

POC tutorial #2: Ontology Development

This tutorial will run automatically in Quicktime.
To run the tutorial at your own pace use the
internal controllers within the tutorial.

-  button goes to next slide
-  button goes to previous slide
-  button goes to last slide
-  button goes to previous slide

Ontology Development

- A. What are the organizing principles?**
- B. How are terms defined?**
- C. How are terms related to each other?**
- D. What is a directed acyclic graph (DAG)?**
- E. Elements and attributes of terms**
- F. What is the “True Path Rule” ?**
- G. Species specificity: problems and solutions**
- H. How are ontologies maintained?**

What are the organizing principles?

Keep it simple: strive for a robust extensible structure, rather than comprehensiveness.

Where possible, rely on synonyms (equivalence of terms) rather than creating a new term.

The criterion for creating an anatomy term include: location, morphology, derivation and spatial/positional organization.

Include species specific terminology to accommodate annotation and biological accuracy (i.e. maintain the true path rule).

All terms must be defined.

How are terms defined?

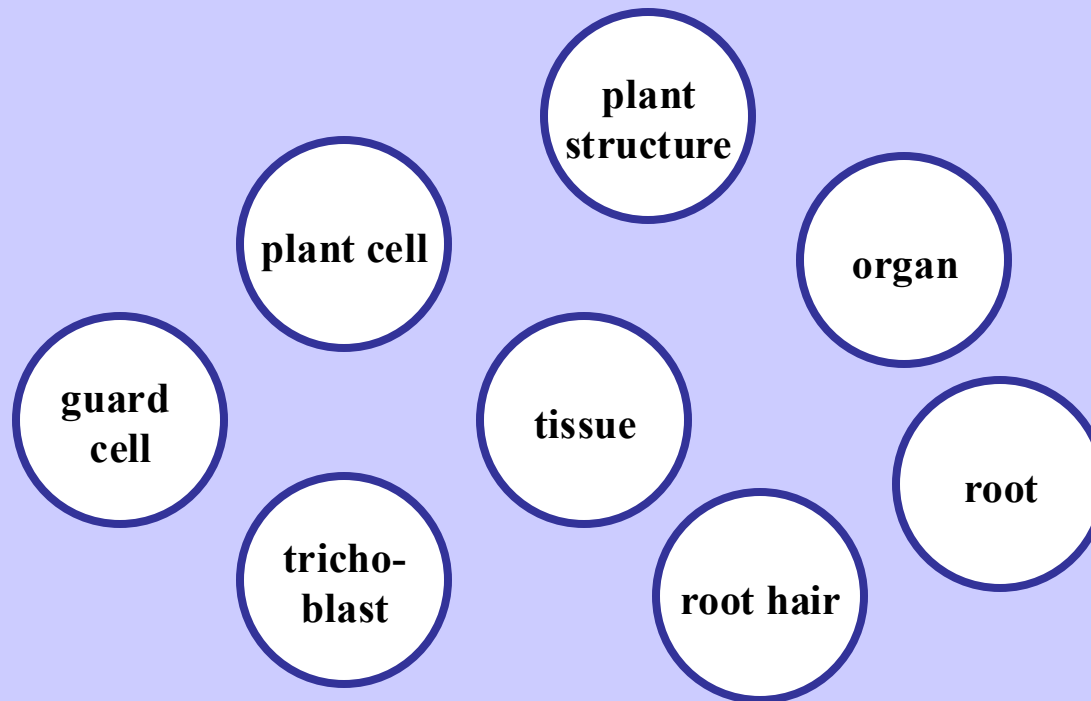
The precise definition of terms is critical to the integrity of the ontologies.

Definitions are obtained primarily from standard references such as textbooks and glossaries. Definitions may be taken verbatim from references or modified for clarity or to reflect common usage. Most definitions come from [Plant Anatomy](#) (K.Esau) and the [Angiosperm Phylogeny website](#) Missouri Botanical Garden).

How are terms related to each other?

