



**HAL**  
open science

## **An integrated information system dedicated to oak genomics and genetics**

Joëlle Amselem, Nicolas Francillonne, Célia Michotey, Thomas Letellier,  
Jean-Marc Aury, Corinne da Silva, Sébastien Duplessis, François Ehrenmann,  
Sébastien Faye, Christine Gaspin, et al.

► **To cite this version:**

Joëlle Amselem, Nicolas Francillonne, Célia Michotey, Thomas Letellier, Jean-Marc Aury, et al.. An integrated information system dedicated to oak genomics and genetics. PAG XXVI Plant and Animal Genome Conference, Jan 2018, San Diego, United States. pp.27 slides. hal-01839185

**HAL Id: hal-01839185**

**<https://hal.science/hal-01839185v1>**

Submitted on 5 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## An Integrated Information System Dedicated to Oak Genomics and Genetics

The Largest Ag-Genomics Meeting in the World  
**PLANT & ANIMAL GENOME CONFERENCE 2017**  
 www.intpag.org  
 Digital Tools and Resources Session 1  
 Saturday, January 13, 2018

joelle.amesem@inra.fr

**GnpIS: A Genetic and genomic Information System**  
<https://urgi.versailles.inra.fr/gnpis>

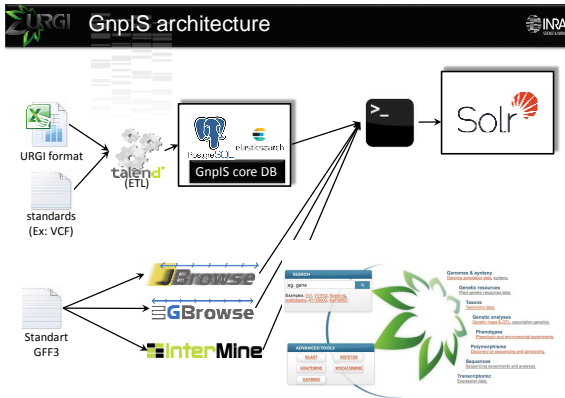
**Database tool**  
 GnpIS: an information system to integrate genetic and genomic data from plants and fungi

Example: *Vitis* **VITIS** database, **Arabidopsis**, **AC10000**, **Arabidopsis**

**SEARCH**  
 eg. gene

**ADVANCED TOOLS**  
 BLAST, GRAPHEME, OAKMINE, REPETOR, WHATISMINE

- Genomes & synteny
- Genetic resources
- Taxons
- Genetic analyses
- Phenotypes
- Polymerphisms
- Sequences
- Transcriptomic



**Quercus robur genome in numbers**

- Physical size: 1.5 Gb (Kremer et al. 2007)
- 12 pairs of chromosomes
- genetic size: 933 cM
  - 12 linkage groups from 5.5k SNP-based genetic linkage maps (Bodénès et al. 2016)
- Assembly and annotation
  - Diploid version (2N): 8,827 scaffolds covering 1.45 Gb
  - Haploid 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 et al. under review

**Structural annotation pipelines**

**Transposable Elements**  
 REPET package  
<https://urgi.versailles.inra.fr/tools/REPET>  
 TEdenovo, TEannot

**Genes prediction**  
 Eugene  
<http://eugene.toulouse.inra.fr/>

**Long Non Coding RNA**  
 Alignment, Repeat detection, Repeat calling, Repeat filtering, Repeat annotation

**Non Coding RNA prediction and annotation**  
 Genome sequence, tRNA subfamily, Structural/Functional annotation, tRNA subfamily, tRNA subfamily, tRNA subfamily

