10,6
Bourran_2000_2002_QTL7_Delta.F	Qrob_Chr09	9	v_5944_442	s_1BA1PC_866	23.51	10,96	35,74	lod	4.1466	0.041
Bourran2_2003_QTL11_peak_Bud_burst_3P	Qrob_Chr09	9	s_1CGP2H_273	v_15801_330	27,16	4,16	48,16	lod	2,3	5,1
Bourran2_2003_QTL13_peak_Bud_burst_A4	Qrob_Chr09	9	s_1AP8MN_635	s_1A3QQ_692	18,18	10,88	25,88	lod	3,4	7,2
Bourran2_2004_QTL12_peak_Bud_burst_3P	Qrob_Chr09	9	s_1BDO6G_250	s_1AB3AM_496	34,31	9,31	44,31	lod	3,6	7,6
Bourran2_2004_QTL14_peak_Bud_burst_A4	Qrob_Chr09	9	s_1BY6BQ_440	s_1AOIKO_756	16,83	10,33	22,33	lod	3,8	9



QTLTerm :

Qtl Name	**Bourran_2000_2002_QTL7_Delta.F**	Chromosome Name	**Qrob_Chr09**
Linkage Group	**9**	Prox Marker	**v_5944_442**
Dist Marker	**s_1BA1PC_866**	Position QTL	**23.51**
Pos One	**10,96**	Pos Two	**35,74**
Test Type	**lod**	Test Value	**4.1466**
R 2	**0.041**	QTL Detection Name	**Delta.F_2000-2002**

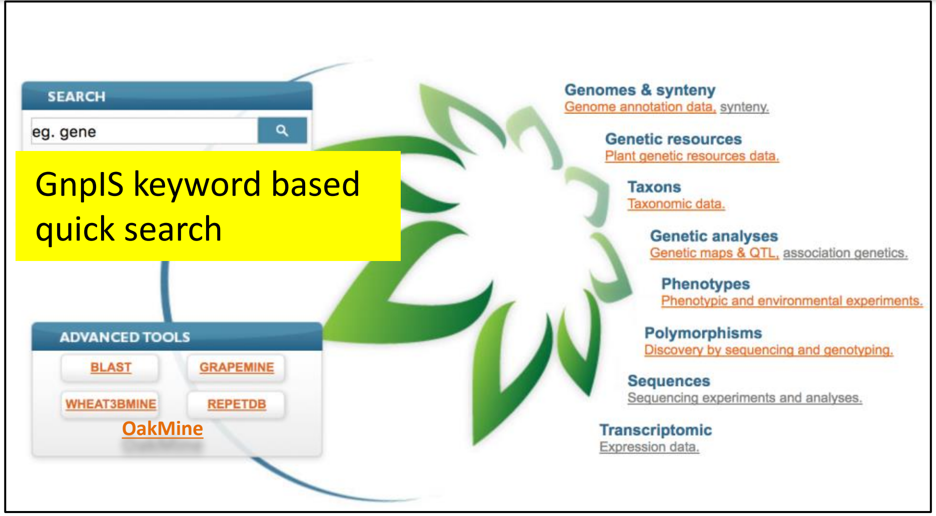
QTL infos

609 Protein

Protein Identifier	Organism . Name	Score	Score Type	Protein Description	Alias (in v1)	Code Enzyme	Gene Prediction Quality
Qrob_P0013790.2	Quercus robur	0.0	egn	(M=4) K08150 - MFS transporter, SP family, solute carrier family 2 (myo-inositol transporter), member 13			validated
Qrob_P0013800.2	Quercus robur	100.0	egn	(M=1) K14516 - ethylene-responsive transcription factor 1			validated
Qrob_P0013940.2	Quercus robur	100.0	egn	(M=1) K07359 - calcium/calmodulin-dependent protein kinase kinase [EC:2.7.11.17]		EC:2.7.11.1	validated
Qrob_P0013950.2	Quercus robur	100.0	egn	(M=1) PTHR13847//PTHR13847:SF166 - FAD NAD BINDING OXIDOREDUCTASES // SUBFAMILY NOT NAMED			validated
Qrob_P0013970.2	Quercus robur			(M=108) PF00847 - AP2 domain	Qrob_P0236220.1		manual_v
Qrob_P0013990.2	Quercus robur	48.1	egn	(M=2) 3.4.19.9 - Gamma-glutamyl hydrolase.		EC:3.4.19.9	validated
Qrob_P0014010.2	Quercus robur	30.1	egn	(M=2) 3.4.19.9 - Gamma-glutamyl hydrolase.		9.9	validated
Qrob_P0014040.2	Quercus robur	99.0	egn	(M=2) 1.8.4.12 - Peptide-methionine (R)-S-oxide reductase.		.12	validated
Qrob_P0014150.2	Quercus robur	100.0	egn	(M=5) PTHR22936:SF15 - gb def: Drosophila melanogaster C			validated
Qrob_P0014160.2	Quercus robur	98.9	egn				validated
Qrob_P0014170.2	Quercus robur	100.0	egn	(M=2) PTHR32295:SF11 - IQ-DOMAIN 24 PROTEIN (PTHR32295:SF11)			validated
Qrob_P0014180.2	Quercus robur	100.0	egn				validated
Qrob_P0014190.2	Quercus robur	100.0	egn	(M=4) PF00234 - Protease inhibitor/seed storage/LTP family			validated
Qrob_P0014200.2	Quercus robur	93.1	egn	(M=4) PTHR10108:SF753 - METHYLTRANSFERASE PMT17-RELATED (PTHR10108:SF753)			validated
Qrob_P0014210.2	Quercus robur	97.6	egn	(M=4) PTHR10108:SF753 - METHYLTRANSFERASE PMT17-RELATED (PTHR10108:SF753)			validated

Genes overlapped by this QTL

Several ways to query data



SEARCH

eg. gene

ADVANCED TOOLS

BLAST | GRAPEMINE | WHEAT3BMINE | REPETDB | OakMine

Genomes & synteny
Genome annotation data, synteny.

Genetic resources
Plant genetic resources data.

Taxons
Taxonomic data.

Genetic analyses
Genetic maps & QTL, association genetics.

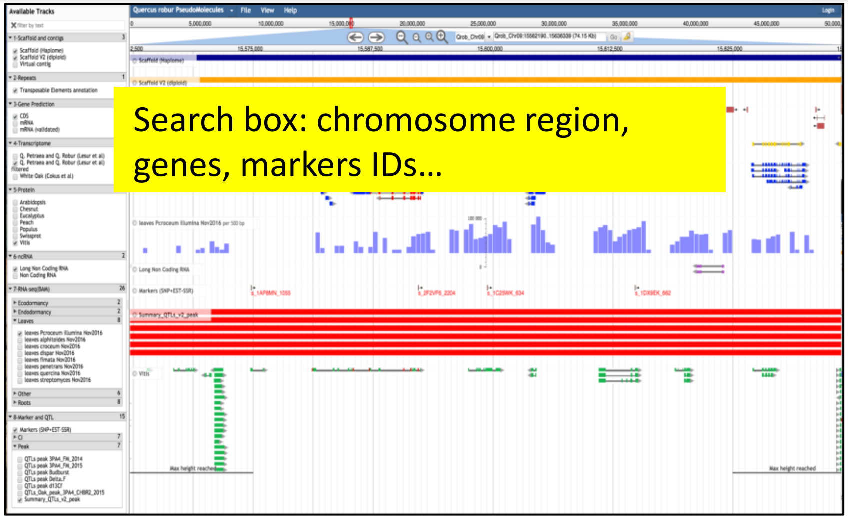
Phenotypes
Phenotypic and environmental experiments.

Polymorphisms
Discovery by sequencing and genotyping.

Sequences
Sequencing experiments and analyses.

Transcriptomic
Expression data.