**Terms are related to each other as children to parents.
Each child term can have one or more parents.**

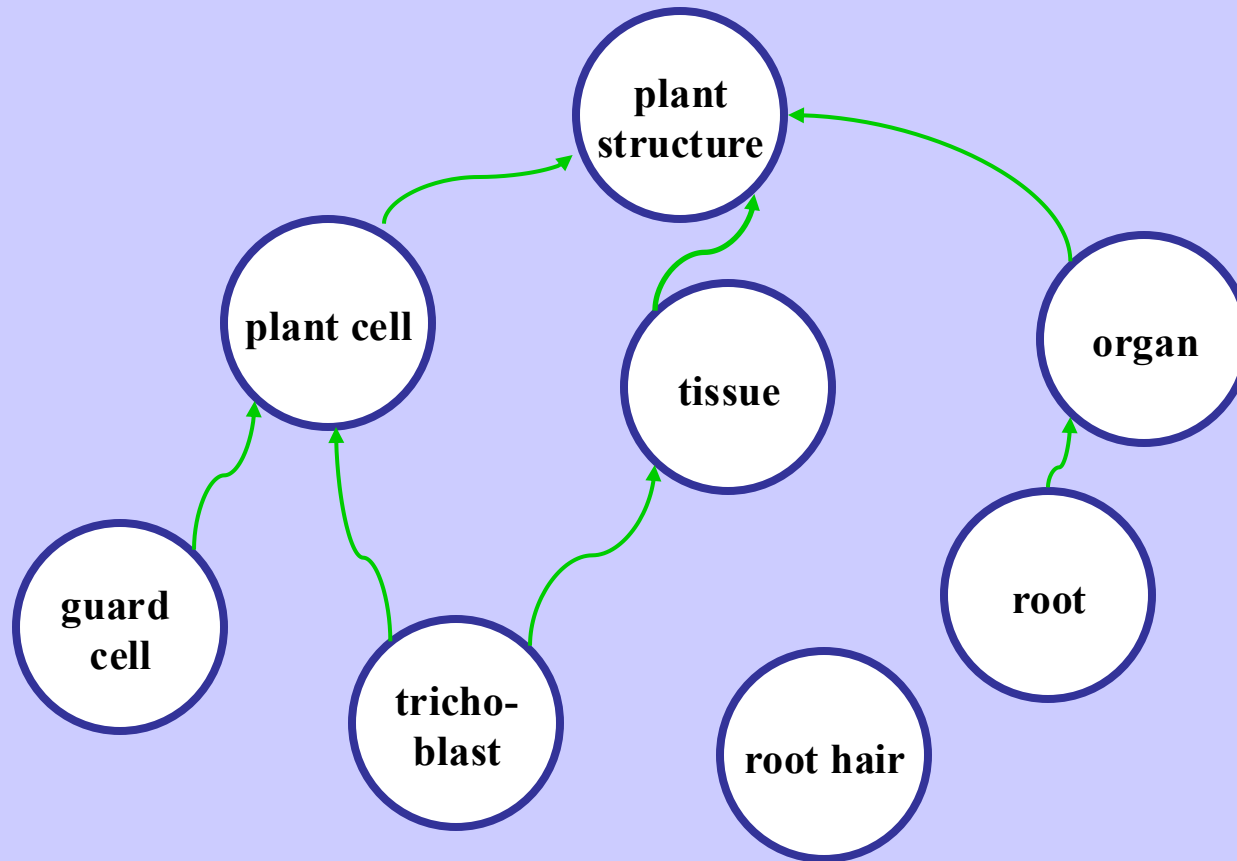
There are three basic types of child-parent relationships used in the plant ontologies, which are illustrated in the following graph.



