An Integrated Information System Dedicated to Oak Genomics and Genetics

Digital Tools and Resources

Session 1
Saturday, January 13, 2018

GnpIS: A Genetic and genomic Information System

GnpIS architecture

GnpIS core DB

URGI format standards (Ex: VCF)

GBrowse

InterMine

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/

Functional annotation pipeline

Blast similarities

KOG

rpsBLAST

Conserved domains

Protein domain identification

InterProScan

Protein definition* based on results of

EC Number

Orthology

Gene ontology

KEGG

*script from David Goodman, personal communication
Oak resources

Markers, QTLs
Genetic maps
Phenotypes (in progress)

Navigating oak genomic and genetic data

Pseudomolecules … to JBrowse

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

Jbrowse overview and cross references

OakMine portal

OakMine_PM1N central object: the gene card
functional annotation: cross references to external DBs

- Gene Ontology
- Pfam
- KEGG
- Panther
- Interpro

Gene ontology parent terms

Other genes sharing the same ontology term

Interoperability with other features in the database

- QTL infos
- Genes overlapped by this QTL

Several ways to query data

- Graph keyword-based quick search
- Advanced search using Query the builder
- Search for genes in genome regions
- Most search

Use list to store and/or improve complex queries

- Oak assemblies
- Oak coding sequences
- 134 proteins
- 66 proteins
- Save results in list
- Domain Description contains "xylanase"
- Most Description contains "xylanase"
Interoperability between Genomics and Genetics

From Gbrowse to Markers in GnpIS

- List of QTL overlapping this marker
- Link to the QTL card
- List of markers overlapped by this QTL
- Link to a genetic map card

Open the Genetic map with BioMercator (java webstart)
From GnpIS QTL card to JBrowse

Acknowledgments

C. Plomion
C. Klopp
J. Salse
F. Martin
J.M. Aury
C. Gaspin
P. Faivre-
rampant
LabEx
Genoak
Treepeace
S. Duplessis
F. Ehrenmann
J.C. Leplé
T. Leroy
I. Lesur
C. Bodénès
G. Le Provost
A. Kremer
F. Murat
C. Da Silva
K. Labadie
Faye S
O.Rué

Funding

H. Quesneville
N. Francillonne
C. Michotey
T. Letellier
J. Amselem
R. Flores