Knowtator, a framework to annotate phenotype-genotype relationships relevant to Arabidopsis leaf growth and development

Pierre Hilson

pierre.hilson@versailles.inra.fr

AGRON OMICS

SIXTH FRAMEWORK PROGRAMME

INRA

ijpb
Institut Jean-Pierre Bourgin

AGRON-OMICS

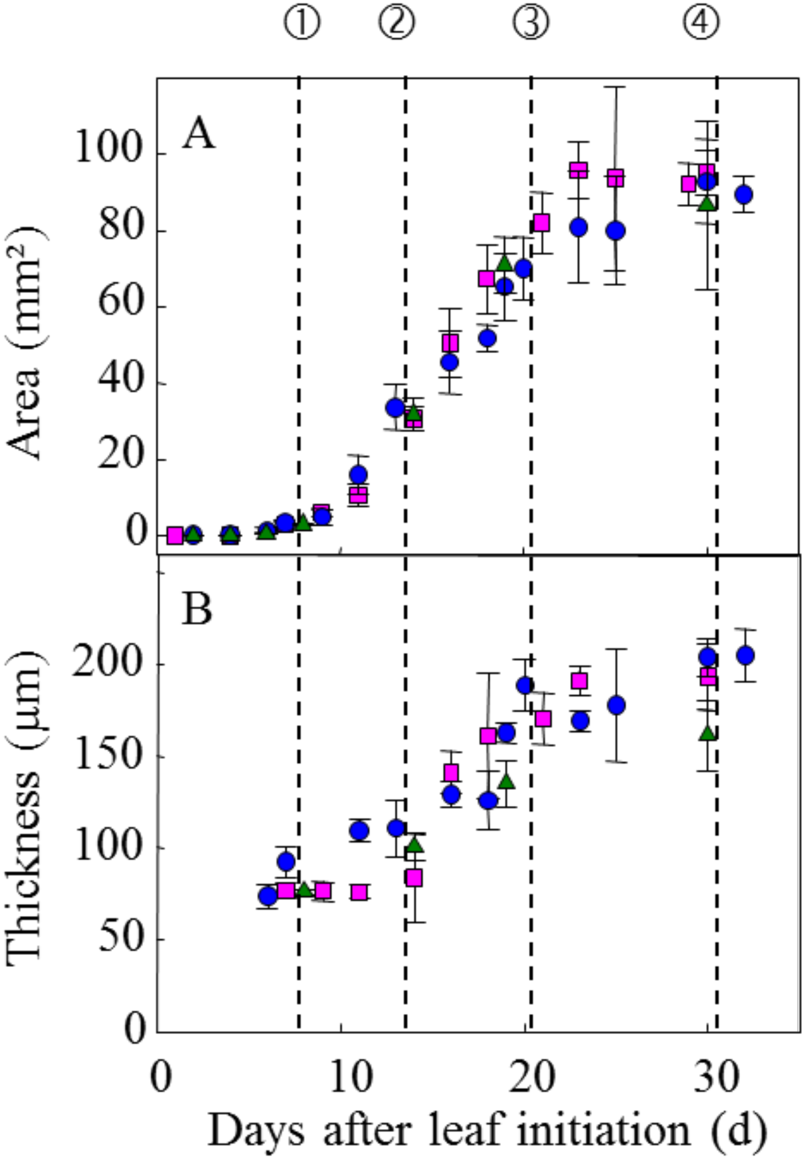
= **A**rabidopsis **GRO**wth **N**etwork integrating **OMICS technologies**

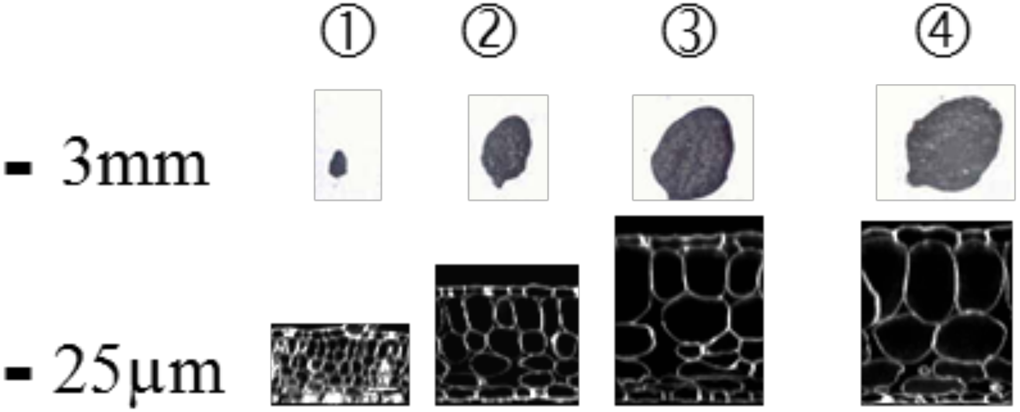
Focus: **leaf growth**

2. Development of tools for the research community

- **AGRONOMICS1 tiling array** covering both genome strands
- **pep2pro** for comprehensive proteome data analysis, **MASCP Gator** for integration and visualization of Arabidopsis proteomics data, ORFeome resources and cloning resources
- Protein-protein interaction networks- **Arabidopsis Interactome Mapping**
- Improved methods for protein localization studies
- Enzymatic and metabolic networks involved in biomass production
- Plant structure visualization by high-resolution X-ray computed tomography
- Data mining and integration, **CORNET , Arabidopsis Reactome**

http://www.agron-omics.eu/





Stage characterised by
(Boyes identifiers in SOW)

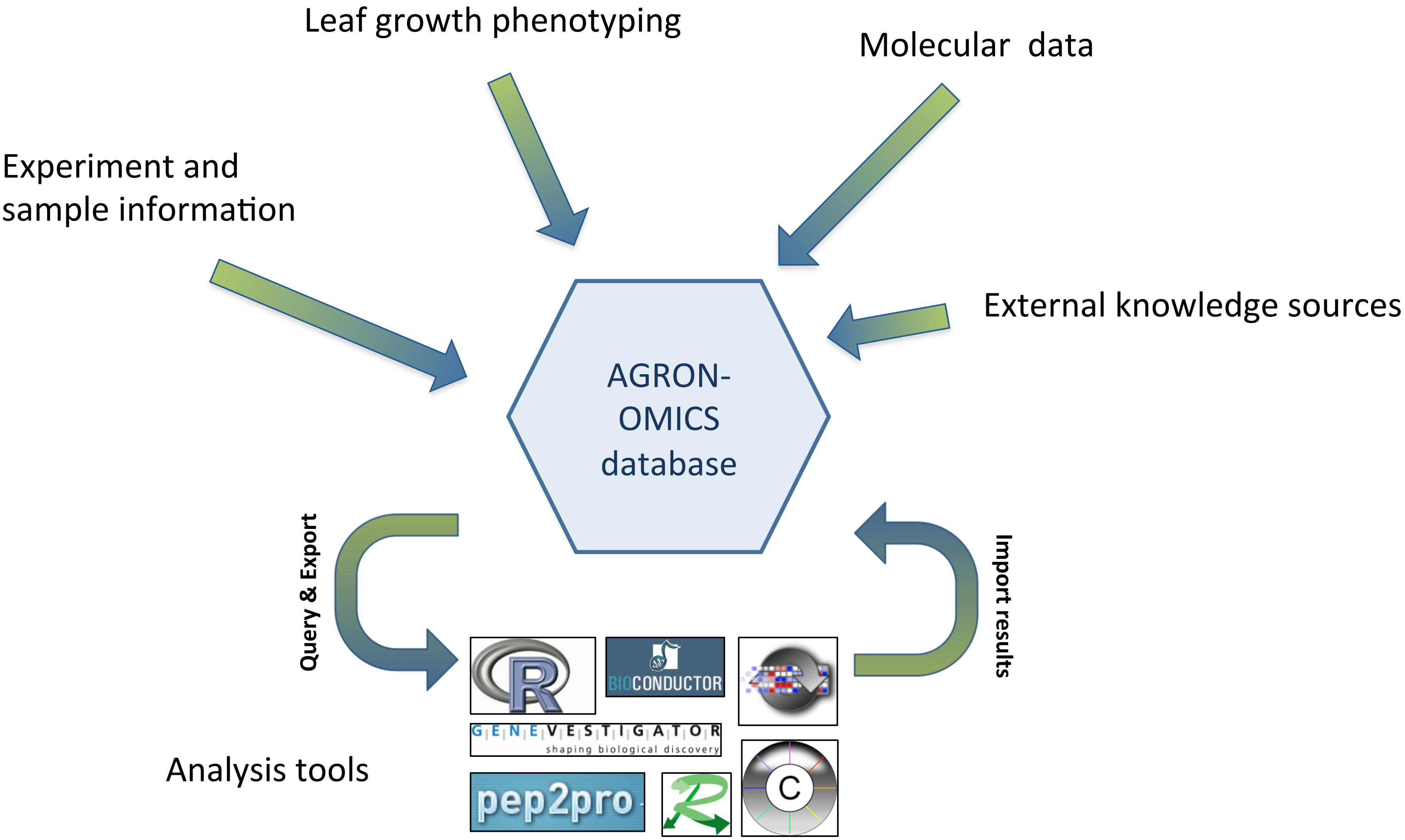
Stage 1
active cell division (1.07)

Stage 2
rapid cell expansion (1.11)

Stage 3
decreasing cell expansion (1.18)

Stage 4
mature leaf, no expansion and well before senescence (> 1.28)

The AGRON-OMICS database



Leaf growth phenotyping

Molecular data

Experiment and
sample information

External knowledge sources

AGRON-
OMICS
database

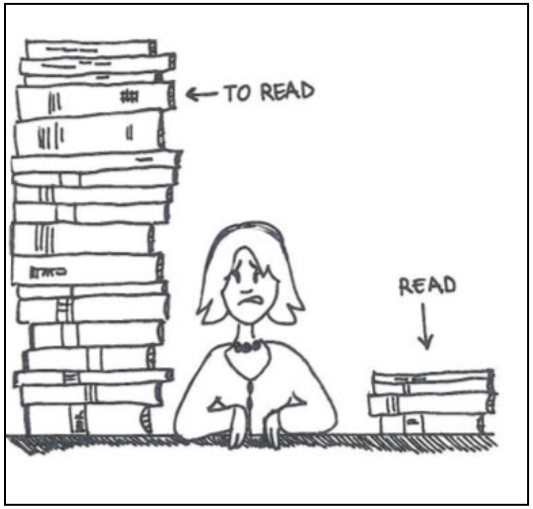
Query & Export

Import results

Analysis tools

Sean Walsh, ETH Zurich

Need to access data in primary literature



There are no good comprehensive literature databases for plant research.

