PCP
PLANT & CELL PHYSIOLOGY

Special Focus Issue – Databases

The Plant Ontology as a Tool for Comparative Plant Anatomy and Genomic Analyses

Laurel Cooper[1,16,17], Ramona L. Walls[2,16,17], Justin Elser[1,17], Maria A. Gandolfo[3,17], Dennis W. Stevenson[2,17], Barry Smith[4,17], Justin Preece[1], Balaji Athreya[1], Christopher J. Mungall[5], Stefan Rensing[6], Manuel Hiss[6], Daniel Lang[7], Ralf Reski[7,8], Tanya Z. Berardini[9], Donghui Li[9], Eva Huala[9], Mary Schaeffer[10,11], Naama Menda[12], Elizabeth Arnaud[13], Rosemary Shrestha[14], Yukiko Yamazaki[15] and Pankaj Jaiswal[1,17,*]

[1]Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, Corvallis, OR 97331-2902, USA
[2]New York Botanical Garden, 2900 Southern Blvd., Bronx, NY 10458-5126, USA
[3]L.H. Bailey Hortorium, Department of Plant Biology, Cornell University, 412 Mann Library Building, Ithaca, NY 14853, USA
[4]Department of Philosophy, University at Buffalo, 126 Park Hall, Buffalo, NY 14260, USA
[5]Berkeley Bioinformatics Open-Source Projects, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mailstop 64-121, Berkeley, CA 94720, USA
[6]Faculty of Biology and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Schänzlestr. 1, D-79104 Freiburg, Germany
[7]Plant Biotechnology, Faculty of Biology, University of Freiburg, Germany
[8]FRIAS - Freiburg Institute for Advanced Studies, University of Freiburg, Freiburg, Germany
[9]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA
[10]Agriculture Research Services, United States Department of Agriculture, Columbia, MO 65211, USA
[11]Division of Plant Sciences, Department of Agronomy, University of Missouri, Columbia, MO 65211, USA
[12]Boyce Thompson Institute for Plant Research, 533 Tower Road, Ithaca, NY 148533, USA
[13]Bioversity International, via dei Tre Denari, 174/a, Maccarese, Rome, Italy
[14]Genetic Resources Program, Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico
[15]Center for Genetic Resource Information, National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan
[16]These authors contributed equally to this work.
[17]These authors contributed equally to the development of the Plant Ontology.
*Corresponding author: E-mail,: jaiswalp@science.oregonstate.edu; Fax, +1-541-737-3573.
(Received October 14, 2012; Accepted November 23, 2012)

The Plant Ontology (PO; http://www.plantontology.org/) is a publicly available, collaborative effort to develop and maintain a controlled, structured vocabulary ('ontology') of terms to describe plant anatomy, morphology and the stages of plant development. The goals of the PO are to link (annotate) gene expression and phenotype data to plant structures and stages of plant development, using the data model adopted by the Gene Ontology. From its original design covering only rice, maize and Arabidopsis, the scope of the PO has been expanded to include all green plants. The PO was the first multispecies anatomy ontology developed for the annotation of genes and phenotypes. Also, to our knowledge, it was one of the first biological ontologies that provides translations (via synonyms) in non-English languages such as Japanese and Spanish. As of Release #18 (July 2012), there are about 2.2 million annotations linking PO terms to >110,000 unique data objects representing genes or gene models, proteins, RNAs, germplasm and quantitative trait loci (QTLs) from 22 plant species. In this paper, we focus on the *plant anatomical entity* branch of the PO, describing the organizing principles, resources available to users and examples of how the PO is integrated into other plant genomics databases and web portals. We also provide two examples of comparative analyses, demonstrating how the ontology structure and PO-annotated data can be used to discover the patterns of expression of the *LEAFY* (*LFY*) and terpene synthase (*TPS*) gene homologs.

Keywords: Bioinformatics • Comparative genomics • Genome annotation • Ontology • Plant anatomy • Terpene synthase.

Abbreviations: BFO, Basic Formal Ontology; CARO, Common Anatomy Reference Ontology; CL, Cell Ontology; FMA, Foundational Model of Anatomy; GO, Gene Ontology; GCP, Generation Challenge Program; ID, identifier; MGCSC, Maize Genetics Cooperation Stock Center; NASC, European Arabidopsis Stock Centre; OBO, Open Biological and

Plant Cell Physiol. 54(2): e1(1–23) (2013) doi:10.1093/pcp/pcs163, available online at www.pcp.oxfordjournals.org
© The Author 2012. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://pcp.oxfordjournals.org/ at University of Missouri-Columbia on September 30, 2013

Biomedical Ontologies; OBOF, Open Biomedical Ontologies flat file format; PO, Plant Ontology; QTL, quantitative trait locus; RO, Relation Ontology; PATO, Phenotypic Quality Ontology; SGN, Sol Genomics Network; SVN repository, Subversion repository; TPS, terpene synthase; TAIR, The Arabidopsis Information Resource; URI; Uniform Resource Identifier; URL, Uniform Resource Locator; OWL, web ontology language; OWL-DL, Web ontology language sublanguage named for its correspondence with descriptive logics.

Introduction

Analyses of vast data sets from genetic and genomic studies have the potential to improve our understanding of species evolution, development and the molecular basis of traits of economic relevance. To realize this potential, plant scientists must be able to connect the spatial and temporal expression patterns of genes and gene products to their molecular functions, their roles in biological processes and gene–gene interactions. Associating qualitative and quantitative phenotypes derived from mutants and breeding populations with the functional and expression aspects of the genome helps to identify candidate genes and regions of the genome that may be associated with traits of interest. Sequenced genomes are available for an ever-growing number of Viridiplantae species ranging from algae, e.g. *Volvox carteri* (Prochnik et al. 2010) and *Chlamydomonas reinhardtii* (Merchant et al. 2007), and bryophytes, e.g. *Physcomitrella patens* (Rensing et al. 2008), to many angiosperms, such as *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), *Populus trichocarpa* (Tuskan et al. 2006) and *Oryza sativa* (Goff et al. 2002, Yu et al. 2002). This now makes it possible to connect genotype to phenotype for intraspecific genetic diversity comparison and also allows interspecific comparison of gene expression, phenotypes and functions of genes and gene family members.

Effective interspecific comparisons at the genome scale demand a common vocabulary (ontology), structured in a way that permits computer-aided reasoning about relationships among entities of different sorts. Ontologies have become indispensable tools for data curation and analysis in the life sciences (Blake and Bult 2006, Jensen and Bork 2010). Basically, an ontology is a structured vocabulary that provides a set of terms to describe the types of entities within a given domain and the relationships among these entities. Terms from an ontology are associated with genes or gene products through annotation (or 'tagging') of data with ontology labels. Because the same term names are used to annotate diverse bodies of data, the results can then be used to serve integration and analysis across multiple studies or species. For example, a user can compare genes expressed in a soybean (legume) pod with those expressed in a silique of an Arabidopsis plant. Though defined differently in a species-specific context, both pod and silique are synonyms of *fruit* in the Plant Ontology (PO) and it may be of interest to investigate what makes a pod different from a silique or how they are similar (note: throughout the paper, ontology terms and relations are printed in *italics*). The PO organizes the conventional knowledge, such as that about types of *fruit*, into a common structured vocabulary that alerts a researcher (and also a computer) that both pod and silique share similar characteristics of the PO term *fruit*.

Widespread use of ontologies in the life sciences began with the development of the Gene Ontology (GO) in the late 1990s. Recognizing that many genes and proteins are conserved in most or all living cells, developers of the GO made the first significant effort to develop a unified vocabulary to describe the attributes of gene products in species-neutral fashion (Ashburner et al. 2000, Gene Ontology Consortium 2012). The GO Consortium developed a standard protocol for annotating genes with ontology terms, laying the foundation for the first serious effort to unify molecular and cell biology in a computationally useful way, thereby radically improving the process of computationally driven functional annotation and comparative analysis of genes and gene products.

