
Plant Structure Ontology (PSO) — A Morphological and Anatomical Ontology of Flowering Plants

Katica Ilic, Seung Y. Rhee, Elizabeth A. Kellogg and Peter F. Stevens

Summary. The Plant Structure Ontology (PSO) is a controlled vocabulary of anatomy and morphology of a generic flowering plant, developed by the Plant Ontology Consortium (POC). The main goal of the POC was to reduce the problem of heterogeneity of terminology used to describe comparable object types in plant genomic databases. PSO provides a standardized set of terms describing anatomical and morphological structures pertinent to flowering plants during their normal course of development. Created as a tool for annotation of gene expression patterns and description of phenotypes across angiosperms, PSO is intended for plant genomics databases and broad plant genomic research community. Currently, this ontology encompasses diverse angiosperm taxa; further development will include new model organisms and important crop species. This chapter describes the rationales for creating PSO and discusses the guiding principles for its development and maintenance. The content of the PSO and the ontology browsing functionalities are outlined. The PSO can be browsed and downloaded at www.plantontology.org.

2.1 Introduction

Terminology-based application ontologies, or controlled vocabularies [21], have become increasingly important tools in biological and medical fields. This is largely due to two factors: they facilitate standardization of terminology of a given domain, and they allow for acquisition, integration and computation of large amount of biological information (i.e., data annotated to terms in the ontology). The best known and most widely used application ontology in biology, the Gene Ontology (GO), was initiated by a few model organism databases several years ago [9, 10]. Over the years it has become an established standard for describing functional aspects of genes and gene products and is used by a number of genomic databases, as well as by the research community at large. GO was the first generic controlled vocabulary that described three well-defined and distinct biological domains - cellular component, biological process and molecular function. As each of the three aspects is taxon-independent, that is, applicable to any given gene in any organism – GO has succeeded in facilitating consistent functional characterization of gene products in

many species, spanning all kingdoms.

Since GO does not describe morphological and anatomical structures above the level of a cell, anatomical controlled vocabularies have been created for animal model organisms, e.g., fruit fly [8], mouse [4, 11], zebrafish [24], and humans [12]. Anatomical vocabularies were developed for a few plant species too, such as *Arabidopsis* [3], maize [25] and cereals [26]. Plant anatomical ontologies were either species-specific (*Arabidopsis* and *Zea* vocabularies) or applicable to a small number of closely related cereal crops (e.g., Cereal Plant Anatomy Ontology). No attempts were made to map the existing plant ontologies to each other, conceivably due to apparent variation in nomenclature and different organizational principles on which these vocabularies were built. Following the GO paradigm and embracing the idea of a generic, standardized terminology that would ultimately encompass many flowering plants, and that would allow for comparison across species, the Plant Ontology Consortium (POC) developed the first controlled vocabulary of anatomy and morphology of flowering plants, the Plant Structure Ontology (PSO) [13]. The primary goal of the POC was to create a shared descriptive set of terms that can be consistently applied across many angiosperms, and be used to associate and compare gene expression data and phenotypic descriptions across several plant genomic databases.

In this chapter we describe the PSO. This ontology represents the morphological-anatomical aspect of Plant Ontology (PO); the temporal aspect, Plant Growth and Development Stages Ontology, is described elsewhere [19]. This chapter primarily focuses on why and how we developed PSO, its content and comparison with other anatomical ontologies and basic ontology browsing functionalities. The applications of PSO as a tool for functional annotations are demonstrated briefly with examples from plant genomic databases. Possible future directions and further development of this ontology are briefly discussed at the end.

2.2 Objectives and Scope of Plant Structure Ontology

To our knowledge, PSO is currently the only morphological-anatomical ontology in the public domain that is pertinent to more than one organism. The initial public release of the PSO (July 2004) integrated existing species-specific ontologies for *Arabidopsis*, maize and rice; subsequent releases have encompassed terms for other cereal crops (Triticeae), Fabaceae, Solanaceae, and a number of terms for *Populus* (poplar), a recently sequenced model woody plant. The long term goal of the POC is to keep expanding the PSO by adding terms for other angiosperms, keeping pace with whole-genome sequencing efforts and large-scale functional genomics projects. Development and active maintenance of the PSO eliminate the need for creating species-specific anatomical ontologies for each plant whose genome sequence will be determined by large scale genome sequencing projects [e.g., tomato (*Lycopersicon esculentum* L.), potato (*Solanum tuberosum* L.), barrel medic (*Medicago trun-*

catula L.) and grapevine (*Vitis vinifera* L.)].

The main practical purpose of this ontology is to provide a standardized, biologically sound and computationally tractable set of terms describing a distinct domain (i.e., plant structure) as a tool for facilitating annotation of genes and germplasms in angiosperms. Therefore, PSO can be characterized as an annotation-centric controlled vocabulary. The level of detail in PSO is determined pragmatically – it is limited (i.e., very granular terms are generally excluded) but should be sufficient to make possible the description of tissue samples and experimental data, such as mRNA expression patterns, protein localization, description of mutant phenotypes and natural variants. PSO is not designed as a botanical glossary or as a vocabulary for taxonomy for use in taxonomic databases. This is because descriptors (attributes) of the component terms are, to a large extent, avoided in the PSO. Also, PSO does not address phylogeny of angiosperms and is neutral on the questions of organ homology in different angiosperm clades.

