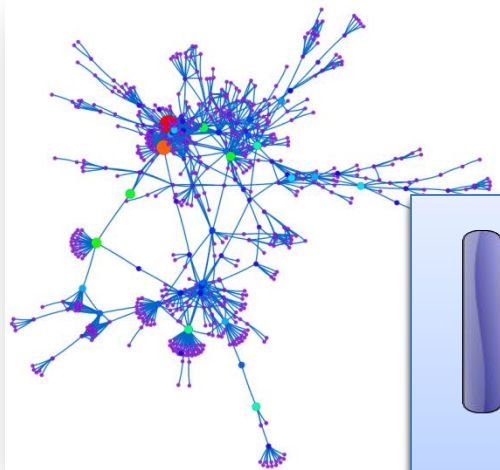# Functional Genome Annotation

Creating Genomic Networks with Ondex
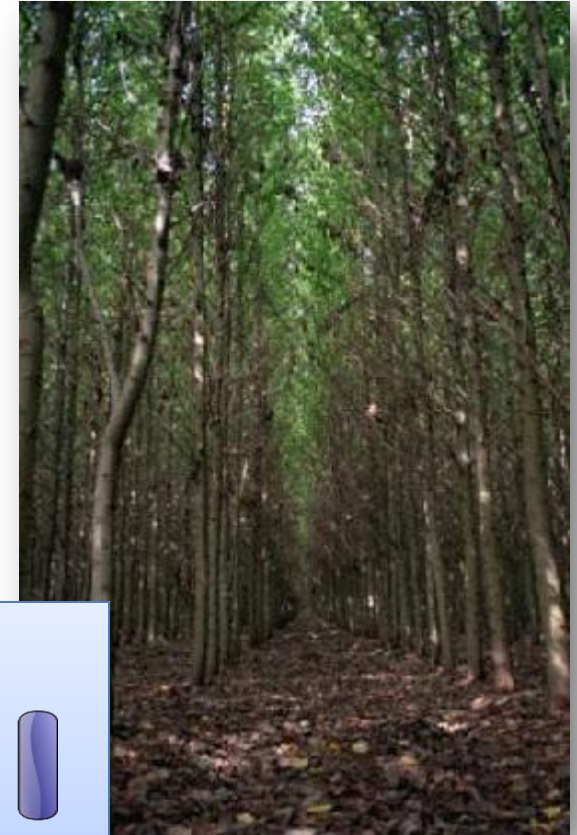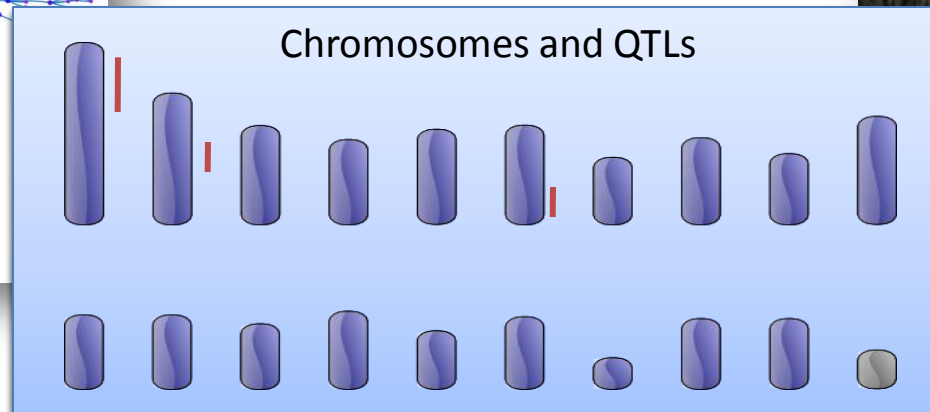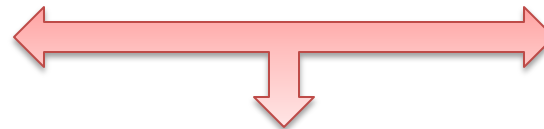
Keywan Hassani-Pak