AIM: To capture and mine knowledge from literature

PREMISES:

- Papers about leaf growth and development
- Information collected with biomedical ontologies
- Structured statements that can be simplified and imported into relational databases

TASKS:

- Develop a method (test library, 174 papers)
- Annotate papers with volunteers (adding 109 papers)
- Merge with public molecular resources
- Deploy query and visualization tools

Further rules and restrictions

- Only Arabidopsis
- Only leaf data (also partly cotyledon, meristem, embryo)
- Restricted to the *Results* section
- Exclude reviews
- Information structure, one-to-one relations

What sort of information to record?

Relation	Example
Phenotype	The rot3-2 allele causes enlarged leaf blades
Gene expression	ANT mRNA accumulated in leaf
Feature	AtCPL2 contains one dsRNA-binding domain
DNA-protein interaction	ARF2 ... bound to the promoter region of GH3.1
Genetic interaction	hyl1 ... appeared to suppress the as2 phenotypes
Protein-protein interaction	AN3 interacted strongly with ... AtGRF9
Process	RHL2 ... involved during endocycles
Regulation of gene expression	AtCPL1 ... negative regulators of RD29A expression
Regulation of process	AN3 ... promoting ... cell proliferation
Regulation of phenotype	PHABULOSA ... influence leaf shape

Structured statement

Phenotype	The reduced leaf area in the *hub1-1* mutant was confirmed by morphological measurements of the fully expanded leaves 1 and 2

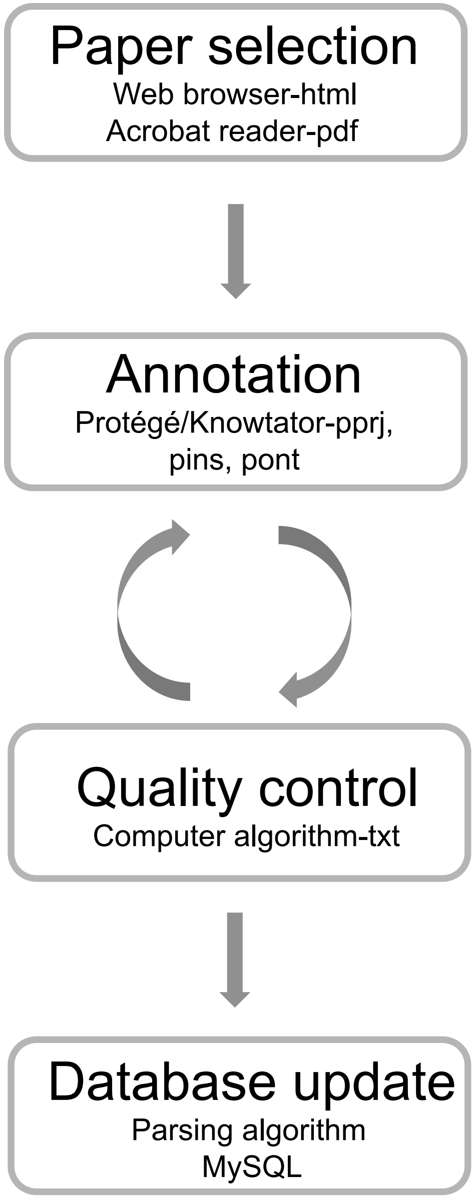
Slot	Original text	Ontology
Developmental stage	fully expanded leaves	3 leaf fully expanded_PO:0001053
Factuality		
Genotype	hub1-1	mutated gene_MI:0804 RDO4 HUB1_AT2G44950 loss of function_APO:0000011 homozygous diploid _APO:0000229
Growth condition		
Localisation		
Methodology		
Plant part	leaf	leaf_PO:0025034
Process		
Property	area	area_PATO:0001323
Value	reduced	decreased area_PATO:0002058

Biomedical ontologies

Defined terms
Avoiding redundancy and confusion
Established parent-child relationships
Community endorsed

Ontology	Acronym	URL	Reference
BRENDA tissue / enzyme source	BTO	http://www.brenda-enzymes.info	Gremse *et al.* 2011
Gene Ontology	GO	http://www.geneontology.org/	Ashburner *et al.* 2000
Molecular Interaction	MI	http://psidev.sf.net http://psidev.sourceforge.net/molecular_interactions/xml/doc/user/index.html	Hermjakob *et al.* 2004
Phenotype, Attribute and Trait Ontology	PATO	http://obofoundry.org/wiki/index.php/PATO:Main_Page	
Plant environmental conditions	EO	http://www.gramene.org/plant_ontology/ontology_browse.html#eo	Liang *et al.* 2008
Plant Ontology	PO	http://www.plantontology.org/	Jaiswal *et al.* 2005
The Arabidopsis Information Resource	TAIR	http://arabidopsis.org/	Lamesch *et al.* 2012

Annotation flow chart



Paper selection

Web browser-html

Acrobat reader-pdf

Annotation

Protégé/Knowtator-pprj,

pins, pont

Quality control

Computer algorithm-txt

Database update

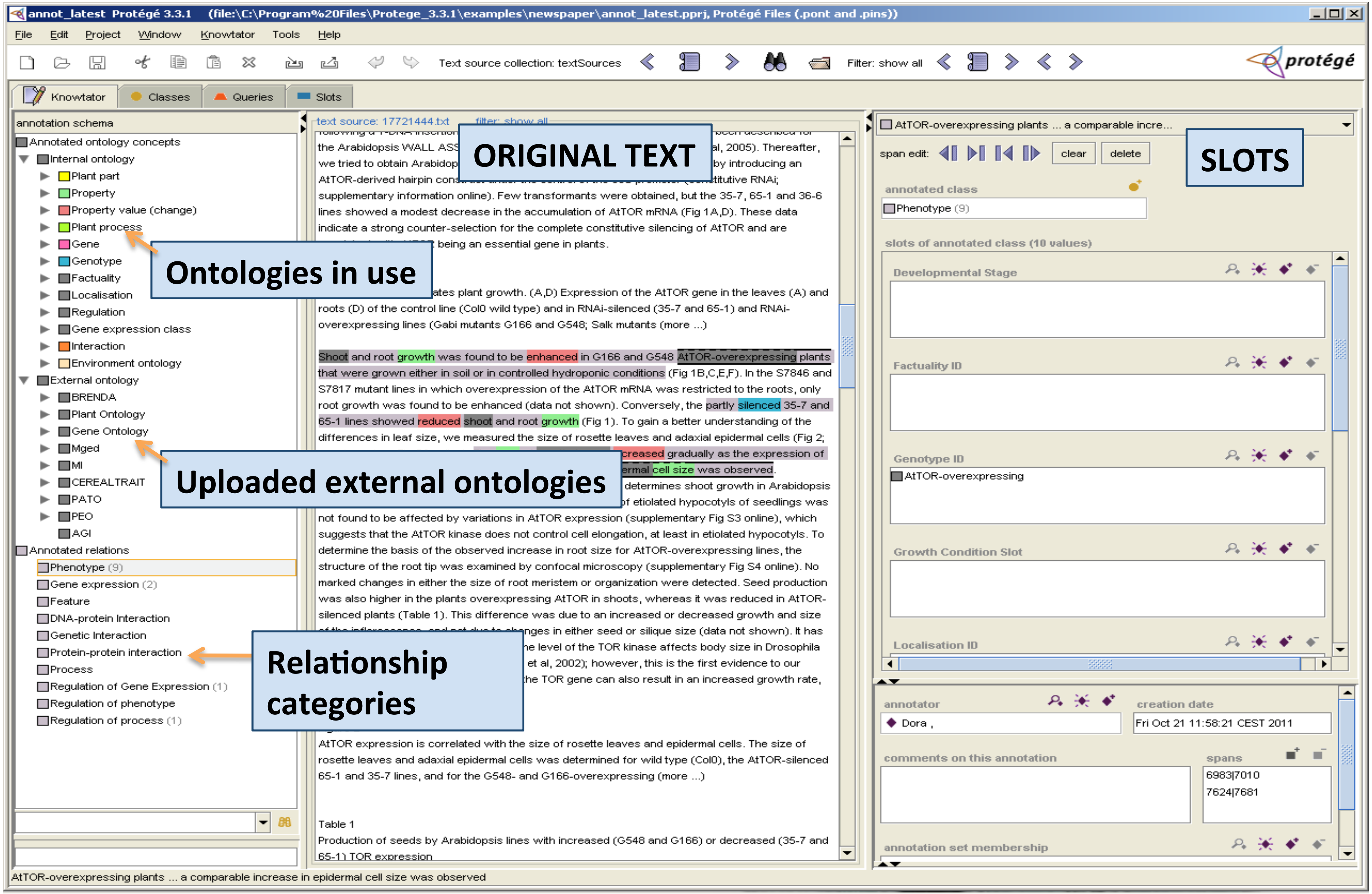
Parsing algorithm

MySQL

Knowtator, a custom annotation interface

- “general-purpose text annotation tool”
- is a plug-in to Protégé, free, open-source platform to construct domain models and knowledge-based applications with ontologies
- flexible, can be adapted to the project
- easy to import and handle existing ontologies
- export to xml format
- small project files
- easy to share
- mistakes can be corrected
- relatively user friendly

Knowtator interface



the Arabidopsis WALL ASS … al, 2005). Thereafter, we tried to obtain Arabidop … by introducing an AtTOR-derived hairpin cons … stitutive RNAi; supplementary information online). Few transformants were obtained, but the 35-7, 65-1 and 36-6 lines showed a modest decrease in the accumulation of AtTOR mRNA (Fig 1A,D). These data indicate a strong counter-selection for the complete constitutive silencing of AtTOR and are … being an essential gene in plants.

… ates plant growth. (A,D) Expression of the AtTOR gene in the leaves (A) and roots (D) of the control line (Col0 wild type) and in RNAi-silenced (35-7 and 65-1) and RNAi-overexpressing lines (Gabi mutants G166 and G548; Salk mutants (more ...)