2.3 Organizing Principles of the PSO

At the inception of collaborative work on the ontology, a set of organizing principles for the PSO was established by the POC. While keeping in mind the main practical purpose of this ontology (gene and germplasm annotations), the POC members agreed on the crucial importance of preserving biological accuracy of the descriptions and relationships of the domain of knowledge this ontology encompasses, that is, anatomical and morphological structures of flowering plants during the normal course of their development, from zygote to an adult organism. A virtual, generic flowering plant would consist of anatomical and morphological parts that have been described in a range of angiosperm species, and so would comprise terms applicable to diverse angiosperms. The decision was made first to create an extensible backbone for the ontology, initially encompassing only a few species, namely *Arabidopsis*, rice and maize [18]. We would then proceed gradually to add terms for other angiosperms, keeping up with whole-genome sequencing projects and expanding the POC to include individual research communities involved in these large genomic projects. Retrospectively, the lessons we learned from each of the three predecessors of the PSO were invaluable, and have greatly influenced our decisions on the principles and design for the PSO. The following shortcomings in the anatomical ontologies created by the three plant genomic databases were discovered:

- Terms with no definitions or with definitions that were difficult to understand;
- Terms referring to a developmental stage rather than anatomical structure (such as *seedling* or *tetrad of microspores*);
- A number of non-botanical, mainly agronomical or colloquial terms (such as *whole plant*, *crown* or *shank*);
- Terms that did not describe plant structure but rather some qualitative features and/or descriptors (e.g., *leaf blade color*, *ligule appearance* or *ligule consistency*).

Based on our analysis of these three plant ontologies, several decisions were made in the initial stages of development of the PSO:

1. Every term in the PSO would be defined as concisely as possible.
2. Non-botanical and crop-specific agronomical terms would be avoided as much as possible.
3. Terms describing developmental stages would be excluded from the PSO.
4. Attributes of terms would be avoided.
5. Most importantly, synonymy would be used whenever possible, to group species-specific terms.

We also established criteria of what would constitute a valid term in PSO. Terms in PSO are morphological and anatomical structures of a flowering plant, from a cell to the whole plant level. Unlike in botanical or taxonomical glossaries, descriptors (also called attributes or qualifiers) are intentionally omitted. For example, the term *leaf margin* exists in the PSO (PO:0020128), but leaf margin shapes, such as dentate or serrate are not included. Very few exceptions have been made, such as cases where positional attributes of terms are included to accommodate accurate gene annotations (for instance, terms like *terminal bud*, PO:0004713, and *axillary bud*, PO:0004709). Each term in the PSO has a term name, a unique identifier, i.e., an accession number that always starts with the PO prefix followed by seven digits (PO:nnnnnnn), a textual definition, and a specified relationship to at least one other term. While term names can be modified to some extent (only if absolutely necessary), and may or may not be unique in the PO, the accession number associated to each term is always unique - it does not change or get reassigned to another term. Textual definitions are brief and sufficiently broad to reflect the position of a term in the ontology. They are often adopted from standard references, such as textbooks [7] and glossaries [Angiosperm Phylogeny Website (APWeb) (URL: <http://www.mobot.org/MOBOT/research/APweb/>)]. These definitions are conventional botanical definitions; at present, they do not follow formal ontological rules such as transitivity or reflexivity [22].

One of the most important organizational principles in the PSO is the use of synonymy. Extensive use of synonyms was acknowledged to be critical to keep the PSO relatively straightforward and easy to understand, and also to avoid problems with term multiplication (see below). Therefore, we chose a generic form of a plant organ as a term name, while various specific types of that organ were created as its synonyms. This was particularly effective for the *inflorescence* and *fruit* nodes. Since both structures occur in a range of morphological forms in angiosperms, instead of creating multiple terms for different specific types of inflorescence and fruit (such as the several terms for fruit types in the original Cereal Plant Anatomy Ontology [26], we decided to introduce a single generic term for each entity, and place all specific types of inflorescence and fruit as their synonyms (Fig. 2.1). Although this practice of creating synonyms in PO differs from the botanical usage of synonymy, to some extent, it is similar to the GO concept of narrower synonyms (see URL: <http://www.geneontology.org/GO.usage.shtml#synonyms>). As a result, the hi-

erarchy of some higher-level nodes was considerably simplified, and excessive term multiplication was, to some degree, alleviated. Also, users could search the PSO using either generic terms, for instance, fruit, or its taxon-specific synonyms, silique, caryopsis or kernel.

fruit

Accession: PO:0009001

Ontology: plant structure

Synonyms:

- achene
- berry
- capsule
- caryopsis
- circumcissile capsule
- cypsela
- drupe
- follicle
- grain
- kernel
- legume
- loculicidal capsule
- lomentum
- nut
- pod
- pome
- poricidal capsule
- schizocarp
- septicidal capsule
- septifragal capsule
- silicula
- siliqua
- silique

Definition:

The seed-bearing structure in angiosperms, formed from the ovary after flowering.