GnpIS keyword based quick search

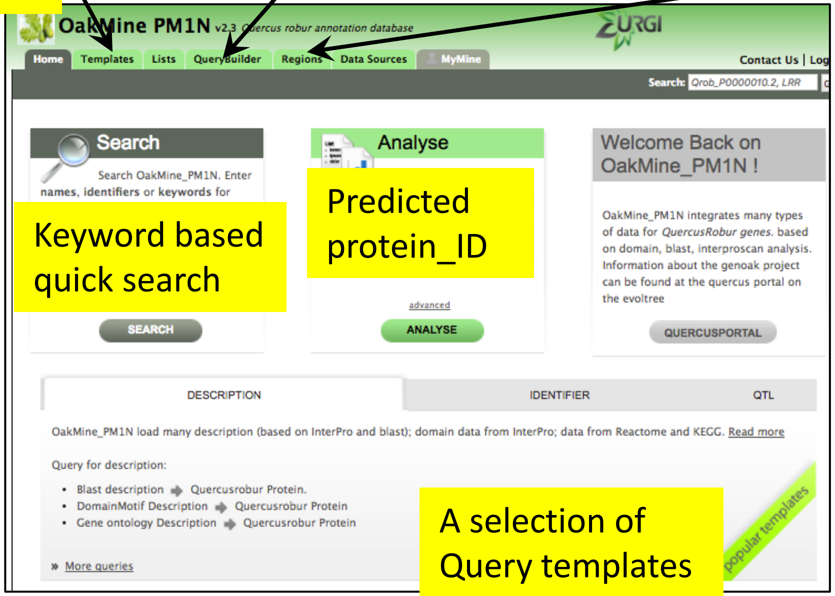


Search box: chromosome region, genes, markers IDs…

Query templates

Advanced search using Query the builder

Search for genes in genome regions



Search: Qrob_P0000010.2, LRR

Search

Search OakMine_PM1N. Enter names, identifiers or keywords for

Analyse

Welcome Back on OakMine_PM1N !

OakMine_PM1N integrates many types of data for *QuercusRobur genes.* based on domain, blast, interproscan analysis. Information about the genoak project can be found at the quercus portal on the evoltree

OakMine_PM1N load many description (based on InterPro and blast); domain data from InterPro; data from Reactome and KEGG. Read more

Query for description:

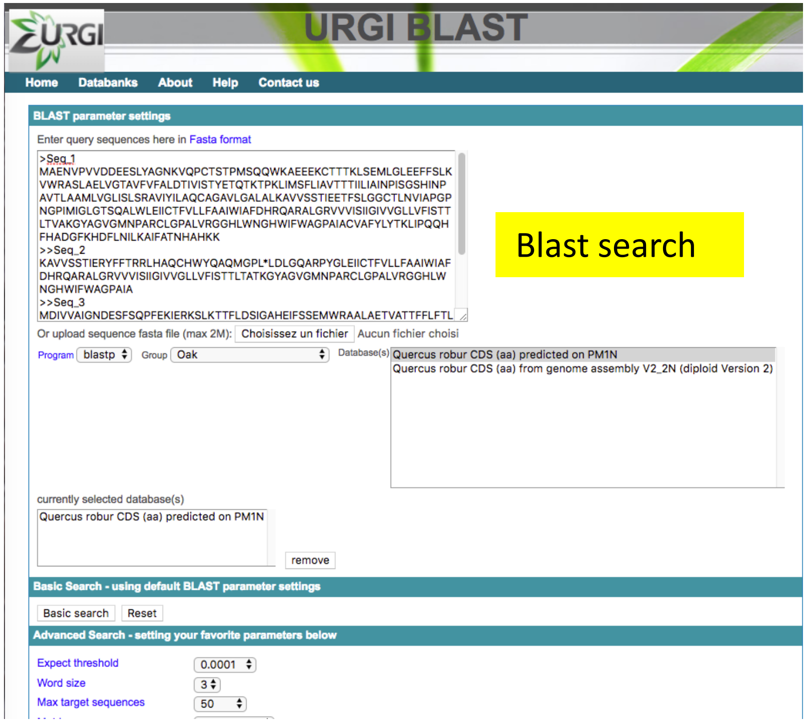
- Blast description ➡ Quercusrobur Protein.
- DomainMotif Description ➡ Quercusrobur Protein
- Gene ontology Description ➡ Quercusrobur Protein

» More queries

Keyword based quick search

Predicted protein_ID

A selection of Query templates



BLAST parameter settings

Enter query sequences here in Fasta format

```
>Seq_1
MAENVPVVDDEESLYAGNKVQPCTSTPMSQQWKAEEEKCTTTKLSEMLGLEEFFSLK
VWRASLAELVGTAVFVFALDTIVISTYETQTKTPKLIMSFLIAVTTTIILIAINPISGSHINP
AVTLAAMLVGLISLSRAVIYILAQCAGAVLGALALKAVVSSTIEETFSLGGCTLNVIAPGP
NGPIMIGLGTSQALWLEIICTFVLLFAAIWIAFDHRQARALGRVVVISIIGIVVGLLVFISTT
LTVAKGYAGVGMNPARCLGPALVRGGHLWNGHWIFWAGPAIACVAFYLYTKLIPQQH
FHADGFKHDFLNILKAIFATNHAHKK
>>Seq_2
KAVVSSTIERYFFTRRLHAQCHWYQAQMGPL*LDLGQARPYGLEIICTFVLLFAAIWIAF
DHRQARALGRVVVISIIGIVVGLLVFISTTLTATKGYAGVGMNPARCLGPALVRGGHLW
NGHWIFWAGPAIA
>>Seq_3
MDIVVAIGNDESFSQPFEKIERKSLKTTFLDSIGAHEIFSSEMWRAALAETVATTFFLFTL
```

Or upload sequence fasta file (max 2M): Choisissez un fichier Aucun fichier choisi

Program: blastp
Group: Oak
Database(s): Quercus robur CDS (aa) predicted on PM1N
Quercus robur CDS (aa) from genome assembly V2_2N (diploid Version 2)

currently selected database(s)
Quercus robur CDS (aa) predicted on PM1N

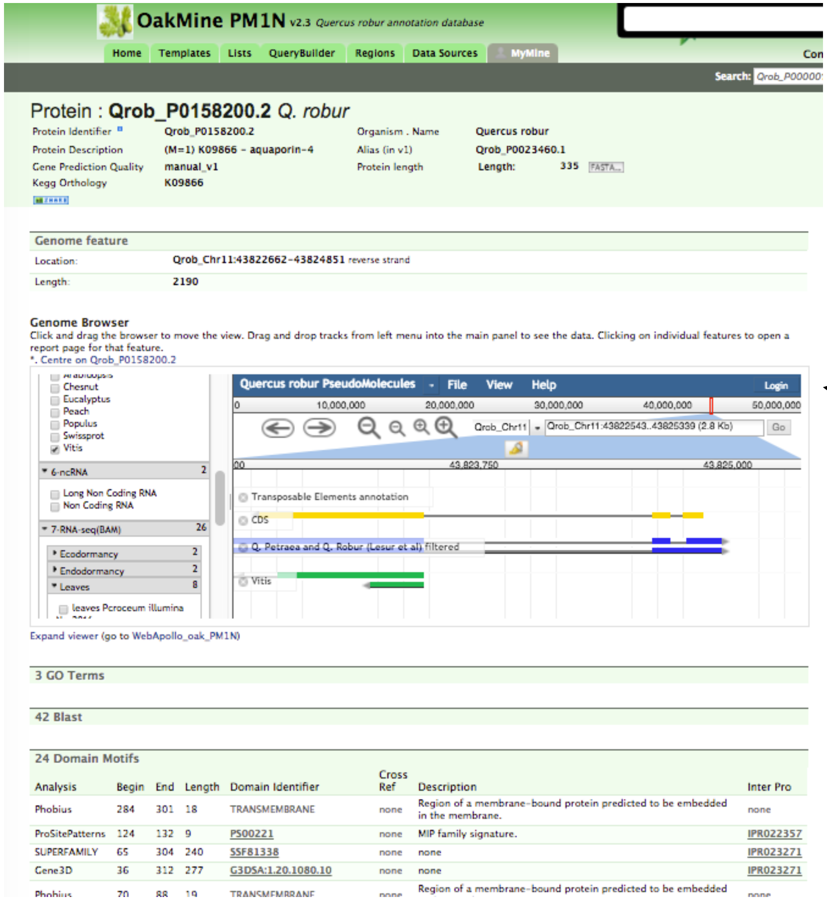
Basic Search - using default BLAST parameter settings

Advanced Search - setting your favorite parameters below

Expect threshold: 0.0001
Word size: 3
Max target sequences: 50

Blast search

Oak assemblies



Protein : Qrob_P0158200.2 *Q. robur*

Protein Identifier	Qrob_P0158200.2	Organism . Name	Quercus robur
Protein Description	(M=1) K09866 - aquaporin-4	Alias (in v1)	Qrob_P0023460.1
Gene Prediction Quality	manual_v1	Protein length	Length: 335 FASTA...
Kegg Orthology	K09866		

Genome feature

Location:	Qrob_Chr11:43822662-43824851 reverse strand
Length:	2190

Genome Browser

Click and drag the browser to move the view. Drag and drop tracks from left menu into the main panel to see the data. Clicking on individual features to open a report page for that feature.

