



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.

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

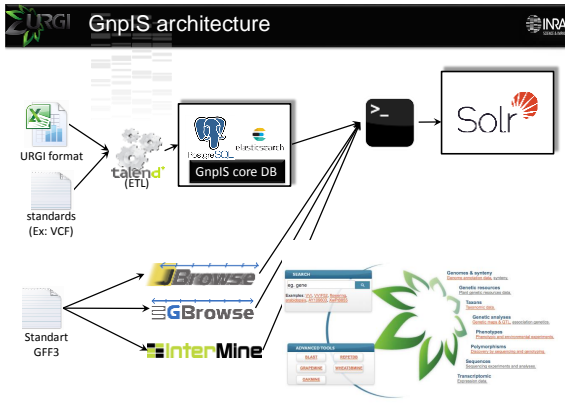
<https://urgi.versailles.inra.fr/gnpis>

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

SEARCH
 eg. gene
 Example: *UGL1*, *VUERC*, *Quercus*, *amblyopaea*, *AV10000*, *QNP255*

ADVANCED TOOLS
 BLAST, REPETITOR, GRAPHEME, WHATISMINE, OAKMINE

- Genomes & synteny: Genome association data, synteny
- Genetic resources: Plant genetic resources data
- Taxons: Taxonomic data
- Genetic analyses: Genetic maps & QTL, association genetics
- Phenotypes: Phenotypic and environmental experiments
- Polymerphisms: Discovery by sequencing and genotyping
- Sequences: Sequencing experiments and analysis
- Transcriptomic: Expression data



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/>

AB initio / Homolog methods, training required

Signal prediction, Content Analysis, Gene Finding

Localizations, Exon prediction, Repeat Prediction

Transcriptome, Proteome, Genomes T

Structural / synteny methods

Long Non Coding RNA

Alignment, Repeat Exclusion, Repeat Calling, Repeat Masking

RNA, Repeat, Repeat, Repeat

Non Coding RNA prediction and annotation

Genome sequence, RNA subtypes, Structural / Functional annotation resources

UniProt (1), UniProt (2), UniProt (3), UniProt (4)

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

Localization Targeting: TMHMM, SignalP, TargetP, Phobius

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

*script from David Goodman, personal communication

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:0006351' (mRNA processing) and 'GO:0006352' (mRNA splicing).
- KEGG:** Shows the gene's location on a metabolic pathway map.
- Pfam:** Shows protein domain annotations, such as 'PF00001' (RNA polymerase).
- panther:** Shows protein family classifications.
- 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:0006351' and 'GO:0006352'.
- Other genes sharing the same ontology term:** A list of other genes in the database that are annotated with the same GO term.

IRGI Interoperability with other features in the database

The screenshot displays the IRGI interface for QTL analysis. 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' 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 are located within 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 interface that allows users to use predefined query templates.
- Search for genes in genome regions:** A search interface that allows users to search for genes within specific genomic regions.
- Keyword based quick search:** A search interface that allows users to search for genes using keywords.
- Predicted protein_ID:** A search interface that allows users to search for genes using predicted protein IDs.
- A selection of Query templates:** A list of predefined query templates that can be used for searching.

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 columns for Protein Name, ID, and various metrics. A pop-up window titled 'Oak assemblies' shows a list of oak assemblies, and another pop-up window titled 'Oak coding sequences' shows 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 columns for Protein Name, ID, and various metrics. A pop-up window titled 'Oak assemblies' shows a list of oak assemblies, and another pop-up window titled 'Oak coding sequences' shows a list of oak coding sequences. Annotations are present on the list, such as '134 proteins' and '66 proteins'.



Interoperability between Genomics and Genetics



Marker: s_2F2VF6_2204

MARKER DETAILS

Name	s_2F2VF6_2204
Species	Quercus
Accession	Quercus_2013_0204
Chromosome	12
Start	11,995,000
End	11,995,000

ASSOCIATED EFFECTS

Effect name	Map name	Linkage group	LOD	Test type	RS	Distance	From	To	Reliability	Link	Number of markers
QTL	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	10.96	39.74	inf	100	1
QTL	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	11.32	39.72	projected	100	1

19

20

QTL results

QTL name	Theme	Test	Trait	Map	Linkage group	LOD	Test type	RS	Distance	From	To	Reliability	Link	Number of markers
Quercus_2013_0204	Quercus	LOD	Inf	12	LOD	inf	4.15	0.04	23.91	10.96	39.74	inf	100	1
Quercus_2013_0204	Quercus	LOD	Inf	12	LOD	inf	4.15	0.04	23.91	11.32	39.72	projected	100	1