Ondex
http://ondex.org

Populus trichocarpa (Poplar)

http://ondex.org

- First tree with fully sequenced genome
- Model plant for bioenergy crops such as Willow
- 4x larger genome than Arabidopsis
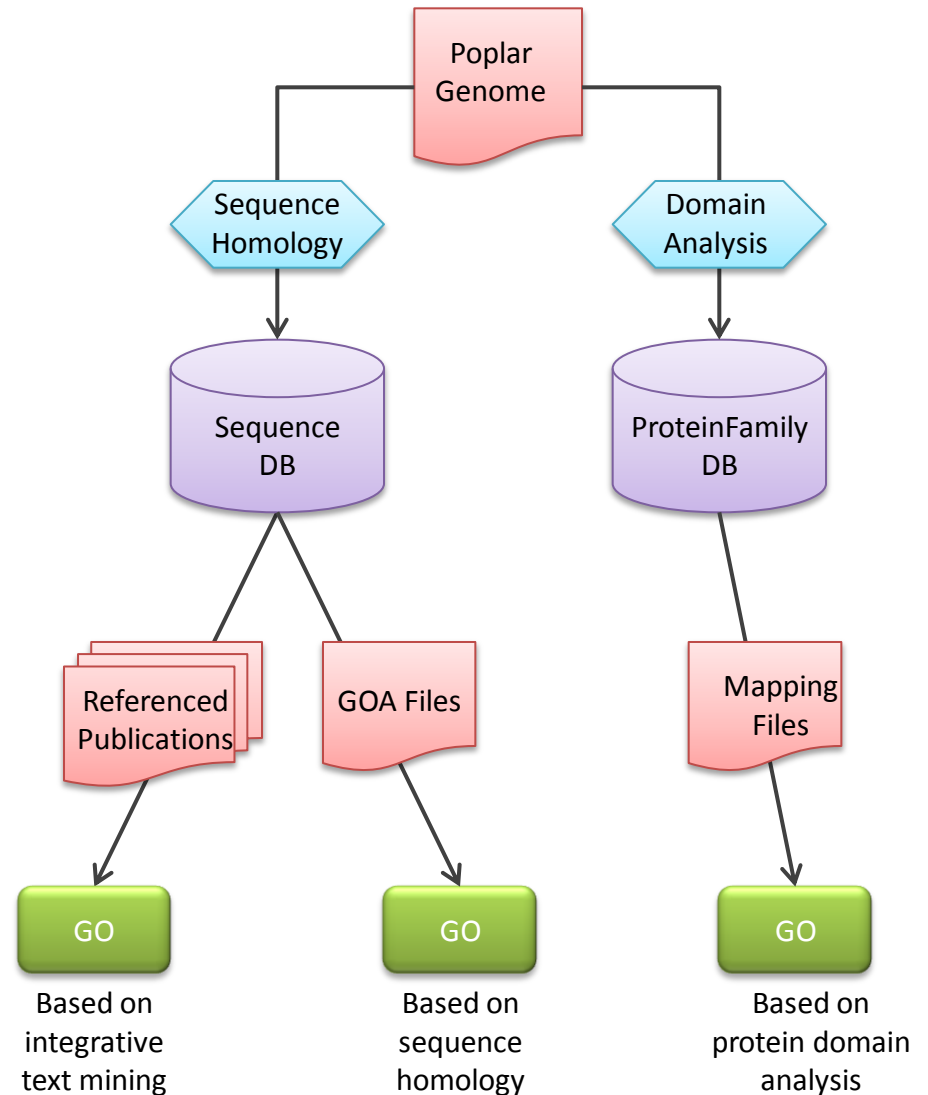- 19 Chromosomes
- 45,555 predicted genes

Underlying Networks

Chromosomes and QTLs

Phenotypes

Re-Annotation of the Poplar genome

Ondex
http://ondex.org

- Data Integration
  - UniProtKB, Pfam
  - GO, GOA, EC
  - MEDLINE

- Method Integration
  - Sequence Similarity
  - Domain Analysis
  - Text Mining

- Automatic function prediction with different levels of evidence

Poplar Genome

Sequence Homology → Sequence DB

Domain Analysis → ProteinFamily DB

Referenced Publications

GOA Files

Mapping Files

GO — Based on integrative text mining

GO — Based on sequence homology

GO — Based on protein domain analysis

BBSRC
bioscience for the future

Integrating Poplar data with…

Ondex
http://ondex.org

UniProt

Magnoliophyta   9,915
    Core eudicotyledons   7,873
        Arabidopsis thaliana (Mouse-ear cress)   7,753
        Vitis vinifera (Grape)   120
    Oryza sativa subsp. japonica (Rice)   2,042
Physcomitrella patens (Moss)   92