Shoot and root growth was found to be enhanced in G166 and G548 AtTOR-overexpressing plants that were grown either in soil or in controlled hydroponic conditions (Fig 1B,C,E,F). In the S7846 and S7817 mutant lines in which overexpression of the AtTOR mRNA was restricted to the roots, only root growth was found to be enhanced (data not shown). Conversely, the partly silenced 35-7 and 65-1 lines showed reduced shoot and root growth (Fig 1). To gain a better understanding of the differences in leaf size, we measured the size of rosette leaves and adaxial epidermal cells (Fig 2; … creased gradually as the expression of … ermal cell size was observed. … determines shoot growth in Arabidopsis … of etiolated hypocotyls of seedlings was not found to be affected by variations in AtTOR expression (supplementary Fig S3 online), which suggests that the AtTOR kinase does not control cell elongation, at least in etiolated hypocotyls. To determine the basis of the observed increase in root size for AtTOR-overexpressing lines, the structure of the root tip was examined by confocal microscopy (supplementary Fig S4 online). No marked changes in either the size of root meristem or organization were detected. Seed production was also higher in the plants overexpressing AtTOR in shoots, whereas it was reduced in AtTOR-silenced plants (Table 1). This difference was due to an increased or decreased growth and size … nges in either seed or silique size (data not shown). It has … he level of the TOR kinase affects body size in Drosophila … et al, 2002); however, this is the first evidence to our … the TOR gene can also result in an increased growth rate,

AtTOR expression is correlated with the size of rosette leaves and epidermal cells. The size of rosette leaves and adaxial epidermal cells was determined for wild type (Col0), the AtTOR-silenced 65-1 and 35-7 lines, and for the G548- and G166-overexpressing (more ...)

Table 1
Production of seeds by Arabidopsis lines with increased (G548 and G166) or decreased (35-7 and 65-1) TOR expression

AtTOR-overexpressing plants ... a comparable incre...

annotated class: Phenotype (9)

Genotype ID: AtTOR-overexpressing

annotator: Dora ,

creation date: Fri Oct 21 11:58:21 CEST 2011

spans:
6983|7010
7624|7681

AtTOR-overexpressing plants ... a comparable increase in epidermal cell size was observed

ORIGINAL TEXT

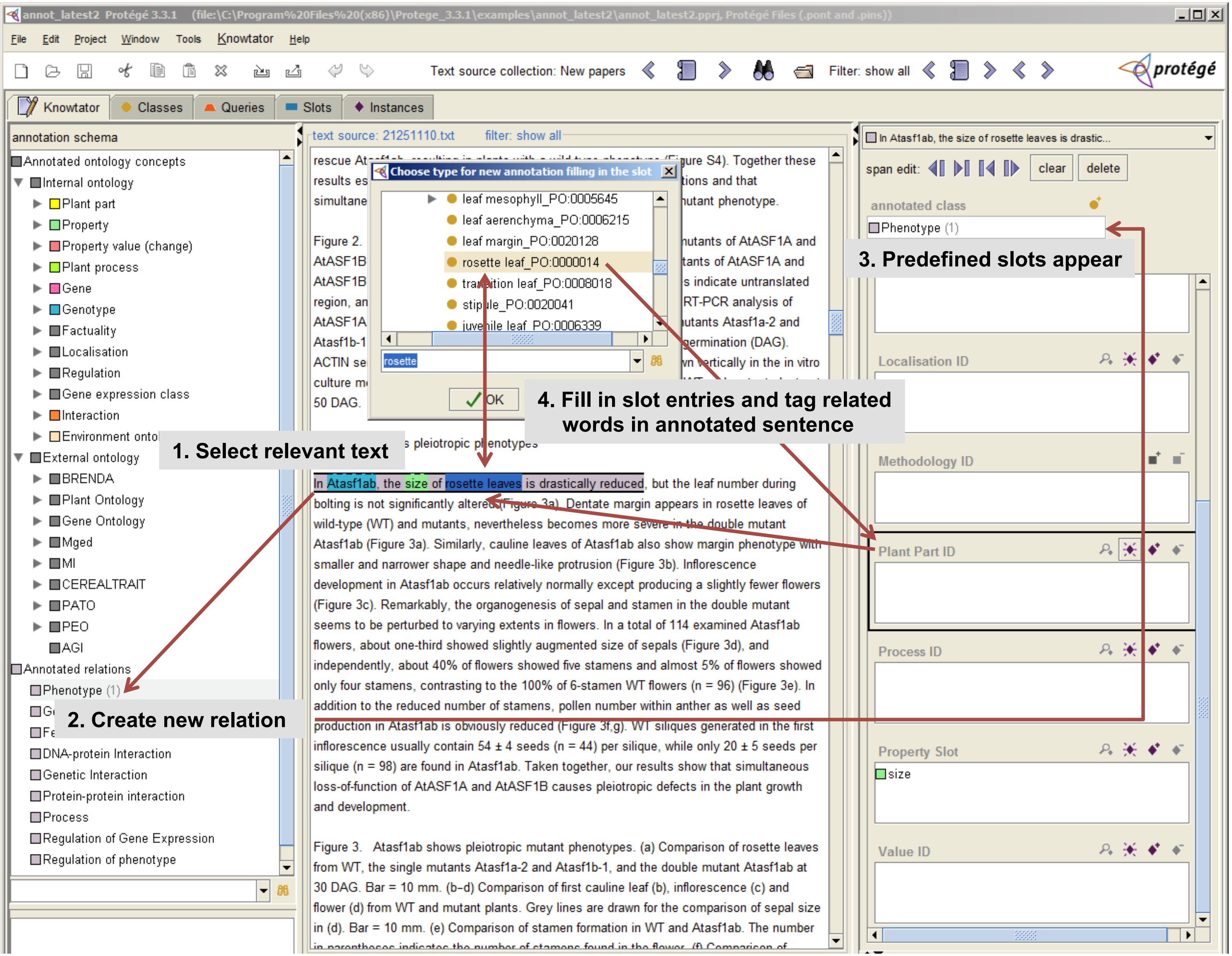
SLOTS

Ontologies in use

Uploaded external ontologies

Relationship categories

Text tagging and annotation



rescue At... (Figure S4). Together these
results es...tions and that
simultane...nutant phenotype.

Figure 2. ...nutants of AtASF1A and
AtASF1B ...tants of AtASF1A and
AtASF1B ...s indicate untranslated
region, an... RT-PCR analysis of
AtASF1A ...nutants Atasf1a-2 and
Atasf1b-1 ...germination (DAG).
ACTIN se...wn vertically in the in vitro
culture m...
50 DAG.

...s pleiotropic phenotypes

In Atasf1ab, the size of rosette leaves is drastically reduced, but the leaf number during bolting is not significantly altered (Figure 3a). Dentate margin appears in rosette leaves of wild-type (WT) and mutants, nevertheless becomes more severe in the double mutant Atasf1ab (Figure 3a). Similarly, cauline leaves of Atasf1ab also show margin phenotype with smaller and narrower shape and needle-like protrusion (Figure 3b). Inflorescence development in Atasf1ab occurs relatively normally except producing a slightly fewer flowers (Figure 3c). Remarkably, the organogenesis of sepal and stamen in the double mutant seems to be perturbed to varying extents in flowers. In a total of 114 examined Atasf1ab flowers, about one-third showed slightly augmented size of sepals (Figure 3d), and independently, about 40% of flowers showed five stamens and almost 5% of flowers showed only four stamens, contrasting to the 100% of 6-stamen WT flowers (n = 96) (Figure 3e). In addition to the reduced number of stamens, pollen number within anther as well as seed production in Atasf1ab is obviously reduced (Figure 3f,g). WT siliques generated in the first inflorescence usually contain 54 ± 4 seeds (n = 44) per silique, while only 20 ± 5 seeds per silique (n = 98) are found in Atasf1ab. Taken together, our results show that simultaneous loss-of-function of AtASF1A and AtASF1B causes pleiotropic defects in the plant growth and development.

Figure 3. Atasf1ab shows pleiotropic mutant phenotypes. (a) Comparison of rosette leaves from WT, the single mutants Atasf1a-2 and Atasf1b-1, and the double mutant Atasf1ab at 30 DAG. Bar = 10 mm. (b–d) Comparison of first cauline leaf (b), inflorescence (c) and flower (d) from WT and mutant plants. Grey lines are drawn for the comparison of sepal size in (d). Bar = 10 mm. (e) Comparison of stamen formation in WT and Atasf1ab. The number in parentheses indicates the number of stamens found in the flower. (f) Comparison of

1. Select relevant text

2. Create new relation

3. Predefined slots appear

4. Fill in slot entries and tag related words in annotated sentence

Monitoring relation consistency

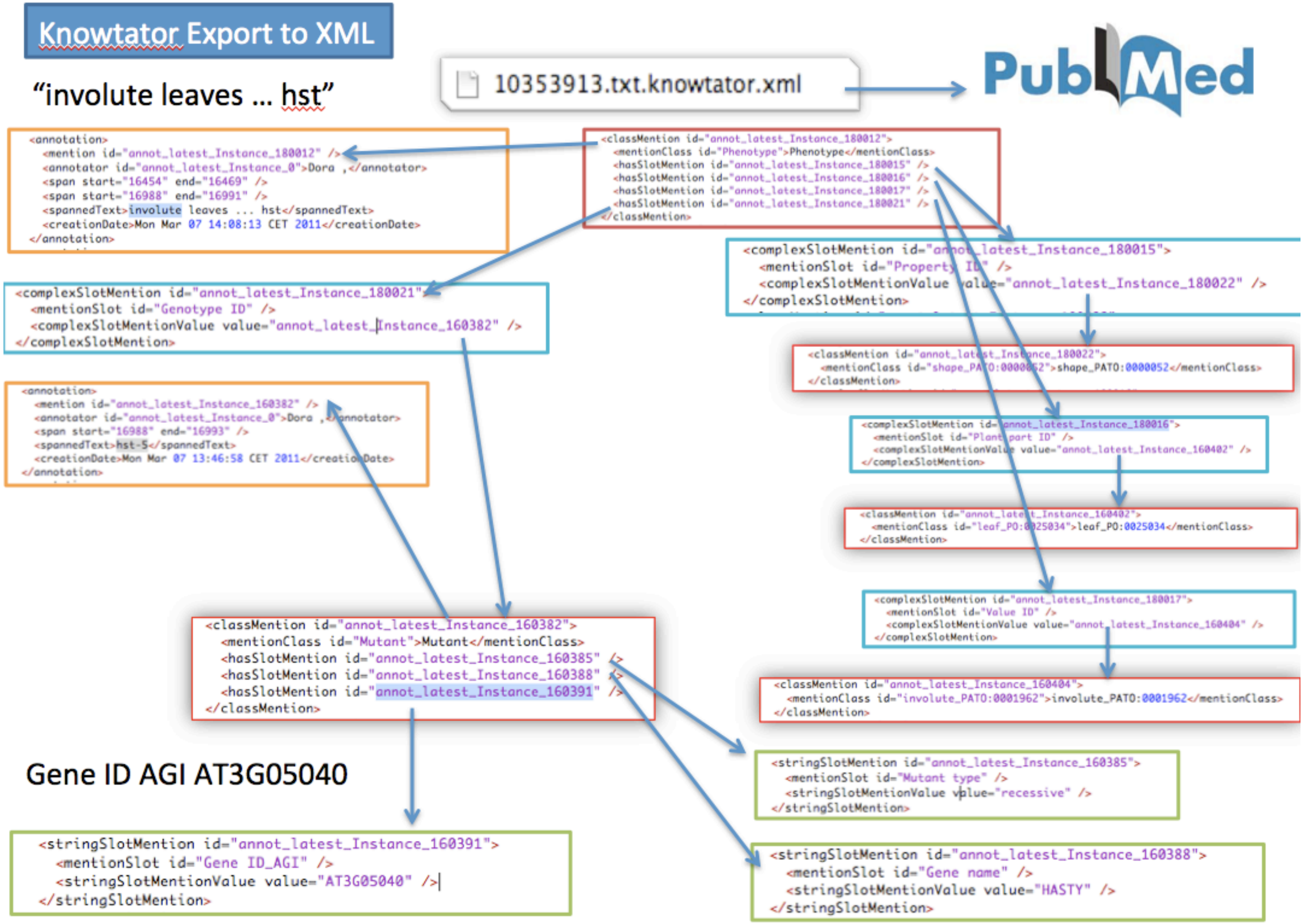
- Rigorous guidelines and training of community curators (hands-on sessions, documentation
- Records quality checked with scripts designed detecting different types of errors
  - completeness of relation annotations (i.e. were required slots filled)
  - consistency of ontology terms
  - report of orphan annotations or seemingly undefined ontology terms
- Logs examined by curators, relations adjusted when necessary