The **is a** relationship

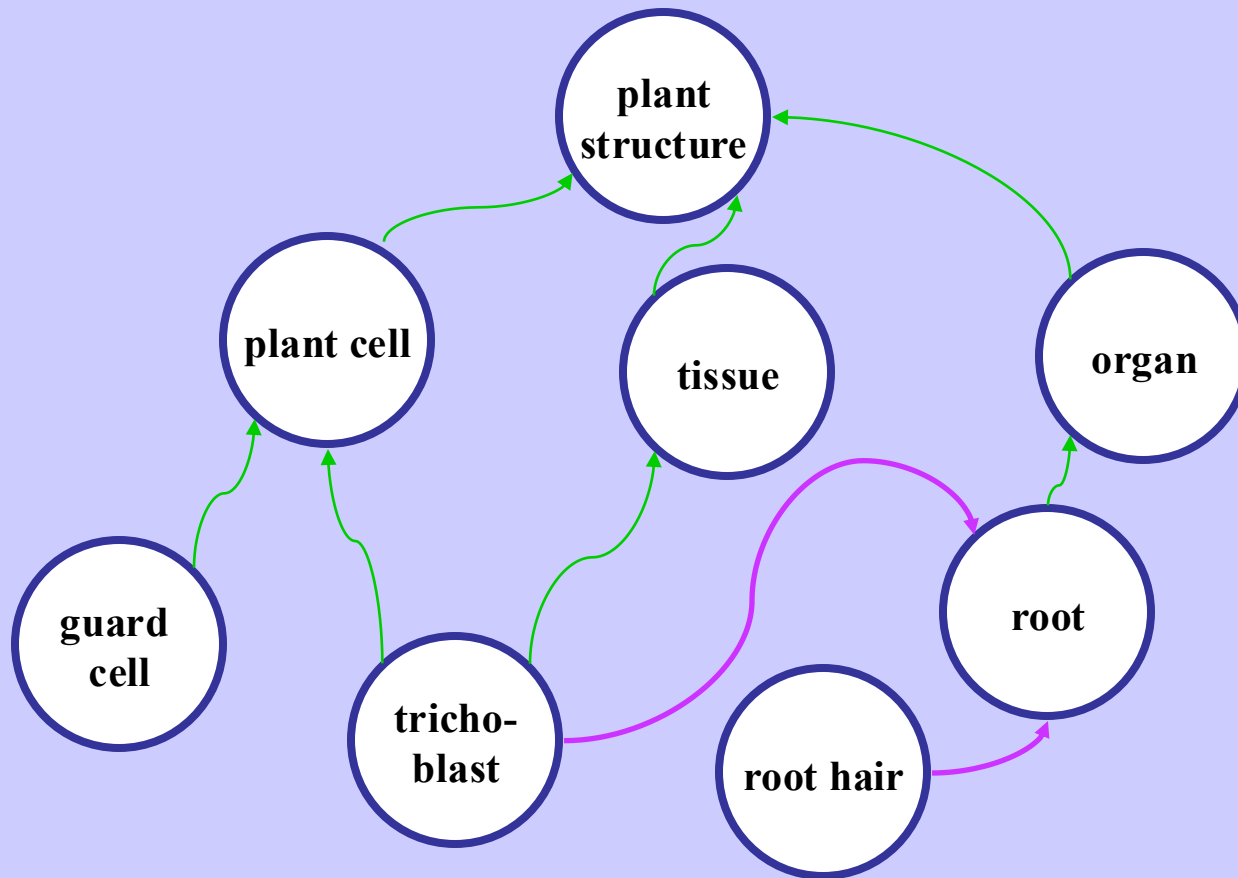
is a is a simple class-subclass relationship.

For example, a trichoblast **is a** plant cell which **is a** plant structure. A root **is a** organ which **is a** plant structure.