Early on, major plant genome sequencing and annotation projects adopted the GO approach for annotating the *A. thaliana* and *O. sativa* genomes (Garcia-Hernandez et al. 2002, Ware et al. 2002, Haas et al. 2003). Researchers soon realized that in order to utilize the full potential of data sets arising from genomic, proteomic, metabolomic and other '-omics' studies, additional controlled vocabularies were needed to describe the anatomical spatial location, temporal growth and developmental stages of plant parts and whole plants. Therefore, as the potential for comparative biology grew, the PO was developed to provide terms that describe flowering plant anatomy and morphology (Ilic et al. 2007) and development stages (Pujar et al. 2006) in model plant species, in order to annotate gene expression and phenotype data sets more accurately (Avraham et al. 2008). For example, the GO *biological process* term *C4 photosynthesis* (GO:0009760) in maize differs from *C3 photosynthesis* (narrow synonym of *reductive pentose-phosphate cycle*; GO:0019253) in a rice plant by localizing and coordinating *carbon fixation* (GO:0015977) in *plastids* (GO:0009536) found in two different cell types. In maize, *C4 photosynthesis* is coordinated between the *mesophyll cell* (PO:0004006) and the cells of the *bundle sheath* (PO:0006023); whereas, in rice, *C3 photosynthesis* occurs only in the *mesophyll cell*. Therefore, if we simply look at the GO annotations of the rice and maize gene products without the context of the mesophyll/bundle sheath cell type specificity provided by the PO, a user will not be able to differentiate the physiological and anatomical significance. The PO makes it possible to extend GO functional annotations to plant molecular biology data, thereby linking known gene functions annotated to GO terms with PO annotations to spatial- and temporal-specific gene product expression and observed phenotypes.

Since the initial development of the PO for the model plant species *A. thaliana*, *O. sativa* and *Zea mays* (Jaiswal et al. 2005, Avraham et al. 2008, Ilic et al. 2008), the scope of the PO project

 © The Author 2012.

has expanded to develop the controlled vocabularies required to annotate anatomy and development stages of all green plants, thus covering a wide array of new plant model species. In its current form, the PO bridges diverse experimental data derived from genetics, molecular and cellular biology, taxonomy, botany and genomics research. The power of the PO lies in its ability to resolve disparities, not only between the various terminologies used by researchers in different genomics projects, but also between the names classically used by different groups of investigators to describe plant anatomy. As such, the PO serves as a common reference ontology of plant structures and development stages.

A recent review of the utility of ontologies to plant science describes the challenges in adopting such a unified approach, as well as the organizing principles behind the development of the PO (Walls et al. 2012a). Here, in contrast, we focus on detailing the composition of the *plant anatomical entity* branch of the PO, its guiding principles for development and expansion, and applications of data annotation, integration and analysis. We provide examples of how the PO is integrated into many other plant genomics databases and web portals, and describe associated online tools for curation and data mining. Furthermore, we demonstrate the power of the PO for comparative plant anatomy and genomics by showing how the PO annotations of the *LEAFY* (*LFY*) and terpene synthase (*TPS*) gene homologs can be explored for inter- and intraspecific comparative analysis. This article describes the PO in reference to Release #18 (July 2012).

Components and Features of the Plant Ontology

In its original form, describing anatomy and growth stages for monocots and dicot plants (primarily *A. thaliana*, *Z. mays* and *O. sativa*), the PO (Avraham et al. 2008) was the first multispecies anatomy ontology among the various biological ontologies. The multispecies anatomy ontologies that have been developed since then include the Teleost Anatomy Ontology (Dahdul et al. 2010), and Uberon, a multispecies anatomy ontology primarily covering metazoans (Mungall et al. 2012). By providing both species-neutral terminology and references to taxon-specific terminology for the respective taxonomic kingdoms, the PO and Uberon enable research that compares anatomy, development and phenotypes across species. However, developing such ontologies presents challenges due to a diversity of phenotypic characters and anatomy contributed by the evolution of species and their adaptation to different environments. Such challenges are minimal in the development of ontologies that cover a single species or group of closely related species. Encompassing the diversity of anatomy and morphology found in green plants is particularly challenging, because green plants are one of the few groups in which structures found in the gametophytic phase of the life cycle are similar to those found in the sporophytic life cycle phase. For example, *non-vascular leaves* (phyllids) are found in the gametophytic phase in bryophytes and the similar structure *vascular leaf* is found in the sporophytic phase of the vascular plant life cycle. The following sections describe in detail how the PO is organized, with emphasis on anatomy and morphology, as encompassed by the ontology term *plant anatomical entity* and its child terms. They include descriptions of some of the specific plant structures that are included in the PO to accommodate a wide variety of plant species, a discussion of ontology design practices and examples of how the PO and the annotated data sets can be used for comparative analyses.

Organization of the Plant Ontology

The PO follows the ontology standards set forth by the Open Biological and Biomedical Ontologies (OBO) Foundry initiative (Smith et al. 2007). The PO can be represented as a graph or tree (e.g. **Fig. 1**), consisting of nodes that correspond to the PO terms, joined by edges representing relationships among the terms (Smith et al. 2005). Each node in such an ontology graph consists of a standard or preferred name (often referred to as a 'term'), a scientifically correct definition with appropriate references, a list of synonyms, e.g. exact, narrow, broad or related synonyms, or foreign language synonyms, as described in Walls et al. (2012a) and, most importantly, a unique alphanumeric identifier (e.g. PO:0025034 for *leaf*) which is used to form a Uniform Resource Identifier (URI). Terms are related to one another by relationships such as *is_a* or *part_of*, as described below. Every term has at least one *is_a* relationship to a parent term.

The PO consists of two branches, each with a topmost or 'root' term—*plant anatomical entity* and *plant structure development stage*, respectively. Each PO branch is organized hierarchically by means of the *is_a* (or subclass of) relation, by appropriately placing it under a single root term. The *plant anatomical entity* branch, which is the focus of this paper, describes morphological and anatomical structures such as *plant organ*, *whole plant* and *plant cell*, while the root term *plant structure development stage* describes the stages of development of *plant structures* (including the *whole plant*). A more detailed discussion of the *plant structure development stage* branch is the topic of a future paper.

Plant anatomical entity. Plant anatomical entity and its child terms (**Fig. 1**) are organized as a structural anatomy ontology, in which all child terms are defined in terms of structure, including spatial information, rather than function. In addition, a number of definitions include a reference to the ontogenic development lineage. The use of the *develops_from* relation (Smith et al. 2005) (**Fig. 1**, **Table 1**) acknowledges the intrinsic link between a structure and its ontogenic predecessor parent structure, e.g. *fruit develops_from gynoecium*. The PO largely follows the Foundational Model of Anatomy (FMA) (Rosse and Mejino 2003) in defining terms structurally. Nevertheless, the PO includes comments describing the common functions