Relations produced by reference annotator highly consistent and complete 174 curated articles, the quality control script reported on average only **3.1** missing slots reported per article (not expected to be zero, missing textual info)
19,267 required slots in total, on average of 111 per article; 2.8% missing slots

Relations encoded originally by twelve community annotators contained on average **4.9** missing slots per article, dropped to **2.8**

Knowtator output format

Protégé/Knowtator exports XML representing objects and pointers to objects



Knowtator Export to XML

“involute leaves … hst”

10353913.txt.knowtator.xml

PubMed

```
<annotation>
  <mention id="annot_latest_Instance_180012" />
  <annotator id="annot_latest_Instance_0">Dora ,</annotator>
  <span start="16454" end="16469" />
  <span start="16988" end="16991" />
  <spannedText>involute leaves ... hst</spannedText>
  <creationDate>Mon Mar 07 14:08:13 CET 2011</creationDate>
</annotation>
```

```
<classMention id="annot_latest_Instance_180012">
  <mentionClass id="Phenotype">Phenotype</mentionClass>
  <hasSlotMention id="annot_latest_Instance_180015" />
  <hasSlotMention id="annot_latest_Instance_180016" />
  <hasSlotMention id="annot_latest_Instance_180017" />
  <hasSlotMention id="annot_latest_Instance_180021" />
</classMention>
```

```
<complexSlotMention id="annot_latest_Instance_180015">
  <mentionSlot id="Property ID" />
  <complexSlotMentionValue value="annot_latest_Instance_180022" />
</complexSlotMention>
```

```
<complexSlotMention id="annot_latest_Instance_180021">
  <mentionSlot id="Genotype ID" />
  <complexSlotMentionValue value="annot_latest_Instance_160382" />
</complexSlotMention>
```

```
<classMention id="annot_latest_Instance_180022">
  <mentionClass id="shape_PATO:0000052">shape_PATO:0000052</mentionClass>
</classMention>
```

```
<annotation>
  <mention id="annot_latest_Instance_160382" />
  <annotator id="annot_latest_Instance_0">Dora ,</annotator>
  <span start="16988" end="16993" />
  <spannedText>hst-5</spannedText>
  <creationDate>Mon Mar 07 13:46:58 CET 2011</creationDate>
</annotation>
```

```
<complexSlotMention id="annot_latest_Instance_180016">
  <mentionSlot id="Plant part ID" />
  <complexSlotMentionValue value="annot_latest_Instance_160402" />
</complexSlotMention>
```

```
<classMention id="annot_latest_Instance_160402">
  <mentionClass id="leaf_PO:0025034">leaf_PO:0025034</mentionClass>
</classMention>
```

```
<complexSlotMention id="annot_latest_Instance_180017">
  <mentionSlot id="Value ID" />
  <complexSlotMentionValue value="annot_latest_Instance_160404" />
</complexSlotMention>
```

```
<classMention id="annot_latest_Instance_160382">
  <mentionClass id="Mutant">Mutant</mentionClass>
  <hasSlotMention id="annot_latest_Instance_160385" />
  <hasSlotMention id="annot_latest_Instance_160388" />
  <hasSlotMention id="annot_latest_Instance_160391" />
</classMention>
```

```
<classMention id="annot_latest_Instance_160404">
  <mentionClass id="involute_PATO:0001962">involute_PATO:0001962</mentionClass>
</classMention>
```

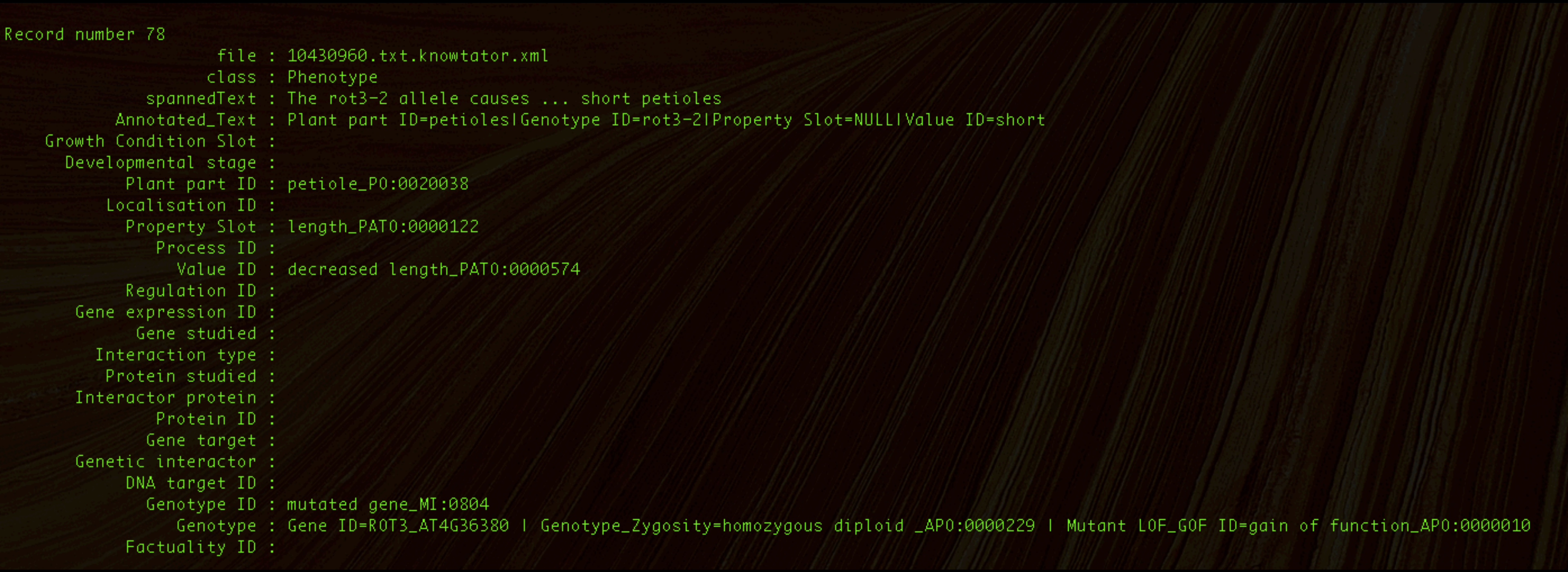
```
<stringSlotMention id="annot_latest_Instance_160385">
  <mentionSlot id="Mutant type" />
  <stringSlotMentionValue value="recessive" />
</stringSlotMention>
```

Gene ID AGI AT3G05040

```
<stringSlotMention id="annot_latest_Instance_160391">
  <mentionSlot id="Gene ID_AGI" />
  <stringSlotMentionValue value="AT3G05040" />
</stringSlotMention>
```

```
<stringSlotMention id="annot_latest_Instance_160388">
  <mentionSlot id="Gene name" />
  <stringSlotMentionValue value="HASTY" />
</stringSlotMention>
```

Parsing the data



```
Record number 78
                     file : 10430960.txt.knowtator.xml
                    class : Phenotype
              spannedText : The rot3-2 allele causes ... short petioles
           Annotated_Text : Plant part ID=petioles|Genotype ID=rot3-2|Property Slot=NULL|Value ID=short
    Growth Condition Slot :
      Developmental stage :
            Plant part ID : petiole_PO:0020038
          Localisation ID :
            Property Slot : length_PATO:0000122
               Process ID :
                 Value ID : decreased length_PATO:0000574
            Regulation ID :
       Gene expression ID :
             Gene studied :
         Interaction type :
          Protein studied :
       Interactor protein :
               Protein ID :
              Gene target :
       Genetic interactor :
            DNA target ID :
              Genotype ID : mutated gene_MI:0804
                 Genotype : Gene ID=ROT3_AT4G36380 | Genotype_Zygosity=homozygous diploid _APO:0000229 | Mutant LOF_GOF ID=gain of function_APO:0000010
            Factuality ID :
```

