



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-01839185>

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.

GnplS

An Integrated Information System Dedicated to Oak Genomics and Genetics

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

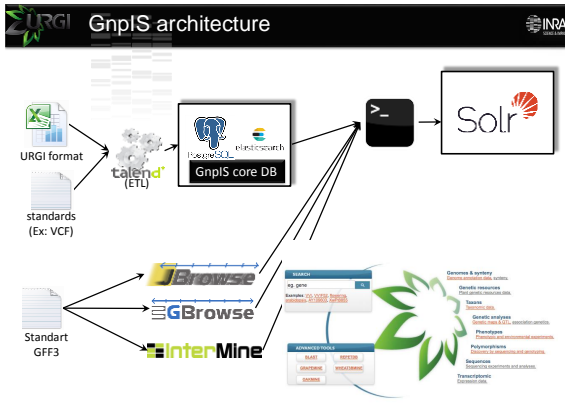
joelle.amesem@inra.fr

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

SEARCH
 eg. gene
 Example: *UGL1*, *VUERC*, *Quercus*, *amblyopaea*, *AV109003*, *QNP3555*

ADVANCED TOOLS
 BLAST, REPETITOR, GRAPHEME, WHATISMINE, OAKMINE

- 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>
 Tdenovo, TEannot

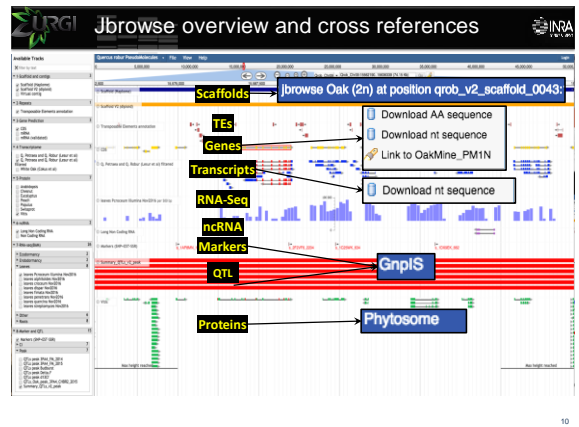
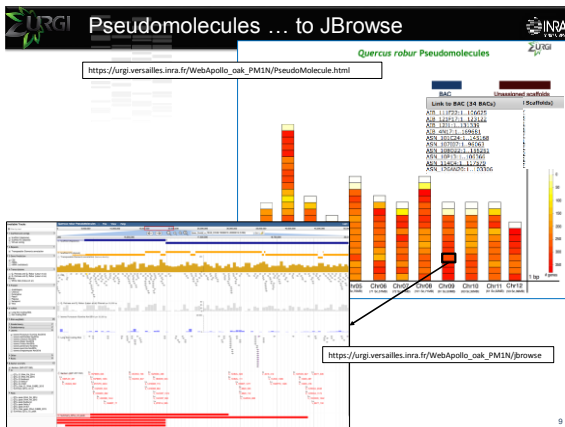
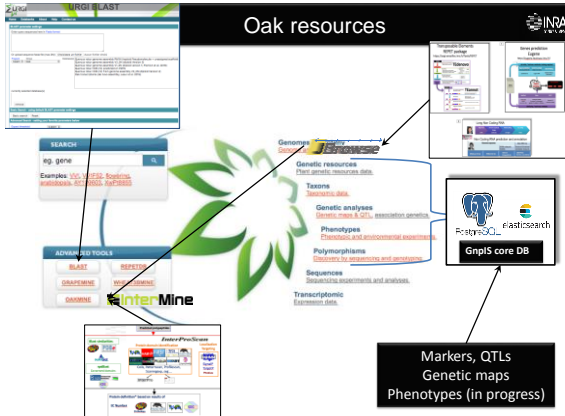
Genes prediction
 Eugene
<http://eugene.toulouse.inra.fr/>

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 resources, tRNA-Plant, tRNA-Plant, tRNA-Plant