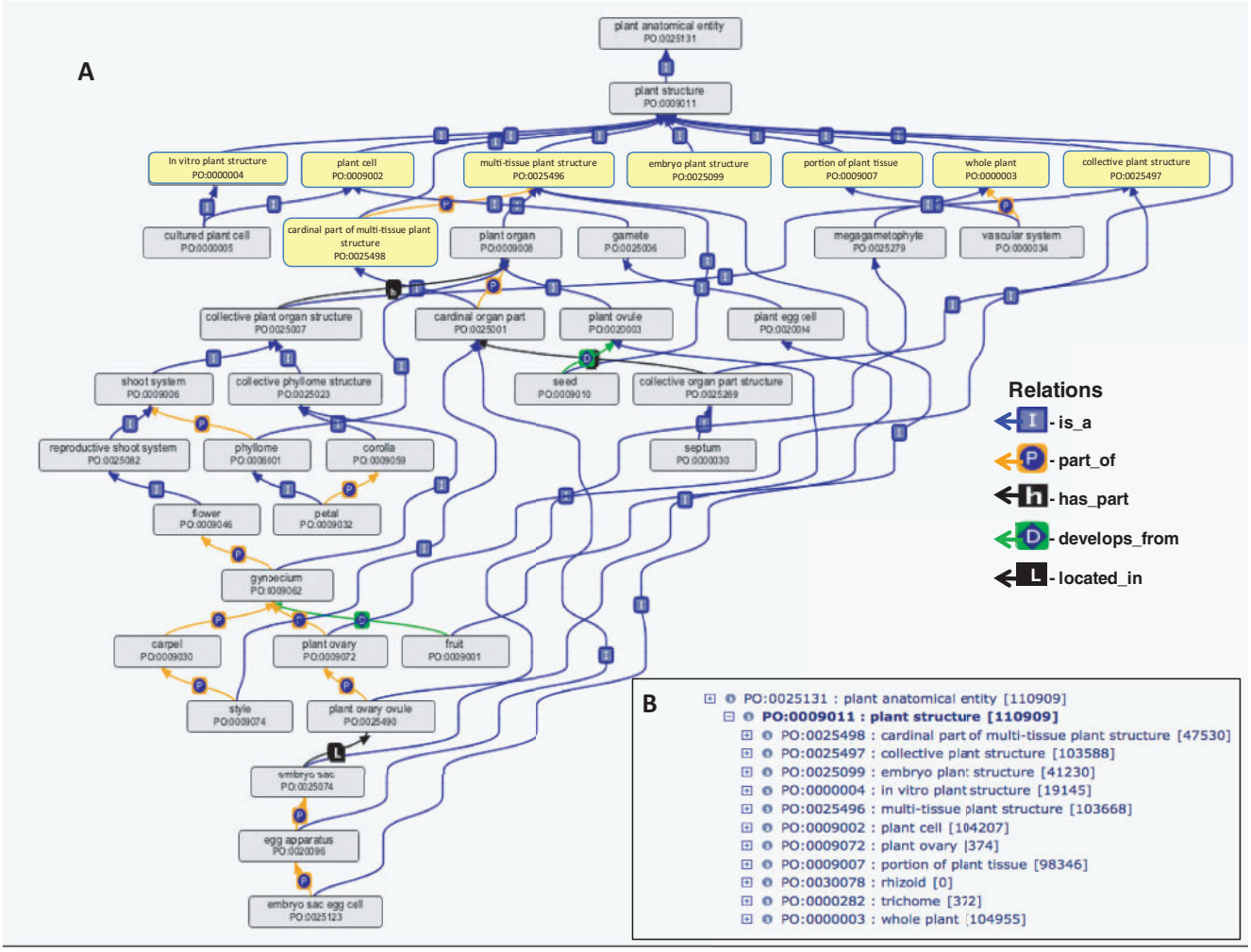

Fig. 1 The term *plant structure* and its children make up the majority of the *plant anatomical entity* branch of the Plant Ontology. (A) Eight of the direct subclasses of plant structure (highlighted in yellow in the tree) are shown with representative child terms and the relationships between them. (B) *Plant structure* is divided into 11 child terms, shown in the tree viewer. The three child terms not shown on the tree are *trichome*, *plant ovary* and *rhizoid*. The ontology diagram was generated using the ontology editor software OBO-Edit (Day-Richter et al. 2007).

of some anatomical entities. For example, a comment states that *xylem* functions in the translocation of water and solutes and, in combination with other portions of vascular tissues, also provides structural support to the *plant axis*, but this statement is not an essential part of the definition of *xylem*. Because it is largely neutral with respect to function as well as homology (Walls et al. 2012a), the PO can be used in many different applications, including other ontologies that aim to model plant function.

Like any graph tree, the nodes that are closer to the root term (towards the top of the tree) are more general terms, compared with the more specific terms that are farther from the root (**Fig. 1**). The direct subclasses of plant structure (highlighted in yellow in the tree, **Fig. 1A**; and in the tree viewer, **Fig. 1B**), along with numerous new mid-level terms, provide the framework into which specific *plant anatomical entities* can be incorporated, allowing the PO to accommodate a diverse range of agronomically important species and emerging plant models for genetic and taxonomic studies. This format allows the *plant anatomical entity* branch of the PO to serve as reference plant anatomy ontology for all plants, to which species and/or other specific vocabularies can be mapped. The definitions of nearly all previously existing high-level terms (those in the first two levels below the root terms) of *plant anatomical entity* have been modified, and several new ones have been added (see **Fig. 1** and **Table 2**). Although many of these terms will probably never be used directly by data annotators (e.g. gene expression would not be annotated directly to *collective plant structure*, but instead to one of its child terms, such as *shoot system* or *perianth*), these high-level categories are essential for ontology maintenance and logical reasoning. The processes of integrating new mid- to lower level terms and improving existing definitions, driven by the addition of new plant models, are described below.

Table 1 Relations in the Plant Ontology

Relation	Meaning	Transitive	Example(s)	No. of assertions
A is_a B	Every instance of *A* is an instance of *B*.	True	*stem is_a shoot axis, epidermis is_a portion of plant tissue*	1,606
A part_of B	Every instance of *A* is a part of some instance of *B*.	True	*stem internode part_of stem, epidermal cell part_of epidermis*	736
A has_part B	Every instance of *A* has some instance of B as a part.	True	*inflorescence has_part flower, meristem has_part meristematic cell*	41
A derives_by_manipulation_from B	(i) *A* is a type of *in vitro plant structure*, (ii) *A* exists at a point in time later than *B*, from which it was created through human manipulation, and (iii) *A* inherited a biologically significant portion of its matter from *B*.	False	*cultured leaf cell dervies_by_manipulation_from leaf*	2
A develops_from B	Either *A* and *B* are *plant cells*, and the lineage of *B* can be traced back to *A*; or *A* and *B* are *plant structures* made of cells, and the majority of cells in *B* develop from cells in *A*	True	*apical hook develops_from hypocotyl, trichoblast develops_from epidermal initial*	117
A adjacent_to B	Every instance of *A* is adjacent to (in contact with or in spatial proximity to) some *B*.	False	*anther wall middle layer adjacent_to anther wall endothecium*	11
A participates_in B	Every instance of *plant anatomical entity A participates_in* some instance of *plant structure development stage B*.	False	*paraphysis participates_in gametophyte development stage*	27
A has_participant B	Every instance of *plant structure development stage A* has some instance of *plant anatomical entity B* as a participant.	False	*seed trichome development stage has_participant seed trichome*	13
A is located_in B	*A* is a *plant anatomical entity* that is part of one organism, *B* is a *plant anatomical entity* that is part of another organism, and *A* is *located_in B*	True	*embryo sac located_in plant ovary ovule*	1

A and *B* represent ontology terms in the PO. The number of assertions (or times a relation is used) in the PO is provided in the last column (based on the July 2012 Release: http://www.plantontology.org/docs/release_notes/index.html). For a more detailed description of the relations, see the Relations Wiki page: (http://wiki.plantontology.org/index.php/Relations_in_the_Plant_Ontology).

Table 2 Child terms of *plant structure* of the *plant anatomical entity* branch of the PO

***Plant structure* child terms**[a]	**Identifier**	**Examples of child terms**
plant cell	PO:0009002	*embryo plant cell, gamete, ground tissue cell, plant spore, plant egg cell, embryo sac egg cell, archegonium egg cell*
portion of plant tissue	PO:0009007	*columella, dehiscence zone, placenta, portion of embryo tissue, portion of ground tissue, primordium, meristem*
embryo plant structure	PO:0025099	*plumule, scutellum, suspensor, embryo hypocotyl*
in vitro plant structure	PO:0000004	*cultured plant callus, cultured plant cell, cultured plant embryo, microspore-derived cultured plant embryo*
whole plant	PO:0000003	*plant embryo, plant spore, thallus, megagametophyte, microgametophyte, plant zygote*
cardinal part of multi-tissue plant structure	PO:0025498	*cardinal organ part, hilum, seed funicle, arilloid, fruit distal end*
>cardinal organ part	PO:0025001	*stalk, stigma, raphe, leaf apex, sporangium theca, leaf lamina, plant axis differentiation zone, organ margin*
collective plant structure	PO:0025497	*collective plant organ structure, collective organ part structure*
>collective organ part structure	PO:0025269	*fruit operculum, pappus, septum, pseudostem*
>collective plant organ structure	PO:0025007	*root system, shoot system, collective phyllome structure*
multi-tissue plant structure	PO:0025496	*plant organ, seed, fruit*
>plant organ	PO:0009008	*plant axis, shoot axis, plant gametangium, petal, phyllome, floral organ, carpel, plant ovule*
>seed	PO:0009010	(has only *part_of* children, e.g. *hilum, seed funicle, arilloid*)
>fruit	PO:0009001	(has only *part_of* children, e.g. *fruit distal end*)
rhizoid	PO:0030078	*epidermal rhizoid, protonemal rhizoid*
trichome	PO:0000282	*seed trichome, glandular trichome, multicellular trichome, shoot axis trichome*
plant ovary	PO:0009072	n/a—has only *part_of* children, e.g. *ovary wall, plant ovary ovule*

[a]All these terms are direct *is_a* children of *plant structure*, except for those indicated with a '>' symbol, which are direct *is_a* children of the term above.