*. Centre on Qrob_P0158200.2

Expand viewer (go to WebApollo_oak_PM1N)

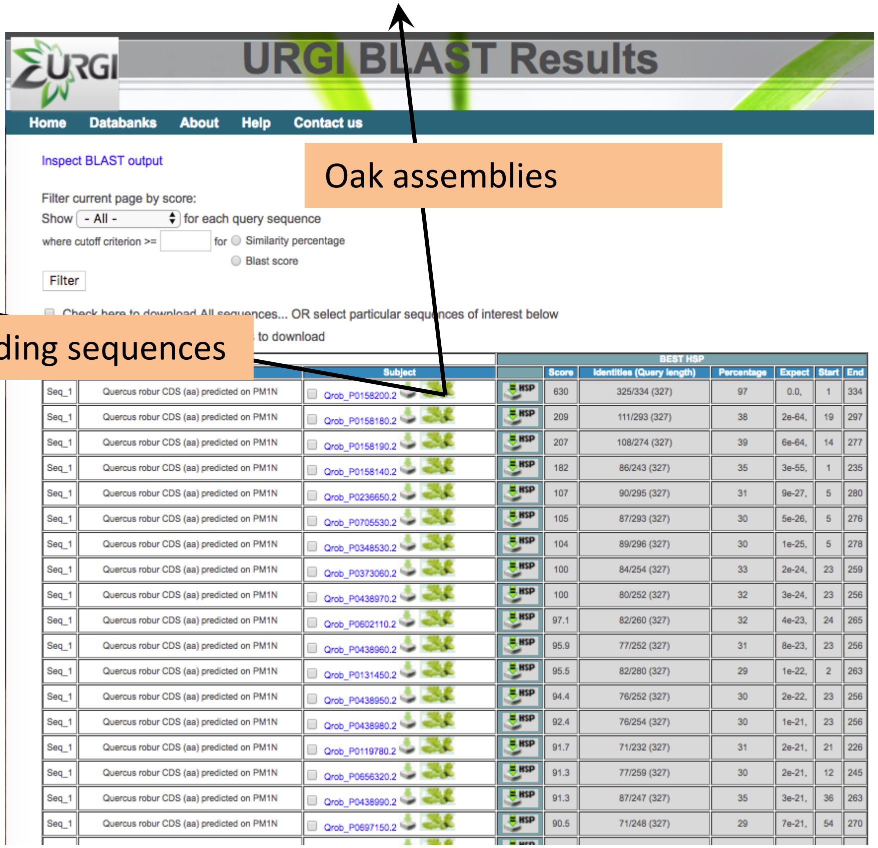
3 GO Terms

42 Blast

24 Domain Motifs

Analysis	Begin	End	Length	Domain Identifier	Cross Ref	Description	Inter Pro
Phobius	284	301	18	TRANSMEMBRANE	none	Region of a membrane-bound protein predicted to be embedded in the membrane.	none
ProSitePatterns	124	132	9	PS00221	none	MIP family signature.	IPR022357
SUPERFAMILY	65	304	240	SSF81338	none	none	IPR023271
Gene3D	36	312	277	G3DSA:1.20.1080.10	none	none	IPR023271
Phobius	70	88	19	TRANSMEMBRANE	none	Region of a membrane-bound protein predicted to be embedded	none

Oak coding sequences



Inspect BLAST output

Filter current page by score:

Show - All - for each query sequence

where cutoff criterion >= for ○ Similarity percentage ○ Blast score

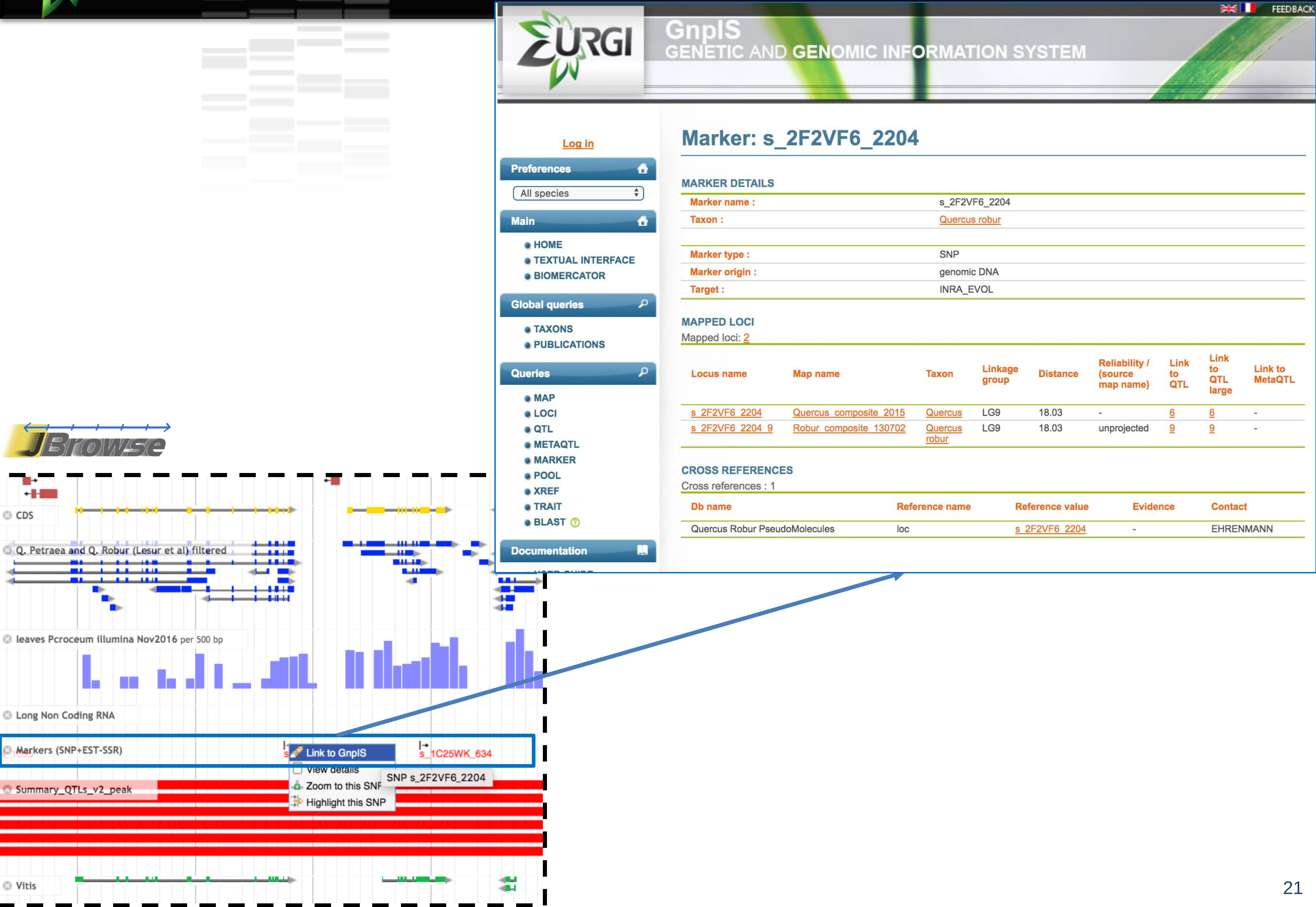
Filter

☐ Check here to download All sequences... OR select particular sequences of interest below to download