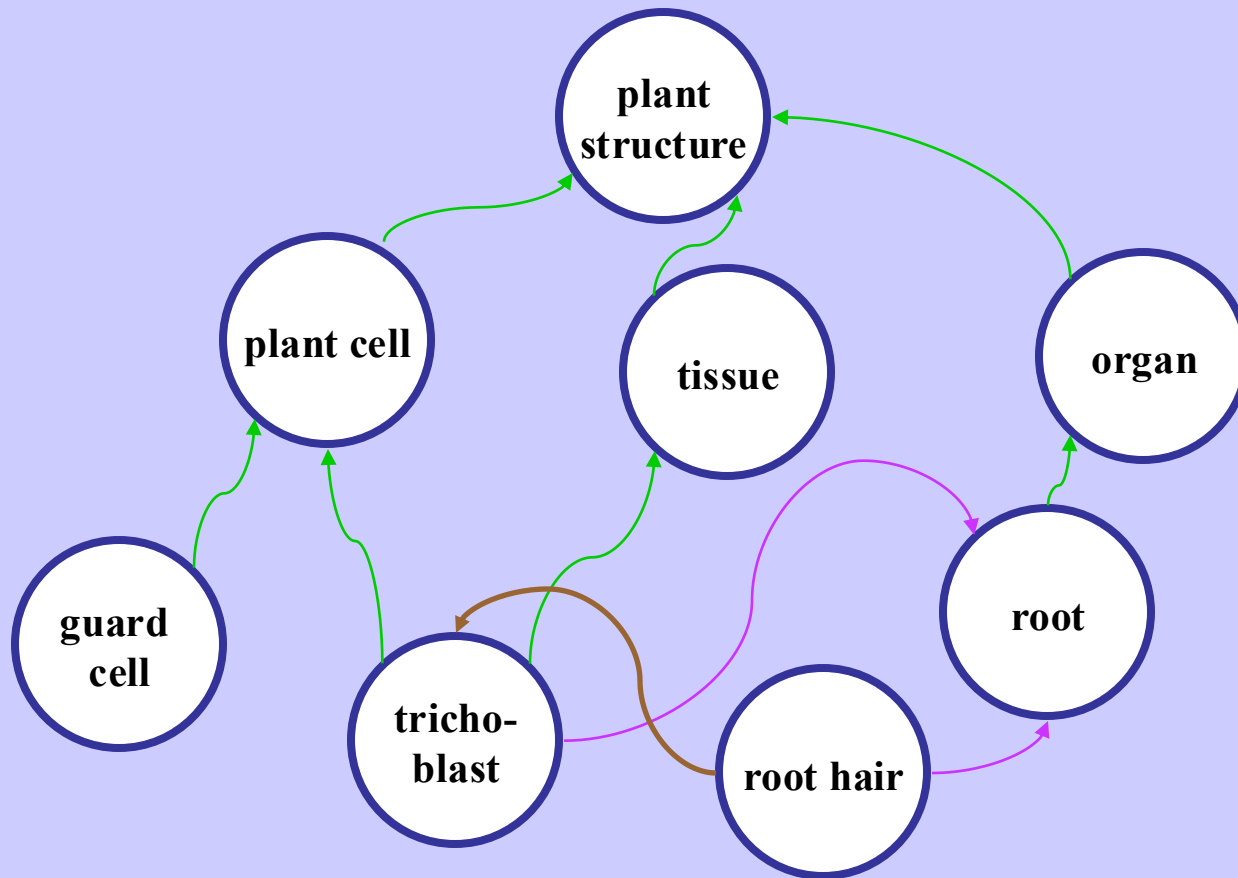
The **part of** Relationship

It indicates a subpart/part relationship within a tissue or organ. Used in a non-restrictive manner. An example would be root hair **part of** root; root hair is always part of a root, but not all roots have root hair.



The **develops from** Relationship

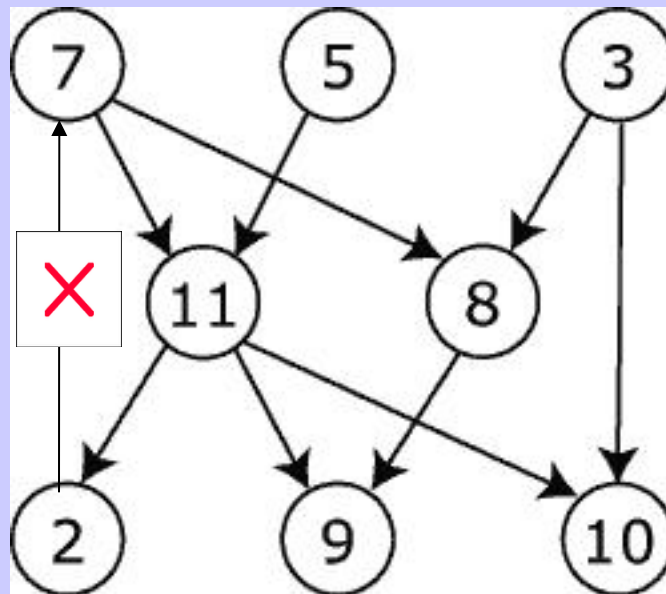
It indicates that cell/tissue/organ develops from its parent term. Implies both, develops from and a more indirect relationship, derive from. For example, the root hair **develops from** the trichoblast which is a plant cell which is a plant structure.



What is a directed acyclic graph (DAG)?