Comment: None

Fig. 2.1. Term detail page for *fruit* (PO:0009001), showing multiple synonyms of the generic term, all of which are particular fruit types in angiosperms. From the top, encircled are specific fruit types for Solanaceae (including tomato), rice, maize and Brassicaceae (including *Arabidopsis*).

Terms in PSO (as in GO) are linked in a hierarchical network structure called a Directed Acyclic Graph (DAG) (URL: <http://www.nist.gov/dads/HTML/directAcycGraph.html>). The DAG structure of the ontology allows a term to be placed in multiple branches of the ontology - a term can have multiple parents and consequently have relationships to more than one parental (higher-level) node [1]. Parent-child relationships are connecting links between the nodes in the ontological hierarchy. However, a general term cannot appear as a child of a more specific (lower level) term. This is because the relationships in the DAG structure are directional; no path

can start and end at the same node. Most importantly, the position of any given term in the ontology and the type of its relationship to a parental term conveys information about that term beyond its name and textual definition (see examples in text).

The PSO has three types of parent-child relationships, as illustrated in Fig. 2.2. The relationships **is_a** and **part_of** (Fig. 2.2a and Fig. 2.2b) are adopted from GO, and are the principal relationships used in PSO. The third relationship, **develops_from** (Fig. 2.2c), is an additional type, a modification of **derived_from** that is sometimes used in anatomical ontologies [22]. The **is_a** relationship represents a generalized relationship where a child is a subclass or a type of its parent. The **is_a** relationship is transitive. For example, a *cambial initial* **is_a** (a type of) *initial cell*, which **is_a** *meristematic cell*. Therefore, a *cambial initial* **is_a** *meristematic cell* (Fig. 2.2a). Mainly for computational purposes (e.g., automated reasoning and error checking), each term in the ontology is required to have at least one parent with an **is_a** relationship. This specifies that every term in PSO is at least a subclass of *plant structure*. Several terms with missing **is_a** relationship were detected in PSO by using the software tool Obol [17], and we are currently trying to assign a parent with an **is_a** relationship to every term that is missing such a relationship. The **part_of** relationship represents a component or subset relationship. In this ontology, **part_of** is used in a non-restrictive manner, where a parent does not have to be composed of all of its children. However, the child must be a **part_of** the parent to exist in the ontology. For instance, *stamen* is necessarily a **part_of** *androecium*, which is a **part_of** *flower*; therefore, whenever a stamen exists, it is part of a flower, but not all flowers have stamens (Fig. 2.2b). Like the **is_a** relationship, the **part_of** relationship is transitive. The **develops_from** relationship indicates that structure A develops from structure B, meaning that the structure A begins to exist at the same time as structure B ceases to exist. Unlike the other two relationship types, **develops_from** is not transitive. For example, a guard cell develops as a result of asymmetric division of a guard mother cell. Therefore, in the ontology, *guard cell* **develops_from** *guard mother cell*, which in turn **develops_from** the *epidermal initial* (a type of meristematic cell), as shown in Fig. 2.2c. However a guard cell does not develop from a meristematic cell. Consequently, annotations to the term *guard cell* or *guard mother cell* should not be propagated up in the ontology tree to any of their parental terms.

2.4 Content of the PSO

An important decision in the design of the PSO was the organization of the top-level nodes. The goal was to make a robust and extensible backbone of the ontology, which would allow regular updating without a need for significant changes of the top-level hierarchy of the ontology. The PSO describes flowering plant structures spanning cell types to organs and organ systems, from zygote to adult organism and including both sporophytic and gametophytic generations. Hence, the high level nodes in the ontology are *plant cell*, *tissue*, *organ*, *sporophyte* and *gametophyte* (Fig. 2.3a). Because of the need for more accurate gene annotations, an additional top level node,

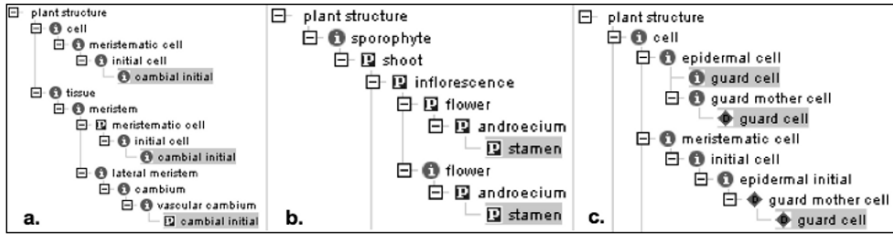


Fig. 2.2. Relationship types in PSO: **a.** The relationship **is_a** - a term is a subclass of its parent, i.e., *cambial initial* **is_a** *initial cell*, which **is_a** *meristematic cell*. **b.** relationship type, **part_of**, e.g., *stamen* is **part_of** *androecium*, which is **part_of** *flower*, therefore, *stamen* is **part_of** *flower*. **c.** Non-transitive, temporal relationship **develops_from** is used to make derivation assertion, e.g., *guard cell* **develops_from** *guard mother cell*, which in turn **develops_from** *epidermal initial*.

in vitro cultured cell, *tissue* and *organ* was added. The term *whole plant* was also introduced at the top level of the PSO (it was previously used in annotations by the three databases). This term is not a botanical term and is intentionally left without children terms. Thus, we recommend that the term *whole plant* is used as a last option, only when precise annotation to any other term in the PO is not possible.