**Functional annotation pipeline**

Predicted polypeptides → InterProScan

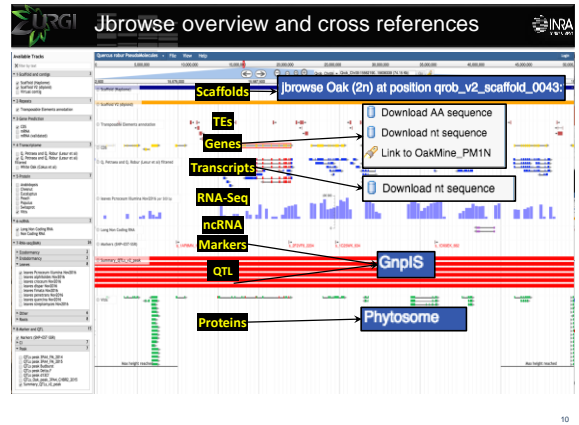
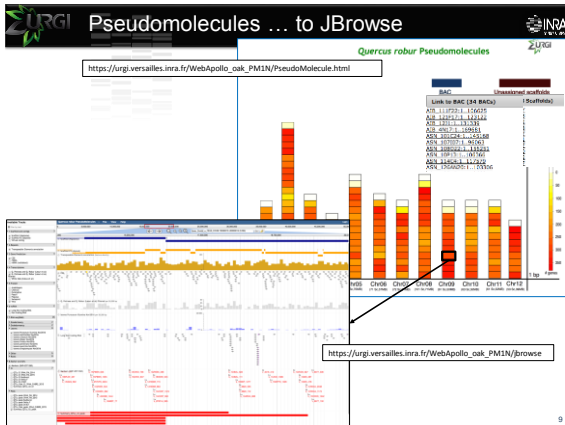
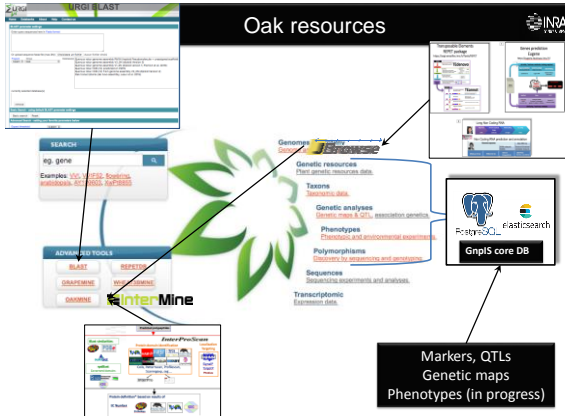
**Blast similarities**  
 PDB, swissprot, rpsBlast, Conserved domains, InterPro, KOG

**Protein domain identification**  
 HAM, PIRSF, JIGR, PANTHER, PRINTS, ProDom, SMART, Superfamily, Coils, PatternsScan, Profilescan, Scanregexp, seg ...

**Localization Targeting**  
 TMHMM, SignalP, TargetP, Phobius

**Protein definition\* based on results of**  
 EC Number, Orthology, PANTHER, KOG

\*script from David Goodman, personal communication



**OakMine portal**

Search OakMine\_PM1N. Enter names, identifiers or keywords for genes, proteins, pathways, ontology terms, etc. (e.g. LRR, glycolysis).

Analyse: Enter a list of identifiers.

Welcome Back on OakMine\_PM1N!

DESCRIPTION IDENTIFIER QTL

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

**OakMine\_PM1N central object: the gene card**

Protein: Qrob\_P0153160.2 q. robur

Location: qrob\_v2\_scaffold\_00493

Genome features

Gene structure: Exons, Introns, UTRs, etc.

Integrated JBrowse

Functional annotation

QTLs overlapping this gene

### IRGI Functional annotation : cross references to external DBs

The screenshot displays the IRGI interface for functional annotation. It features a table of gene annotations with columns for Gene Name, ID, and various metrics. Several pop-up windows are overlaid on the table, providing detailed information from external databases:
 

- Gene Ontology:** Shows GO terms associated with the gene, such as 'GO:0006918' (cellular response to hypoxia).
- KEGG:** Shows KEGG pathway maps and gene names.
- Pfam:** Shows protein domain families and their locations within the protein.
- panther:** Shows protein families and clusters.
- Interpro:** Shows multiple domain annotations from various databases.

### IRGI Interoperability with other features in OakMine

The screenshot shows the OakMine PMIN interface for a specific GO term: 'GO:0006378 mRNA polyadenylation GO'. The interface includes:
 

- GO Term Details:** A detailed description of the GO term and its associated gene.
- Gene ontology parent terms:** A list of parent GO terms, such as 'GO:0006377 mRNA processing'.
- Other genes sharing the same ontology term:** A list of other genes that share the same GO term, including their names and accession numbers.