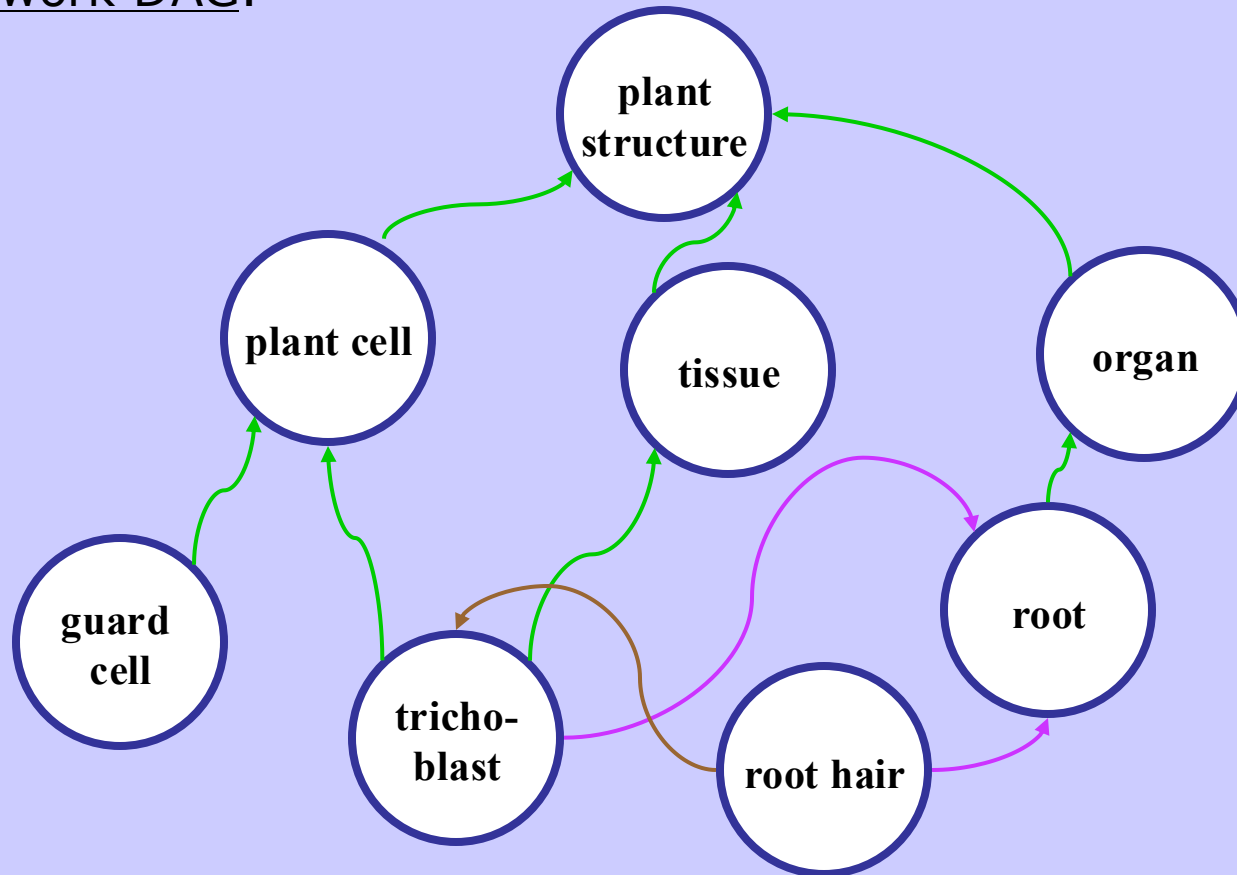
A DAG is a collection of ordered nodes (e.g. parent-child) and edges (e.g. relationships) that flows in a specific direction. In the ontologies, nodes are terms.

A path through the nodes cannot cycle, or double back on itself.



If every child node has no more than one parent node, then the DAG is a tree. If at least one child node has two parents, the DAG is a network.

The plant ontology, like the Gene Ontology can be represented as a network DAG.



Tree view in "AmiGo browser"

```

+ all ( 15814 )
  + PO:0009011 : plant structure ( 15734 )
    + PO:0009008 : organ ( 6129 )
      + PO:0009005 : root ( 2127 )
        + PO:0006036 : root epidermis ( 61 )
          + PO:0000256 : root hair ( 34 )
          + PO:0000262 : trichoblast ( 34 )
          + PO:0000256 : root hair ( 34 )
        + PO:0009002 : plant cell ( 294 )
          + PO:0004013 : epidermal cell ( 174 )
            + PO:0000256 : root hair ( 34 )
            + PO:0000262 : trichoblast ( 34 )
            + PO:0000256 : root hair ( 34 )
          + PO:0004010 : meristematic cell ( 96 )
        + PO:0009003 : sporophyte ( 13200 )
          + PO:0009005 : root ( 2127 )
            + PO:0006036 : root epidermis ( 61 )
              + PO:0000256 : root hair ( 34 )
              + PO:0000262 : trichoblast ( 34 )
              + PO:0000256 : root hair ( 34 )
          + PO:0009007 : tissue ( 1571 )
            + PO:0009014 : dermal tissue ( 277 )
              + PO:0005679 : epidermis ( 277 )
                + PO:0004013 : epidermal cell ( 174 )
                  + PO:0000256 : root hair ( 34 )
                  + PO:0000262 : trichoblast ( 34 )
                  + PO:0000256 : root hair ( 34 )
                + PO:0006036 : root epidermis ( 61 )
                  + PO:0000256 : root hair ( 34 )
                  + PO:0000262 : trichoblast ( 34 )
                  + PO:0000256 : root hair ( 34 )