Plant cell types are included in the PSO, but subcellular structures are not, since GO describes these in the cellular component ontology. The POC has made efforts to eliminate overlaps of PO with other bio-ontologies under the Open Biomedical Ontologies (OBO) umbrella (<http://obo.sourceforge.net>), and has been looking for the best solutions to eliminate an apparent overlap of the *plant cell* node in PSO with the Cell Ontology [2].

At the top-level of PSO, terms under *plant cell* (PO:0009002) and *tissue* (PO:0009007) nodes constitute all cell and tissue types found in flowering plants. Many cell and tissue types are located in different plant organs and organ systems, during different stages in development (often with slight modifications). In such cases, introduction of a number of more granular terms in the ontology was necessary, with corresponding terms in every organ (and position within the organ) where the cells or tissues are located. To help avoid a massive proliferation of terms in PSO, which would make ontology navigation and browsing very difficult, (discussed in more detail by Ilic et al. [13]), a decision was made to instantiate cell and tissue terms on a selective basis – only as a response to annotation requirements. Exceptions were made in cases where a cell or tissue type was not present everywhere in a plant, but was localized to specific organs, such as *stem periderm* or *leaf mesophyll*.

At the top level of PO, parallel to the node *plant structure* (PO:0009011) is the *obsolete_plant_structure* node. A term that has been removed from the ontology is never permanently deleted. Instead, the term and its assigned identifier (accession)

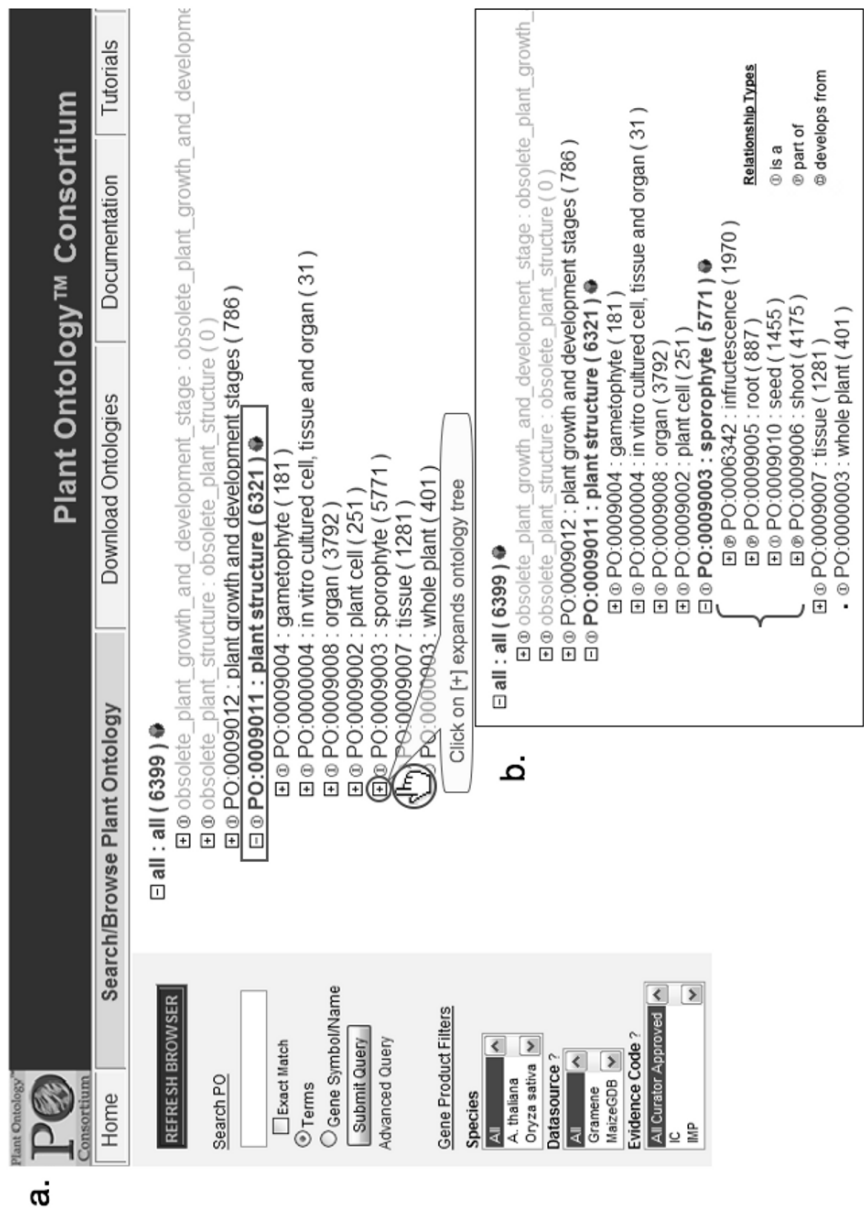


Fig. 2.3. a. A screenshot of the ontology browser - top nodes of the PSO. Clicking on the [+] or [-] sign in front of a term vertically expands or collapses the ontology tree, respectively. **b.** Expanded *sporophyte* node is indicated by curly bracket. A mouse click on a term itself opens a term detail page (see Fig. 2.1). Numbers in parentheses next to the term name indicate the number of the annotations for unique object types associated to a term (including annotations to all children terms). Relationship type icons are shown at the bottom right.