### IRGI Interoperability with other features in the database

The screenshot displays the IRGI interface for QTL information. It features a table of QTLs with columns for QTL Name, Chromosome Name, Linkage Group, and various metrics. A pop-up window titled 'QTL info' is overlaid, showing details for a specific QTL. Another pop-up window titled 'Genes overlapped by this QTL' is overlaid, showing a list of genes that overlap with the QTL.

### IRGI Several ways to query data

The screenshot shows the OakMine PMIN interface with several query methods highlighted:
 

- Gngits keyword based quick search:** A search box for keywords related to chromosomes, regions, genes, and markers.
- Advanced search using Query templates:** A search box for advanced queries using templates.
- Search for genes in genome regions:** A search box for genes in specific genome regions.
- Keyword based quick search:** A search box for keywords.
- Predicted protein ID:** A search box for predicted protein IDs.
- A selection of Query templates:** A list of query templates.

### IRGI Use list to store and/or improve complex queries

The screenshot shows the OakMine PMIN interface with a list of proteins and oak assemblies. The list includes:
 

- 134 proteins:** A list of 134 proteins.
- 66 proteins:** A list of 66 proteins.
- Oak assemblies:** A list of oak assemblies.
- Oak coding sequences:** A list of oak coding sequences.

### IRGI Use list to store and/or improve complex queries

The screenshot shows the OakMine PMIN interface with a list of proteins and oak assemblies. The list includes:
 

- 134 proteins:** A list of 134 proteins.
- 66 proteins:** A list of 66 proteins.
- Oak assemblies:** A list of oak assemblies.
- Oak coding sequences:** A list of oak coding sequences.



# Interoperability between Genomics and Genetics

**From Gbrowse to Markers in GnpIS**

**ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM**

Marker: s\_2F2VF6\_2204

**MARKER DETAILS**

Marker name: s\_2F2VF6\_2204  
 Species: Sus scrofa  
 Chromosome: SSC11  
 Position: 10,900,000  
 Length: 100 bp  
 Reference: Ensembl  
 Type: SNP

**ASSOCIATED EFFECTS**

Effect number: 1  
 Effect name: s\_2F2VF6\_2204  
 Reference name: s\_2F2VF6\_2204  
 Reference value: 1  
 Evidence: Ensembl

**CROSS REFERENCES**

DB name: Reference name: Reference value: Evidence: Contact

19

20

**ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM**

Marker: s\_2F2VF6\_2204

**QTL results**

RESULTS

#	QTL name	Theme	Test	Test statistics	Map	Linkage group	LOD	Test type	Test value	RS	Distance	From %	Reliability (Genetics)	Link to map	Link to QTL
1	Breeder_2015_QTL_1_Breeder_2015	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	1.1	1.2	18.18	18.18	14.30	preliminary	1	1
2	Breeder_2015_QTL_2_Breeder_2015	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	4.1	3.8	23.51	11.32	35.74	preliminary	2	2
3	Neofunctionalised_2015_annotated_and_valued_QTL_2204	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	3.1	3.0	26.91	16.0	47.0	preliminary	3	3
4	Breeder_2015_QTL_4_B	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	5.8	12.4	36.28	2.8	35.4	preliminary	4	4
5	CHIBP_2015_QTL_5_B	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	3.8	3.7	36.97	3.1	45.6	preliminary	5	5
6	Breeder_2015_QTL_6_B	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	4.9	10.8	35.91	5.1	45.6	preliminary	6	6

**List of QTL overlapping this marker**

**ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM**

**QTL card**

QTL name: Breeder\_2015\_QTL\_2\_Breeder\_2015

QTL location: Data E: Quantitative

Experimentation: Breeder

Trait name: Data E

**ASSOCIATED EFFECTS**

Effect number: 1

**ASSOCIATED ASSIGNMENTS**

Assignment number: 2

Map name	Linkage group	LOD	Test type	Test value	RS	Distance	From %	Reliability (Genetics)	Link to map	Link to QTL
Breeder_2015_QTL_2_Breeder_2015	LOD	inf	1.1	1.2	18.18	14.30	preliminary	1	1	
Breeder_2015_QTL_2_Breeder_2015	LOD	inf	4.1	3.8	23.51	11.32	35.74	preliminary	2	2

