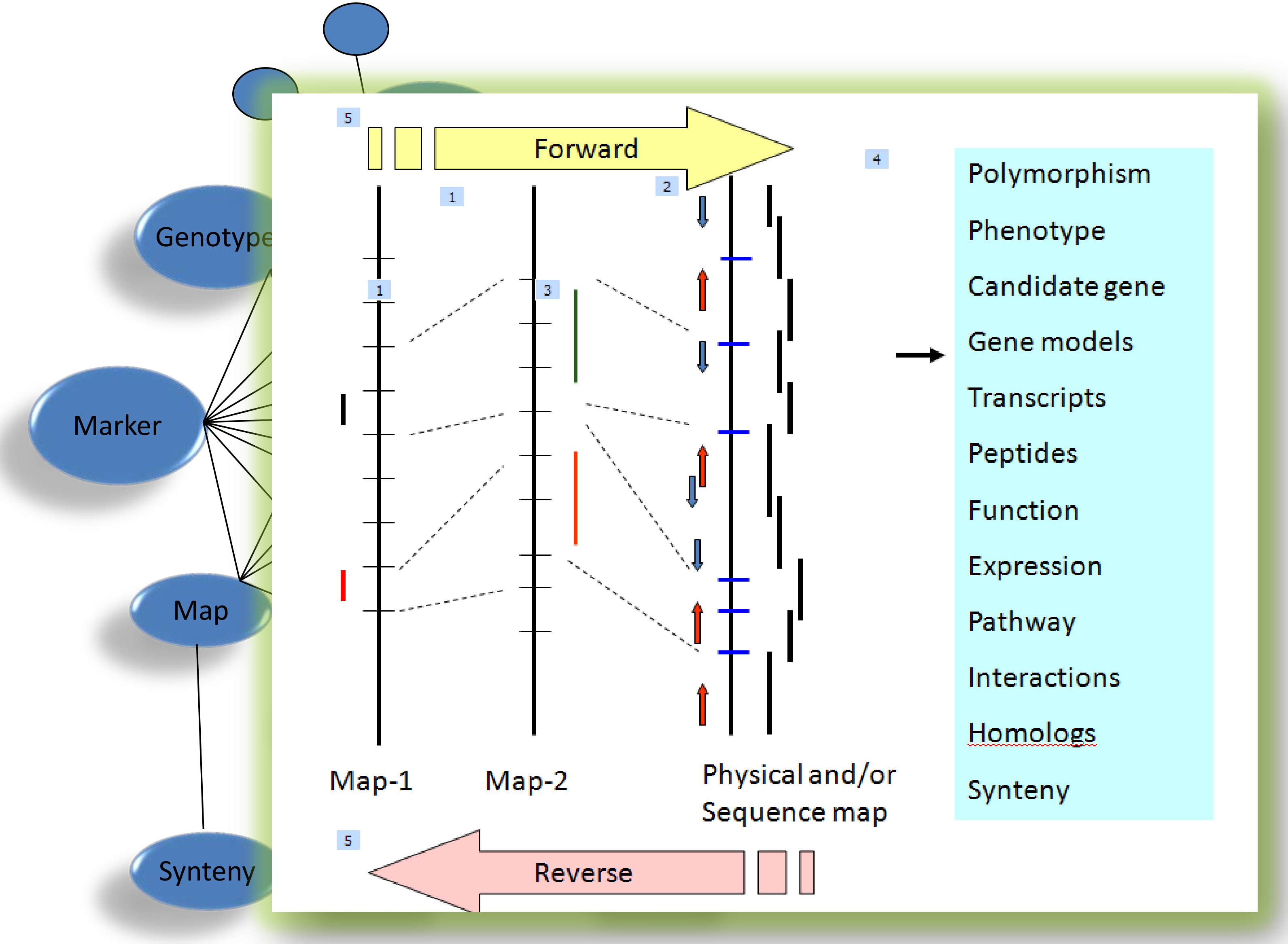
Plant and Trait Ontology Current Status

Pankaj Jaiswal

Oregon State University

Plant Ontologies for Agronomic Traits
EBI, UK
December 8, 2011

Comparative and Translational Genomics



Polymorphism

Phenotype

Candidate gene

Gene models

Transcripts

Peptides

Function

Expression

Pathway

Interactions

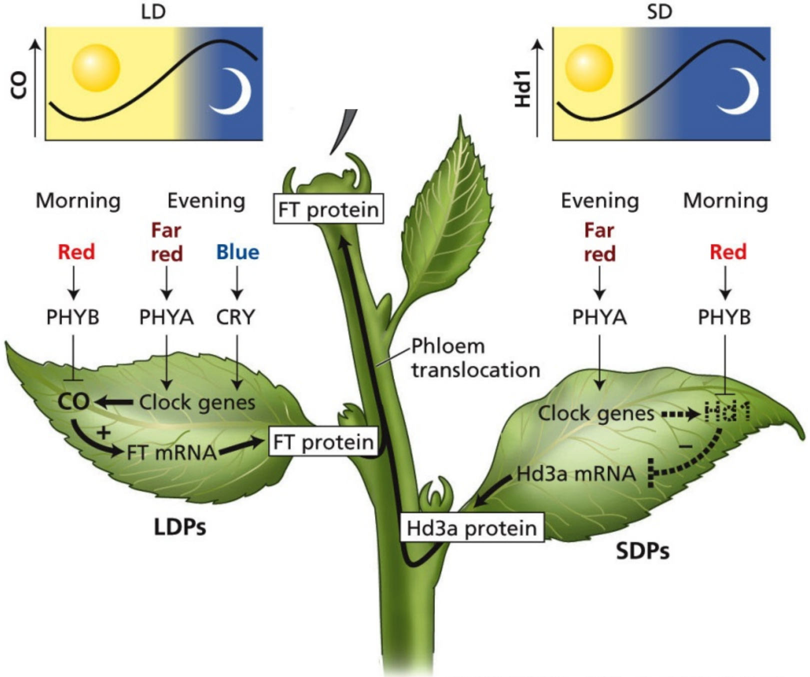
Homologs

Synteny

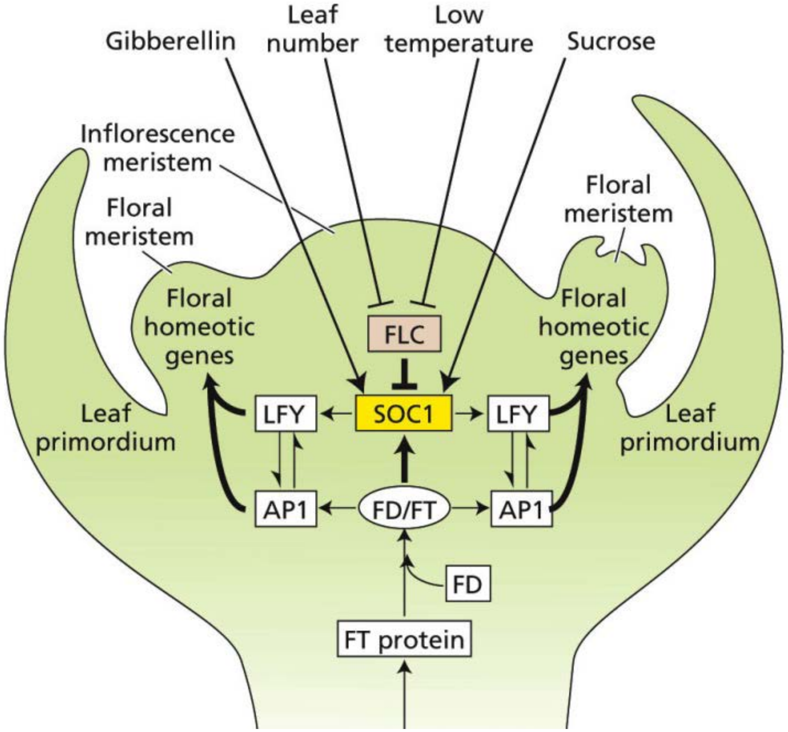
Photoperiod and the process of flowering

Arabidopsis
long-day
flowering

Rice
Short-day
flowering



PLANT PHYSIOLOGY, Fourth Edition, Figure 25.33 (Part 1) © 2006 Sinauer Associates, Inc.