are kept in the ontology file for the record. The definition is always prefaced by the word with OBSOLETE. Because many obsoleted terms in PSO are valid botanical terms, the Comment section always provides an explanation as to why a term was removed (e.g. term *filiform apparatus* was made obsolete because it depicted a sub-cellular structure).

2.5 Comparison to other Anatomical Ontologies

Three plant anatomical ontologies that paved the way to the PSO have been superseded by it and are no longer actively maintained. Shortly after the PSO was publicly released, TAIR and the Gramene database retired their respective anatomical ontologies for *Arabidopsis* and cereals, respectively, and began using the PSO. The original *Zea mays* Plant Structure Ontology [25] has been partially integrated with the PSO terms, and both controlled vocabularies are currently in use by the MaizeGDB.

The *Arabidopsis* anatomical ontology [3] had just over 300 terms that were pertinent to the model core eudicot plant, *Arabidopsis thaliana*, with an organization of the top-level nodes that was similar to the other plant species-specific ontologies. The *Zea mays* Plant Structure ontology [25] had 136 terms that described anatomical and morphological structures of maize, a member of Poaceae. The Cereal Plant Anatomy Ontology [26] had 360 terms, mainly describing anatomical structure of cereal crop species (also Poaceae). However, it also contained many terms that were specific to core eudicotyledonous families, such as Asteraceae, Fabaceae, Cucurbitaceae and others. For example, unlike the two other vocabularies, the Cereal Plant Anatomy Ontology contained several terms describing different fruit types in angiosperms. Fruit types were divided into main classes: dry and fleshy, and also into monocarpous and syncarpous fruits. Each specific fruit type was a separate term in this ontology. After assessing the scope of terms that would be required for each fruit type, and also taking into consideration that a number of fruit part terms needed to be added for each fruit type, we realized that the increase in the total number of terms under the fruit node in the PSO would grow exponentially, resulting in a massive term proliferation. Our solution to this problem was to create multiple synonyms of a single generic term *fruit* – this became one of the founding principles for the PSO.

Unlike vertebrate anatomical ontologies, the three original plant anatomical ontologies were never mapped to each other. Because of historical differences in the ways these ontologies were constructed, term-to-term mapping of the three controlled vocabularies would be difficult at best, and a number of terms would be left unmapped. Furthermore, since the contributing databases were each committed to adopting the PSO, and retiring their own anatomical ontologies, the mapping of species-specific ontologies to PSO was not necessary.

Compared to animal anatomical ontologies or to GO, PSO is a relatively small ontology, comprising just 727 terms (in the release PO_0906) that are descendants

of the root node, *plant structure* (PO:0009011). There are 384 (or 53%) leaf terms, also called terminal nodes (the most specific terms with no children terms below), and 342 (47%) interior nodes (terms with children). Currently, PSO also has 304 synonyms assigned to 149 terms. The length of the longest path (i.e., from the root node to the leaf node in the path) is 15 nodes, while an average ontology depth is 5 nodes. More detailed analysis of the ontology structure is provided elsewhere [13].

2.6 Search, Browse and Download PSO

POC has shared software resources with GO and adopted the GO database schema and software infrastructure for storing, editing (OBO-editor and its preceding version, DAG-editor, <http://www.oboedit.org>) and displaying ontologies and annotations (ontology browsing tool, AmiGO). AmiGO is a web-based tool for searching and browsing ontologies and associated data and is freely available, open source software (http://www.geneontology.org/GO.tools.shtml#in_house). The browser has been slightly modified to suit specific requirements of the PO and its association files. Some browse and search functionalities of the PO AmiGO are displayed in Fig. 2.3a, which shows the top-level nodes in the screenshot of the browser. The *obsolete_plant_structure* node (grayed out) contains terms that are no longer in use in PSO, and are kept in the ontology file for historical and record keeping purposes. This node is placed at the same level as the PSO root node, *plant structure* (PO:0009011).