**QTL results**

RESULTS

#	QTL name	Theme	Test	Test statistics	Map	Linkage group	LOD	Test type	Test value	RS	Distance	From %	Reliability (Genetics)	Link to map	Link to QTL
1	Breeder_2015_QTL_1_Breeder_2015	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	1.1	1.2	18.18	14.30	14.30	preliminary	1	1
2	Breeder_2015_QTL_2_Breeder_2015	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	4.1	3.8	23.51	11.32	35.74	preliminary	2	2
3	Neofunctionalised_2015_annotated_and_valued_QTL_2204	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	3.1	3.0	26.91	16.0	47.0	preliminary	3	3
4	Breeder_2015_QTL_4_B	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	5.8	12.4	36.28	2.8	35.4	preliminary	4	4
5	CHIBP_2015_QTL_5_B	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	3.8	3.7	36.97	3.1	45.6	preliminary	5	5
6	Breeder_2015_QTL_6_B	Quantitative	SN	Per Link	Genetic map 2015	LOD	inf	4.9	10.8	35.91	5.1	45.6	preliminary	6	6

**Link to the QTL card**

21

22

**ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM**

**Map card**

MAP DETAILS

Map name: Quantitative\_2015

Map type: genetic

Map date: 2015/12/15

Theme: Quantitative

Provenance: Quantitative\_2015

Submitted by: LINDOUBEN Françoise

Genotype file: LINDOUBEN Françoise

Image file: Map details linkage mappings and distribution of recombination fraction maps in the job archive

Publication: Map details linkage mappings and distribution of recombination fraction maps in the job archive

Linkage group number: 12

Last number: 5089

QTL number: 122

Projected QTL number: 122

**ASSOCIATED EFFECTS**

Effect number: 2

**ASSOCIATED ASSIGNMENTS**

Assignment number: 2

Map name	Linkage group	LOD	Test type	Test value	RS	Distance	From %	Reliability (Genetics)	Link to map	Link to QTL
Breeder_2015_QTL_2_Breeder_2015	LOD	inf	4.1	3.8	23.51	11.32	35.74	preliminary	2	2
Quantitative_2015	LOD	inf	4.1	3.8	23.51	11.32	35.74	preliminary	2	2

**Link to a genetic map card**

**List of markers overlapped by this QTL**

23

**ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM**

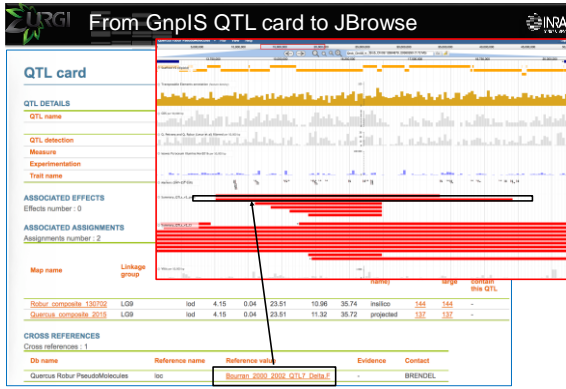
**Genetic map with BioMercator (java webstart)**

Linkage group number: 12

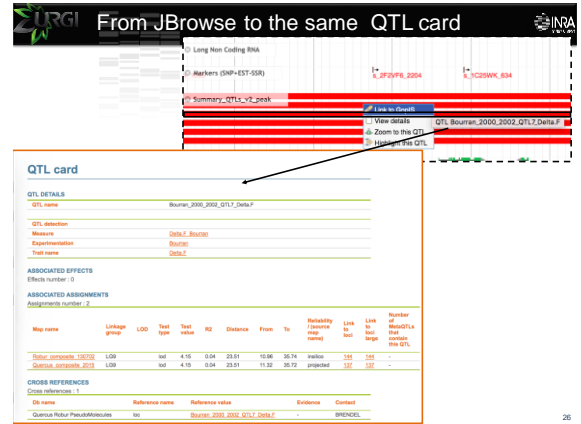
Last number: 5089

QTL number: 122

Projected QTL number: 122





25



26

## Acknowledgments

J. Ansellem

R. Flores

H. Quesneville

C. Michotey

N. Francillonne

T. Letellier

C. Da Silva

K. Labadie

Faye S.

C. Plomion

F. Ehrensmann

J.C. Leprie

T. Leroy

C. Boudonès

G. Le Provost

A. Kremer

P. Falvère-ranipant

S. Duplessis

F. Martin

F. Murat

J. Salse

C. Klopp

C. Gaignon

O. Rué

**Funding**