The root term *plant anatomical entity* has three immediate child terms: (i) *portion of plant substance*; (ii) *plant anatomical space*; and (iii) *plant structure*. In order to follow the OBO Foundry guidelines on anatomy ontologies and to keep the *plant anatomical entity* organization consistent with other biomedical ontologies, these three second-level terms correspond to terms from the Common Anatomy Reference Ontology (CARO) (Haendel et al. 2007), which are in turn modeled on the FMA (Rosse and Mejino 2003). For example, the definition of *portion of plant substance* is 'A portion of organism substance that is or was part of a plant'. This is based upon the definition of *portion of organism substance* (CARO:0000004) and, thus, it prevents having to redefine the concepts and allows the user to make broad comparisons across annotated data sets in diverse species. Similarly, the definition of *plant anatomical space* is based upon the definition of *anatomical space* (CARO:0000005). For the user's convenience, the CARO terms and definitions are provided in the comment field of the respective PO term pages. The term *portion of plant substance* (see 'Design Practices and Naming Conventions' below for an explanation of *portion of*) consists of child terms that describe entities that are substances rather than structures, such as *plant cuticle*, *cuticular wax* and *cutin*. *Plant anatomical space* represents pores or other spaces that are part of a plant and surrounded by one or more anatomical structures. They are distinguished from arbitrary spaces, e.g. between adjacent leaves, in that they are generated by developmental, morphogenetic or other physiological processes. Examples of *plant anatomical space* include: *hydathode pore*, *stomatal pore*, *stomium*, *axil* and *canal*. The following section describes *plant structure* in more detail.

Plant structure and its child terms. The child terms of *plant structure* make up the largest group of *plant anatomical entity* terms (**Fig. 1**, **Table 2**). Based on *anatomical structure* from the CARO (Haendel et al. 2007) and the FMA (Rosse and Mejino 2003), a *plant structure* in the PO includes the organism itself (*whole plant*) as the largest anatomical structure, while the smallest is a *plant cell*. As a best practice to avoid redundancy among ontologies, subcellular plant structures are represented in the *cellular component* branch of the GO (Gene Ontology Consortium 2012).

The broad category of *plant structure* includes familiar plant parts such as *leaf*, *stem*, *flower*, *fruit* and *seed*, and also any *in vitro plant structure* that was derived from a plant part. *Plant structure* has 11 child terms, some of which—such as *plant cell*, *portion of plant tissue*, *embryo plant structure* and *whole plant*—are intuitively understandable by most plant biologists. Others, such as *collective plant structure* or *multi-tissue plant structure*, are less intuitive but are needed in order to ensure that the PO provides a complete and logically well-structured set of definitions for all the terms in the PO. They allow the ontology to support the widest possible interspecific comparisons of plant structures, make it easier to browse the ontology tree and aid in checking for errors.

One important child term of plant structure is *plant organ* (**Fig. 2**, **Table 2**), which is defined as 'A multi-tissue plant structure that is a functional unit, is a proper part of a whole plant, and includes portions of tissues of at least two different types that derive from a common developmental pathway.' Some examples of plant organs are: *plant axis*, *coleoptile*, *coleorhiza*, *plant gametangium*, *sporangium*, *phyllome* and

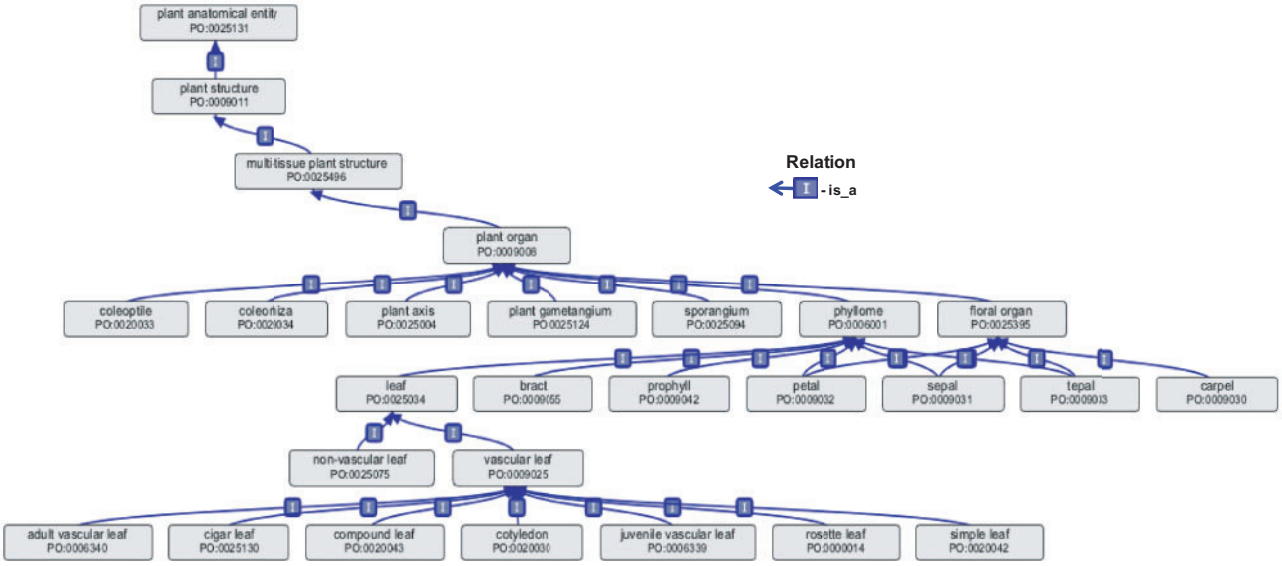

Fig. 2 *Plant organ* is a *multi-tissue plant structure* that encompasses *plant axis*, the various types of *phyllomes* and *floral organs*, along with other structures. Child terms of *phyllome* include *leaf*, *bract* and *prophyll*, as well as the *floral organs*: *petal*, *sepal* and *tepal*. The term *leaf* is the parent term to both *vascular leaf* and *non-vascular leaf*. The ontology diagram was generated using the ontology editor software OBO-Edit (Day-Richter et al. 2007).

floral organ. *Plant axis* includes any axial *plant organ*, i.e. organs that make up the roughly linear axes of a plant, such as *root* and *shoot axes*. The child term *shoot axis* includes structures such as *stem*, *branch* and *rhizome*. The term *phyllome*, widely used for leaf/leaf-like organs, is defined as 'A lateral *plant organ* produced by a *shoot apical meristem*.' Its child terms include *leaf*, *bract* and *prophyll*, as well as the floral organs *petal*, *sepal* and *tepal* (**Fig. 2**).

One of the challenges inherent in describing the anatomy of all plants is resolving issues where the same name is used to describe different plant structures. For example, the term *leaf* is commonly used to describe the *vascular leaf* structure found in angiosperms, gymnosperms and ferns, as well as the similar leaf-like non-vascular structure called a phyllid found in bryophytes. In order to differentiate the vascular and non-vascular types of leaf structures, we defined the general parent term *leaf* and created two child terms, *non-vascular leaf* (synonym: phyllid) and *vascular leaf* (**Fig. 2**). The term *non-vascular leaf* has no *is_a* child terms. It does have several *part_of* children, which are exclusively recognized in non-vascular leaves, such as the *alar cell* (not shown in **Fig. 2**), found at the base of a *non-vascular leaf* adjacent to where the *leaf* attaches to the *stem*, and *costa* or *non-vascular leaf midvein*. A number of child terms that are common to both vascular and non-vascular leaves are *part_of* children of the parent term *leaf*, e.g. *leaf margin*, *leaf apex* and *leaf stomatal complex* (not shown in **Fig. 2**). Some subtypes of *vascular leaf* described by the PO are *adult vascular leaf*, *cigar leaf* (as in banana plants), *compound leaf*, *cotyledon*, *juvenile vascular leaf*, *rosette leaf* and *simple leaf*. Together, all these terms share the common properties of the parent term *vascular leaf*, but, because of their individual characteristics and prevalence in the plant science literature, it was important to create specific child terms for them. A computational reasoner applied to PO-annotated data would be able to make inferences that any *part_of vascular leaf* is also *part_of* some instance of *leaf*.