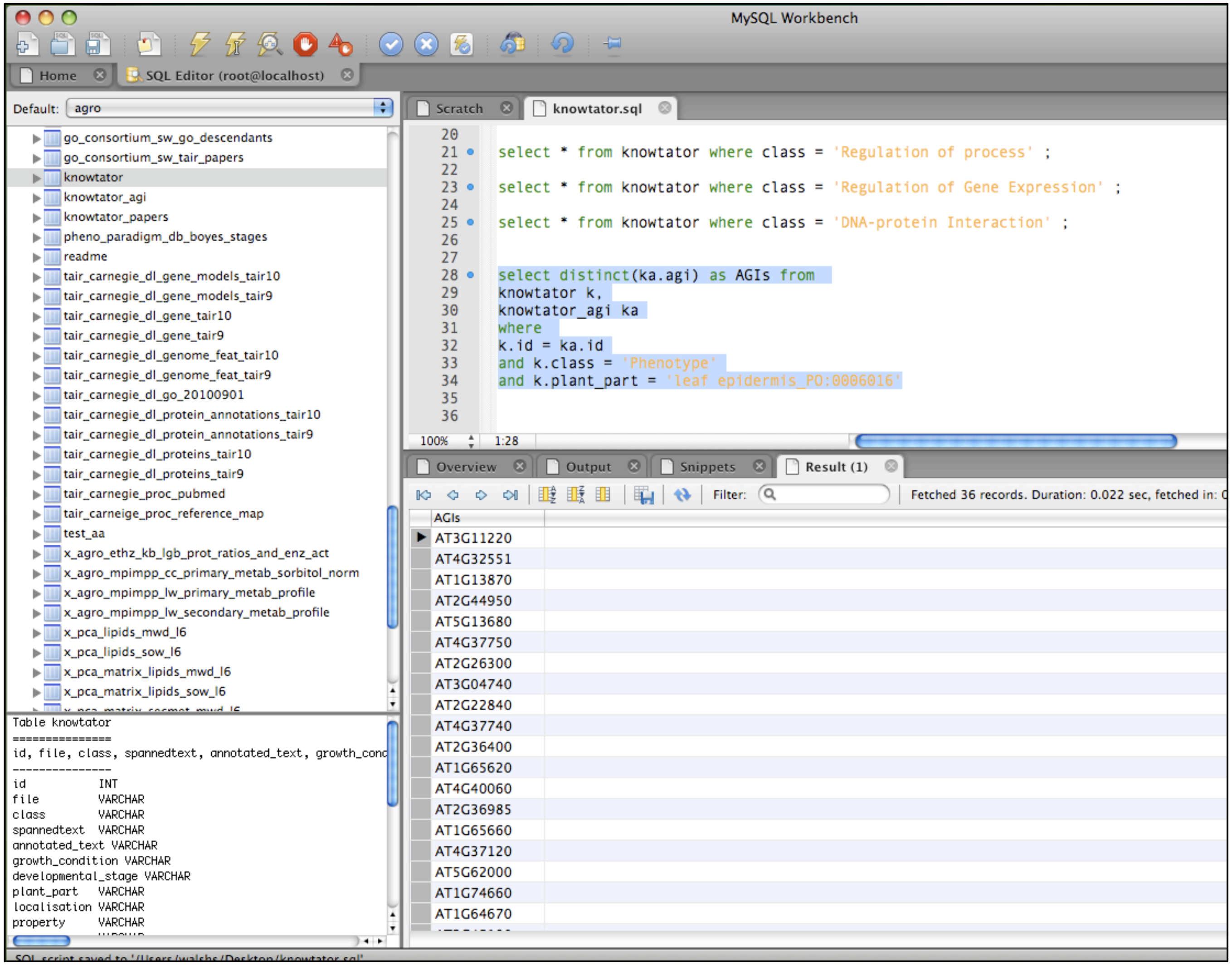
Parsing the data

Information type	Record
file	10430960.txt.knowtator.xml
class	Phenotype
SpannedText	The rot3-2 allele causes ... short petioles
span	10725|10749,10776|10790
Annotated_Text	Plant part ID=petioles|Genotype ID=rot3-2|Property Slot=NULL|Value ID=short
Developmental stage	
Factuality	
Genotype ID	mutated gene_MI:0804
Mutant info	Genotype_Zygosity=homozygous diploid _APO:0000229|Mutant LOF_GOF ID=gain of function_APO:0000010
AGI	Gene ID=ROT3_AT4G36380
Growth condition	
Localisation	
Methodology	
Plant part	petiole_PO:0020038
Process	
Property	length_PATO:0000122
Value	decreased length_PATO:0000574

KnownLeaf MySQL database

Now in a position to begin asking questions of the data as a whole

e.g. give list of AGIs involved in leaf epidermal phenotypes



```
go_consortium_sw_go_descendants
go_consortium_sw_tair_papers
knowtator
knowtator_agi
knowtator_papers
pheno_paradigm_db_boyes_stages
readme
tair_carnegie_dl_gene_models_tair10
tair_carnegie_dl_gene_models_tair9
tair_carnegie_dl_gene_tair10
tair_carnegie_dl_gene_tair9
tair_carnegie_dl_genome_feat_tair10
tair_carnegie_dl_genome_feat_tair9
tair_carnegie_dl_go_20100901
tair_carnegie_dl_protein_annotations_tair10
tair_carnegie_dl_protein_annotations_tair9
tair_carnegie_dl_proteins_tair10
tair_carnegie_dl_proteins_tair9
tair_carnegie_proc_pubmed
tair_carneige_proc_reference_map
test_aa
x_agro_ethz_kb_lgb_prot_ratios_and_enz_act
x_agro_mpimpp_cc_primary_metab_sorbitol_norm
x_agro_mpimpp_lw_primary_metab_profile
x_agro_mpimpp_lw_secondary_metab_profile
x_pca_lipids_mwd_l6
x_pca_lipids_sow_l6
x_pca_matrix_lipids_mwd_l6
x_pca_matrix_lipids_sow_l6
x_pca_matrix_secmet_mwd_l6
```

```
Table knowtator
===============
id, file, class, spannedtext, annotated_text, growth_conc
---------------
id           INT
file         VARCHAR
class        VARCHAR
spannedtext  VARCHAR
annotated_text VARCHAR
growth_condition VARCHAR
developmental_stage VARCHAR
plant_part   VARCHAR
localisation VARCHAR
property     VARCHAR
```

```sql

select * from knowtator where class = 'Regulation of process' ;

select * from knowtator where class = 'Regulation of Gene Expression' ;

select * from knowtator where class = 'DNA-protein Interaction' ;


select distinct(ka.agi) as AGIs from
knowtator k,
knowtator_agi ka
where
k.id = ka.id
and k.class = 'Phenotype'
and k.plant_part = 'leaf epidermis_PO:0006016'


```

Fetched 36 records. Duration: 0.022 sec, fetched in: 0

AGIs
AT3G11220
AT4G32551
AT1G13870
AT2G44950
AT5G13680
AT4G37750
AT2G26300
AT3G04740
AT2G22840
AT4G37740
AT2G36400
AT1G65620
AT4G40060
AT2G36985
AT1G65660
AT4G37120
AT5G62000
AT1G74660
AT1G64670

SQL script saved to '/Users/walshs/Desktop/knowtator.sql'

Overview of KnownLeaf database content

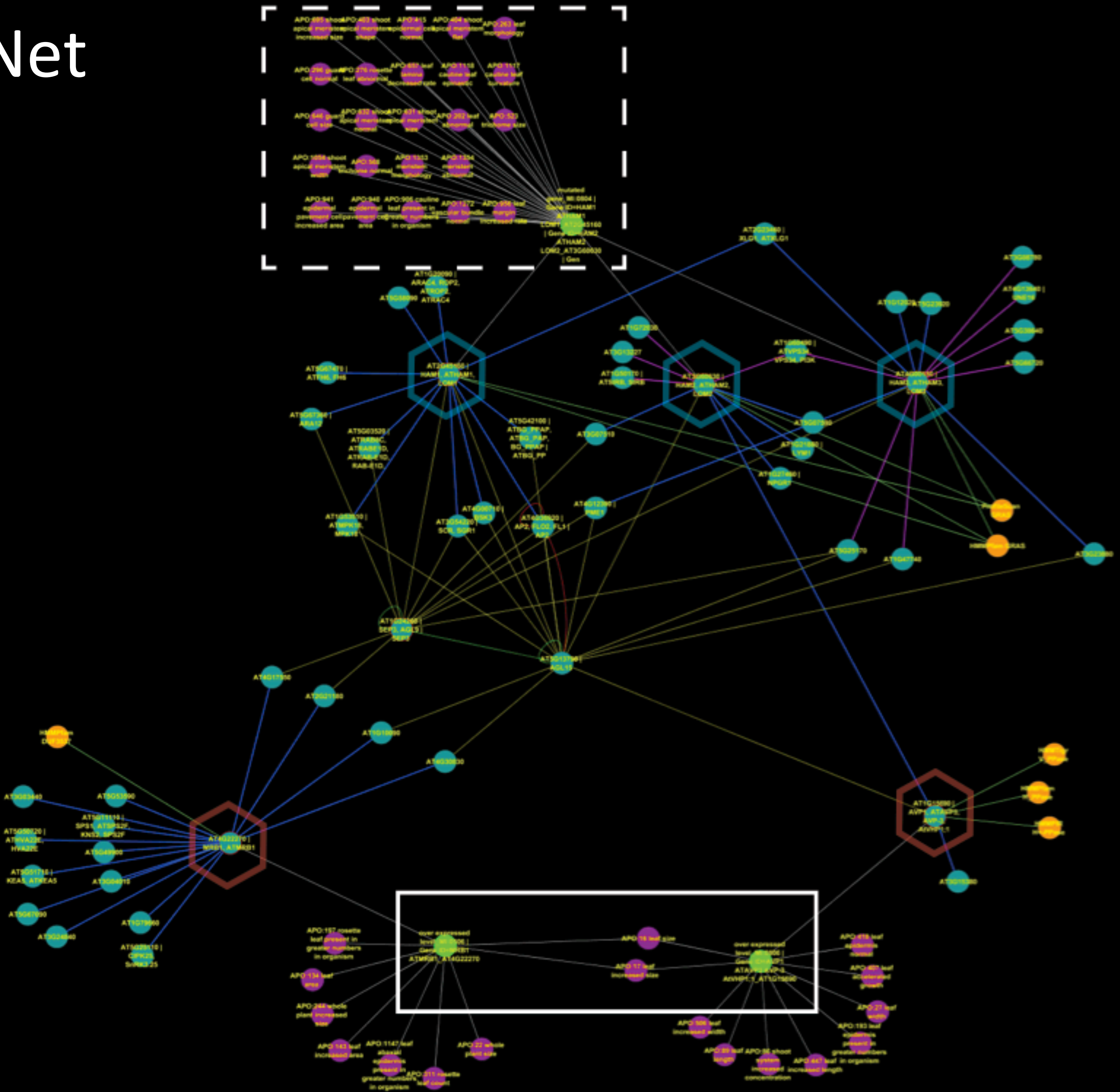
Relation category	# AGI	# unique AGI	Ratio
Phenotype	5608	381	14.72
Gene expression	4767	704	6.77
Genetic interaction	658	186	3.54
Feature	462	175	2.64
Protein-protein interaction	310	121	2.56
Process	235	140	1.68
Regulation of gene expression	204	70	2.91
Regulation of process	178	85	2.09
DNA-protein interaction	92	47	1.96
Regulation of phenotype	20		
Total	12534		