GWAS and QTL studies

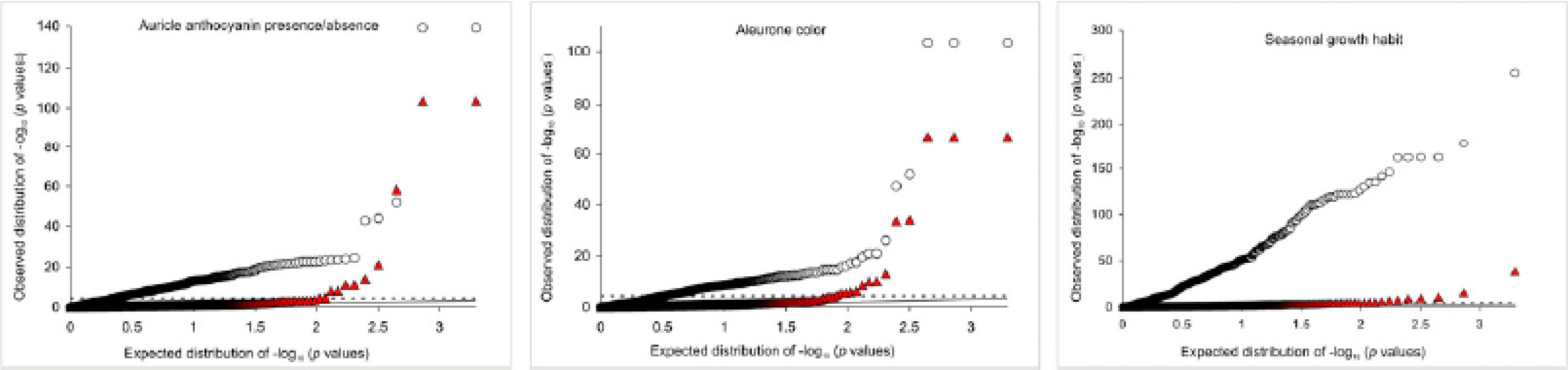
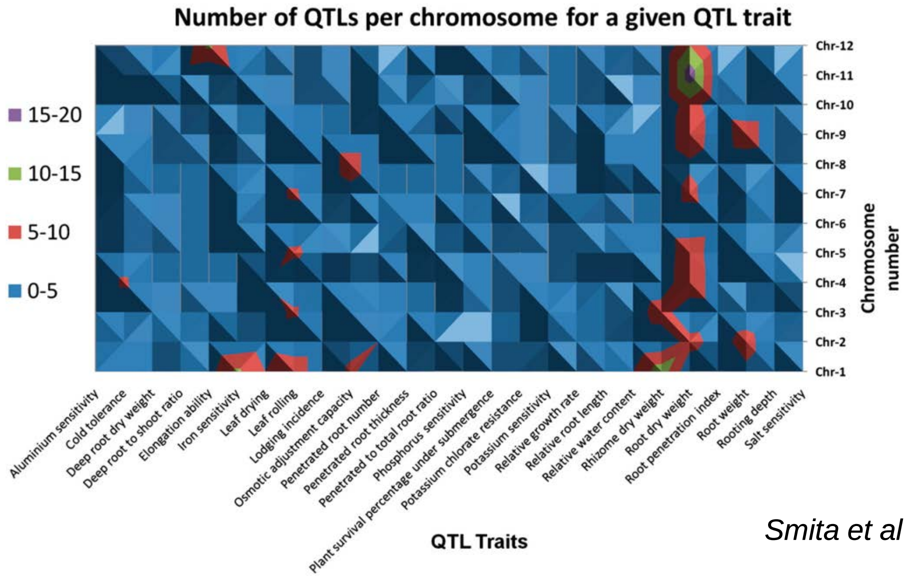


Fig. S4. Quantile–quantile plots for traits returning significant associations after GWA analysis. Expected vs. observed P values are plotted for naive (circles) and mixed model corrected (triangles) analyses. The $x = y$ line (solid) and Bonferroni corrected $P = 0.05$ significance thresholds (dashed line) are indicated.

Cockram et al (2010) PNAS



Smita et al (2011) doi: 10.1093/database/bar037

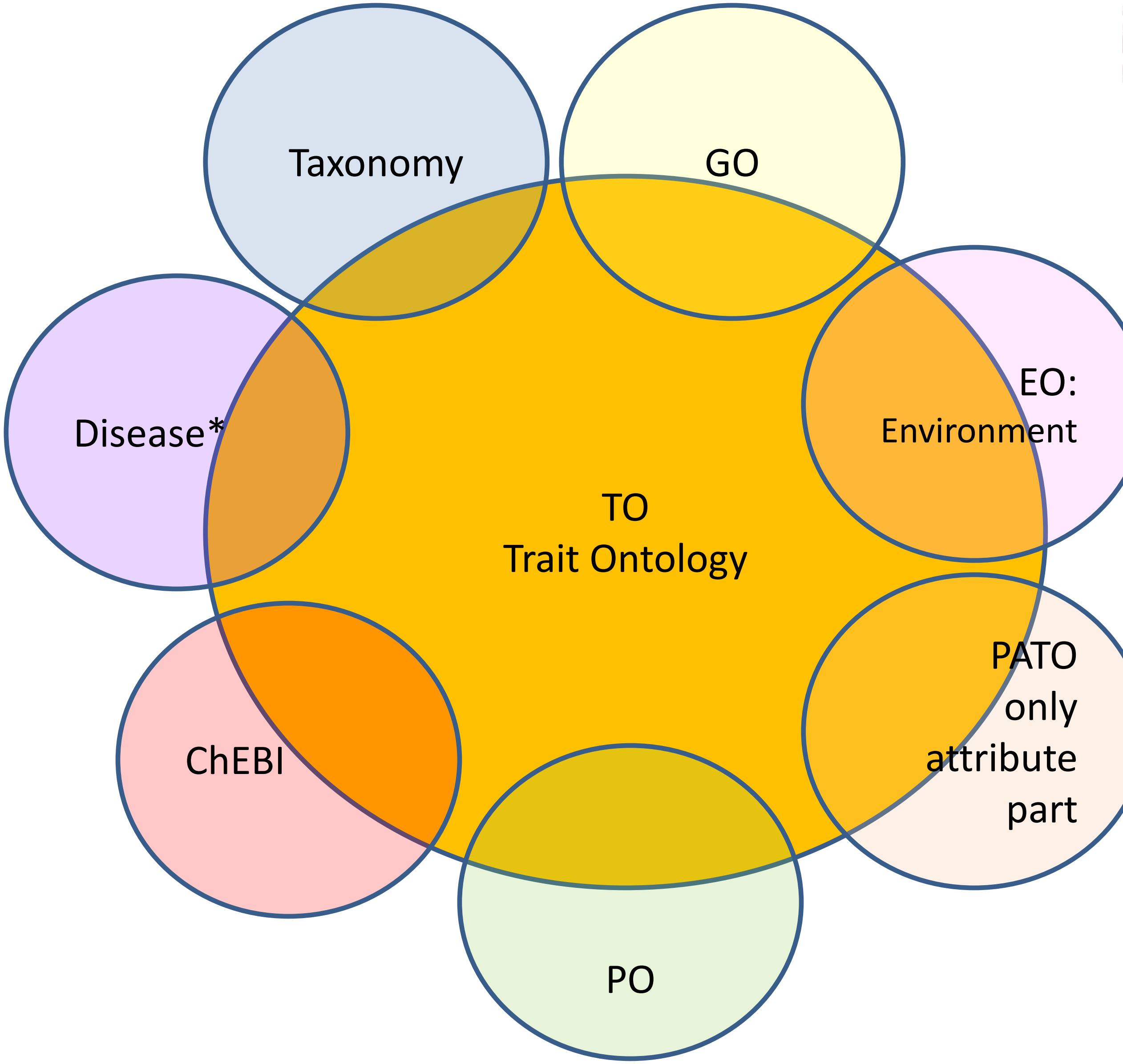
Trait vs Phenotype

- Entity+Attribute = Trait [**observable**]
  E.g. Leaf(PO) + color (PATO-A) = Leaf color(TO)