In vitro plant structures. In order to maintain logical simplicity, many anatomy ontologies deal exclusively with *in vivo* structures (Dahdul et al. 2010, Yoder et al. 2010, Mungall et al. 2012). However, because the use of *in vitro* culture is so prevalent in plant sciences, there is a need to annotate gene expression for *in vitro* plant structures. Thus it was important to include *in vitro plant structure* as a direct child term of *plant structure* in the PO. This presented a challenge, however, because every *in vitro plant structure* can be classified in at least two ways. For example, an *in vitro plant cell* is both a *plant cell* and an *in vitro plant structure*. Ideally, each *in vitro plant structure* would be included only as a direct *is_a* child of the respective *in vivo plant structure* (e.g. an *in vitro plant cell* as a child of *plant cell*). However, the information that the structure in question was grown *in vitro* would be lost. In order to capture this information, an exception was made to the rule of single inheritance for PO (i.e. each term must have exactly one *is_a* parent). In other words, some *in vitro plant structure* terms were assigned two *is_a* parent terms (e.g. *culture plant cell is_a plant cell* and *is_a in vitro plant structure*).

Development and Expansion of the Plant Ontology

Compliance with OBO Foundry principles

As mentioned above, the PO was created and developed in accordance with the principles of the OBO Foundry (http://www.obofoundry.org; Smith et al. 2007) to ensure interoperability with ontologies created in other life science domains. The OBO Foundry is a collaboration among science-based ontology developers that aims to establish a set of best practices for ontology development, with the goal of creating a suite of orthogonal, interoperable reference ontologies in the biomedical domain (Smith et al. 2007). The PO aims to follow the OBO Foundry principles (http://obofoundry.org/crit.shtml) such as having a unique identifier space, clearly delineated content that is orthogonal to other OBO Foundry ontologies and textual definitions for all terms.

One of the accepted principles of the OBO Foundry is that the ontologies have a unique identifier (ID). All the term IDs in the PO are prefixed by 'PO:' and include a seven-digit, zero-padded integer. No other ontology in the OBO suite of ontologies is allowed to use the PO designation, thereby ensuring that the term identifiers are unique. This allows the PO to exist alongside all the other ontologies, and if a user sees the 'PO' designation, you always know it is from the Plant Ontology. The PO ID corresponds to a universally unique Uniform Resource Locator (URL; http://purl.obolibrary.org/obo/PO_XXXXXXX). These URLs are resolvable via the Ontobee website (http://www.ontobee.org/index.php).

To ensure compatibility with other OBO Foundry ontologies, the top level (root) terms in the PO are defined on the basis of the Basic Formal Ontology (BFO) (Grenon et al. 2004, Smith 2012). The BFO is an upper-level ontology that is used to support domain ontologies developed for scientific research. There are currently >100 ontology projects using BFO as common upper-level framework, including the ontologies within the OBO Foundry (Grenon et al. 2004, Arp and Smith 2008). The BFO does not contain physical, chemical, biological or other terms that would fall within the domain of specific fields of inquiry. Instead, it provides a context for organizing the knowledge within those domains.

Textual definitions and is_a completeness. Another accepted principle of the OBO foundry is that all terms in the ontology must have a textual definition. All terms in the PO have textual (human-readable) definitions. The long-term goal of the PO is for all the definitions in the ontology to be logically structured in a way that promotes both consistent formulation of the definitions and automatic reasoning. All definitions are structured as Aristotelian definitions (Rosse and Mejino 2003), which means that they are of the genus–differentia

form illustrated as follows and discussed further in Walls et al. (2012a).

plant egg cell = def. A *gamete* (PO: 0025006) (genus) produced by an *archegonium* (PO: 0025126) or an *embryo sac* (PO: 0025074) (differentia).

OBO best practices require that all terms beneath the root should have *is_a* parents. While it is not strictly necessary to provide such parents from within the ontology, doing so ensures that the ontology is self-contained and makes it possible to formulate definitions consistently for all terms using the genus–differentia format. Other important OBO Foundry principles to which the PO adheres are the requirements that the ontology be openly available and that there is a consistent versioning system. More details can be found at the OBO Foundry Principles Page (http://obofoundry.org/crit.shtml).

Relations used in the Plant Ontology. More than just a list of terms, an ontology represents relationships among the entities to which its different terms refer. The asserted relational connections between the nodes of the ontology can be used for multiple purposes, including ontology navigation and enhancement of queries across annotation data. The PO utilizes relationship assertions of seven types in addition to the basic *is_a* and *part_of* relations, namely: *has_part*, *derives_by_manipulation_from*, *develops_from*, *adjacent_to*, *participates_in*, *has_participant* and *located_in* (**Table 1**). Formal, logical definitions of these relations can be found in the Relation Ontology (RO; Smith et al. 2005). The meanings of *participates_in* and *has_participant* used in the PO are more restrictive than the RO definitions. The relation *derives_by_manipulation_from* is a special case of the RO relation *derives_from*. The PO maintains a Wiki page describing the relations in much more detail (http://wiki.plantontology.org/index.php/Relations_in_the_Plant_Ontology). Where possible, the PO uses the OWL version of the RO: http://code.google.com/p/obo-relations/), which is a descendant of the Smith et al. (2005) RO, and itself makes use of BFO relations. A new version of BFO is currently under development (http://code.google.com/p/bfo/), and in the future the relations will be incorporated in a single file with BFO (Smith 2012).

Design practices and naming conventions. The design of the PO follows the OBO Foundry principles and guidelines (http://www.obofoundry.org/crit.shtml), as well as best practices of the ontology community, as described in Walls et al. (2012a). The particular needs of describing *plant anatomical entities* dictate several additional practices that are described in more detail below.

To ensure biologically correct definitions and consistent use of terms in annotation, a number of nomenclature rules were followed in the development of the *plant anatomical entity* branch of the PO. First, term names for some common plant parts such as cell, tissue, organ, zygote and embryo are prefixed with 'plant' and thus referred to as *plant cell*, *portion of plant tissue*, *plant organ*, etc. This helps to differentiate them from terms of the same name in other non-plant ontologies and vocabularies, and ensures that their meaning is accurately reflected outside the context of plants. Secondly, following the practice laid down in the FMA (Rosse and Mejino 2003), several terms use the prefix 'portion of' in their names, e.g. *portion of plant tissue*, and many of its child terms. Although such use of the 'portion of' phrase is not part of the standard language of biologists, it is important as a means of distinguishing between the physical object that is a *portion of plant tissue* and a description of the corresponding tissue type. Many tissue types do not have 'portion_of' in their term name, because their single-word names are widely used and already imply a physical entity rather than a description (e.g. *epidermis*). In such cases, more specific names were added as exact synonyms (e.g. *portion of epidermal tissue*). Definitions are written to make it clear when a term is referring to some arbitrary *portion of tissue* or to the maximal *portion of tissue* in some given *plant structure*.

Finally, the use of the words 'cardinal' and 'proper' have specific meanings in the context of the PO and other ontologies. The use of 'cardinal' in the term name *cardinal organ part* refers to the fact that these are biologically meaningful and not arbitrary parts of a *plant organ*. The word 'proper' is used in the PO, as in mereology (the study of parts and wholes) (Schulz et al. 2005), to denote the non-reflexive form of a relationship. When one *plant anatomical entity* is defined as being a 'proper part' of another, this refers to the fact that the first entity is a genuine subpart of the second, thus falling short of being identical. This distinction is important because the *part_of* relation (as defined by the RO) is reflexive, so special cases when it is not meant to be reflexive must be specified.

Interactions with other ontologies

The PO collaborates with a number of other ontologies, especially with the GO (Gene Ontology Consortium 2012), the well-established ontology widely used for the annotation of gene product function. Following OBO Foundry ontology principles, the PO strives for orthogonality between the domains of GO and those of the PO. The GO is made up of three branches: *molecular function*, *biological process* and *cellular component*. The domain of the *plant anatomical entity* branch of the PO includes *plant structures* ranging from the *plant cell* and larger, while the parts of a *plant cell*, for instance the *chloroplast*, are described in the *cellular component* branch of the GO.