For browsing the PSO, a click on the [+] or [-] sign in front of a term vertically expands or collapses the ontology tree, respectively (Fig. 2.3a). Each horizontal line corresponds to a term (i.e., node), and consists of an icon depicting the specific relationship (edge), term accession number, term name, and the number of associated annotations in parentheses. The icon represents the relationship between the term and its immediate parental term, e.g., *shoot* is a **part_of** *sporophyte*, which **is_a** *plant structure* (Fig. 2.3b). The accession number starts with the PO prefix, identifying the Plant Ontology database. Next to the term name, the number in parentheses specifies the number of annotations of unique object types, i.e., genes and germplasms, which are associated directly to the term and also to its children terms. A mouse click on any term name opens a new page, the term detail page. This page consists of two parts. On the top, elements of a term are displayed, such as accession, aspect of the ontology, synonyms, definition, comment and term lineage (see example of the term *fruit* in Fig. 2.1), and the second part contains a list of annotations associated to the term, i.e., gene and germplasm annotations (see annotations to the term *embryo* in Fig. 2.4).

The output of a query can be filtered using several options. If a user clicks on the term *embryo* (PO:0009009), and is primarily interested in mutant phenotypes in *Arabidopsis*, maize and rice, he or she can apply species-specific filters, choosing particular species under Gene Product Filters, and selecting a specific Evidence Code (Fig.

2.3a, left-hand panel, and also top of Fig. 2.4, Filter Associations). The Evidence Code is the type of experimental or computational evidence used to support the annotation. For example, the evidence code IMP (*Inferred from Mutant Phenotype*) is used to support a phenotype annotation. The resulting term detail page will have a list of all IMP annotations for *Arabidopsis*, maize and rice, with several hyperlinks which provide quick access to additional information about each annotation entry (for details see Fig. 2.4).

To search the PSO, the left-hand panel on the browser offers a choice of searching for terms (and synonyms) or gene symbols. When searching for terms in the PSO, users are advised to apply the Ontology filter, choosing the Plant Structure aspect; otherwise the search result page contains hits to both aspects of the Plant Ontology, the Plant Structure (PSO) and Plant Growth and Developmental Stages. For example, a search with the term *mesophyll* (not shown) resulted in 9 terms in the PSO that contained word mesophyll in a term name or in a synonym, including the term *leaf mesophyll*. Clicking on the tree icon right below the check box in the first column, multiple term lineages for the *leaf mesophyll* in the PSO will be displayed. This term is placed under *tissue* (as a type of a parenchyma tissue), and also under *organ* and *sporophyte* nodes (as a part of a leaf). Children terms of the term *leaf mesophyll* can be viewed by expanding the ontology tree downward, after clicking on the [+] sign in front of the term accession (PO:0005645). The filtering options described previously for browsing options are also available for searches in the Advanced Queries in AmiGO, and also on the term detail page (see top of Fig. 2.4, Filter Associations is encircled).

The PSO ontology files (in OBO and GO flat file formats) and annotations from each contributing database are freely available and can be downloaded from the POC Concurrent Versions System (CVS) repository (<http://www.plantontology.org/download/database/>). Users can also download copies of the entire POC database as MySQL dumps from POCs CVS repository (<http://www.plantontology.org/download/database/>). The POC continues to maintain the CVS repository as the central place for the plant ontology files, associations, and mapping files. The Plant Ontology Browser accesses the MySQL database located at Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. The structure of the POC database, CVS repository and the main features of the web site are described in more detail elsewhere [14].

2.7 Application of Plant Structure Ontology

PSO has been used for annotations of gene expression data and phenotype descriptions of mutants and natural variants in plant genomic databases such as TAIR [20], Gramene [14], NASC (URL: <http://arabidopsis.info>) and MaizeGDB [16], and also in databases specializing in large scale gene expression data such as GENEVESTIGATOR [27] NASCArrays [6] and ArrayExpress (H. Parkinson, personal communication). As of August 31, 2006, the POC database has over 4,400 unique object

Filter Associations

Datasource Evidence Code Species

☐ All ☐ All Curator Approved ☐ All
☐ TAIR ☐ IMP ☐ A. thaliana
☐ Gramene ☐ IGI ☐ Oryza sativa

Submit Query

Check/Uncheck All Get Detailed View Submit

Symbol	Information	Source	Evidence	Reference
<input type="checkbox"/> embryo ACT11	ACTIN-11, gene from <i>Arabidopsis thaliana</i>	TAIR	IDA	PMID:9037165
<input type="checkbox"/> ADG1	ADP GLUCOSE PYROPHOSPHORYLASE SMALL SUBUNIT 1, gene from <i>Arabidopsis thaliana</i>	TAIR	IDA	PMID:15598655
<input type="checkbox"/> agr1	agravitropism-1, gene from <i>Oryza sativa</i>	Gramene	IMP	GR_ref:1477
<input type="checkbox"/> dek ⁺ -MS4160	gene from <i>Zea mays</i>	MaizeGDB	IMP	MaizeGDB:935747
<input type="checkbox"/> dek ⁺ -MS6214	defective kernelMS6214, gene from <i>Zea mays</i>	MaizeGDB	IMP	MaizeGDB:935747
<input type="checkbox"/> dek1	defective kernel1, gene from <i>Zea mays</i>	MaizeGDB	IMP	MaizeGDB:935747
<input type="checkbox"/> dek3	defective kernel3, gene from <i>Zea mays</i>	MaizeGDB	IMP	MaizeGDB:935747

Hyperlink to the contributing database

Hyperlink to publication source for annotation

Fig. 2.4. Ontology browser term detail view for the term *embryo* (PO:0009009), showing selected annotations for this term.

types (genes and germplasms) annotated with PSO terms amounting to over 10,000 annotations. These annotations were contributed by TAIR, Gramene, MaizeGDB and NASC.