- [Entity+Attribute+Value] = Phenotype [**observed**]
  e.g. Leaf(PO) + color (PATO-A) + yellow (PATO-V) = Leaf color yellow [EAV model:old]

- [Entity+(Attribute+Value)] = Phenotype [**observed**]
  e.g. Leaf(PO) + color yellow (PATO-AV) = Leaf color yellow
  [EA model:NEW]

Traits are the visible markers from multiple dimensions



Taxonomy

GO

Disease*

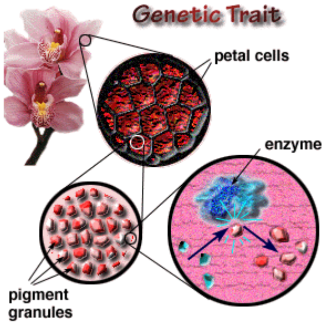
EO:
Environment

TO
Trait Ontology

PATO
only
attribute
part

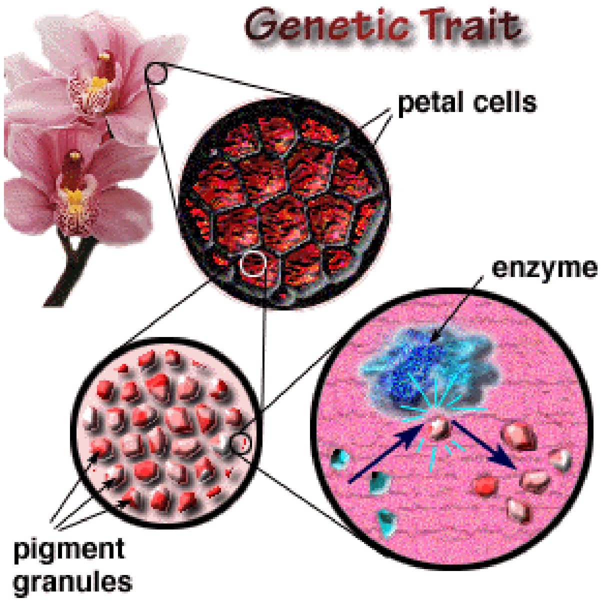
ChEBI

PO



Trait Ontology

- It’s an ontology of Plant traits (characteristics/observable)
- Precomposed trait terms
- Includes aspects of:
    - Anatomy trait
    - Growth and development (phenology) trait
    - Biochemical trait
    - Stress (abiotic and biotic stress) trait
    - Agronomic quality trait
    - Yield trait
    - Metabolomic trait
- Does not include phenotype term



Pre Composition Approach-1

```
[Term] id: TO:0000372 name: amylose to amylopectin ratio
namespace: plant_trait_ontology
intersection_of: PATO:0000025  !composition
intersection_of: towards CHEBI:28102  ! amylose
intersection_of: relative_to CHEBI:28057  ! amylopectin
```

```
[Term] id: TO:0000188 name: drought sensitivity
namespace: plant_trait_ontology
intersection_of: PATO:0000085  ! sensitivity
intersection_of: towards EO:0007404 ! drought environment
```

```
[Term] id: TO:0000227 name: root length
intersection_of: PATO:0000122  ! Length
intersection_of: inheres_in PO:0009005  ! root
```

Pre Composition Approach-1

Disease resistance phenotypes

[Term] id: TO:0000323 name: stem rot disease resistance

def: "Causal agent: Magnaporthe salvinii (Nakataea sigmoidea, Sclerotium oryzae), and Helminthosporium sigmoideum var. irregulare. Symptoms: dark lesions develop on the stems near the water line. Small, dark bodies (sclerotia) develop, weaken the stem and cause lodging."

*Easy to define if we had an orthogonal ontology of plant diseases and infectious agents. The definition pattern would be: A <TO:resistance> which *towards* <InfectiousAgent>*

Assay-specific terms

Example: root dry weight

It is unclear how to proceed with these. No action was taken for these.

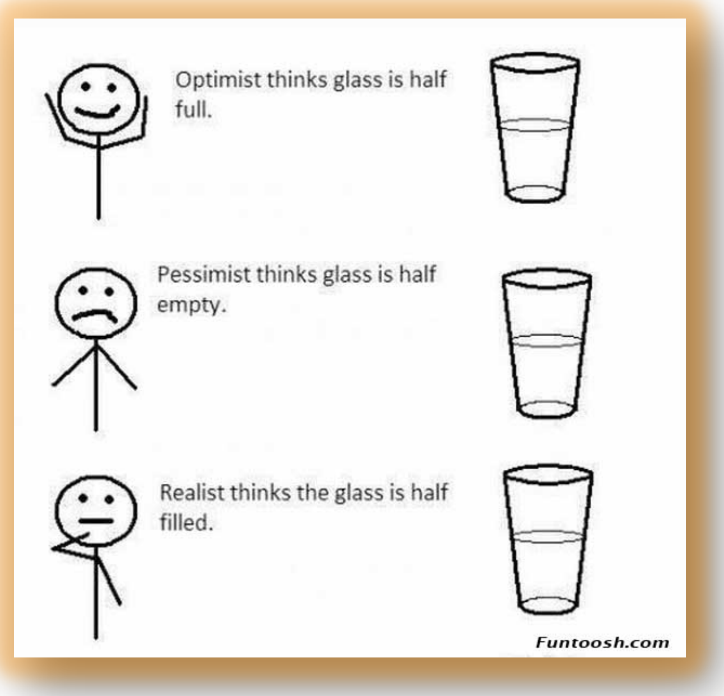
Conjunctive terms

Example: lemma and palea related traits

Example: lemma and palea pubescence

a pubescence which *inheres_in* the lemma and *inheres_in* the palea

Why TO Has No Observed Phenotype Values?



$$C = \begin{array}{c|cccccccccccccccc|}
 & A & B & C & D & E & F & G & H & I & J & K & L & M & N & O & P \\
A & 4 & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
B & 3 & 4 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 2 & 2 & 4 & 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
D & 2 & 2 & 3 & 4 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
E & 1 & 1 & 1 & 1 & 4 & 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
F & 1 & 1 & 1 & 1 & 3 & 4 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
G & 1 & 1 & 1 & 1 & 2 & 2 & 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
H & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 3 & 2 & 2 & 1 & 1 & 1 & 1 \\
J & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 2 & 2 & 1 & 1 & 1 & 1 \\
K & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 4 & 3 & 1 & 1 & 1 & 1 \\
L & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 3 & 4 & 1 & 1 & 1 & 1 \\
M & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 4 & 3 & 2 & 2 \\
N & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 3 & 4 & 2 & 2 \\
O & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 4 & 3 \\
P & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 4
\end{array}.$$

Scale-1	Scale-2	Scale-3	Scale-4
0	Resistant/tolerant (R)	Susceptible (S)	R/S
2-3	Moderate R	Moderate S	-
5	Normal (no change)	Normal	Moderate
7-8	Moderate S	Moderate R	
9	S	R	S/R

Do we benefit from TO ?