The GO branch *biological process* encompasses many terms that describe processes that occur during plant development, e.g. flower morphogenesis (GO:0048439) and seed germination (GO:0009845). As far as possible, the GO plant development terms are composed using terms from Mungall et al. (2011). For example, the GO biological process term *shoot system development*:

G0:0022621 ! shoot system development:

Equivalent: GO: 0048856 ! anatomical structure development and RO: 0002296 ! *results in* development of some 'PO: 0009006 ! *shoot system*

PO and GO are working together to align these two ontologies systematically through an ongoing process of suggesting new terms and modifications of existing plant-specific GO terms through the GO SourceForge tracker (https://sourceforge.net/tracker/?func=add&group_id=36855&atid=440764). In the future, the GO intends to use PO in combination with TermGenie (http://go.termgenie.org/), a template-based, reasoner-assisted ontology term generation tool, for creation of new plant-related terms (Chris Mungall, personal communication).

Arabidopsis annotations to GO terms are developed by the TAIR (Berardini et al. 2004, Lamesch et al. 2012) through their curation pipeline and are added to the PO database at each PO release through an automated pull process from the GO FTP site (ftp://ftp.geneontology.org/pub/go/gene-associations/). Recent advances in gene annotation efforts in other plants such as rice (Hamada et al. 2011, Nagamura et al. 2011, Sakurai et al. 2011), barley (Mochida et al. 2011), maize (Sekhon et al. 2011, Kakumanu et al. 2012) and *Physcomitrella* (Lang et al. 2005, Rensing et al. 2007, Rensing et al. 2008, Wolf et al. 2010, Timmerhaus et al. 2011), among many others, are contributing to the body of knowledge about plant gene functional annotations, but, since these annotations are not yet cross-referenced to PO terms, this information is not yet available through the PO database.

Similar to the GO approach above, developers of the PO and the Trait Ontology (TO) (Jaiswal et al. 2002, Yamazaki and Jaiswal, 2005) are working to align these two ontologies. It is being accomplished by creating cross-references in the TO to PO terms and their qualities or attributes from the Phenotypic Quality Ontology (PATO; Gkoutos et al. 2004). For example, the trait *leaf color* (TO:0000326) is referenced as PO *leaf* (PO:0025034) bearing the quality *color* (PATO:0000014) (Pankaj Jaiswal, personal communication).

To maintain orthogonality, the PO re-uses terms from existing ontologies in definitions wherever appropriate. As described above and in Walls et al. (2012a), many *plant anatomical entity* terms draw on the CARO (Haendel et al. 2007). Although the CARO structural classification is based on the FMA, a human anatomy ontology (Rosse and Mejino 2003), many of its terms are defined broadly enough to encompass plants. Ongoing discussions with CARO curators ensure the continued compatibility of the CARO and the PO, and enhance the possibilities for comparative research across eukaryotes.

The PO term *plant cell* presents a special example of interactions among OBO Foundry ontologies. It is an important principle that ontologies in the OBO Foundry should have clearly specified and delineated content that is orthogonal to other OBO Foundry ontologies (http://www.obofoundry.org/crit.shtml). The GO term *cell* is a child term of *cellular component*, and the definition of *plant cell* in the PO references *cell* in the GO as its parent term. However, most organisms, including plants, have cells of specialized types that are considered an essential part of their anatomy. To standardize descriptions of cell types across species, the Cell Ontology (CL) was developed as the reference ontology for the representation of in vivo cell types from all biology (Meehan et al. 2011). Previously, the CL contained its own parallel hierarchy of plant cell terms that were cross-referenced with the PO. This, however, created serious problems in maintaining two parallel ontologies. Therefore, it was decided that the CL would import the *plant cell* term and all its child terms from the PO and retain the original PO identifiers, relationships and definitions. This allows maintenance of terms for *plant cell* types to remain within the control of plant experts, but provides for cross-ontology interoperability.

Enriching *plant anatomy entity* terms for all plants

Since April 2009, the *plant anatomical entity* branch of the PO has grown from 808 terms to 1,203, a 49% increase, and from describing nine plant species to the current 22 species (http://www.plantontology.org/docs/release_notes/index.html). All terms have text definitions, with many refinements of those from the initial project. During this period of time, the scope and amount of genomics data represented in the PO have increased from about 45,000 data objects (genes, mRNA, proteins, etc.) annotated in 2009 to more than 110,000 data objects in 2012 (**Table 3**). These data representations result in about 2.2 million individual annotations, or links between PO terms and the genomic data, as many of the data objects are annotated to more than one PO term.

Two of the major challenges in developing the PO are (i) the need to define high-level terms in such a way that they are appropriate for all instances in all taxa and (ii) dealing with differences in vocabulary usage among groups working on different taxa. The process of expanding the coverage and enriching the PO to provide new terms for plant anatomical entities is highly collaborative, involves many different database groups and user communities (below and **Table 4**) and is continuously evolving. Such collaborative developments help to ensure that PO terms and definitions can be used across different taxa.

The sections below detail four collaborative projects, which resulted in term enrichment and expanding the *plant anatomical entity* branch of the PO. The PO SourceForge tracker (http://sourceforge.net/tracker/?group_id=76834&atid=835555) is the main avenue for new term requests and/or modifications and collaborations for larger scale projects. Outreach workshops and presentations have been held at national and international conferences, and in-house workshops are held with specific groups of domain experts such as wood anatomists (Lens et al. 2012). For more information, see the PO

Table 3 Sources and types of data objects in the Plant Ontology database

Type of data	Plant species	Source	No. of annotated data objects
Genes and gene products	*A. thaliana*, *Gossypium hirsutum*, *Fragaria vesca*, *P. patens*, *O. sativa*, *Z. mays*, Solanaceae spp.	TAIR, AgBase, Jaiswal lab, Rensing lab and cosmoss, Gramene, PO, MaizeGDB, SGN[a]	92,393
Germplasm	*A. thaliana*, *Z. mays*, Solanaceae spp.	NASC[b], MGCSC[c], SGN	10,009
QTL	*O. sativa*	Gramene	8,558
Total			110,960

[a] Sol Genomics Network
[b] European Arabidopsis Stock Centre
[c] Maize Genetics Cooperation Stock Center
More detailed statistics of the database contents and annotations can be viewed on the PO Release Page (http://plantontology.org/docs/release_notes/archive.html).

Table 4 List of some of the databases and web sites that utilize and/or contribute data to the Plant Ontology

Name	Web address	Reference
AgBase	http://www.agbase.msstate.edu/	McCarthy et al. (2011)
ARTADE2DB	https://database.riken.jp/sw/	Iida et al. (2011)
Biological Linked Open Database (BioLOD)	http://biolod.org/	Makita et al. (2009)
BRENDA	http://www.brenda-enzymes.info/	Sohngen et al. (2011)
cosmoss	http://www.cosmoss.org/	Lang et al. (2005)
Crop Ontology	http://www.cropontology.org/	Shrestha et al. (2010)
Gene Ontology	http://www.geneontology.org/	Gene Ontology Consortium (2012)
Gramene	http://www.gramene.org	Jaiswal (2011)
Genevestigator	https://www.genevestigator.com/gv/plant.jsp	Hruz et al. (2008)
MaizeGDB	http://www.maizegdb.org/	Lawrence et al. (2007), Schaeffer et al. (2011)
OryzaBase	http://www.shigen.nig.ac.jp/rice/oryzabaseV4/	Yamazaki and Jaiswal (2005), Yamazaki et al. (2010)
plantco.de	http://plantco.de/	Not available
PLEXdb	http://www.plexdb.org	Wise et al. (2008)
Sol Genome network (SGN)	http://solgenomics.net/tools/onto/index.pl	Bombarely et al. (2011)
Superfamily	http://supfam.cs.bris.ac.uk/SUPERFAMILY/index.html	Wilson et al. (2009)
SoyBase	http://soybase.org/	Nelson et al. (2010)
The *Arabidopsis* Information Resource (TAIR)	http://*Arabidopsis*.org/	Lamesch et al. (2012)
TOMATOMA	http://tomatoma.nbrp.jp/plantOntology/plantOntology.jsp	Saito et al. (2011)
VirtualPlant	http://virtualplant.org	Katari et al. (2010)
VphenoDBS	http://vphenodbs.rnet.missouri.edu/	Green et al. (2011) Harnsomburana et al. (2011)

outreach page (http://wiki.plantontology.org/index.php/POC_Outreach_Events).

