

The Plant Ontology: A Common Reference Ontology for Plants

Ramona L. Walls*¹, Laurel D. Cooper*², Justin Elser², Dennis W. Stevenson¹, Barry Smith³, Chris Mungall⁴, Maria A. Gandolfo⁵, Pankaj Jaiswal²

1. The New York Botanical Garden, Bronx, NY, 2. Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, 3. Department of Philosophy, University at Buffalo, NY, 4. Lawrence Berkeley National Lab, Berkeley, CA, 5. Department of Plant Biology, Cornell University, Ithaca, NY

The Plant Ontology (PO) (<http://www.plantontology.org>) (Jaiswal et al., 2005; Avraham et al., 2008) was designed to facilitate cross-database querying and to foster consistent use of plant-specific terminology in annotation. As new data are generated from the ever-expanding list of plant genome projects, the need for a consistent, cross-taxon vocabulary has grown. To meet this need, the PO is being expanded to represent all plants. This is the first ontology designed to encompass anatomical structures as well as growth and developmental stages across such a broad taxonomic range. While other ontologies such as the Gene Ontology (GO) (The Gene Ontology Consortium, 2010) or Cell Type Ontology (CL) (Bard et al., 2005) cover all living organisms, they are confined to structures at the cellular level and below. The diversity of growth forms and life histories within plants presents a challenge, but also provides unique opportunities to study developmental and evolutionary homology across organisms.

This paper addresses how this challenge is being met within the Plant Structure Ontology (Ilic et al., 2007; 2008) of the PO. Upper level terms have been revised to ensure applicability to the entire plant kingdom, allowing comparative studies across distantly related taxa. At the same time, many new clade-specific terms are being introduced at lower levels, allowing precise annotation of genomics and genetics datasets for individual species. The PO intentionally excludes descriptions of alternative phenotypes for different structures (except as synonyms), and definitions are written to describe the canonical form based on structural characteristics. References to function also fall outside the scope of the PO.

As part of the general movement in biomedical ontologies (Gessler et al., 2009; Mungall et al., 2010), the PO is following OBO foundry principles (Smith et al., 2007) to integrate the PO with other ontologies and to improve compatibility with the semantic web. Collaborations with other consortia insure that terms are used consistently and ontologies are orthogonal. As new terms are added or old terms updated, term names and definitions are written to be readable by automated reasoners. Whenever applicable, PO curators include cross-references to other databases such as the Phe-

notypic Attributes and Trait Ontology (PATO) or GO in term definitions. Similar to ontologies developed for animal anatomy, the PO relies on *is_a*, *part_of*, and *develops_from* relations (Mungall et al. 2010).

The PO currently contains links to annotations for expressed genes and phenotypes from databases such as [TAIR](#), [Gramene](#), [NASC](#) and [MaizeGDB](#). Several more datasets and sources will be added in the near future. Image libraries are being created and linked to terms to provide reference images for plant structures along with the definitions.

The PO is unique for its combination of broad taxonomic coverage and clade-specific terms. The multi-level approach of generic terms with more specific children allows instances of known homology among structures to be grouped together, but provides enough independence so that the function, expression, phenotype, and orthology of genes from any two taxonomic groups can be annotated separately. Ultimately, the consistent use of cross products and logical definitions will allow users to search for annotations for terms that span ontologies or do not exist in any current ontology.

REFERENCES

- Avraham, S. et al. (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucl. Acids Res.*, **36**, D449-454.
- Bard, J. et al. (2005) An ontology for cell types. *Genome Biology*, **6**, R21.
- Gessler, D. et al. (2009) SSWAP: A Simple Semantic Web Architecture and Protocol for semantic web services. *BMC Bioinformatics*, **10**, 309.
- Ilic, K. et al. (2007) The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant. *Plant Physiol.*, **143**, 587-599.
- Ilic, K. et al. (2008) Plant Structure Ontology (PSO)— A Morphological and Anatomical Ontology of Flowering Plants. In, *Anatomy Ontologies for Bioinformatics*.
- Jaiswal, P. et al. (2005) Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comp Funct Genom*, **6**, 388-397.
- Mungall, C. et al. (2010) Integrating phenotype ontologies across multiple species. *Genome Biology*, **11**, R2.
- Smith, B. et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, **25**, 1251-1255.
- The Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. *Nucl. Acids Res.*, **38**, D331-335.