Functional annotation of genes and gene products can be described as a process of extracting information about gene function at the molecular level, its biological role(s), protein localization, and its spatial/temporal expression patterns during plant development [3]. Manual annotations are made by biologists - scientific curators who either extract the information from published literature, or record phenotype descriptions directly by observing plants (natural variants and mutants) in the greenhouse or field. Based on the type of collected information, curators make short statements by creating gene-to-term associations [3, 5]. Any object type can have multiple controlled vocabulary terms associated to it; however, each gene-to-term association is a separate annotation entry (Fig. 2.4). Indispensable components of each annotation entry are the unique identifier of the object type (gene or germplasm/stock), an appropriate, most granular controlled vocabulary term that describes the object type, a database reference number of the original paper (or another type of source) from which the annotation is extracted, and finally, a specific evidence code - a defined type of experimental or computational evidence that was used to support the annotation. Details on evidence codes and evidence descriptions can also be found online (URL: http://www.plantontology.org/docs/otherdocs/evidence_codes.html). More details on the process of literature curation using controlled vocabularies can be found elsewhere [3, 5].

The POC database is set up as a gateway through which the data curated using PO terms can be easily acquired and downloaded; researchers can quickly retrieve annotation data for multiple species. For example, a user who is interested in all genes that have mutant phenotypes affecting embryo development and all genes that are expressed in the embryo, can search for the term *embryo* in the AmiGO browser and will quickly retrieve all gene annotations and phenotypic descriptions associated to this term (and its children terms) for *Arabidopsis*, rice and maize, as shown in Fig. 2.4. Hyperlinks to the individual databases that contributed the annotation provide a quick access to additional information about annotated genes and germplasms. More detailed description of applications of PSO is provided by Ilic et al. [13].

2.8 Conclusion and Future Directions

PSO has been in the public domain since 2004, and has been actively maintained by the POC. Ontologies are work in progress and this ontology is far from being finished; rather, it is still in the early phase of development. As a long term goal, we envision PSO as a continuously expanding ontology that will gradually encompass many angiosperms. The ultimate measure of the success of this ontology will depend directly on how widely it is used by plant genomic databases as well as bench scien-

tists.

The team that created PSO consisted of database curators and plant scientists, all experimental biologists by training. As expected, our main focus was the biological content of the ontology and the very pragmatic goal of providing a practical framework for annotations. The formal philosophical ontological rules and software implementations have been, to a large extent, neglected. The range of problems this approach created is similar to those described for some other informal bio-ontologies such as GO [23]. In the next phase, the POC will need to make an effort to become more compliant with formal ontology rules by applying formal ontological characteristics for term definitions such as transitivity and reflexivity [22], creating **is_a** relationship type for every term in the PSO, and improving other aspects of this ontology to make it computable by automated reasoners [17].

A persistent problem with term multiplication [13] has limited our ability to further expand the PSO by adding new terms for other angiosperm species. New visualization and ontology-editing software, and perhaps a different approach in designing broad anatomical ontologies, will be needed to reflect the modularity of biological reality more accurately and comprehensively, and to allow PSO to be more accessible and easier to comprehend for end-users with limited ontological expertise.

Acknowledgements

We greatly acknowledge Pankaj Jaiswal, Felipe Zapata, Leszek Vincent and Shuly Avraham for their invaluable participation in development of Plant Structure Ontology, and also Anuradha Pujar, Mary Schaeffer, Leonore Reiser, Doreen Ware, Susan McCouch and Lincoln Stein for their contribution. We also thank POC collaborators, Quentin Cronk, Rex Nelson, Naama Menda, Victoria Carollo and William Friedman for their participation in the ontology development. We are grateful to numerous researchers and curators who have reviewed the Plant Ontologies, who are listed individually online (URL: http://www.plantontology.org/docs/otherdocs/acknowledgment_list.html). We acknowledge the Gene Ontology Consortium for software infrastructure and technical support. This project is supported by National Science Foundation Grant (No. 0321666) to the Plant Ontology Consortium.

References

1. J. Bard. Ontologies: formalizing biological knowledge for bioinformatics. *BioEssays*, 25:501–506, 2003.
2. J.B. Bard, S.Y. Rhee, and M. Ashburner. An ontology for cell types. *Genome Biology*, 6(R21), 2005.
3. T.Z. Berardini, S. Mundodi, L. Reiser, E. Huala, M. Garcia-Hernandez, P. Zhang, L.A. Mueller, J. Yoon, A. Doyle, G. Lander, N. Moseyko, D. Yoo, I. Xu, B. Zoeckler, M. Montoya, N. Miller, D. Weems, and S.Y. Rhee. Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol*, 135:1–11, 2004.