Pfam

GO
the Gene Ontology

Pfam 23.0 (July 2008, 10340 families)   26976 terms, 98.7% with definitions
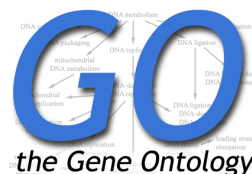
MEDLINE 1010001001001 10
0101010100101 01
1010010001010 10
U.S. National Library of Medicine

MEDLINE 2009 contains 17,764,826 records

BBSRC
bioscience for the future

Integrated Poplar Network
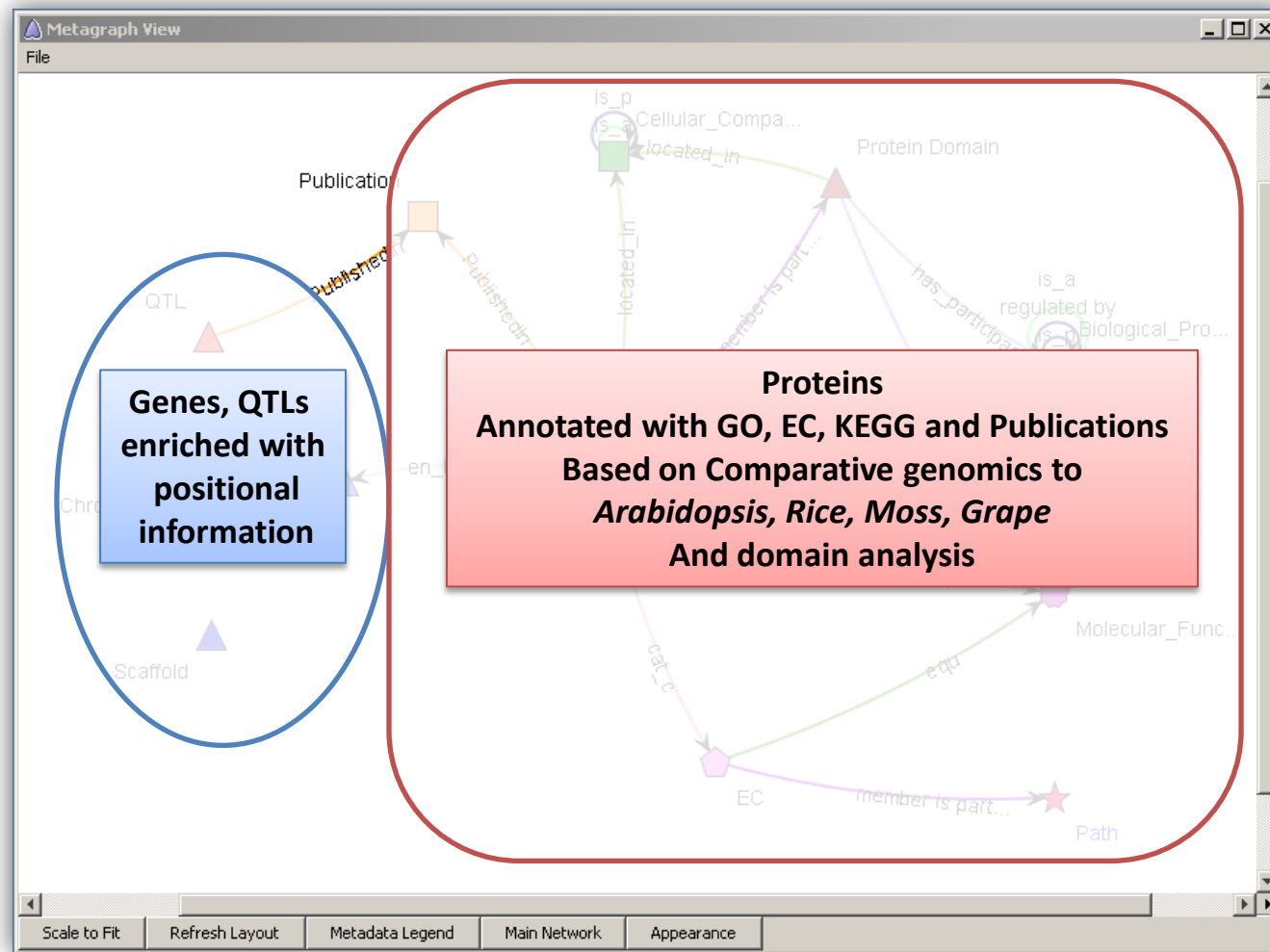
Data integration was done in Ondex