- **Standard EQ Terms**

The standard PT term is a pre-coordinated EQ term; this can easily be defined as:

A <Q> *which* inheres_in an <E>

i.e. a quality carried by a bearer entity

- **Benefits**

– PATO definitions of qualities can be shared and reused

– The oboedit reasoner can keep TO in sync with PO/GO/ChEBI, Taxonomy, EO and PATO, and perform automatic DAG placement of terms

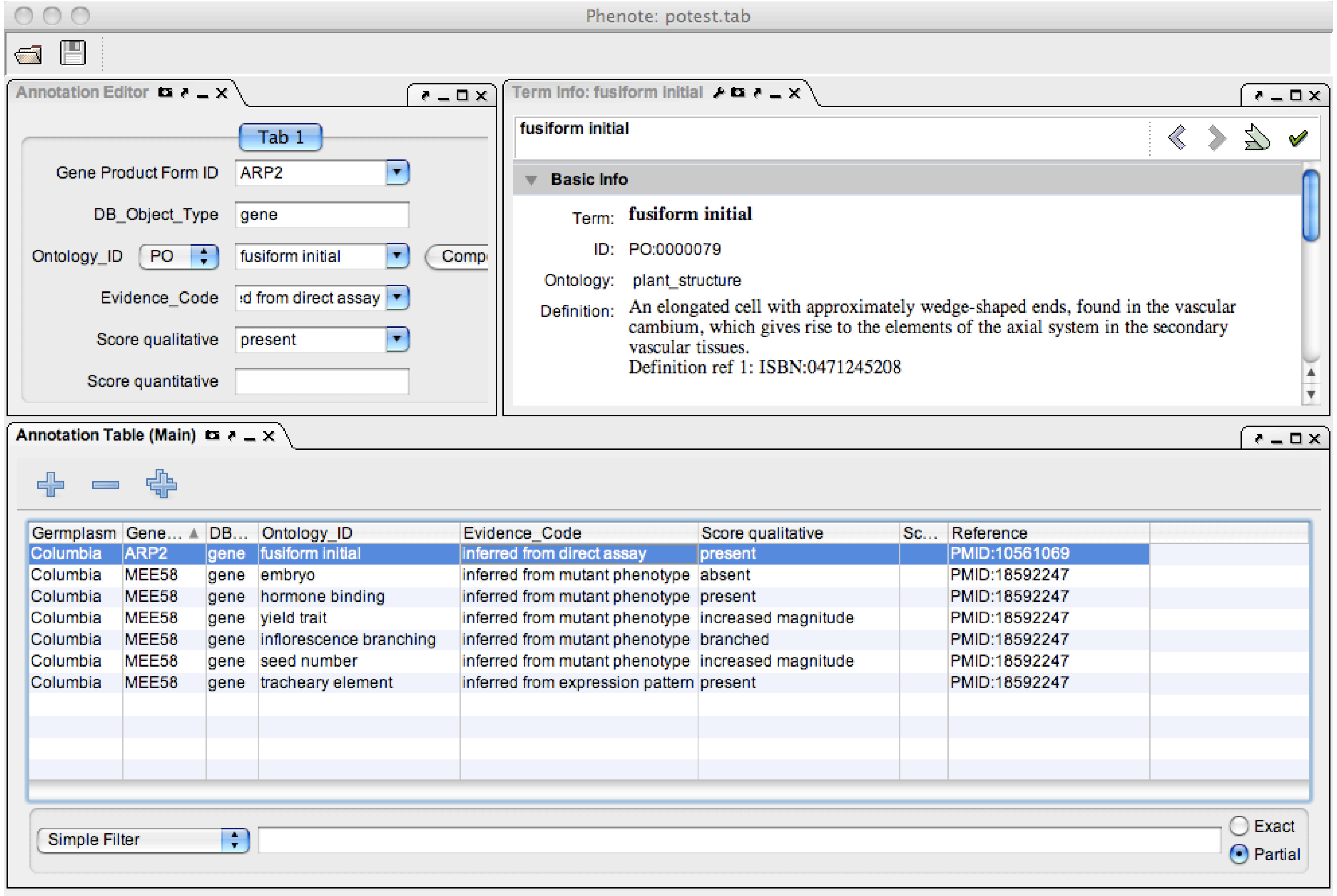
– PATO terms can be used to query TO-annotated phenotypes

– PO/GO/ChEBI, Taxonomy, EO terms can be used to query PT-annotated phenotypes

– Users/curators do not have to deal with post composition of terms with about 100K entity terms from different knowledge domains.

Phenote Annotation tool

With Two Configurations : Advanced and ‘Smart’



Gene Product Form ID: ARP2
DB_Object_Type: gene
Ontology_ID: PO fusiform initial
Evidence_Code: :d from direct assay
Score qualitative: present
Score quantitative:

fusiform initial

Basic Info

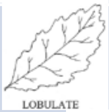
Term: **fusiform initial**
ID: PO:0000079
Ontology: plant_structure
Definition: An elongated cell with approximately wedge-shaped ends, found in the vascular cambium, which gives rise to the elements of the axial system in the secondary vascular tissues.
Definition ref 1: ISBN:0471245208

Germplasm	Gene... ▲	DB...	Ontology_ID	Evidence_Code	Score qualitative	Sc...	Reference
Columbia	ARP2	gene	fusiform initial	inferred from direct assay	present		PMID:10561069
Columbia	MEE58	gene	embryo	inferred from mutant phenotype	absent		PMID:18592247
Columbia	MEE58	gene	hormone binding	inferred from mutant phenotype	present		PMID:18592247
Columbia	MEE58	gene	yield trait	inferred from mutant phenotype	increased magnitude		PMID:18592247
Columbia	MEE58	gene	inflorescence branching	inferred from mutant phenotype	branched		PMID:18592247
Columbia	MEE58	gene	seed number	inferred from mutant phenotype	increased magnitude		PMID:18592247
Columbia	MEE58	gene	tracheary element	inferred from expression pattern	present		PMID:18592247

○ Exact
● Partial

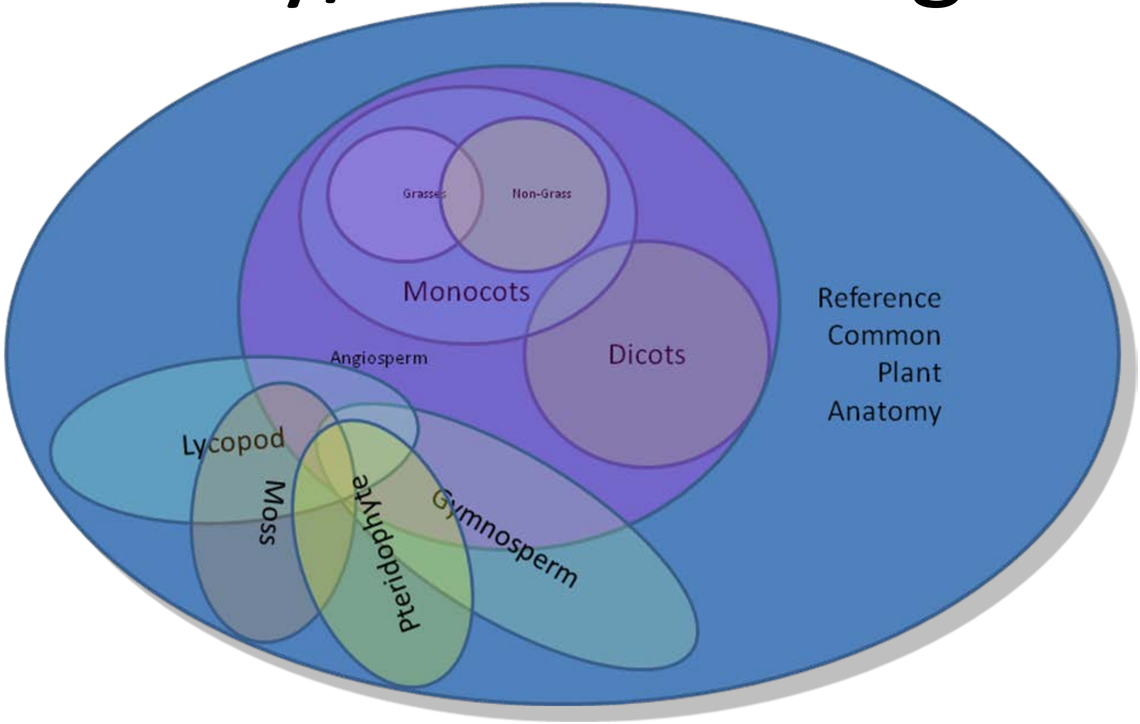
Custom Phenote tool (phenote.org) allows inputs by authors. Advanced configuration allows savvy users to build cross-products for phenotype annotations.