QTL details

QTL name: Quercus_2013_0204_QTL_2204

QTL location: 12:11,995,000

QTL test: LOD

QTL trait: Inf

QTL map: Quercus_2013_0204

QTL linkage group: LOD

QTL test type: LOD

QTL RS: inf

QTL distance: 11.32

QTL from: 10.96

QTL to: 39.72

QTL reliability: inf

QTL link: 100

List of QTL overlapping this marker

QTL card

QTL DETAILS

QTL name: Quercus_2013_0204_QTL_2204

QTL location: 12:11,995,000

QTL test: LOD

QTL trait: Inf

QTL map: Quercus_2013_0204

QTL linkage group: LOD

QTL test type: LOD

QTL RS: inf

QTL distance: 11.32

QTL from: 10.96

QTL to: 39.72

QTL reliability: inf

QTL link: 100

QTL number of markers: 1

ASSOCIATED EFFECTS

Effect name	Map name	Linkage group	LOD	Test type	RS	Distance	From	To	Reliability	Link	Number of markers
QTL	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	10.96	39.74	inf	100	1
QTL	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	11.32	39.72	projected	100	1

ASSOCIATED ASSIGNMENTS

Assignment number	Map name	Linkage group	LOD	Test type	RS	Distance	From	To	Reliability	Link	Number of markers
1	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	10.96	39.74	inf	100	1
2	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	11.32	39.72	projected	100	1

Link to the QTL card

Map card

MAP DETAILS

Map name: Quercus_2013_0204

Map type: genetic

Map date: 2013/12/19

Theme: Quercus

Population: Quercus_2013_0204

Submitted by: IRTV-UMR1066-Genetics

Genotype file: IRTV-UMR1066-Genetics

Image file: IRTV-UMR1066-Genetics

Publication: High-density linkage mapping and distribution of recombination hotspots in the oak genome

Linkage group number: 12

Last number: 5089

QTL number: 122

Projected QTL number: 122

ASSOCIATED EFFECTS

Effect name	Map name	Linkage group	LOD	Test type	RS	Distance	From	To	Reliability	Link	Number of markers
QTL	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	10.96	39.74	inf	100	1
QTL	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	11.32	39.72	projected	100	1

ASSOCIATED ASSIGNMENTS

Assignment number	Map name	Linkage group	LOD	Test type	RS	Distance	From	To	Reliability	Link	Number of markers
1	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	10.96	39.74	inf	100	1
2	Quercus_2013_0204	LOD	inf	4.15	0.04	23.91	11.32	39.72	projected	100	1

List of markers overlapped by this QTL

Genetic map

Map name: Quercus_2013_0204

Map type: genetic

Map date: 2013/12/19

Theme: Quercus

Population: Quercus_2013_0204

Submitted by: IRTV-UMR1066-Genetics

Genotype file: IRTV-UMR1066-Genetics

Image file: IRTV-UMR1066-Genetics

Publication: High-density linkage mapping and distribution of recombination hotspots in the oak genome

Linkage group number: 12

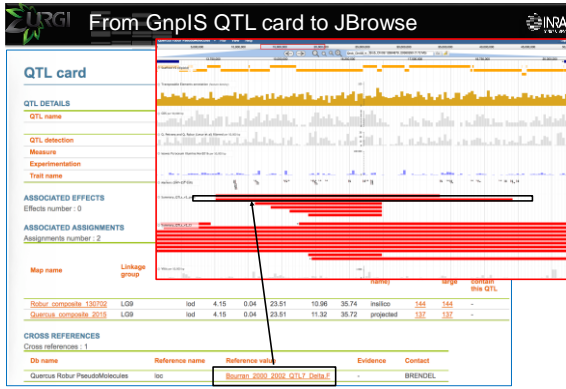
Last number: 5089

QTL number: 122

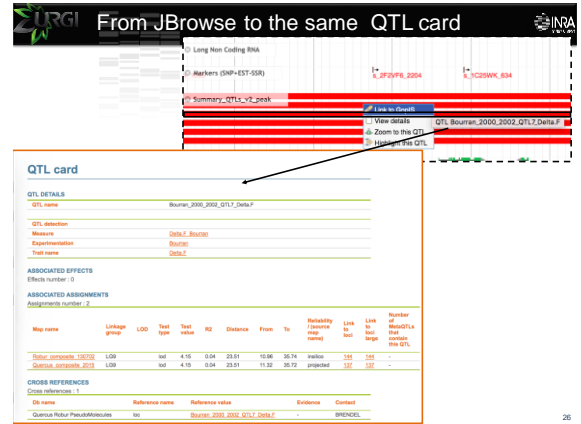
Projected QTL number: 122

Open the Genetic map with BioMercator (java webstart)

23



25



26

Acknowledgments

Researchers: J. Ansellem, H. Quesneville, N. Francillonne, T. Letellier, R. Flores, C. Michotey, F. Ehrensmann, J.C. Leprie, T. Leroy, C. Boudonès, G. Le Provost, A. Kremer, C. Plomion, P. Falvère-ranipant, S. Duplessis, F. Martin, F. Murat, J. Salse, C. Da Silva, K. Labadie, Faye S, JM Aury, C. Kloug, C. Gaignon, O. Rué.

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