Context Filter

Poplar proteins involved in Tryptophan metabolism

Example1

Genomic View in OVTK

Ondex
http://ondex.org

Genomic Layout displays chromosomes, genes and QTLs

Chromosomal regions and QTLs can be selected

Example 2 and 3

BBSRC
bioscience for the future

Network View in OVTK

Genes of interest
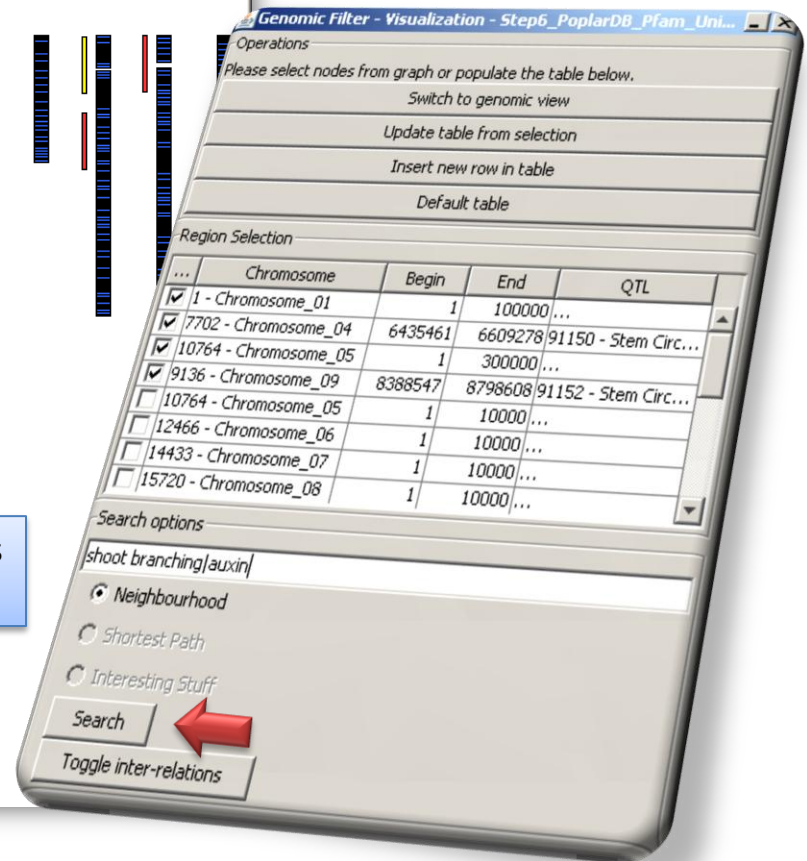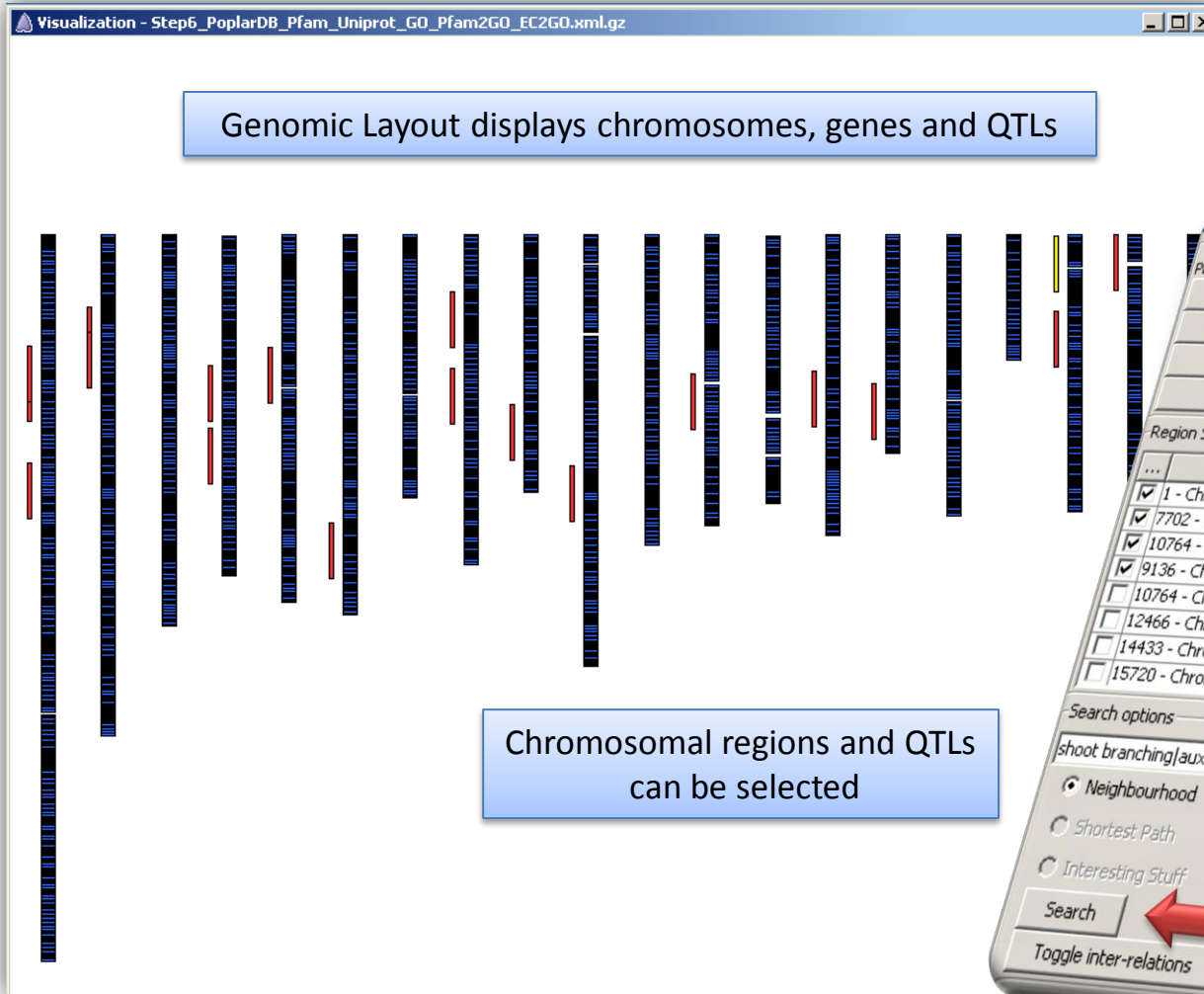
Enriched protein annotation network
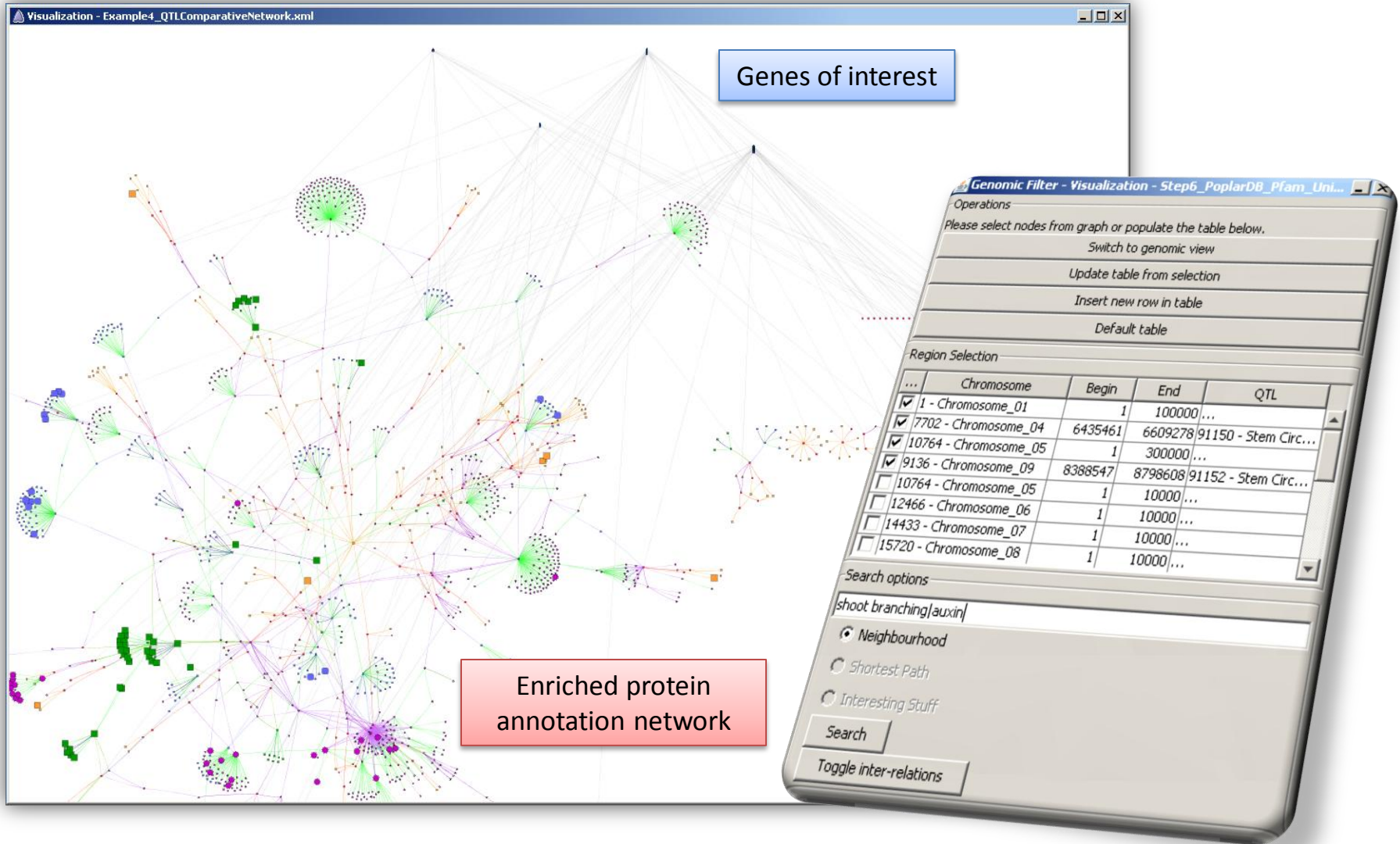
Exploring protein neighbourhoods