```

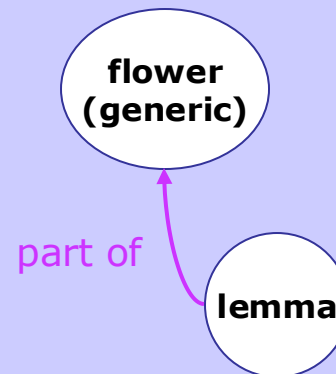
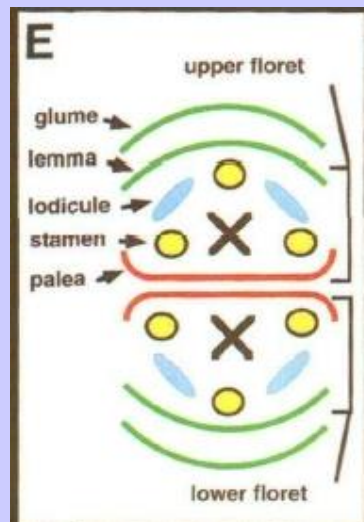
What is the true path rule?

The true path rule states that the path from any node (term) all the way to the top node of the tree must be biologically correct.

When violations of the true path rule are detected the structure of the ontology must be modified.

Example: Maize lemmas

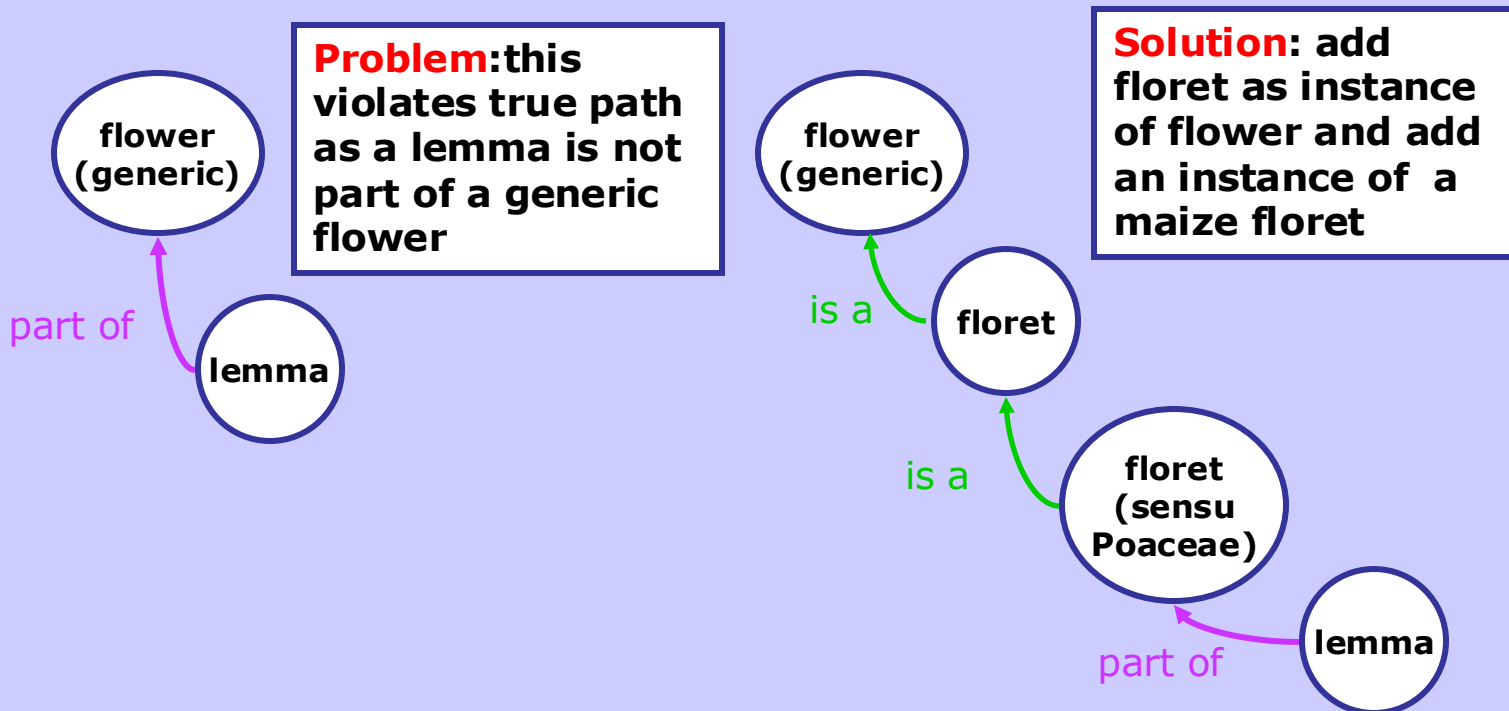
For example, a lemma is a type of bract that is a part of a maize floret but is not present in other flowers.