Flora of North America Glossary. A significant source of new terms and synonyms for existing terms was a collaboration with the curators of the Flora of North America Glossary (http://huntbot.andrew.cmu.edu/hibd/departments/DB-INTRO/IntroFNA.shtml), which resulted in the addition of 333 new synonyms and 143 unique new term requests (Walls et al. 2012a, Walls et al. 2012b). The list of mappings between the PO and the FNA can be downloaded from the PO Subversion (SVN) repository (http://palea.cgrb.oregonstate.edu/viewsvn/Poc/trunk/mapping2po/FNAglossary2po.txt?view=log) and the list of new terms and synonyms can be downloaded from Source Forge (http://sourceforge.net/tracker/index.php?func=detail&aid=3376762&group_id=76834&atid=835555).

Solanaceae and other tuber-bearing plants. Although the PO has been developed as a species-neutral ontology for plants, certain specific introductions and annotation requirements from new species, such as those bearing tubers, challenged the concept of neutrality. Detailed revisions were made to the *plant anatomical entity* term tuber and its *is_a* and *part_of* children, at the request of the Sol Genome Network

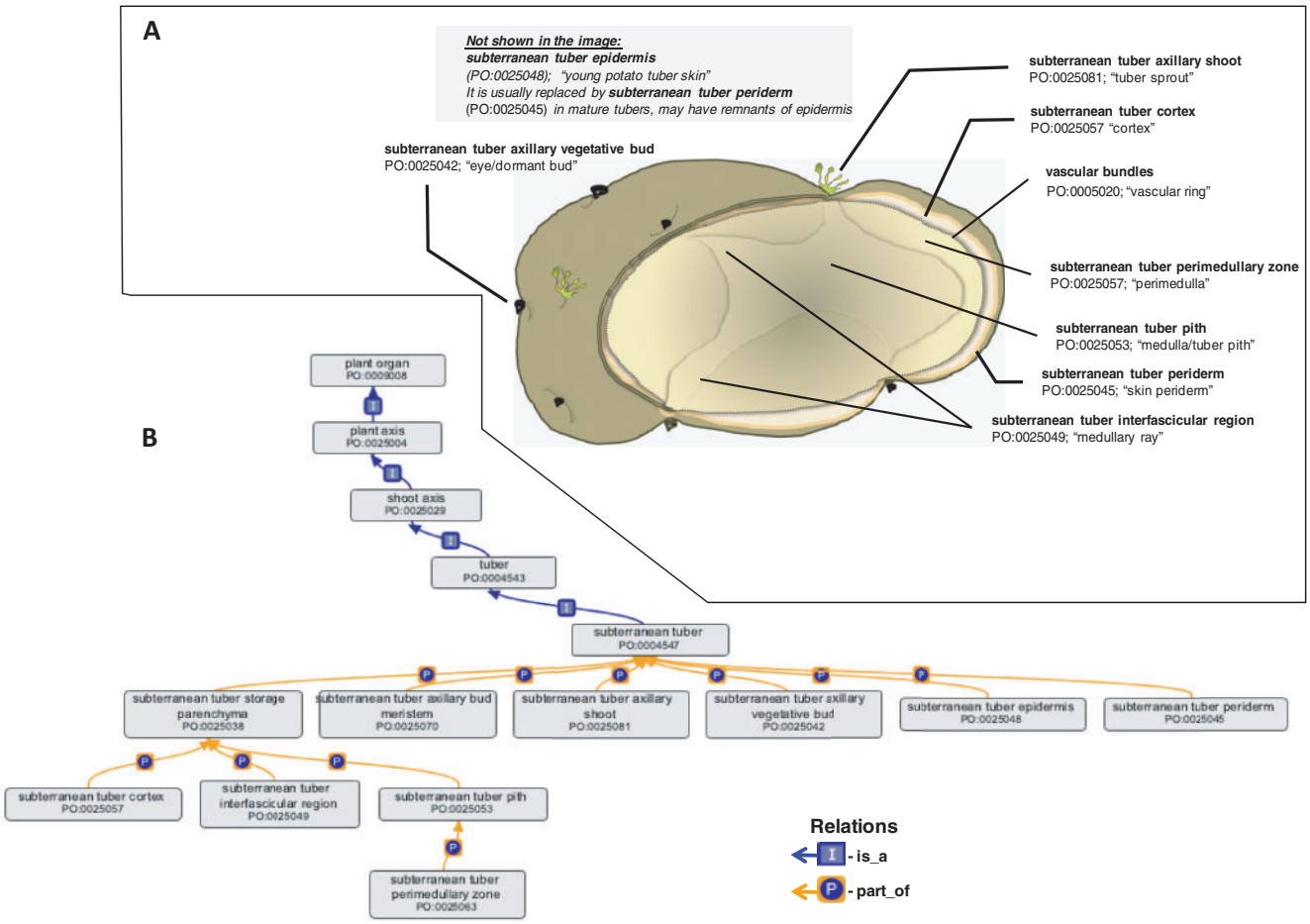

Fig. 3 The terms in the *plant anatomical entity* branch of the PO describe *plant structures* specific to a certain species, while remaining species-neutral. PO terms are supplemented with species-specific synonyms that allow users such as plant breeders to maintain their own vocabulary and relate their terms to the PO hierarchy. (A) An example of using PO to annotate species-specific structures such as the potato tuber anatomy. The parts of any *subterranean tuber* can be described using the general PO terms in the ontology diagram. It also shows that in the PO these terms have potato-specific synonyms. (B) The ontology graph showing the organization of various PO terms that are *part_of subterranean tuber* ontology term. The ontology diagram was generated using the ontology editor software OBO-Edit (Day-Richter et al. 2007).

(SGN; Bombarely et al. 2011) (**Fig. 3**). The revision of the term *tuber* demonstrates how the PO can be used to describe the parts of a complex structure in a species-independent manner, and yet still accurately describe agronomically important crop plants of interest to plant breeders. A number of new terms were created to allow specific annotations of potato tuber structures, but they were added in a way that does not limit their use exclusively to potatoes, i.e. using species-neutral primary names with narrow synonyms that are specific to potatoes. For example, 'potato eye' is a narrow synonym of *subterranean tuber axillary vegetative bud*. Many of the PO terms describing the parts of the *subterranean tuber* are child terms of *portion of plant tissue*. This applies, for example, to *subterranean tuber epidermis* (synonym: young potato tuber skin), *subterranean tuber periderm* (synonym: mature potato tuber skin) and *subterranean tuber pith* (synonym: water core). The use of synonyms such as 'young potato tuber skin' permits ontology builders to maintain strict naming conventions, while allowing plant breeders to search for the terms they need using familiar phraseology. At the same time, the use of species-neutral primary names makes the ontology useful for groups working on other species as well as supporting interspecies comparisons. For example, the *tuber* terms that were added to the PO for potatoes can be applied to *Dioscorea* species (yams) with no modifications. These revisions facilitate research and annotation of the spatial- and temporal-specific profiles of expressed genes determined in the recently sequenced genome of the potato (Potato Genome Sequencing Consortium 2011), one of the world's most important, non-grain food crops.

Physcomitrella patens and non-seed plants. Sequencing of the *P. patens* genome (Rensing et al. 2008) has facilitated the creation of many new expression data sets for *P. patens*, the annotation of which created a need for PO terms to describe plant structures and development stages found in mosses. This was necessary, for example, for comparing the gene functions and

processes essential for various non-vascular plant structures found in mosses with those of the functional and structural homologs found in angiosperms. PO developers worked with researchers from the Rensing lab (http://plantco.de/) and the *Physcomitrella* model species database (cosmoss; http://www.cosmoss.org/) to incorporate anatomical terms for *P. patens* into the PO. The cosmoss curators suggested 63 new *plant structure* terms (**Supplementary Table S1**), along with suggestions for definitions, references and mappings to the PO. In order to integrate the non-angiosperm terms, an additional 44 terms describing the anatomy of bryophytes, lycophytes (club and spike mosses) and pteridophytes (ferns) were added at the same time, to support these taxonomic clades. Many of the new terms, e.g. *seta*, *peristome* and *gametophore*, are found not only in *P. patens* but also throughout the mosses and other bryophytes, and some even in vascular plants (e.g. *rhizoid*, *exothecium* or *archesporial cell*). In keeping with the objective that the PO should be species-neutral, some of the term names and definitions suggested by cosmoss were modified slightly to ensure that they would be applicable to any plant in which the corresponding structure is found (**Supplementary Table S1**).