4. A. Burger, D. Davidson, and R. Baldock. Formalization of mouse embryo anatomy. *Bioinformatics*, 20:259–67, 2004.
5. J.I. Clark, C. Brooksbank, and J. Lomax. It's all GO for plant scientists. *Plant Physiol*, 138:1268–1278, 2005.
6. D. J. Craigon, N. James, J. Okyere, J. Higgins, J. Jotham, and S. May. NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res*, 32:D575–7, 2004.
7. K. Esau. *Anatomy of Seed Plants*. Wiley & Sons, Inc., 2nd edition, 1977.
8. FlyBase Consortium. The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res*, 30:106–8, 2002.
9. Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.
10. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res*, 11:1425–1433, 2001.
11. T.F. Hayamizu, M. Mangan, J.P. Corradi, J.A. Kadin, and M. Ringwald. The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol*, 6(R29), 2005.
12. A. Hunter, M.H. Kaufman, A. McKay, R. Baldock, M.W. Simmen, and J.B. Bard. An ontology of human developmental anatomy. *J Anat*, 203:347–55, 2003.
13. K. Ilic, E.A. Kellogg, P. Jaiswal, F. Zapata, P.F. Stevens, L. Vincent, S. Avraham, L. Reiser, A. Pujar, M.M. Sachs, N.T. Whitman, S. McCouch, M. Schaeffer, D.H. Ware, L. Stein, and S.Y. Rhee. The plant structure ontology - a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiology*, 143:587–599, 2007.
14. P. Jaiswal, S. Avraham, K. Ilic, E.A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, D. Ware, and F. Zapata. Plant Ontology (PO): A controlled vocabulary of plant structures and growth stages. *Comp Funct Genom*, 6:388–397, 2005.
15. P. Jaiswal, J. Ni, I. Yap, D. Ware, W. Spooner, K. Youens-Clark, L. Ren, C. Liang, W. Zhao, K. Ratnapu, B. Faga, P. Canaran, M. Fogleman, C. Hebbard, S. Avraham, S. Schmidt, T. M. Cassteven, E.S. Buckler, L. Stein, and S. McCouch. Gramene: a bird's eye view of cereal genomes. *Nucleic Acids Res*, 34:D717–23, 2006.
16. C.J. Lawrence, Q. Dong, M.L. Polacco, T.E. Seigfried, and V. Brendel. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res*, 32:D393–397, 2004.
17. C. Mungall. Obol: Integrating language and meaning in bio-ontologies. *Comp Funct Genom*, 5:509–520, 2004.
18. Plant Ontology Consortium. The Plant Ontology Consortium and Plant Ontologies. *Comp Funct Genom*, 3:137–142, 2002.
19. A. Pujar, P. Jaiswal, E.A. Kellogg, K. Ilic, L. Vincent, S. Avraham, P. Stevens, F. Zapata, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, L. Stein, D. Ware, and S. McCouch. Whole plant growth stage ontology for angiosperms and its application in plant biology. *Plant Physiol*, 142:414–428, 2006.
20. S.Y. Rhee, W. Beavis, T.Z. Berardini, G. Chen, D. Dixon, A. Doyle, M. Garcia-Hernandez, E. Huala, G. Lander, M. Montoya, N. Miller, L.A. Mueller, S. Mundodi, L. Reiser, J. Tacklind, D.C. Weems, Y. Wu, I. Xu, D. Yoo, J. Yoon, and P. Zhang. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res*, 31:224–228, 2003.

21. C. Rosse, A. Kumar, J.L. Mejino(Jr), D.L. Cook, L.T. Detwiler, and B. Smith. A strategy for improving and integrating biomedical ontologies. In *AMIA Annu Symp Proc.*, pages 639–643, 2005.
22. B. Smith, W. Ceusters, B. Klagges, J. Kohler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A.L. Rector, and C. Rosse. Relations in biomedical ontologies. *Genome Biol*, 6(R46), 2005.
23. B. Smith, J. Williams, and S. Schulze-Kremer. The ontology of the gene ontology. In *AMIA Annu Symp Proc*, pages 609–613, 2003.
24. J. Sprague, L. Bayraktaroglu, D. Clements, T. Conlin, D. Fashena, K. Frazer, M. Haendel, D.G. Howe, P. Mani, S. Ramachandran, K. Schaper, E. Segerdell, P. Song, B. Sprunger, S. Taylor, C.E. Van Slyke, and M. Westerfield. The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res.*, 34:D581–5, 2006.
25. P.L. Vincent, E.H. Coe, and M.L. Polacco. *Zea mays* ontology - a database of international terms. *Trends in Plant Sci*, 8:517–520, 2003.
26. Y. Yamazaki and P. Jaiswal. Biological ontologies in rice databases. an introduction to the activities in Gramene and Oryzabase. *Plant Cell Physiol*, 46:63–68, 2005.
27. P. Zimmermann, M. Hirsch-Hoffmann, L. Hennig, and W. Gruissem. GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol*, 136:2621–32, 2004.

Anatomy Ontologies for Bioinformatics

Principles and Practice

Burger, A.; Davidson, D.; Baldock, R. (Eds.)

2008, XVI, 354 p., Hardcover

ISBN: 978-1-84628-884-5