Representation of collected annotations merged with public knowledge sources

- Interactive graph in Cytoscape
- Selected Knowtator relations part of a larger network of objects to interpret and further expand these relations
- Represented edges
  - Co-expression (ATTED-II mutual rank score ≥ 25)
  - Protein-protein interactions (Y2H, AI-1 interactome)
  - Phenotype relations (gene-mutation-phenotype)
- Seeded nodes
  - Genes/proteins annotated via Knowtator (AGI code)
  - Proteins that vary significantly across leaf development (Bärenfaller et al., 2012, Mol Syst Biol)
  - Connected protein/gene (co-expression, PPI)

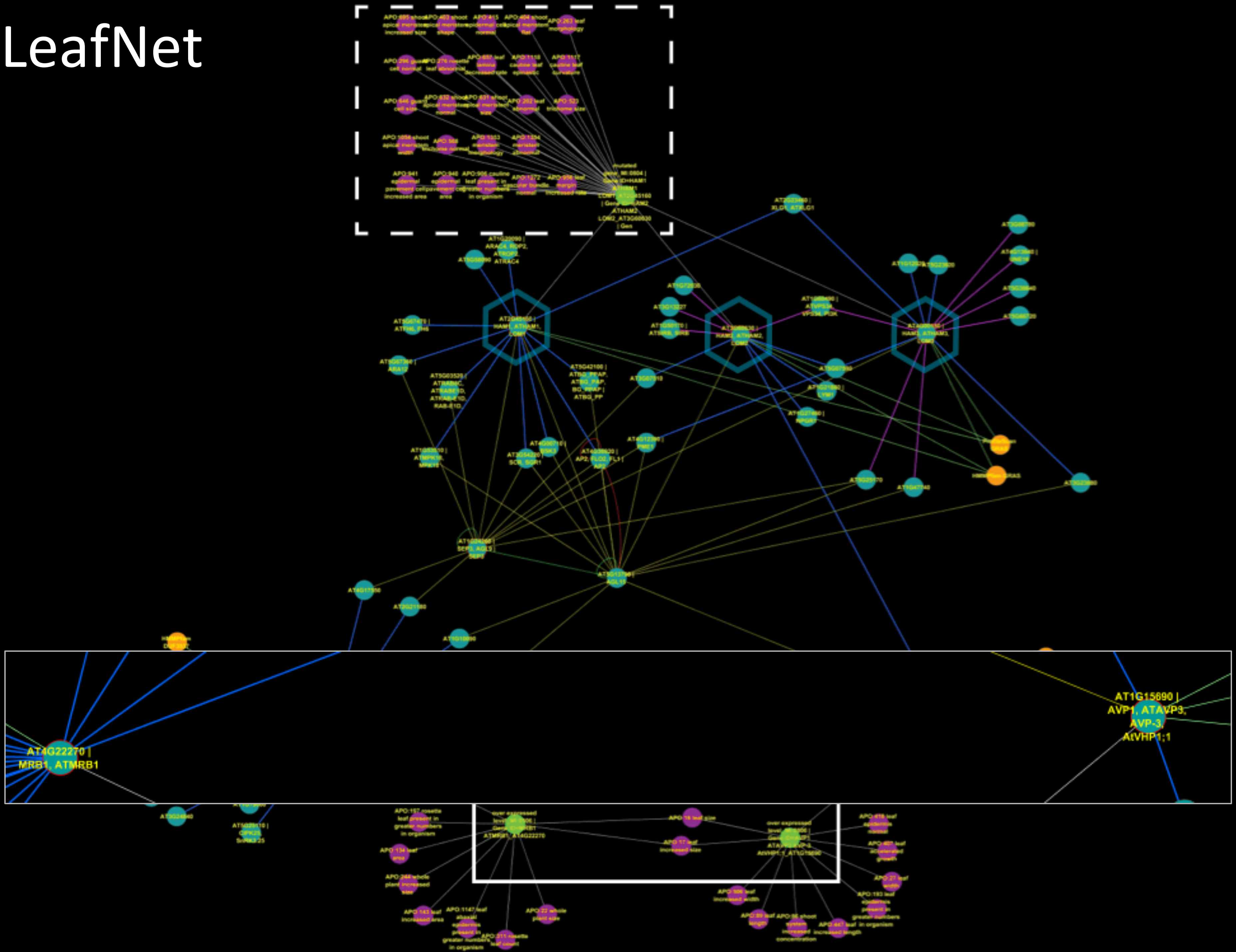
-> LeafNet: 19,055 nodes connected by 39,649 edges

LeafNet

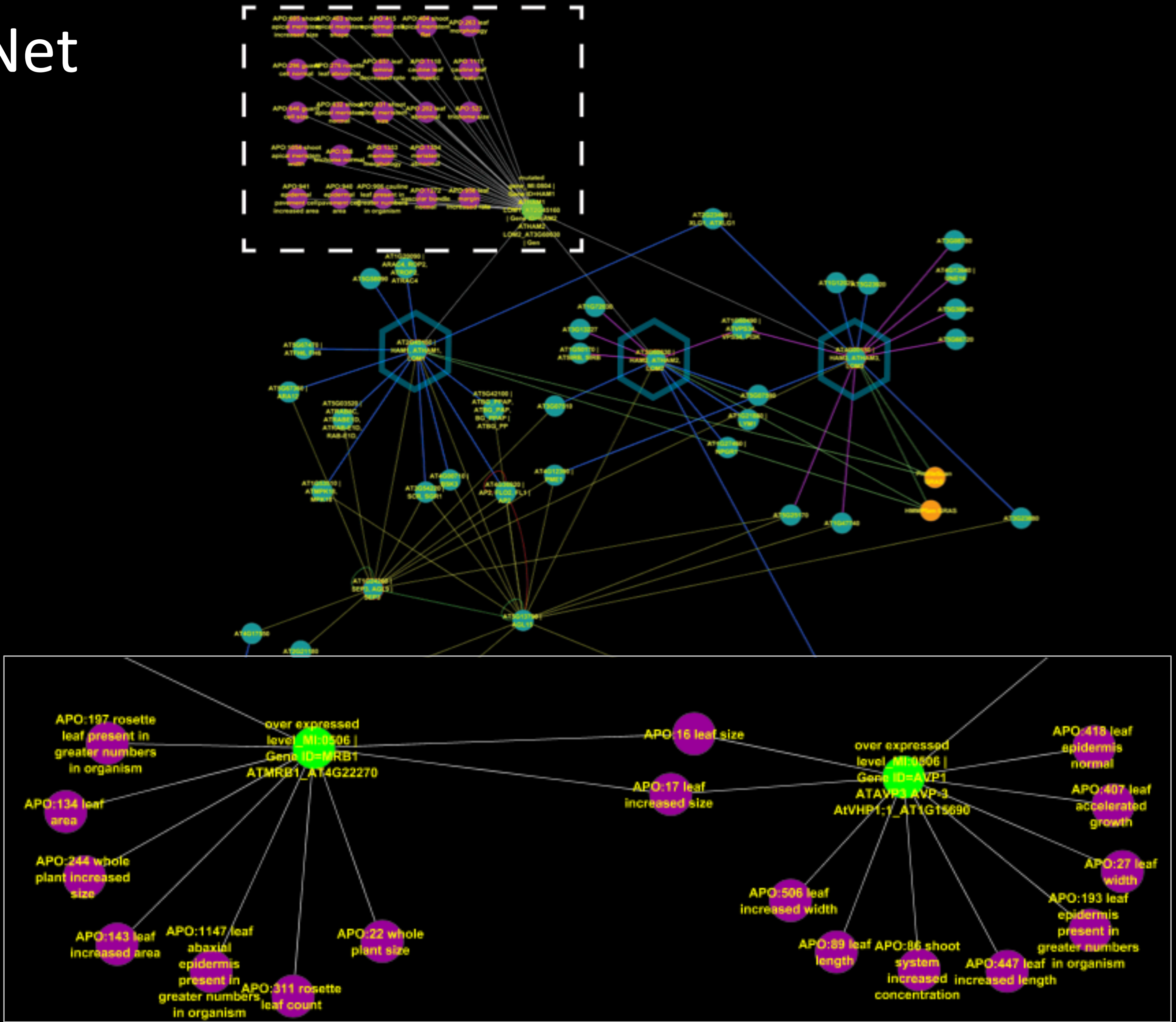


APO:695 shoot apical meristem increased size
APO:403 shoot apical meristem shape
APO:415 epidermal cell normal
APO:404 shoot apical meristem flat
APO:263 leaf morphology
APO:296 guard cell normal
APO:276 rosette leaf abnormal
APO:657 leaf lamina decreased size
APO:1118 cauline leaf epinastic
APO:1117 cauline leaf curvature
APO:646 guard cell size
APO:632 shoot apical meristem normal
APO:631 shoot apical meristem size
APO:262 leaf abnormal
APO:523 trichome size
APO:1056 shoot apical meristem width
APO:568 trichome normal
APO:1333 meristem morphology
APO:1334 meristem abnormal
APO:941 epidermal pavement cell increased area
APO:940 epidermal pavement cell area
APO:906 cauline leaf present in greater numbers in organism
APO:1272 vascular bundle normal
APO:956 leaf margin increased size
mutated gene_MI:0804 | Gene ID=HAM1 ATHAM1 LOM1_AT2G45160 | Gene ID=HAM2 ATHAM2 LOM2_AT3G60630 | Gen
AT1G20090 | ARAC4, ROP2, ATROP2, ATRAC4
AT5G58090
AT2G23460 | XLG1, ATXLG1
AT2G45160 | HAM1, ATHAM1, LOM1
AT3G60630 | HAM2, ATHAM2, LOM2
AT4G00150 | HAM3, ATHAM3, LOM3
AT5G67470 | ATFH6, FH6
AT5G67360 | ARA12
AT5G03520 | ATRAB8C, ATRABE1D, ATRAB-E1D, RAB-E1D
AT1G53510 | ATMPK18, MPK18
AT3G54220 | SCR, SGR1
AT4G00710 | BSK3
AT5G42100 | ATBG_PPAP, ATBG_PAP, BG_PPAP | ATBG_PP
AT4G36920 | AP2, FLO2, FL1 |
AT4G12390 | PME1
AT3G07510
AT1G72630
AT3G13227
AT1G50170 | ATSIRB, SIRB
AT1G60490 | ATVPS34, VPS34, PI3K
AT5G67390
AT1G21880 | LYM1
AT1G27480 | NPGR1
AT5G25170
AT1G47740
AT3G23890
AT2G24200 | SEP3, AGL9 | SEPT
AT5G13790 | AGL15
AT4G17950
AT2G21180
AT1G10090
AT4G33030
AT4G22270 | MRB1, ATMRB1
AT3G63440
AT5G53590
AT5G11110 | SPS1, ATSPS2F, KNS2, SPS2F
AT5G58020 | ATRVA22E, HVA22E
AT5G49900
AT5G51710 | KEA5, ATKEA5
AT3G04015
AT5G67090
AT3G24840
AT1G79960
AT5G25110 | CIPK25, SnRK3.25
AT1G15690 | AVP1, ATAVP3, AVP-3, AtVHP1;1
AT3G15890
over expressed level_MI:0506 | Gene ID=MRB1 ATMRB1_AT4G22270
APO:16 leaf size
APO:17 leaf increased size
over expressed level_MI:0506 | Gene ID=AVP1 ATAVP3 AVP-3 AtVHP1;1_AT1G15690
APO:157 rosette leaf present in greater numbers in organism
APO:134 leaf area
APO:244 whole plant increased size
APO:143 leaf increased area
APO:1147 leaf abaxial epidermis present in greater numbers in organism
APO:311 rosette leaf count
APO:22 whole plant size
APO:506 leaf increased width
APO:89 leaf length
APO:96 shoot system increased concentration
APO:447 leaf increased length
APO:193 leaf epidermis present in greater numbers in organism
APO:27 leaf width
APO:407 leaf accelerated growth
APO:418 leaf epidermis normal