Functional annotation pipeline

Predicted polypeptides → InterProScan → Blast similarities (PDB, SwissProt, rpsBlast) → Protein domain identification (HAM, PIRSF, JIGR, PANTHER, PRINTS, ProDom, SMART) → Localization Targeting (TMHMM, SignalP, TargetP, Phobius) → InterPro → Protein definition* based on results of (EC Number, Orthology, PANTHER, KOG)



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
- DomainMotif Description
- Gene ontology Description

OakMine_PM1N central object: the gene card

Protein: Qrob_P0153160.2 q. robur

Location: qrob_v2_scaffold_00493

Genome features

Gene structure: 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 term and its associated gene products.
- Gene ontology parent terms:** A list of parent GO terms, such as 'GO:0006377 mRNA processing GO'.
- Other genes sharing the same ontology term:** A list of other genes that are annotated with the same GO term.

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 other metrics. A pop-up window titled 'QTL info' provides details for a specific QTL, including its location and associated genes. Another pop-up window titled 'Genes overlapped by this QTL' lists the genes that overlap with the QTL region.

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 builder for complex queries.
- Search for genes in genome regions:** A search for genes within specific genomic regions.
- Keyword based quick search:** A search for specific keywords.
- Predicted protein_ID:** A search for predicted protein IDs.
- A selection of Query templates:** A list of pre-defined query templates.

IRGI Use list to store and/or improve complex queries

The screenshot shows the OakMine PMIN interface for managing query lists. It features a table of queries with columns for Query Name, Query, and other metrics. A pop-up window titled 'Oak assemblies' shows the results of a query, including a list of genes and their associated data.

IRGI Use list to store and/or improve complex queries

The screenshot shows the OakMine PMIN interface for managing query lists. It features a table of queries with columns for Query Name, Query, and other metrics. A pop-up window titled 'Oak assemblies' shows the results of a query, including a list of genes and their associated data. Annotations indicate that 134 proteins and 66 proteins are associated with the query results.



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: 11,200,000
 Length: 100 bp
 Reference: SscRef1

ASSOCIATED EFFECTS

Effect number: 1
 Effect name: BRENDEL
 Reference name: BRENDEL
 Reference value: 1.000000000
 Evidence: BRENDEL

ASSOCIATED ASSIGNMENTS

Assignment number: 1
 Map name: BRENDEL
 Linkage group: BRENDEL
 LOD: 100
 Test type: LOD
 Test value: 100
 R2: 1.00
 Distance: 0
 From: %
 Reliability: 100
 Link to map: BRENDEL

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	R2	Distance	From	%	Reliability (genetic map)	Link to map	Number of markers	Number of QTL
1	BRENDEL_2013_QTL_1	QTL	LOD	100	BRENDEL	BRENDEL	100	LOD	100	1.00	0	%	100	100	100	1	1
2	BRENDEL_2013_QTL_2	QTL	LOD	100	BRENDEL	BRENDEL	100	LOD	100	1.00	0	%	100	100	100	1	1

List of QTL overlapping this marker

ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM

QTL card

QTL name: BRENDEL_2013_QTL_1

QTL location: BRENDEL

Experimentation: BRENDEL

Test name: LOD

ASSOCIATED EFFECTS

Effect number: 1
 Effect name: BRENDEL
 Reference name: BRENDEL
 Reference value: 1.000000000
 Evidence: BRENDEL

ASSOCIATED ASSIGNMENTS

Assignment number: 1
 Map name: BRENDEL
 Linkage group: BRENDEL
 LOD: 100
 Test type: LOD
 Test value: 100
 R2: 1.00
 Distance: 0
 From: %
 Reliability: 100
 Link to map: BRENDEL

QTL results

RESULTS

#	QTL name	Theme	Test	Test statistics	Map	Linkage group	LOD	Test type	Test value	R2	Distance	From	%	Reliability (genetic map)	Link to map	Number of markers	Number of QTL
1	BRENDEL_2013_QTL_1	QTL	LOD	100	BRENDEL	BRENDEL	100	LOD	100	1.00	0	%	100	100	100	1	1
2	BRENDEL_2013_QTL_2	QTL	LOD	100	BRENDEL	BRENDEL	100	LOD	100	1.00	0	%	100	100	100	1	1