		Subject	BEST HSP						
				Score	Identities (Query length)	Percentage	Expect	Start	End
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0158200.2	HSP	630	325/334 (327)	97	0.0,	1	334
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0158180.2	HSP	209	111/293 (327)	38	2e-64,	19	297
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0158190.2	HSP	207	108/274 (327)	39	6e-64,	14	277
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0158140.2	HSP	182	86/243 (327)	35	3e-55,	1	235
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0236650.2	HSP	107	90/295 (327)	31	9e-27,	5	280
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0705530.2	HSP	105	87/293 (327)	30	5e-26,	5	276
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0348530.2	HSP	104	89/296 (327)	30	1e-25,	5	278
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0373060.2	HSP	100	84/254 (327)	33	2e-24,	23	259
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0438970.2	HSP	100	80/252 (327)	32	3e-24,	23	256
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0602110.2	HSP	97.1	82/260 (327)	32	4e-23,	24	265
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0438960.2	HSP	95.9	77/252 (327)	31	8e-23,	23	256
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0131450.2	HSP	95.5	82/280 (327)	29	1e-22,	2	263
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0438950.2	HSP	94.4	76/252 (327)	30	2e-22,	23	256
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0438980.2	HSP	92.4	76/254 (327)	30	1e-21,	23	256
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0119780.2	HSP	91.7	71/232 (327)	31	2e-21,	21	226
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0656320.2	HSP	91.3	77/259 (327)	30	2e-21,	12	245
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0438990.2	HSP	91.3	87/247 (327)	35	3e-21,	36	263
Seq_1	Quercus robur CDS (aa) predicted on PM1N	☐ Qrob_P0697150.2	HSP	90.5	71/248 (327)	29	7e-21,	54	270

Interoperability between Genomics and Genetics

From Gbrowse to Markers in GnpIS



Marker: s_2F2VF6_2204

MARKER DETAILS

Marker name :	s_2F2VF6_2204
Taxon :	Quercus robur
Marker type :	SNP
Marker origin :	genomic DNA
Target :	INRA_EVOL

MAPPED LOCI

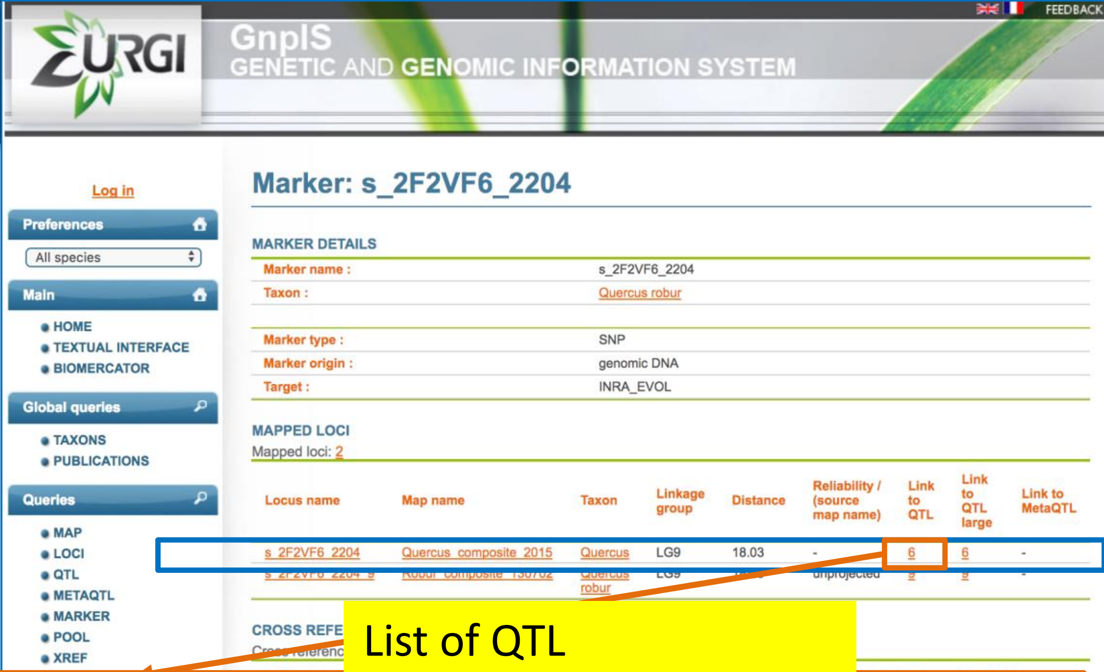
Mapped loci: 2

Locus name	Map name	Taxon	Linkage group	Distance	Reliability / (source map name)	Link to QTL	Link to QTL large	Link to MetaQTL
s_2F2VF6_2204	Quercus_composite_2015	Quercus	LG9	18.03	-	6	6	-
s_2F2VF6_2204_9	Robur_composite_130702	Quercus robur	LG9	18.03	unprojected	9	9	-

CROSS REFERENCES

Cross references : 1

Db name	Reference name	Reference value	Evidence	Contact
Quercus Robur PseudoMolecules	loc	s_2F2VF6_2204	-	EHRENMANN



Marker: s_2F2VF6_2204

MARKER DETAILS

Marker name :	s_2F2VF6_2204
Taxon :	Quercus robur
Marker type :	SNP
Marker origin :	genomic DNA
Target :	INRA_EVOL

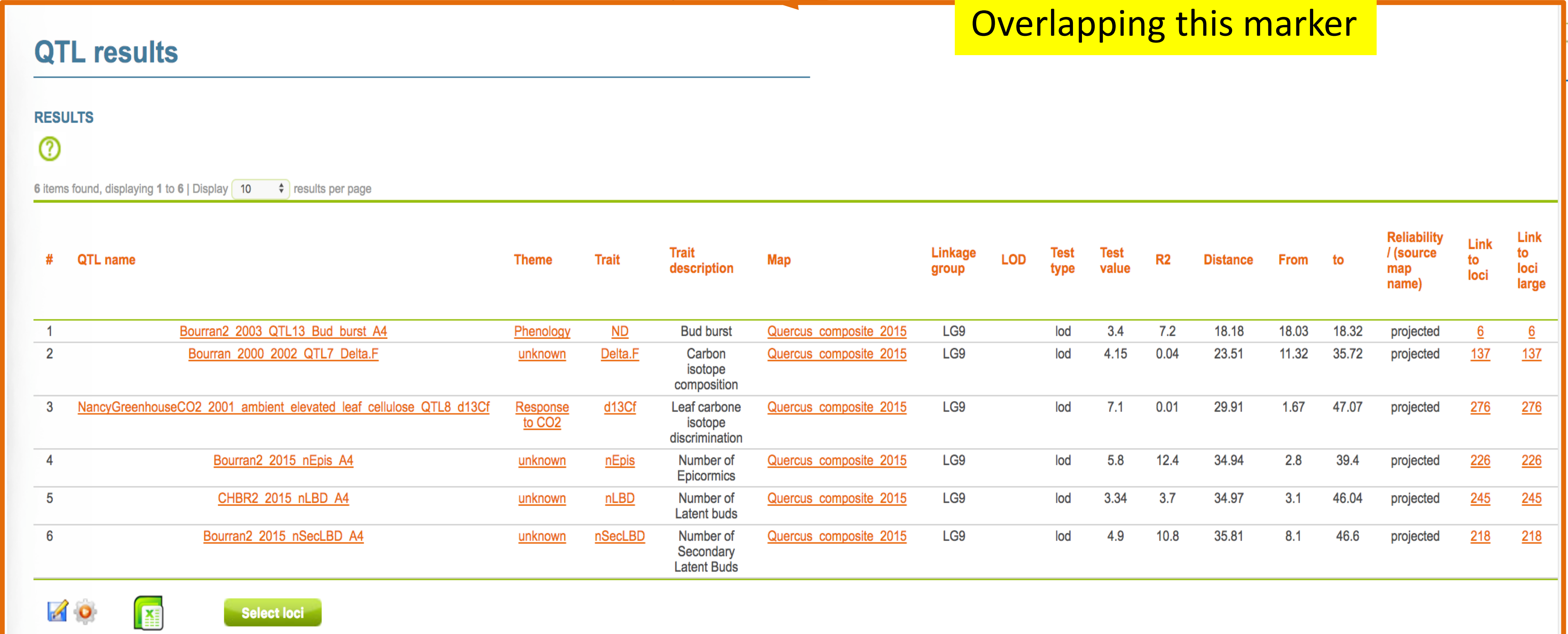
MAPPED LOCI

Mapped loci: 2

Locus name	Map name	Taxon	Linkage group	Distance	Reliability / (source map name)	Link to QTL	Link to QTL large	Link to MetaQTL
s_2F2VF6_2204	Quercus_composite_2015	Quercus	LG9	18.03	-	6	6	-
s_2F2VF6_2204_9	Robur_composite_130702	Quercus robur	LG9	[illegible]	unprojected	9	9	-