Creating Character Matrices

PO: Leaf					
TO: Leaf lamina shape (PO: Leaf lamina x PATO: shape)			**TO: Leaf lamina margin quality** (PO: Lamina margin x PATO: quality)		
Lyrate	☐		Lacerate	☐	
Linear	☐		Laciniate	☐	
Lanceolate	☐		Lobed	☐	
Hastate	☐		Lobulate	☒	
Deltoid	☒		Dentate	☐	

Sample: xxxxxxxxx
Species: xxxxxxxxx
Source: xxxxxxxxxx

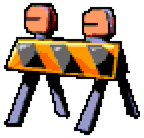
Do we need multiple Plant anatomy/trait ontologies?



Create various version and sub-versions of the PO that may cater to two aspects of the Ontology.

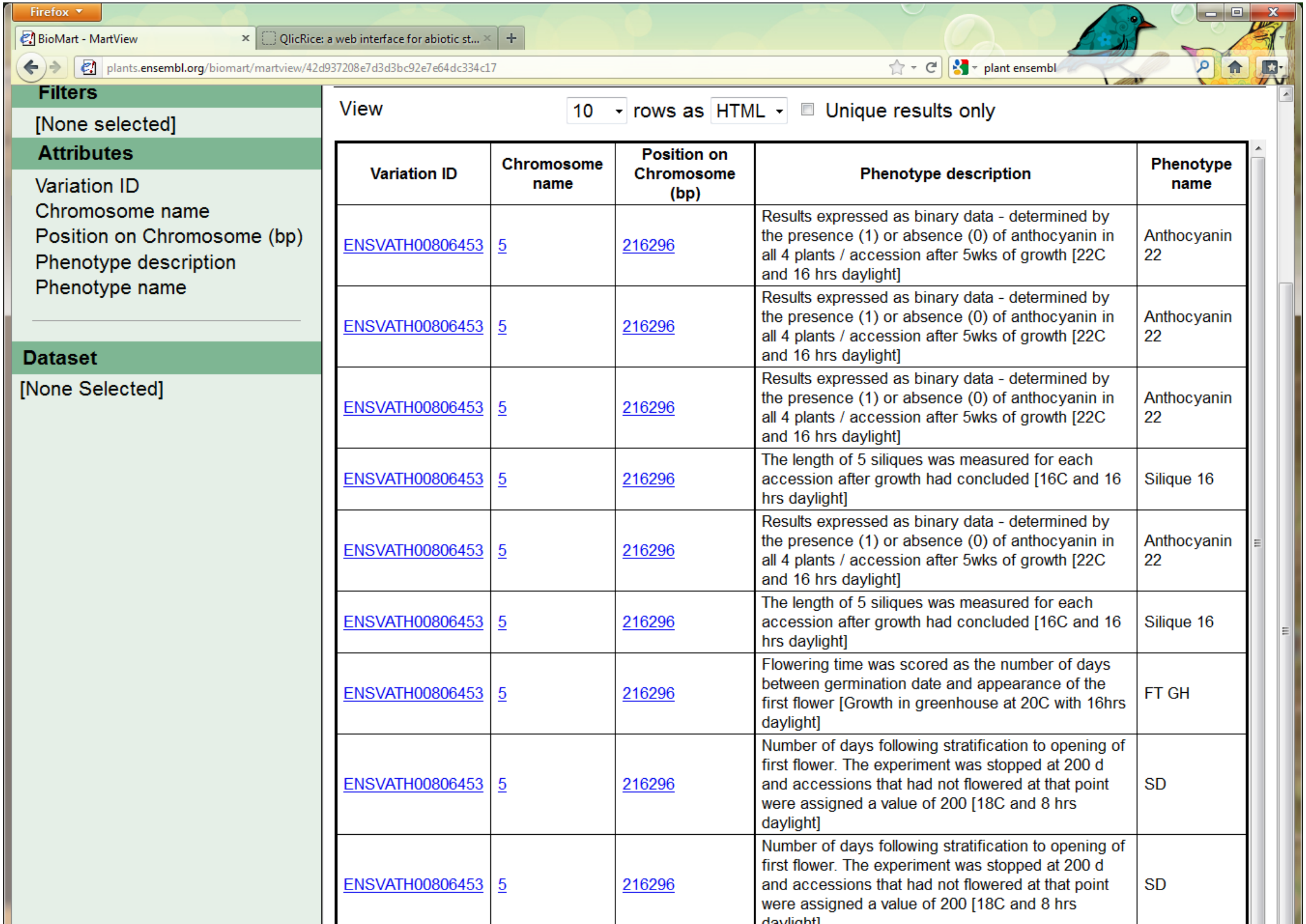
- Common Reference Ontology for Plants (CROP)
- Clade-specific Enriched Ontologies (CLEO)

Challenges



- Encourage species-specific vocabularies to use cross references to PO and TO terms
- Share annotations for robust comparison and hypothesis building
- Shortage of plant-specific phenotypic descriptors in PATO
- Add mapping files for clade oriented trait/phenotype glossaries
- Build consensus on metadata and minimal information guide

Capturing Phenotype Metadata



Filters

[None selected]

Attributes

Variation ID
Chromosome name
Position on Chromosome (bp)
Phenotype description
Phenotype name

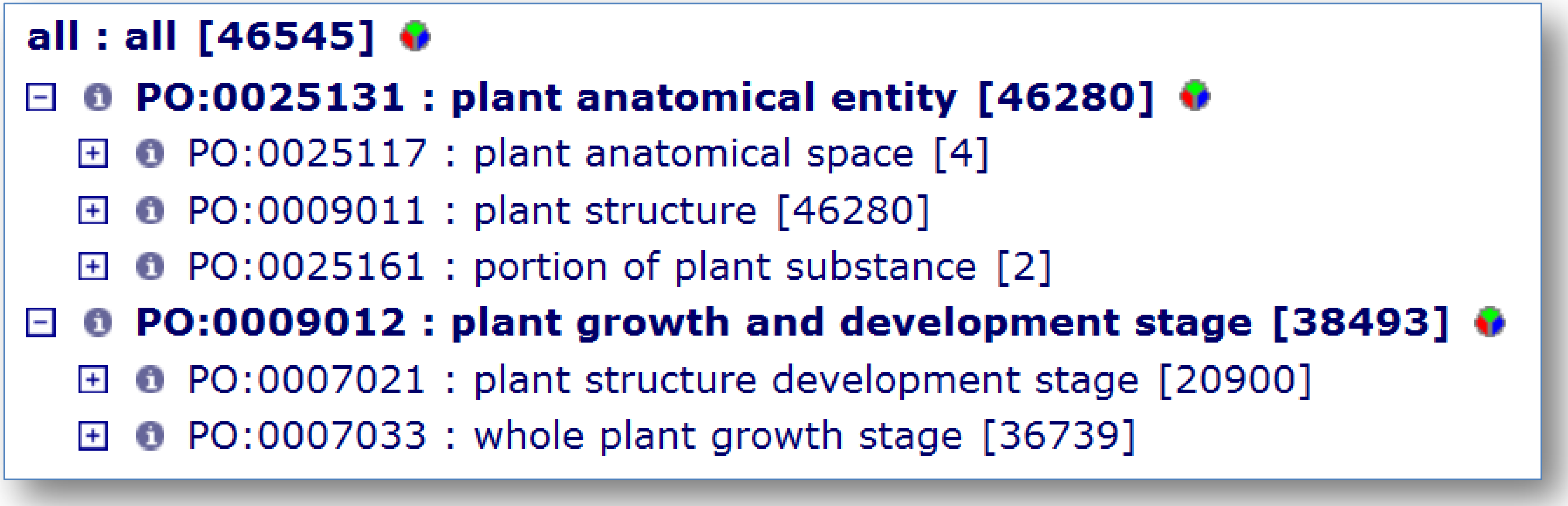
Dataset

[None Selected]

