2

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 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-*