CROSS REFE

List of QTL
Overlapping this marker

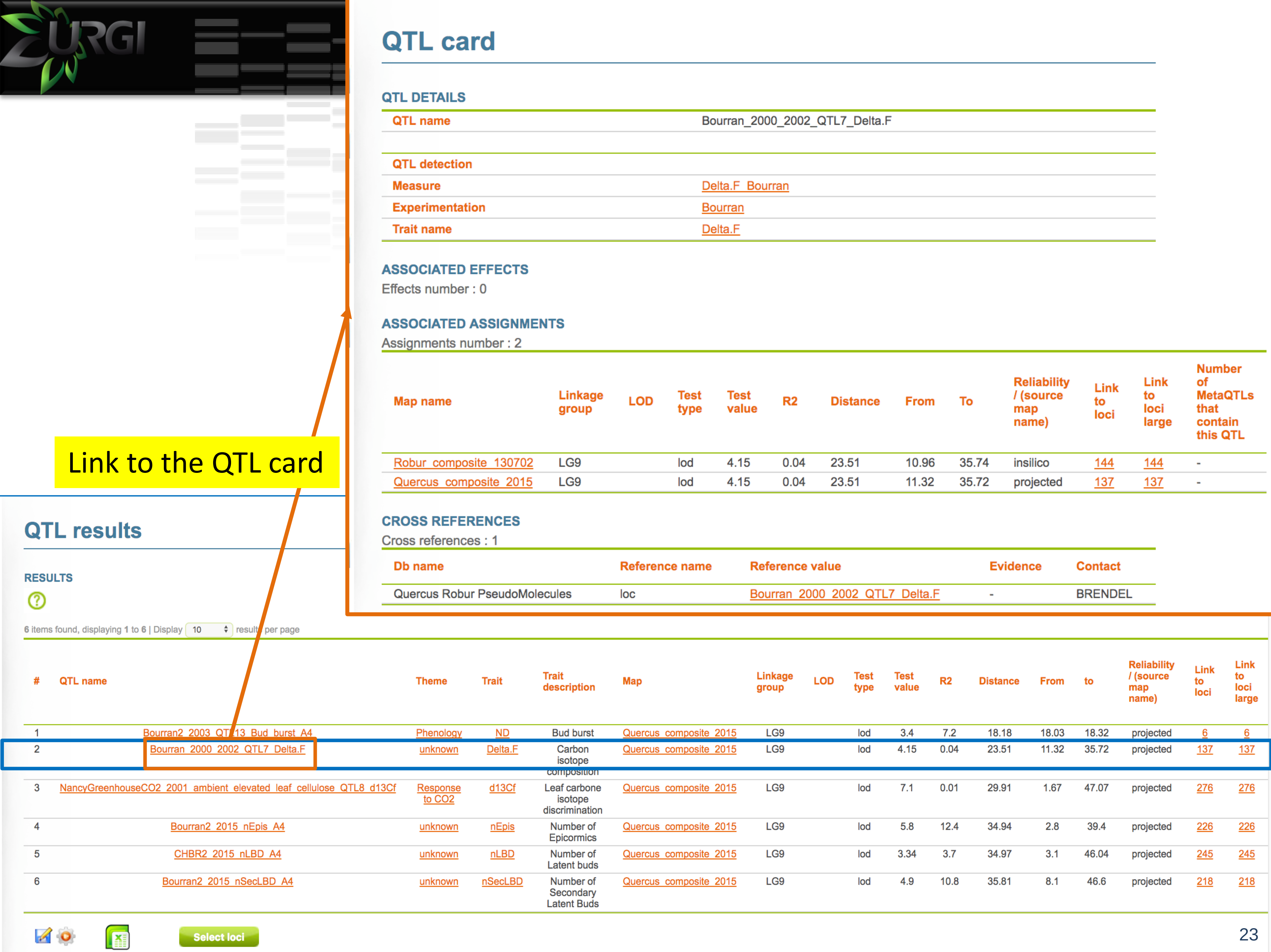


QTL results

RESULTS

6 items found, displaying 1 to 6 | Display 10 results per page

#	QTL name	Theme	Trait	Trait description	Map	Linkage group	LOD	Test type	Test value	R2	Distance	From	to	Reliability / (source map name)	Link to loci	Link to loci large
1	Bourran2_2003_QTL13_Bud_burst_A4	Phenology	ND	Bud burst	Quercus_composite_2015	LG9		lod	3.4	7.2	18.18	18.03	18.32	projected	6	6
2	Bourran_2000_2002_QTL7_Delta.F	unknown	Delta.F	Carbon isotope composition	Quercus_composite_2015	LG9		lod	4.15	0.04	23.51	11.32	35.72	projected	137	137
3	NancyGreenhouseCO2_2001_ambient_elevated_leaf_cellulose_QTL8_d13Cf	Response to CO2	d13Cf	Leaf carbone isotope discrimination	Quercus_composite_2015	LG9		lod	7.1	0.01	29.91	1.67	47.07	projected	276	276
4	Bourran2_2015_nEpis_A4	unknown	nEpis	Number of Epicormics	Quercus_composite_2015	LG9		lod	5.8	12.4	34.94	2.8	39.4	projected	226	226
5	CHBR2_2015_nLBD_A4	unknown	nLBD	Number of Latent buds	Quercus_composite_2015	LG9		lod	3.34	3.7	34.97	3.1	46.04	projected	245	245
6	Bourran2_2015_nSecLBD_A4	unknown	nSecLBD	Number of Secondary Latent Buds	Quercus_composite_2015	LG9		lod	4.9	10.8	35.81	8.1	46.6	projected	218	218



Link to the QTL card

QTL card

QTL DETAILS

QTL name	Bourran_2000_2002_QTL7_Delta.F

QTL detection	
Measure	Delta.F_Bourran
Experimentation	Bourran
Trait name	Delta.F

ASSOCIATED EFFECTS

Effects number : 0

ASSOCIATED ASSIGNMENTS

Assignments number : 2

Map name	Linkage group	LOD	Test type	Test value	R2	Distance	From	To	Reliability / (source map name)	Link to loci	Link to loci large	Number of MetaQTLs that contain this QTL
Robur_composite_130702	LG9		lod	4.15	0.04	23.51	10.96	35.74	insilico	144	144	-
Quercus_composite_2015	LG9		lod	4.15	0.04	23.51	11.32	35.72	projected	137	137	-