View 10 rows as HTML ☐ Unique results only

Variation ID	Chromosome name	Position on Chromosome (bp)	Phenotype description	Phenotype name
ENSVATH00806453	5	216296	Results expressed as binary data - determined by the presence (1) or absence (0) of anthocyanin in all 4 plants / accession after 5wks of growth [22C and 16 hrs daylight]	Anthocyanin 22
ENSVATH00806453	5	216296	Results expressed as binary data - determined by the presence (1) or absence (0) of anthocyanin in all 4 plants / accession after 5wks of growth [22C and 16 hrs daylight]	Anthocyanin 22
ENSVATH00806453	5	216296	Results expressed as binary data - determined by the presence (1) or absence (0) of anthocyanin in all 4 plants / accession after 5wks of growth [22C and 16 hrs daylight]	Anthocyanin 22
ENSVATH00806453	5	216296	The length of 5 siliques was measured for each accession after growth had concluded [16C and 16 hrs daylight]	Silique 16
ENSVATH00806453	5	216296	Results expressed as binary data - determined by the presence (1) or absence (0) of anthocyanin in all 4 plants / accession after 5wks of growth [22C and 16 hrs daylight]	Anthocyanin 22
ENSVATH00806453	5	216296	The length of 5 siliques was measured for each accession after growth had concluded [16C and 16 hrs daylight]	Silique 16
ENSVATH00806453	5	216296	Flowering time was scored as the number of days between germination date and appearance of the first flower [Growth in greenhouse at 20C with 16hrs daylight]	FT GH
ENSVATH00806453	5	216296	Number of days following stratification to opening of first flower. The experiment was stopped at 200 d and accessions that had not flowered at that point were assigned a value of 200 [18C and 8 hrs daylight]	SD
ENSVATH00806453	5	216296	Number of days following stratification to opening of first flower. The experiment was stopped at 200 d and accessions that had not flowered at that point were assigned a value of 200 [18C and 8 hrs daylight]	SD

Plant Ontology



all : all [46545]
- ⊟ **PO:0025131 : plant anatomical entity [46280]**
  - ⊞ PO:0025117 : plant anatomical space [4]
  - ⊞ PO:0009011 : plant structure [46280]
  - ⊞ PO:0025161 : portion of plant substance [2]
- ⊟ **PO:0009012 : plant growth and development stage [38493]**
  - ⊞ PO:0007021 : plant structure development stage [20900]
  - ⊞ PO:0007033 : whole plant growth stage [36739]

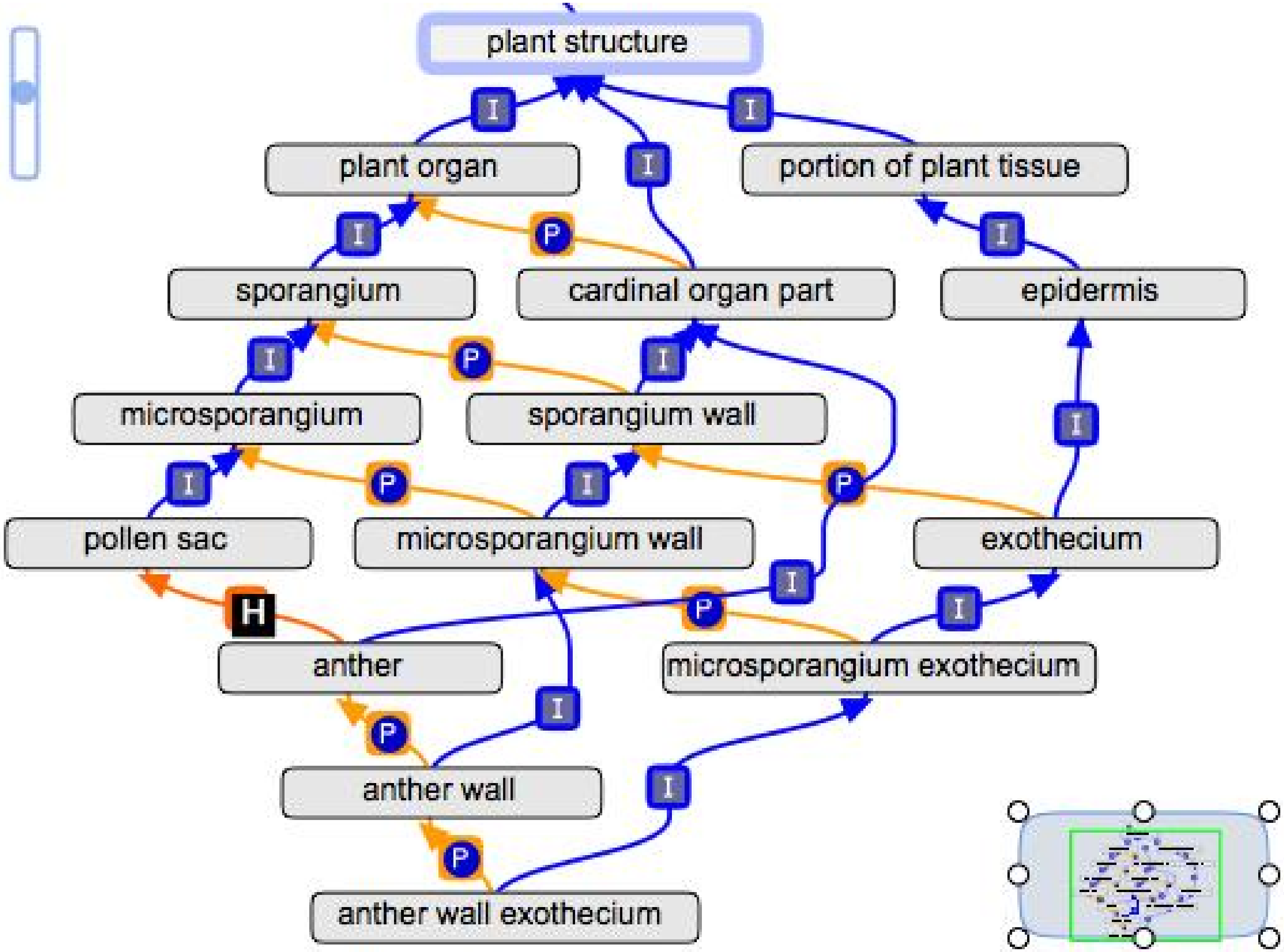
- Has two classes for **Plant Anatomical entities** and **Growth and Development**
- Every term has definition
- Every term has is_a relationship as mandated to have single inheritance (some terms may have multiple for enriched biological reasoning)
- Use multiple relationship types
  - is_a, part_of, has_part, adjacent_to, develops_from, derived_from, participates_in

Vision for Plant Ontology:

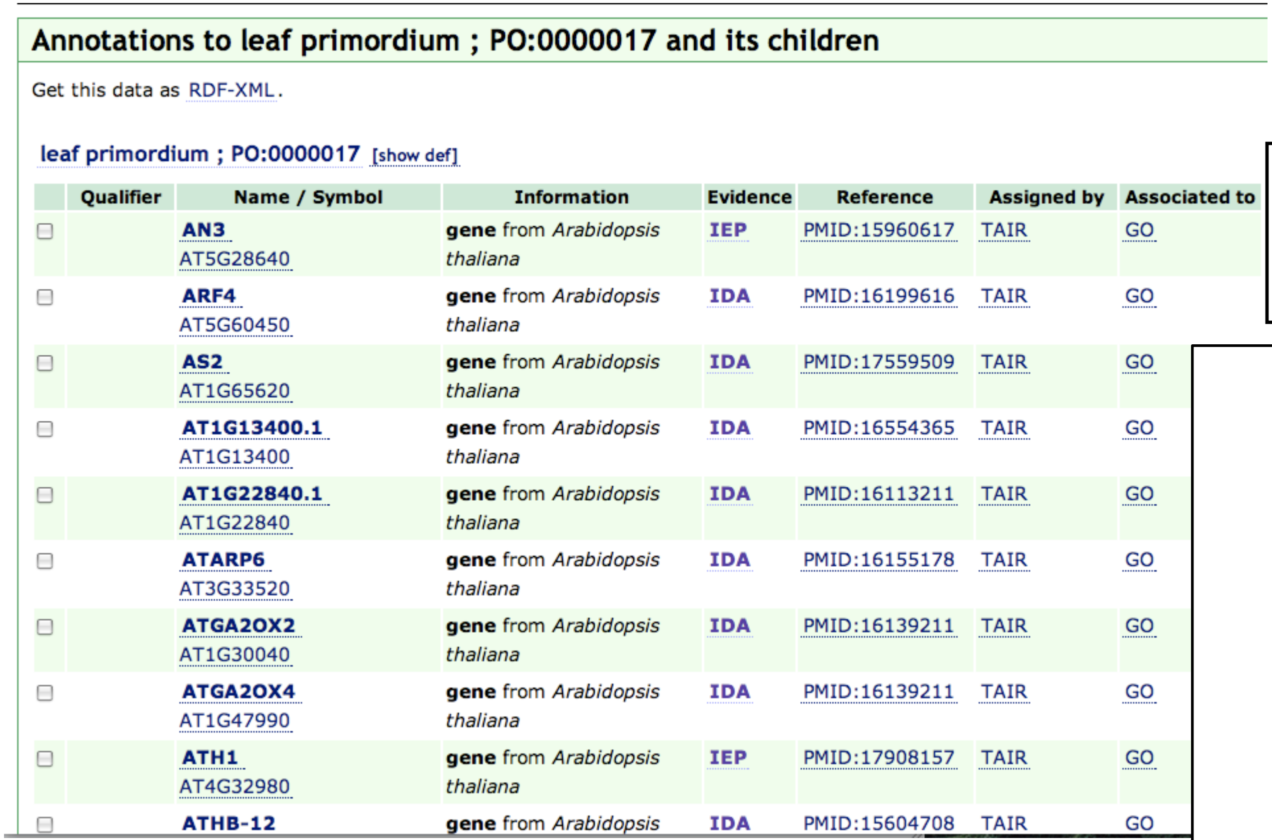
- Encompass **all** plants
- Facilitate consistency in:
    - Annotation of comparative genomics data
    - Cross-database queries
- Develop PO as a reference ontology for plants:
    - Provide mappings to other ontologies in use by plant databases
- Create cross-products to other ontologies such as PATO, to describe phenotypic characters

Source: Laurel Cooper

An example from PO Anatomical Entity



The Plant Ontology facilitates comparative plant genomics, developmental biology and systematics



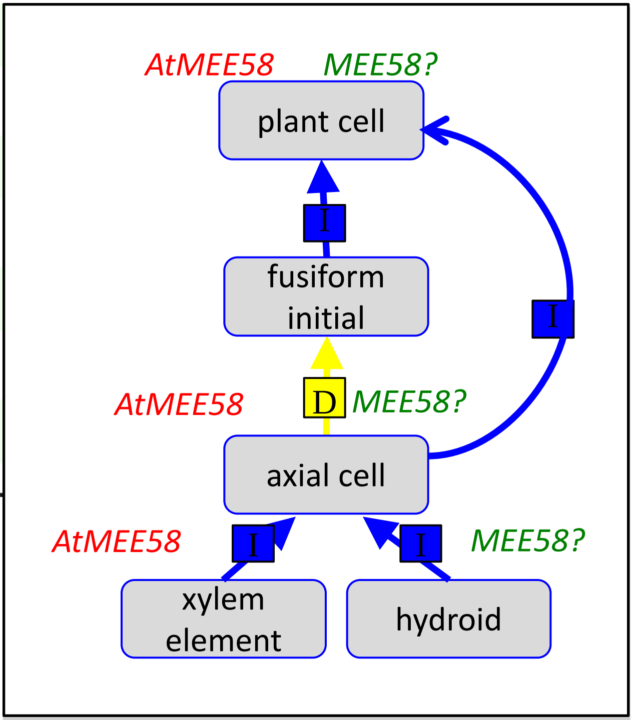
Annotations to leaf primordium ; PO:0000017 and its children

Get this data as RDF-XML.

leaf primordium ; PO:0000017 [show def]

	Qualifier	Name / Symbol	Information	Evidence	Reference	Assigned by	Associated to
☐		**AN3** AT5G28640	**gene** from *Arabidopsis thaliana*	IEP	PMID:15960617	TAIR	GO
☐		**ARF4** AT5G60450	**gene** from *Arabidopsis thaliana*	IDA	PMID:16199616	TAIR	GO
☐		**AS2** AT1G65620	**gene** from *Arabidopsis thaliana*	IDA	PMID:17559509	TAIR	GO
☐		**AT1G13400.1** AT1G13400	**gene** from *Arabidopsis thaliana*	IDA	PMID:16554365	TAIR	GO
☐		**AT1G22840.1** AT1G22840	**gene** from *Arabidopsis thaliana*	IDA	PMID:16113211	TAIR	GO
☐		**ATARP6** AT3G33520	**gene** from *Arabidopsis thaliana*	IDA	PMID:16155178	TAIR	GO
☐		**ATGA2OX2** AT1G30040	**gene** from *Arabidopsis thaliana*	IDA	PMID:16139211	TAIR	GO
☐		**ATGA2OX4** AT1G47990	**gene** from *Arabidopsis thaliana*	IDA	PMID:16139211	TAIR	GO
☐		**ATH1** AT4G32980	**gene** from *Arabidopsis thaliana*	IEP	PMID:17908157	TAIR	GO
☐		**ATHB-12**	**gene** from *Arabidopsis*	IDA	PMID:15604708	TAIR	GO

Cross-taxa comparisons:



>500,000 associations for >1300 terms

Source: Laurel Cooper

Plant Ontology
Taxon Relevance, Synonyms and Xrefs

subsetdef: Angiosperm "Term for angiosperms"

subsetdef: Arabidopsis "Term used for Arabidopsis"

subsetdef: Bryophytes "Term used for mosses, liverworts, and/or hornworts"

subsetdef: Citrus "Term used for citrus"

subsetdef: Gymnosperms "Term used for gymnosperms"

subsetdef: Maize "Term used for maize"

subsetdef: Musa "Terms used for banana"

subsetdef: Poaceae "Term used for grasses"

subsetdef: Potato "Term used for potato“

subsetdef: Solanaceae "Term used for solanaceous plants“

subsetdef: Pteridophytes "Term used for ferns and allies"

subsetdef: reference "reference plant structure term"