Schematic diagram of male florets of maize. Veit et.al. Plant Cell. 1993 Oct;5(10):1205-1215

Maintaining the true path rule

Lemmas are not present in all flowers- therefore it is necessary to create a special instance of a flower - specifically a maize floret.



How does this affect queries?

The path to each parent is true. A query of all genes affecting the generic flower would still return genes affecting the lemma of maize floret. It is possible to find all flower mutations in maize without explicit knowledge of maize-specific terms such as lemma.

```
⊞ all : all ( 15814 )
  ⊞ ① PO:0009011 : plant structure ( 15734 )
    ⊞ ① PO:0009008 : organ ( 6129 )
      ⊞ ① PO:0009034 : floral bract ( 86 )
        ⊞ ① PO:0009037 : lemma ( 66 )
      ⊞ ① PO:0006001 : phyllome ( 4697 )
        ⊞ ① PO:0009055 : bract ( 107 )
          ⊞ ① PO:0009034 : floral bract ( 86 )
            ⊞ ① PO:0009037 : lemma ( 66 )
      ⊞ ① PO:0009003 : sporophyte ( 13200 )
        ⊞ ① PO:0009006 : shoot ( 8551 )
          ⊞ ① PO:0009049 : inflorescence ( 4291 )
            ⊞ ① PO:0009034 : floral bract ( 86 )
              ⊞ ① PO:0009037 : lemma ( 66 )
            ⊞ ① PO:0009046 : flower ( 1605 )
              ⊞ ① PO:0009082 : floret ( 228 )
                ⊞ ① PO:0006318 : floret (sensu Poaceae) ( 228 )
                  ⊞ ① PO:0009037 : lemma ( 66 )
            ⊞ ① PO:0009051 : spikelet ( 1036 )
              ⊞ ① PO:0006318 : floret (sensu Poaceae) ( 228 )
                ⊞ ① PO:0009037 : lemma ( 66 )
          ⊞ ① PO:0006001 : phyllome ( 4697 )
            ⊞ ① PO:0009055 : bract ( 107 )
              ⊞ ① PO:0009034 : floral bract ( 86 )
                ⊞ ① PO:0009037 : lemma ( 66 )
```

Representation of lemma in the plant structure ontology

Elements and attributes of terms

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement in which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

The following section defines the attributes of terms as they are shown in the AmiGO browser. Here, we show the term "inflorescence".

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

① PO:0009006 : shoot

① PO:0009049 : inflorescence

Relationship Types

① is a

① part of

① develops from

Accession

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement in which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

Each term has a unique identifier of that term.

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

① PO:0009006 : shoot

① PO:0009049 : inflorescence

Relationship Types

① is a

① part of

① develops from

Aspect

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement in which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

This refers to the aspect of the Plant Ontology (structure or developmental stage) that includes the term.

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

① PO:0009006 : shoot

① PO:0009049 : inflorescence

Relationship Types

① is a

① part of

① develops from

Synonyms

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement in which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

The synonyms include a variety of alternate forms of the term such as variations, broader/narrower terms, misnomers and equivalent terms.

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

① PO:0009006 : shoot

① PO:0009049 : inflorescence

Relationship Types

① is a

① part of

① develops from

Definition

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement in which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

Definition of the term as used in the Plant Ontologies. Definitions are primarily obtained from text books and glossaries.