CROSS REFERENCES

Cross references : 1

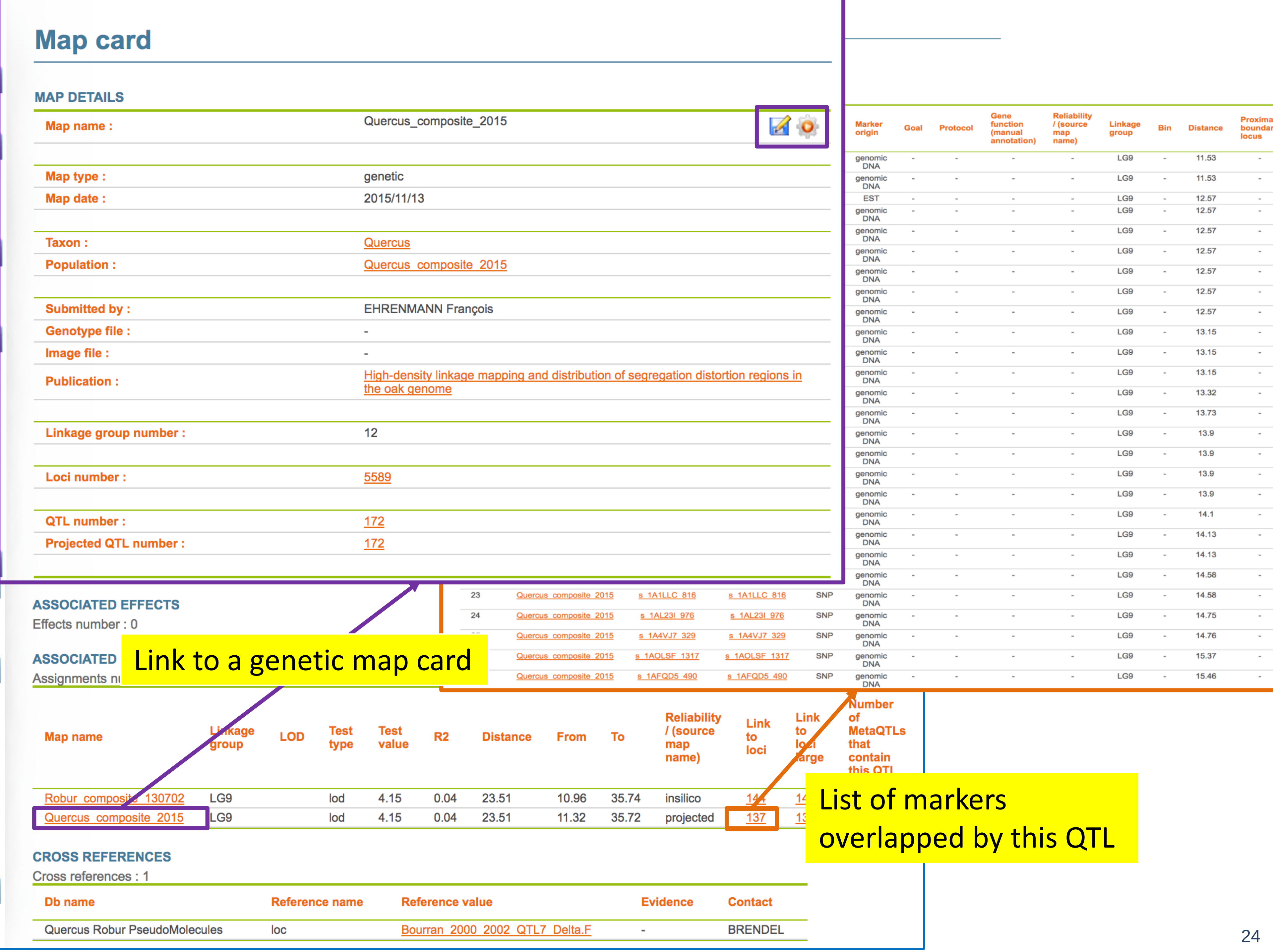
Db name	Reference name	Reference value	Evidence	Contact
Quercus Robur PseudoMolecules	loc	Bourran_2000_2002_QTL7_Delta.F	-	BRENDEL

QTL results

RESULTS

6 items found, displaying 1 to 6 | Display 10 results per page

#	QTL name	Theme	Trait	Trait description	Map	Linkage group	LOD	Test type	Test value	R2	Distance	From	to	Reliability / (source map name)	Link to loci	Link to loci large
1	Bourran2_2003_QTL13_Bud_burst_A4	Phenology	ND	Bud burst	Quercus_composite_2015	LG9		lod	3.4	7.2	18.18	18.03	18.32	projected	6	6
2	Bourran_2000_2002_QTL7_Delta.F	unknown	Delta.F	Carbon isotope composition	Quercus_composite_2015	LG9		lod	4.15	0.04	23.51	11.32	35.72	projected	137	137
3	NancyGreenhouseCO2_2001_ambient_elevated_leaf_cellulose_QTL8_d13Cf	Response to CO2	d13Cf	Leaf carbone isotope discrimination	Quercus_composite_2015	LG9		lod	7.1	0.01	29.91	1.67	47.07	projected	276	276
4	Bourran2_2015_nEpis_A4	unknown	nEpis	Number of Epicormics	Quercus_composite_2015	LG9		lod	5.8	12.4	34.94	2.8	39.4	projected	226	226
5	CHBR2_2015_nLBD_A4	unknown	nLBD	Number of Latent buds	Quercus_composite_2015	LG9		lod	3.34	3.7	34.97	3.1	46.04	projected	245	245
6	Bourran2_2015_nSecLBD_A4	unknown	nSecLBD	Number of Secondary Latent Buds	Quercus_composite_2015	LG9		lod	4.9	10.8	35.81	8.1	46.6	projected	218	218



Map card

MAP DETAILS

Map name :	Quercus_composite_2015
Map type :	genetic
Map date :	2015/11/13
Taxon :	Quercus
Population :	Quercus_composite_2015
Submitted by :	EHRENMANN François
Genotype file :	-
Image file :	-
Publication :	High-density linkage mapping and distribution of segregation distortion regions in the oak genome
Linkage group number :	12
Loci number :	5589
QTL number :	172
Projected QTL number :	172

					Marker origin	Goal	Protocol	Gene function (manual annotation)	Reliability / (source map name)	Linkage group	Bin	Distance	Proxima boundar locus
					genomic DNA	-	-	-	-	LG9	-	11.53	-
					genomic DNA	-	-	-	-	LG9	-	11.53	-
					EST	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	12.57	-
					genomic DNA	-	-	-	-	LG9	-	13.15	-
					genomic DNA	-	-	-	-	LG9	-	13.15	-
					genomic DNA	-	-	-	-	LG9	-	13.15	-
					genomic DNA	-	-	-	-	LG9	-	13.32	-
					genomic DNA	-	-	-	-	LG9	-	13.73	-
					genomic DNA	-	-	-	-	LG9	-	13.9	-
					genomic DNA	-	-	-	-	LG9	-	13.9	-
					genomic DNA	-	-	-	-	LG9	-	13.9	-
					genomic DNA	-	-	-	-	LG9	-	13.9	-
					genomic DNA	-	-	-	-	LG9	-	14.1	-
					genomic DNA	-	-	-	-	LG9	-	14.13	-
					genomic DNA	-	-	-	-	LG9	-	14.13	-
					genomic DNA	-	-	-	-	LG9	-	14.58	-
23	Quercus_composite_2015	s_1A1LLC_816	s_1A1LLC_816	SNP	genomic DNA	-	-	-	-	LG9	-	14.58	-
24	Quercus_composite_2015	s_1AL23I_976	s_1AL23I_976	SNP	genomic DNA	-	-	-	-	LG9	-	14.75	-
	Quercus_composite_2015	s_1A4VJ7_329	s_1A4VJ7_329	SNP	genomic DNA	-	-	-	-	LG9	-	14.76	-
	Quercus_composite_2015	s_1AOLSF_1317	s_1AOLSF_1317	SNP	genomic DNA	-	-	-	-	LG9	-	15.37	-
	Quercus_composite_2015	s_1AFQD5_490	s_1AFQD5_490	SNP	genomic DNA	-	-	-	-	LG9	-	15.46	-