Link to the QTL card

21

22

ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM

Map card

MAP DETAILS

Map name: BRENDEL

Map type: genetic

Map date: 2013/12/13

Theme: QTL

Population: BRENDEL

Submitted by: IRTAD/INRA/FRANCE

Genotype file: BRENDEL

Image file: BRENDEL

Publication: Map details linkage mapping and distribution of quantitative trait loci in the pig genome

Linkage group number: 12

Last number: 5089

QTL number: 122

Projected QTL number: 122

ASSOCIATED EFFECTS

Effect number: 1
 Effect name: BRENDEL
 Reference name: BRENDEL
 Reference value: 1.000000000
 Evidence: BRENDEL

ASSOCIATED ASSIGNMENTS

Assignment number: 1
 Map name: BRENDEL
 Linkage group: BRENDEL
 LOD: 100
 Test type: LOD
 Test value: 100
 R2: 1.00
 Distance: 0
 From: %
 Reliability: 100
 Link to map: BRENDEL

Link to a genetic map card

List of markers overlapped by this QTL

23

ZURGI GnpIS GENETIC AND GENOMICS INFORMATION SYSTEM

Genetic map

Linkage group number: 12

Last number: 5089

QTL number: 122

Projected QTL number: 122

Open the Genetic map with BioMercator (java webstart)

From GnpIS QTL card to JBrowse

QTL card

QTL DETAILS
 QTL name: Bourain_2000_2002_QTL7.Delta.F
 QTL detection: Bourain_2000_2002_QTL7.Delta.F
 Measure: Delta_F_Bourain
 Experimentation: Bourain
 Trait name: Delta_F

ASSOCIATED EFFECTS
 Effects number: 0

ASSOCIATED ASSIGNMENTS
 Assignments number: 2

Map name	Linkage group	LOD	LOI	Test Year	R ²	Distance	From	To	Reliability (average) (%)	Link to map	Number of Rep/QTLs that use this QTL
Robur_composite_190702	LO9	lod	4.15	0.04	23.51	10.96	35.74	trialco	154	154	-
Quercus_composite_2018	LO9	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 PhenolMolecules	inc	Bourain_2000_2002_QTL7.Delta.F	-	BRENDEL

25

From JBrowse to the same QTL card

QTL card

QTL DETAILS
 QTL name: Bourain_2000_2002_QTL7.Delta.F
 QTL detection: Bourain_2000_2002_QTL7.Delta.F
 Measure: Delta_F_Bourain
 Experimentation: Bourain
 Trait name: Delta_F

ASSOCIATED EFFECTS
 Effects number: 0

ASSOCIATED ASSIGNMENTS
 Assignments number: 2

Map name	Linkage group	LOD	LOI	Test Year	R ²	Distance	From	To	Reliability (average) (%)	Link to map	Number of Rep/QTLs that use this QTL
Robur_composite_190702	LO9	lod	4.15	0.04	23.51	10.96	35.74	trialco	154	154	-
Quercus_composite_2018	LO9	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 PhenolMolecules	inc	Bourain_2000_2002_QTL7.Delta.F	-	BRENDEL

26

Acknowledgments

Team Members: J. Ansellem, H. Quesneville, R. Flores, C. Michotey, N. Francillonne, T. Letellier, JM Aury, C. Da Silva, K. Labadie, Faye S, C. Plomion, F. Ehrensmann, JC. Leplé, T. Leroy, C. Boudonès, G. Le Provost, A. Kremer, P. Falvère-ranipant, S. Duplessis, F. Martin, F. Murat, J. Salsie, C. Klopp, C. Gaspin, O. Rué

Funding: INRA, LabEx, ANR, Genoak, erc, Treepeace