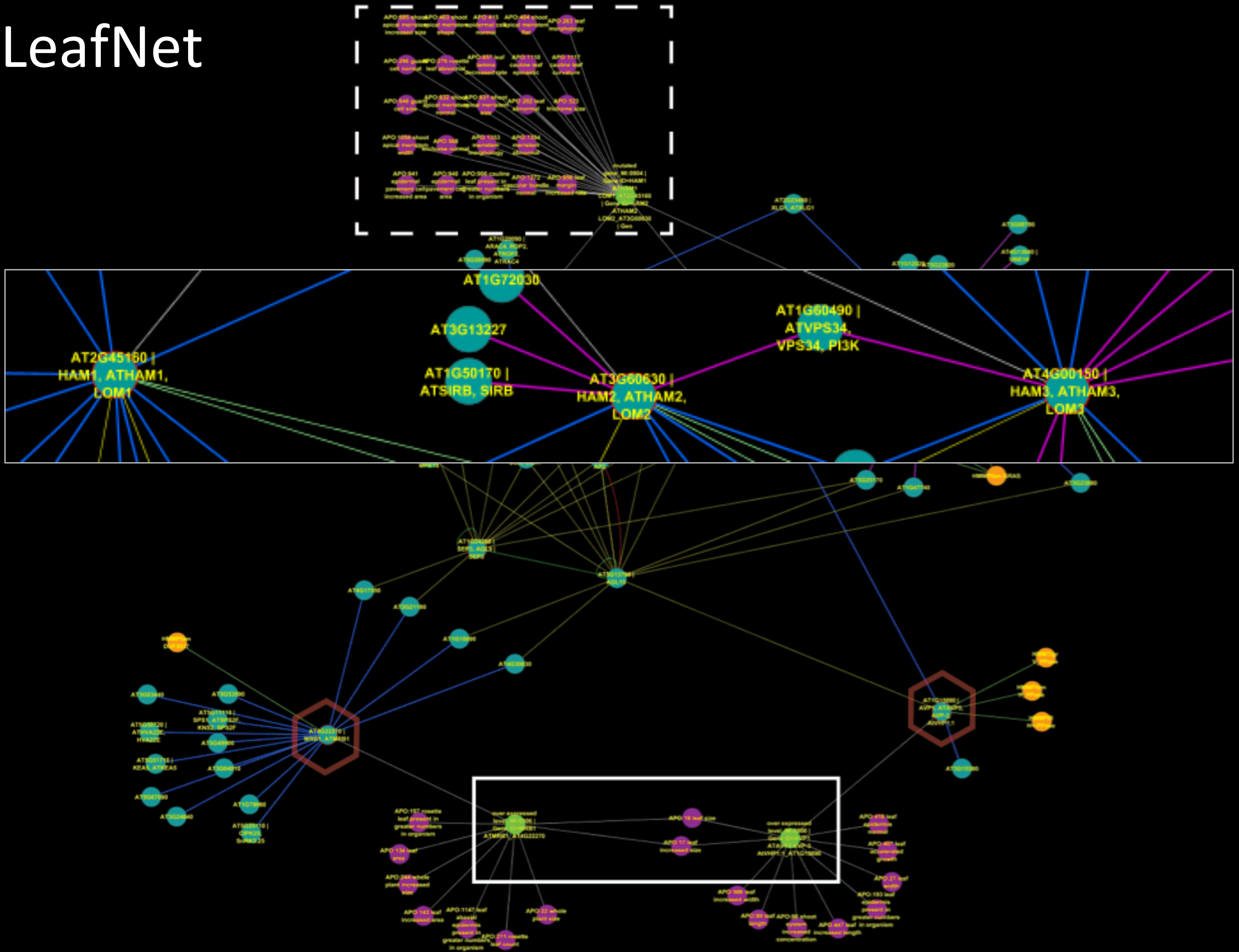
LeafNet


APO:263 leaf morphology
APO:523 trichome size
mutated gene_MI:0804 | Gene ID=HAM1 ATHAM1 LOM1_AT2G45160 | Gene ID=HAM2 ATHAM2 LOM2_AT3G60630 | Gen
AT1G20090 | ARAC4, ROP2, ATROP2, ATRAC4
AT2G45160 | HAM1, ATHAM1, LOM1
AT5G42100 | ATBG_PPAP, ATBG_PAP, BG_PPAP | ATBG_PP
AT4G00710 | BSK3
AT3G54220 | SCR, SGR1
AT4G36920 | AP2, FLO2, FL1 | AP2
AT4G12390 | PME1
AT1G24260 | SEP3, AGL9 | SEP3
AT5G13790 | AGL15
AT4G22270 | MRB1, ATMRB1
AT1G15890 | AVP1, ATAVP3, AVP-3, AtVHP1;1
APO:16 leaf size
APO:17 leaf increased size