Blast 650581 – LAX
E-Value: 3.4E-17
Score: 83.57
Coverage: 27%

Blast 650581 – EN108
E-Value: 9.4E-7
Score: 48.91
Coverage: 20%

PMID:13130077
"LAX and SPA: major regulators of shoot branching in rice."

HMMer: 650581 – HLH
E-Value: 3.4E-7
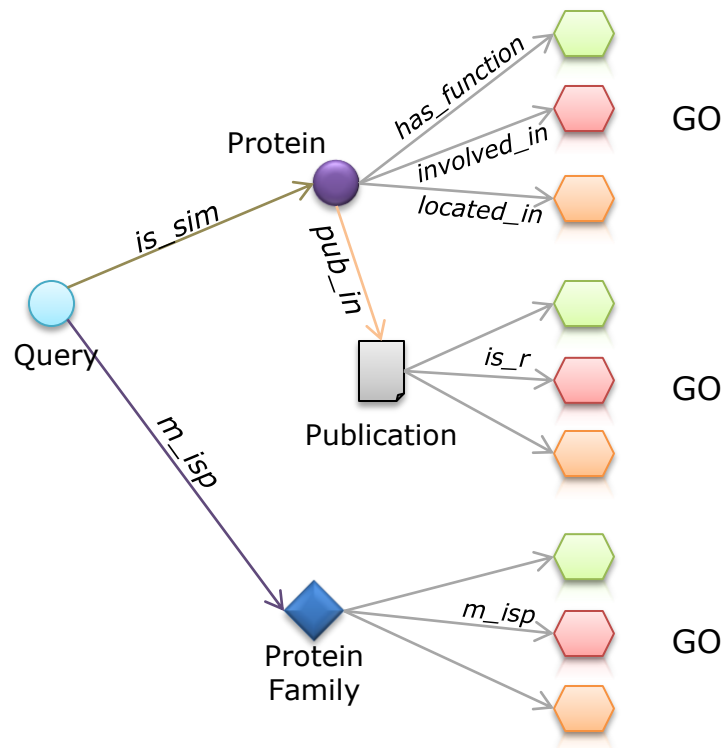Score: 30.0
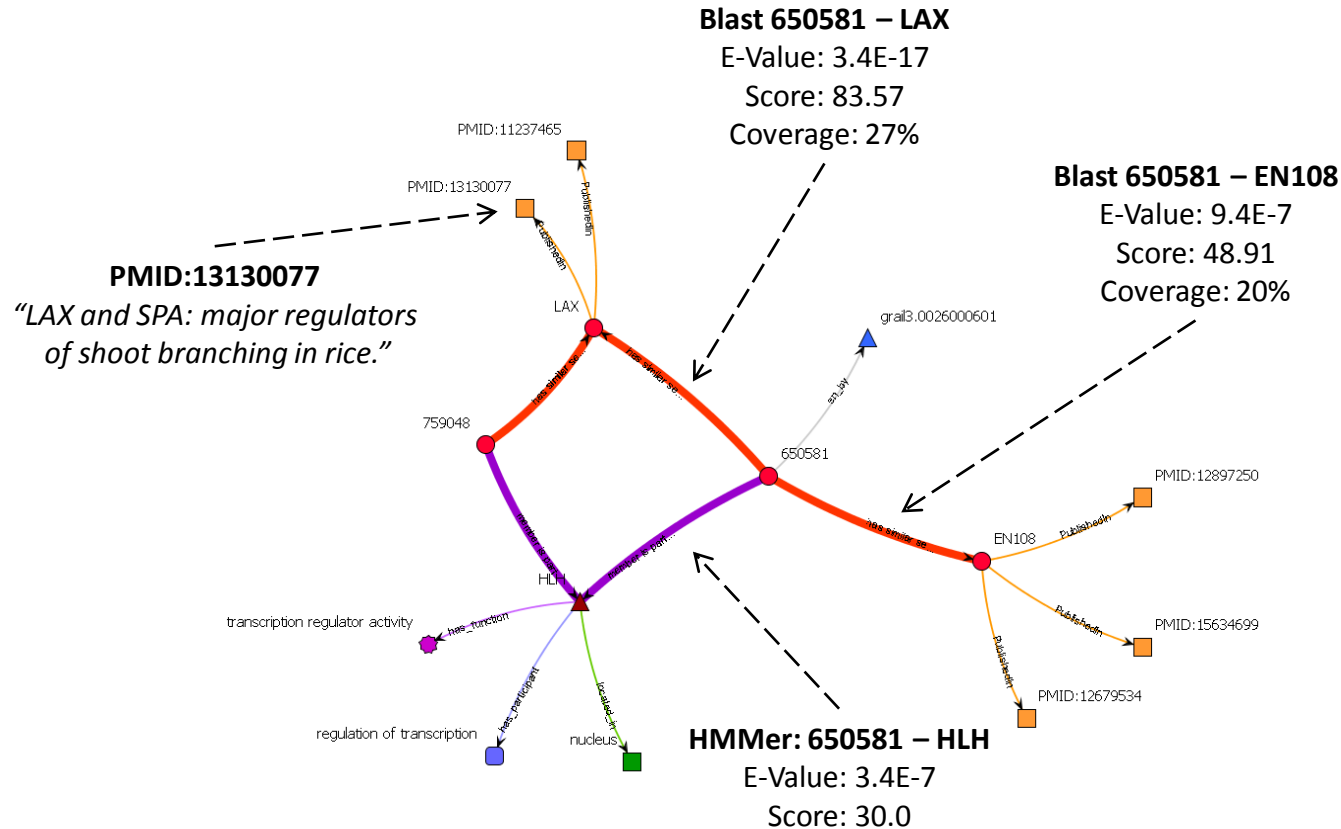
Poplar protein **JGI:650581** has no annotations from JGI
We identified two remote homologs in Rice (**LAX**) and in Arabidopsis (**EN108**),
as well as one protein domain **HLH**
The **LAX** homolog contains evidence to be a major regulator of **shoot branching**
→ **Hypothesis generation**

Mining GO annotations with different evidences

- Assigning functional terms to query sequences from the pool of GO terms gathered in the mapping step

- Take into consideration:
  - Sequence similarity
  - The GO evidence
  - Structure of GO DAG

- Evaluation of different annotation methods