ASSOCIATED EFFECTS

Effects number : 0

ASSOCIATED

Assignments nu

Link to a genetic map card

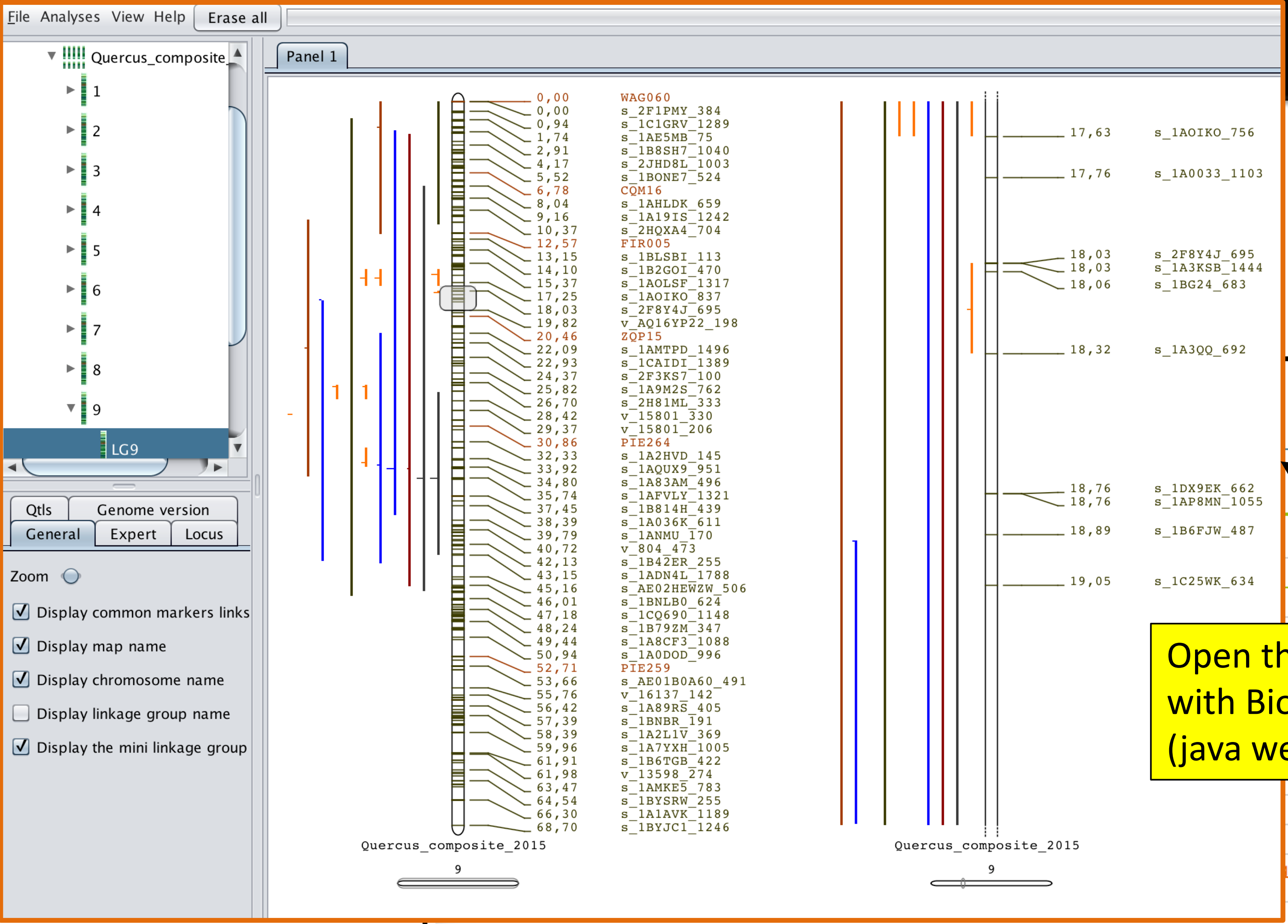
Map name	Linkage group	LOD	Test type	Test value	R2	Distance	From	To	Reliability / (source map name)	Link to loci	Link to loci large	Number of MetaQTLs that contain this QTL
Robur_composite_130702	LG9		lod	4.15	0.04	23.51	10.96	35.74	insilico	14	14	
Quercus_composite_2015	LG9		lod	4.15	0.04	23.51	11.32	35.72	projected	137	13	

List of markers overlapped by this QTL

CROSS REFERENCES

Cross references : 1

Db name	Reference name	Reference value	Evidence	Contact
Quercus Robur PseudoMolecules	loc	Bourran_2000_2002_QTL7_Delta.F	-	BRENDEL



☑ Display common markers links

☑ Display map name

☑ Display chromosome name

☐ Display linkage group name

☑ Display the mini linkage group

Open the Genetic map with BioMercator (java webstart)