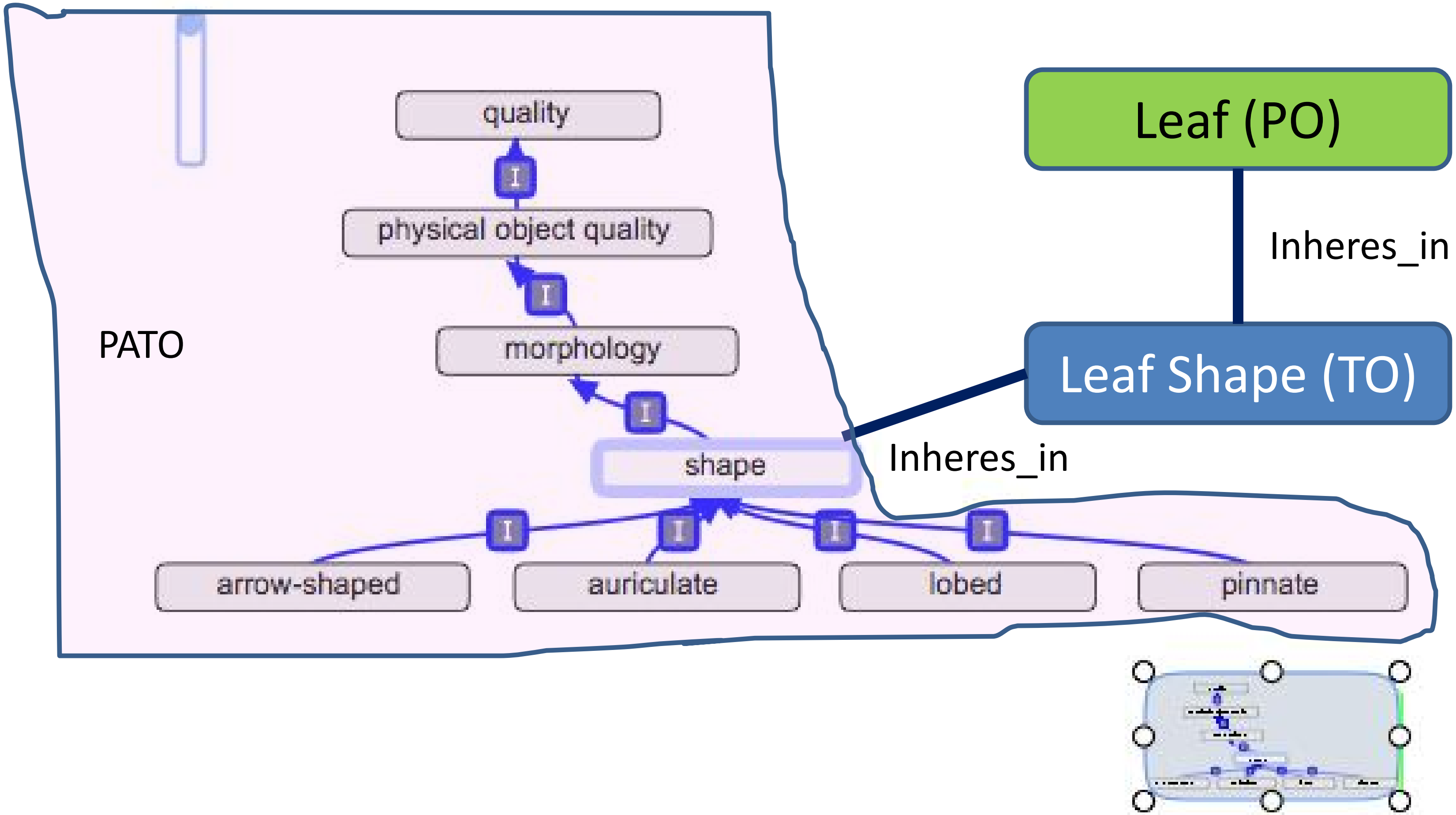
subsetdef: Rice "Term used for rice"

subsetdef: Tomato "Term used for tomato"

subsetdef: TraitNet "Plant Functional Traits“

```
synonymtypedef: Japanese "Japanese synonym" EXACT
synonymtypedef: Plural "Plural" EXACT
synonymtypedef: Spanish "Spanish synonym" EXACT
treat-xrefs-as-is_a: CARO
dbxrf: CL:Cell_type | GO:cellular_component | APweb
```

Examples of shape qualities that are relevant to leaves



PATO

Leaf (PO)

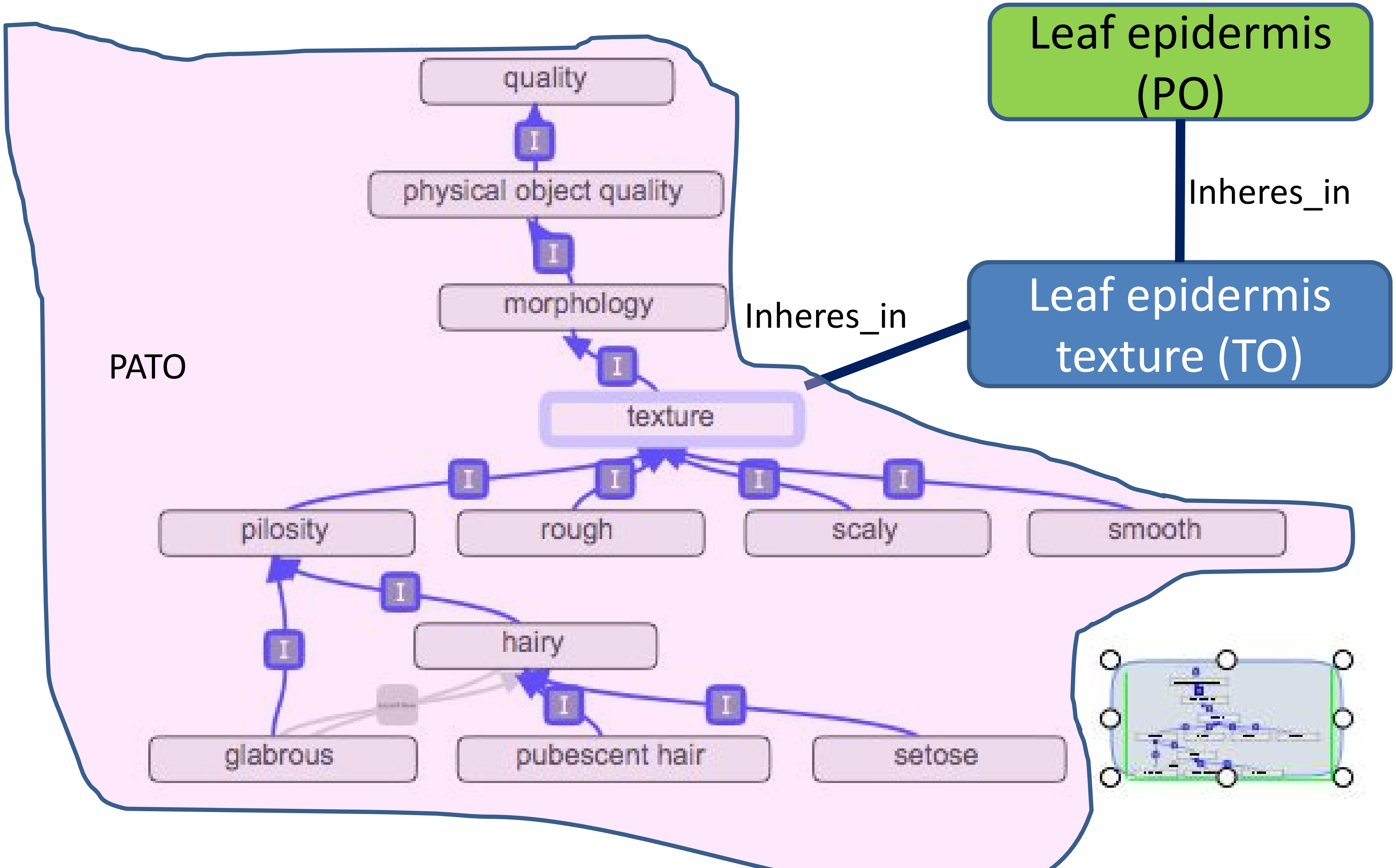
Inheres_in

Leaf Shape (TO)

Inheres_in

Source: Ramona Walls

Examples of textures that are relevant to leaves



PATO

Leaf epidermis (PO)

Inheres_in

Inheres_in

Leaf epidermis texture (TO)

Source: Ramona Walls