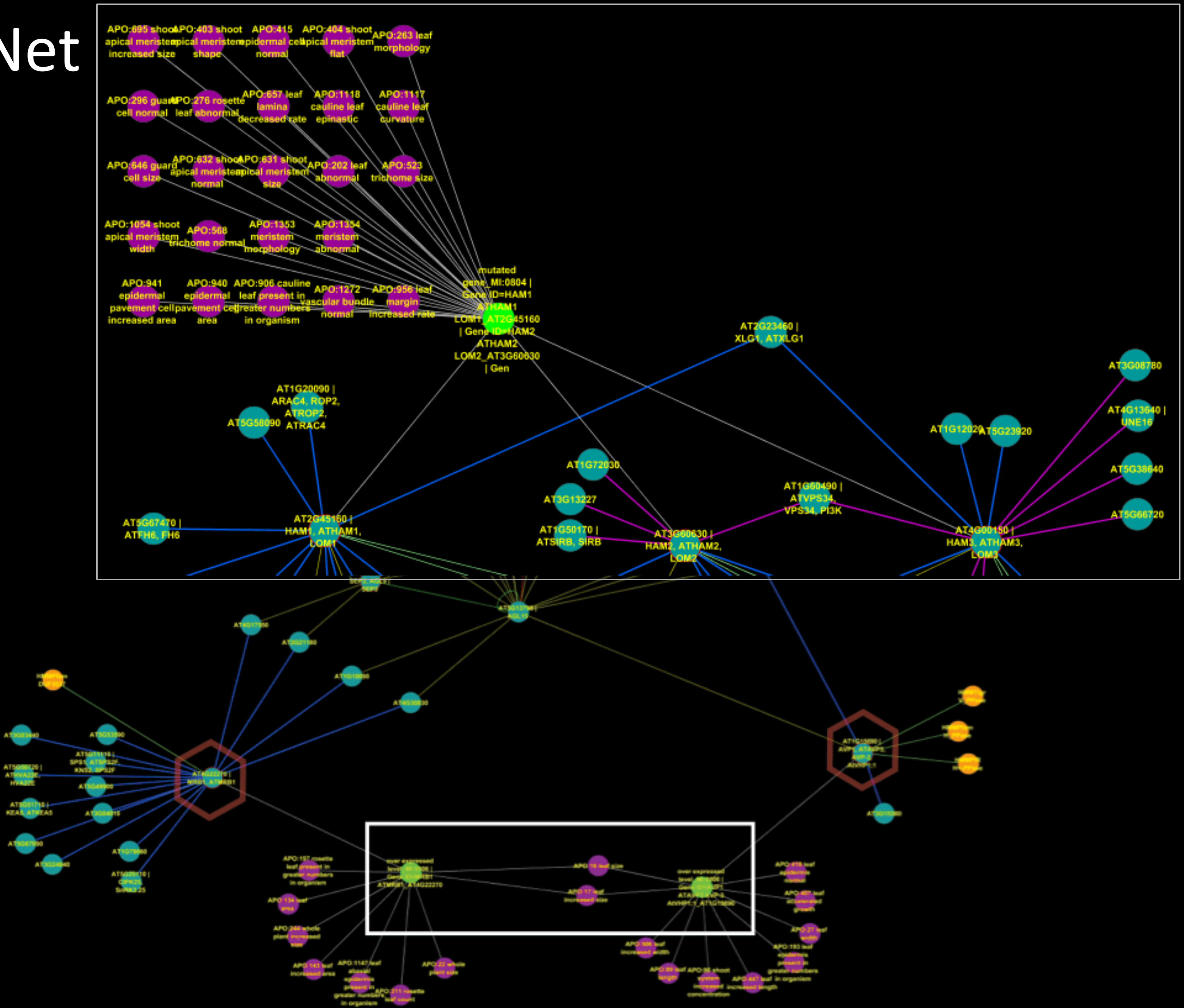
LeafNet


APO:403 shoot apical meristem shape
APO:415 epidermal cell normal
APO:404 shoot apical meristem flat
APO:263 leaf morphology
APO:276 rosette leaf abnormal
APO:1118 cauline leaf epinastic
APO:1117 cauline leaf curvature
APO:632 shoot apical meristem normal
APO:631 shoot apical meristem size
APO:202 leaf abnormal
APO:523 trichome size
APO:1056 shoot apical meristem width
APO:1213 meristem morphology
APO:1334 meristem abnormal
APO:941 epidermal pavement cell increased area
APO:940 epidermal pavement cell area
APO:906 cauline leaf present in greater numbers in organism
APO:1272 vascular bundle normal
APO:956 leaf margin increased rate
mutated gene_MI:0804 | Gene ID=HAM1 ATHAM1 LOM1_AT2G45160 | Gene ID=HAM2 ATHAM2 LOM2_AT3G60630 | Gen
AT1G20090 | ARAC4, ROP2, ATROP2, ATRAC4
AT2G23460 | XLG1, ATXLG1
AT2G45160 | HAM1, ATHAM1, LOM1
AT3G60630 | HAM2, ATHAM2, LOM2
AT5G67470 | ATFH6, FH6
AT5G67360 | ARA12
AT5G03530 | ATRABC, ATRABE1D, ATRAB-E1D, RAB-E1D
AT5G42100 | ATBG_PPAP, ATBG_PAP, BG_PPAP | ATBG_PP
AT1G72830
AT3G13227
AT1G50170 | ATSIRB, SIRB
AT3G07510
AT1G21880 | LYM1
AT1G27460 | NPGR1
AT4G12390 | PME1
AT4G00710 | BSK3
AT3G54220 | SCR, SGR1
AT4G36920 | AP2, FLO2, FL1 | AP2
AT1G53510 | ATMPK18, MPK18
AT1G24260 | SEP3, AGL9 | SEP3
AT5G13790 | AGL15
AT4G17980
AT2G21180
AT1G60490 | ATVPS34, VPS34, PI3K
AT5G25170
AT1G47740
AT3G23880
over expressed level_MI:0506 | Gene ID=MRB1 ATMRB1_AT4G22270
APO:197 rosette leaf present in greater numbers in organism
APO:134 leaf area
APO:244 whole plant increased size
APO:143 leaf increased area
APO:1147 leaf abaxial epidermis present in greater numbers in organism
APO:311 rosette leaf count
APO:22 whole plant size
APO:16 leaf size
APO:17 leaf increased size
over expressed level_MI:0506 | Gene ID=AVP1 ATAVP3 AVP-3 AtVHP1;1_AT1G15690
APO:418 leaf epidermis normal
APO:407 leaf accelerated growth
APO:27 leaf width
APO:193 leaf epidermis present in greater numbers in organism
APO:506 leaf increased width
APO:89 leaf length
APO:86 shoot system increased concentration
APO:447 leaf increased length

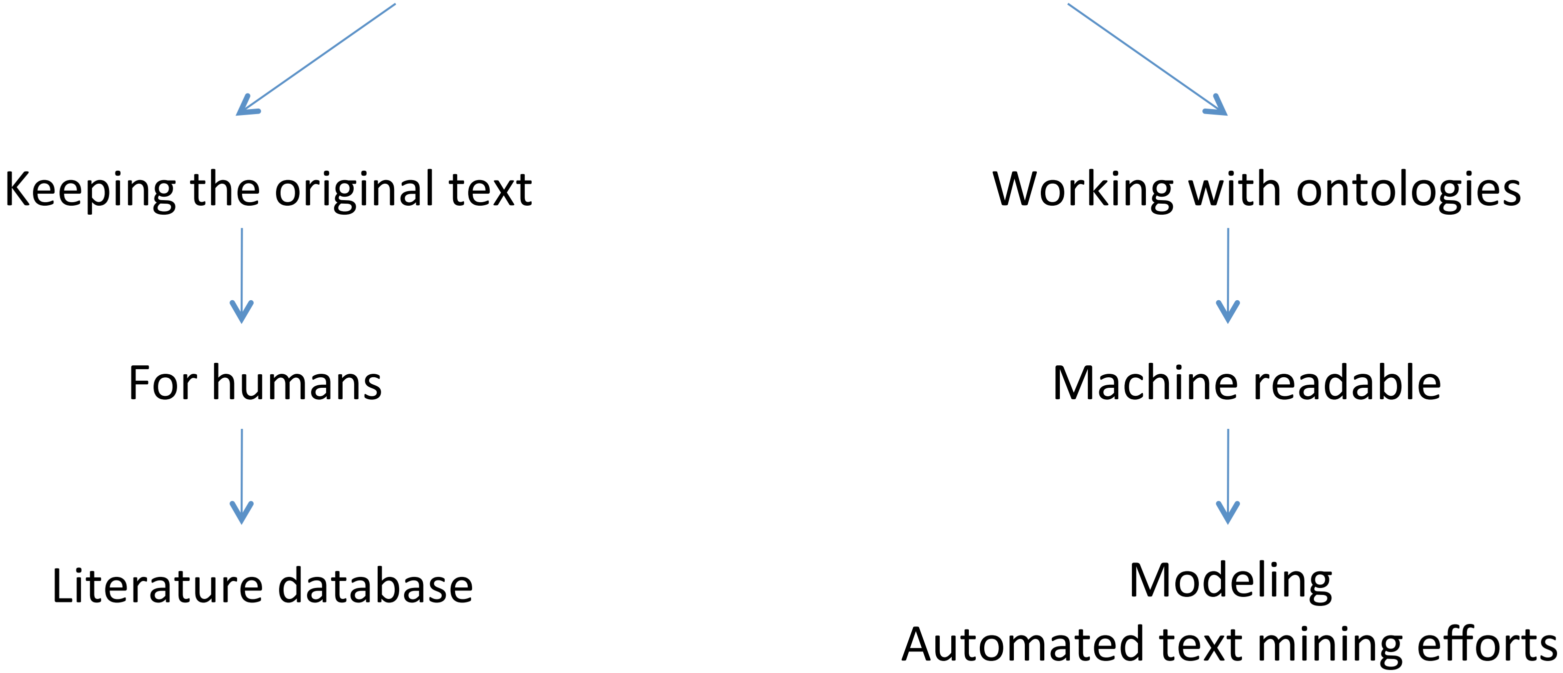
LeafNet


AT1G72030
AT3G13227
AT1G60490 | ATVPS34, VPS34, PI3K
AT2G45160 | HAM1, ATHAM1, LOM1
AT1G50170 | ATSIRB, SIRB
AT3G60630 | HAM2, ATHAM2, LOM2
AT4G00150 | HAM3, ATHAM3, LOM3

LeafNet


APO:695 shoot apical meristem increased size
APO:403 shoot apical meristem shape
APO:415 epidermal cell normal
APO:404 shoot apical meristem flat
APO:263 leaf morphology
APO:296 guard cell normal
APO:276 rosette leaf abnormal
APO:657 leaf lamina decreased rate
APO:1118 cauline leaf epinastic
APO:1117 cauline leaf curvature
APO:646 guard cell size
APO:632 shoot apical meristem normal
APO:631 shoot apical meristem size
APO:202 leaf abnormal
APO:523 trichome size
APO:1054 shoot apical meristem width
APO:568 trichome normal
APO:1353 meristem morphology
APO:1354 meristem abnormal
APO:941 epidermal pavement cell increased area
APO:940 epidermal pavement cell area
APO:906 cauline leaf present in greater numbers in organism
APO:1272 vascular bundle normal
APO:956 leaf margin increased rate
mutated gene_MI:0804 | Gene ID=HAM1 ATHAM1 LOM1_AT2G45160 | Gene ID=HAM2 ATHAM2 LOM2_AT3G60630 | Gen
AT2G23460 | XLG1, ATXLG1
AT3G08780
AT4G13640 | UNE16
AT1G12020
AT5G23920
AT5G38640
AT5G66720
AT1G20090 | ARAC4, ROP2, ATROP2, ATRAC4
AT5G58090
AT1G72030
AT3G13227
AT1G60490 | ATVPS34, VPS34, PI3K
AT1G50170 | ATSIRB, SIRB
AT5G67470 | ATFH6, FH6
AT2G45160 | HAM1, ATHAM1, LOM1
AT3G60630 | HAM2, ATHAM2, LOM2
AT4G00150 | HAM3, ATHAM3, LOM3
over expressed
APO:157 rosette leaf present in greater numbers in organism
APO:134 leaf area
APO:244 whole plant increased size
APO:143 leaf increased area
APO:1147 leaf abaxial epidermis present in greater numbers in organism
APO:211 rosette leaf count
APO:22 whole plant size
APO:16 leaf size
APO:17 leaf increased size
over expressed
APO:27 leaf width
APO:506 leaf increased width
APO:447 leaf increased length

Knowtator is useful for literature annotation



Keeping the original text

For humans

Literature database

Working with ontologies

Machine readable

Modeling

Automated text mining efforts

3 papers/day

Training to use the program: 3 hours intro, few days of practice

Easy to share projects (6 relatively small files + textsources)

Program is free

Engage your community

- Biological systems are complex, many genes/proteins are involved in any given process
- Many scientific articles are published… or will be soon
- Few willing annotators
- Long term benefits obvious, short term investments not so
- Practical tools are required

Perspectives

- Mining of the KnownLeaf database
- Exploration of LeafNet graphs to generate hypotheses
- Exhaustive annotation of all relevant papers in the "leaf growth and development" domain (now about 1/3)
- Implementation of the Leaf Knowtator annotation system in other domains
- Training of machine learning algorithms embedded in automated text mining tools with the Knowtator-generated data set
- All tools and data will soon be made public
  … except spanned text because of copyright restrictions

Contributors

Fabio Fiorani, initiator (Phenote)

Dóra Szakonyi, reference annotator and developer of Leaf Knowtator
Sofie Van Landeghem, text mining expert
Sean Walsh, KnownLeaf database and LeafNet
Pierre Hilson, molecular biologist

Community annotators: Katja Bärenfaller, Lieven Baeyens, Jonas Blomme, Rubén Casanova-Sáez, Stefanie De Bodt, David Esteve-Bruna, Nathalie Gonzalez, Jesper Grønlund, Richard G.H. Immink, Sara Jover-Gil, Asuka Kuwabara, Tamara Muñoz-Nortes, Aalt-Jan van Dijk, David Wilson-Sánchez

PIs: Vicky Buchanan-Wollaston, Gerco C. Angenent, Yves Van de Peer, Dirk Inzé, José Luis Micol, Wilhelm Gruissem,

Researchers groups:
Department of Plant Systems Biology, VIB, Ghent University
Department of Biology, ETH Zurich
División de Genética, Universidad Miguel Hernández, Elche
Warwick Systems Biology Centre, University of Warwick
Plant Research International (PRI), Bioscience, Wageningen

I NEED MORE HELP! DON'T YOU UNDERSTAND?
PK 2006