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

① PO:0009006 : shoot

① PO:0009049 : inflorescence

Relationship Types

① is a

① part of

① develops from

Comments

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement in which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

② PO:0009006 : shoot

② PO:0009049 : inflorescence

**Comments by
curators/developers to provide
clarity or additional information
such as usage.**

Relationship Types

① is a

② part of

③ develops from

Lineage

inflorescence

Accession: PO:0009049

Aspect: plant structure

Synonyms:

cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster

Definition:

The grouping or arrangement of flowers which flowers are borne on a plant.

Comment:

Some plants have only solitary flowers, e.g. Magnolia.

Term Lineage

PO:0009075 : plant ontology

① PO:0009011 : plant structure

① PO:0009003 : sporophyte

② PO:0009006 : shoot

② PO:0009049 : inflorescence

The diagram shows the relationship of the term to all of its parents.

Relationship Types

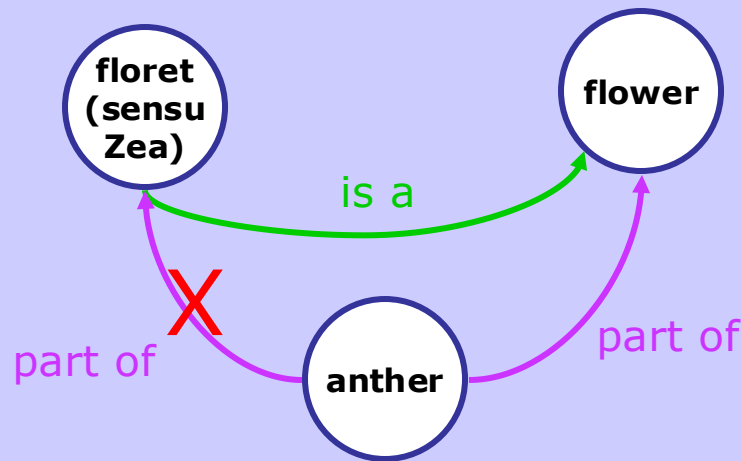
① is a

② part of

③ develops from

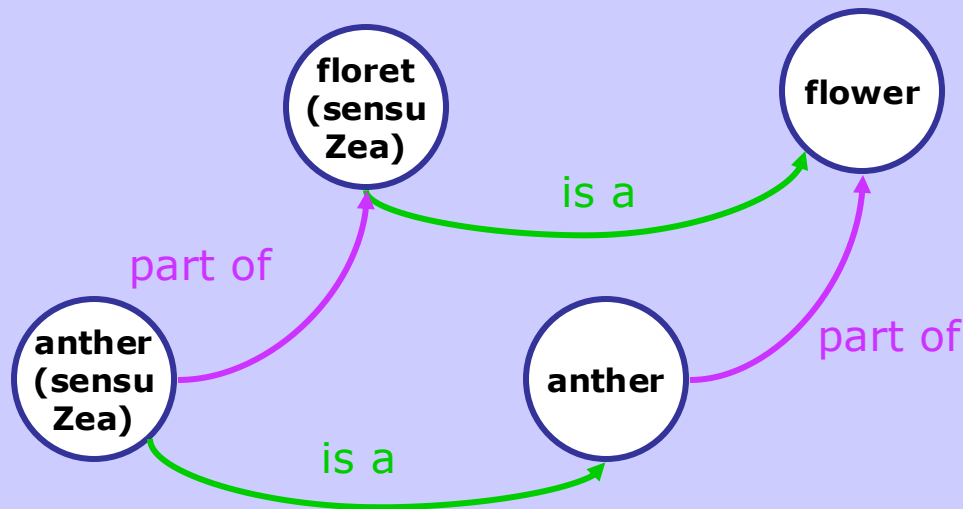
Species-specificity: the problem

In cases where more specific instances of terms (sensu) are created the children terms cannot be generic because this violates the true path rule. An Arabidopsis gene annotated to a generic anther term, should NOT be retrieved in a search for genes expressed in a maize floret.



Species-specificity: the solution

The solution is to create specific (*sensu* Zea) instances for the parts of the maize floret. The new sensu terms are also added as instances of the more generic term, so that a query for mutants affecting the anther will include genes from maize as well as other species.



How are the ontologies maintained?

The ontologies are updated often. The most current versions of the ontologies can be downloaded from the POC CVS repository. The updated ontologies are then used to update the Plant Ontology (AmiGO) browser on a monthly basis.

The ontologies are created and edited by curators using the [DAG Edit ontology editor](#) which is freely available from Sourceforge.

End of tutorial