regation distortion regions in

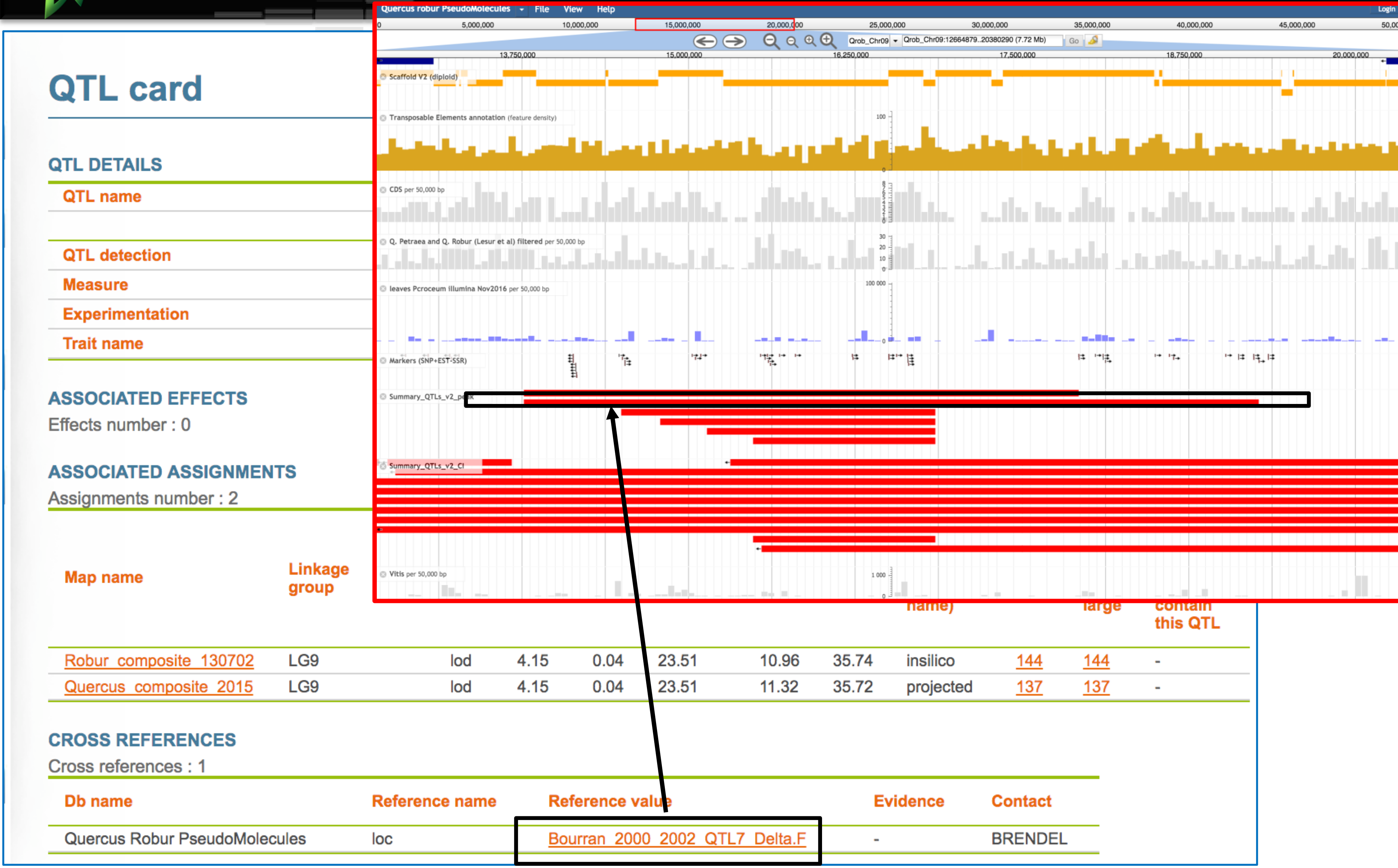
Linkage group number : 12

Loci number : 5589

QTL number : 172

Projected QTL number : 172

From GnpIS QTL card to JBrowse



QTL card

QTL DETAILS

QTL name

QTL detection

Measure

Experimentation

Trait name

ASSOCIATED EFFECTS

Effects number : 0

ASSOCIATED ASSIGNMENTS

Assignments number : 2

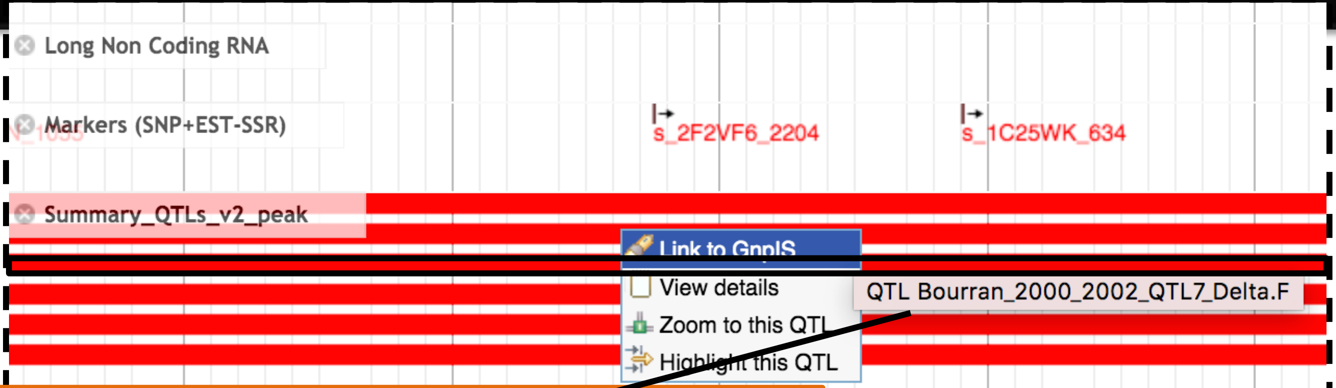
Map name	Linkage group							name)		large	contain this QTL
Robur_composite_130702	LG9	lod	4.15	0.04	23.51	10.96	35.74	insilico	144	144	-
Quercus_composite_2015	LG9	lod	4.15	0.04	23.51	11.32	35.72	projected	137	137	-

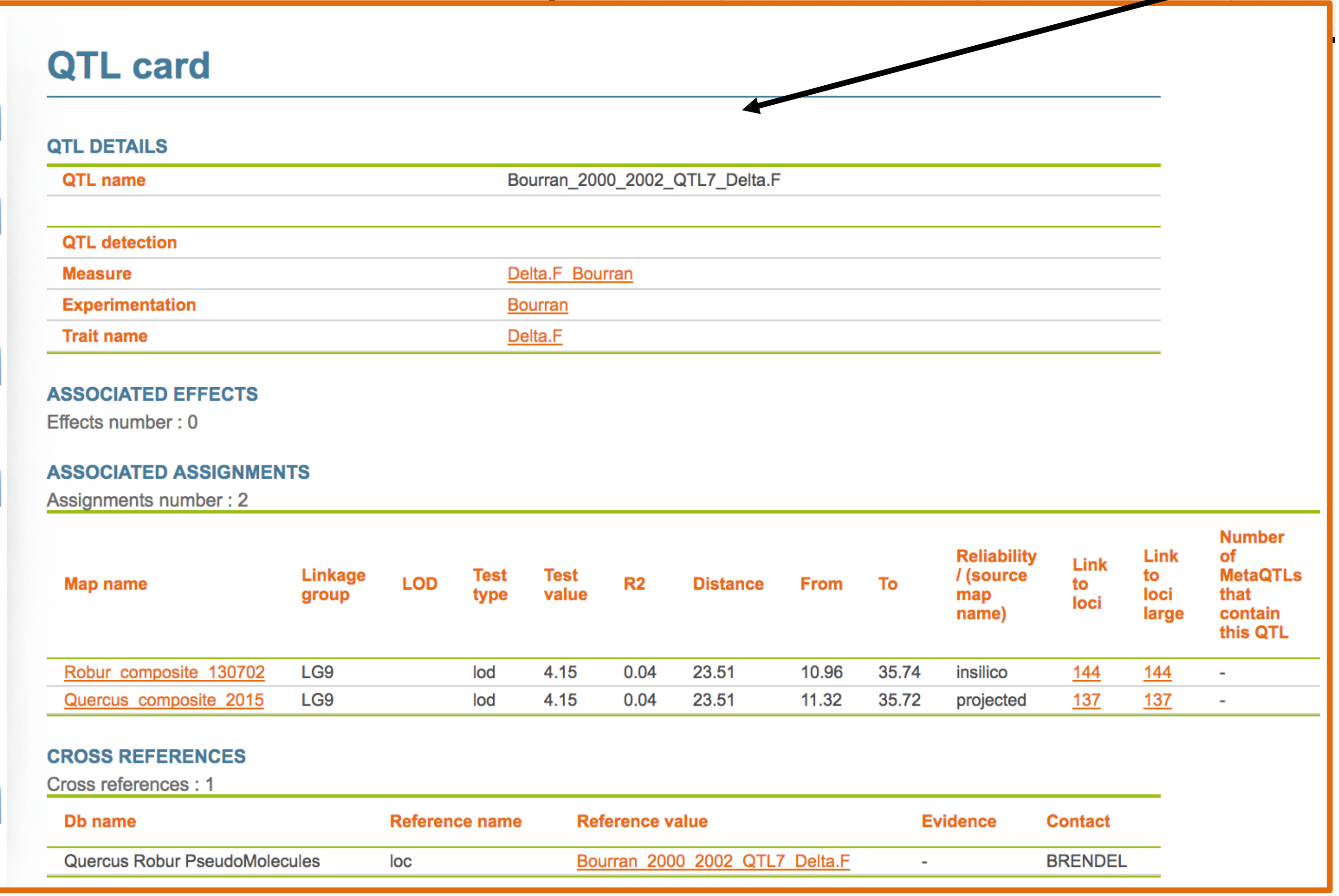
CROSS REFERENCES

Cross references : 1

Db name	Reference name	Reference value	Evidence	Contact
Quercus Robur PseudoMolecules	loc	Bourran_2000_2002_QTL7_Delta.F	-	BRENDEL

From JBrowse to the same QTL card





QTL card

QTL DETAILS

QTL name	Bourran_2000_2002_QTL7_Delta.F
QTL detection	
Measure	Delta.F_Bourran
Experimentation	Bourran
Trait name	Delta.F

ASSOCIATED EFFECTS

Effects number : 0

ASSOCIATED ASSIGNMENTS

Assignments number : 2

Map name	Linkage group	LOD	Test type	Test value	R2	Distance	From	To	Reliability / (source map name)	Link to loci	Link to loci large	Number of MetaQTLs that contain this QTL
Robur_composite_130702	LG9		lod	4.15	0.04	23.51	10.96	35.74	insilico	144	144	-
Quercus_composite_2015	LG9		lod	4.15	0.04	23.51	11.32	35.72	projected	137	137	-

CROSS REFERENCES

Cross references : 1

Db name	Reference name	Reference value	Evidence	Contact
Quercus Robur PseudoMolecules	loc	Bourran_2000_2002_QTL7_Delta.F	-	BRENDEL

Acknowledgments

J Amselem

H Quesneville

N Francillonne

C Michotey

R Flores

T Letellier

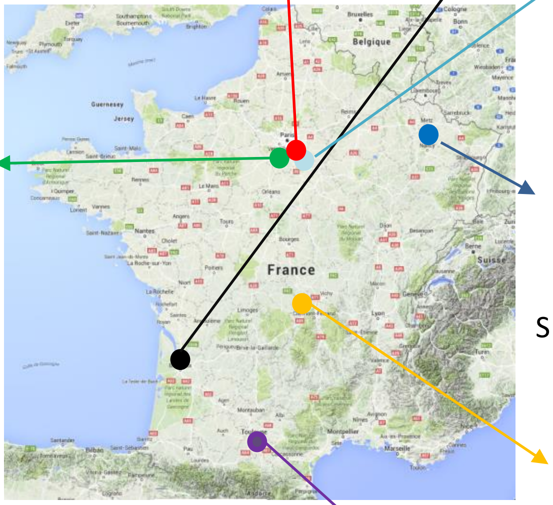
C Da Silva
K Labadie
Faye S

JM Aury

Biodiversité, gènes & communautés

F Ehrenmann
JC Leplé
T Leroy
I Lesur
C Bodénès
G Le Provost
A Kremer

C Plomion



P Faivre-rampant

INTERACTIONS ARBRES-MICROORGANISMES

S Duplessis F Martin

F Murat J Salse

genotoul bioinfo

C Klopp C Gaspin
O Rué

Funding

arbre
LabEx

Agence Nationale de la Recherche ANR
Genoak

erc
Treepeace