Musa spp. (banana and plantain) and other monocots outside the Poaceae family. Banana and plantain (*Musa* spp.) are important tropical fruit crops worldwide. In collaboration with the Generation Challenge Program (GCP; http://www.generationcp.org/) and Bioversity International, 31 new terms were created, and synonyms were added to several existing terms, to accommodate the anatomical descriptions of banana and plantain species that are widely used by plant breeders and collection curators (**Supplementary Table S2**). Similar to the potato tuber terms, many of the structures found in *Musa* are also present in other taxa, particularly in other non-grass monocots. Some terms were already in the PO, and simply required the addition of *Musa*-specific synonyms, e.g. 'male bud' as a synonym for *inflorescence bud*. Examples of some of the terms that were added are *free tepal*, *fused collective tepal structure* and *cigar leaf*.

The Plant Ontology is a Resource for Plant Biologists

Accessing the Plant Ontology terms and annotation data sets

The online PO database provides ontology terms and definitions along with the associated 'annotations' (links) as described by Hill et al. (2008), between the PO terms and data sourced from numerous plant genomics data sets (**Table 3**). PO Release #18 (July 2012) contains about 2.2 million annotations linking PO terms to >110,000 unique data objects representing genes, gene models, proteins, RNAs, germplasm and quantitative trait loci (QTLs). These data are currently contributed by 11 different data sources (**Table 3** and below), primarily collaborating model organism database groups, that cover 22 different plant species. PO curators and researchers at various collaborating database groups work closely to develop the annotation files in the standardized data format (http://plantontology.org/docs/otherdocs/assoc-file-format.html), which are stored in a MySQL database. The database is accessible online (http://plantontology.org/amigo/go.cgi) and also available for download (http://plantontology.org/download/database/).

In some cases, annotation files are a result of special projects devoted to the creation of specific data sets; in others, the creation of annotations results through an ongoing collaboration with more or less regular updates to the data sets housed at the PO. An example of the former is the collaborative project between the Rensing lab (http://plantco.de/), the moss model organism database (cosmoss; http://www.cosmoss.org/) and the PO project. In addition to the new and modified PO terms described above for the moss *P. patens* (see above and **Supplementary Table S1**), we have added some 26,000 gene expression data points for moss anatomy and development, resulting in approximately 82,000 new annotations. Future efforts will include continuing to enrich PO with bryophyte terms and additional gene expression annotations.

Ontology terms and the associated annotation data sets can be accessed through the web browser (Carbon et al. 2009) on the PO home page (http://plantontology.org) (**Fig. 4**) or from any term page. Users can browse for terms or annotation data directly using the tree view, or can 'Search PO' for specific terms or genes of interest. **Fig. 4A** presents an example page for the term *plant egg cell* with the three main panels. The 'Term Information panel' (**Fig. 4B**) contains information about the term such as the term name, accession (ID), any synonyms, the definition and comment. The 'Term Lineage panel' (**Fig. 4C**) shows the location of the term in the PO hierarchy, in either tree format or graphical view. The numbers/counts in parentheses next to the term name is a hyperlink to the data annotations page (**Fig. 4E**) for that term and its direct *is_a* children. These links will take the researcher to the annotation data source for more information (**Fig. 4F**). For example, the term *plant egg cell* and its child terms have 175 annotations to data objects, which in turn are linked out to the source database (TAIR) and to the relevant gene product page in GO, if that information is available (**Fig. 4G**). At the bottom of the term page in the 'External References' panel (**Fig. 4D**) is a link to the SourceForge Tracker entry (https://sourceforge.net/tracker/index.php?func=detail&aid=3030032&group_id=76834&atid=835555) related to that specific term. The user can follow that link to view the history of the term and definition and to make comments or suggestions. In future versions of the PO, many of the term pages will also have links to images of the relevant plant parts (including images specific to particular developmental stages).

The ontology files for download are accessible in two formats: Open Biomedical Ontologies flat file format (OBOF;

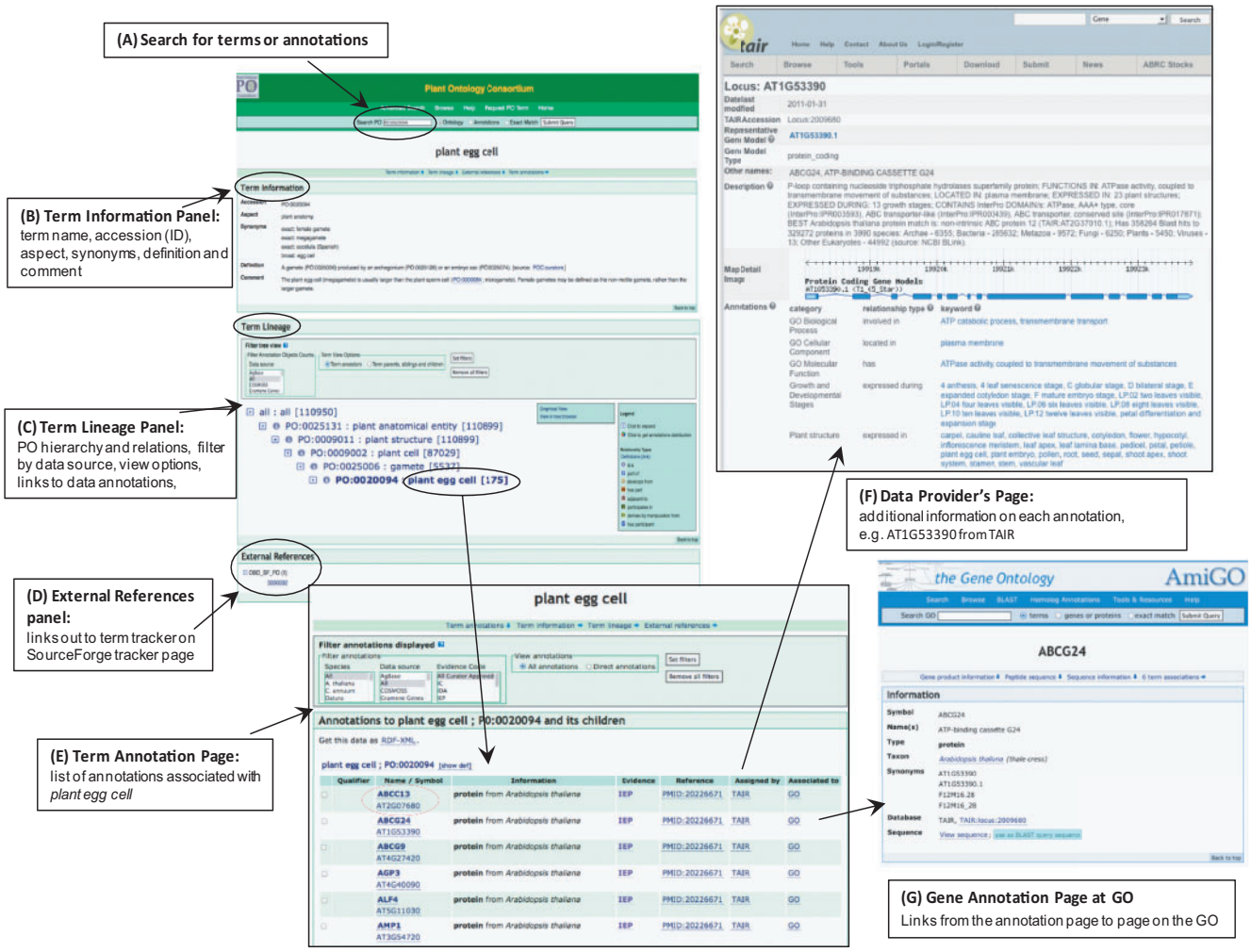

Fig. 4 Accessing Plant Ontology terms and annotation data through the plantontology.org website. (A) The search box at the top of each page is a starting point for finding specific term pages or annotation data, e.g. the page for *plant egg cell* (PO:0020094). (B) The Term Information Panel contains information such as the term name, synonyms, accession of identifier (ID), the definition and any comments. (C) In the Term Lineage Panel, the PO hierarchy and relationships are displayed and can be browsed. The page provides options to view the ontology tree in a graphical tree format and setting filters to query the annotations by species, source provider and/or evidence type. (D) The External References Panel links out to term tracker on SourceForge. (E) Clicking on the number in square brackets links out to the Term Annotation Page showing a list of annotations associated with the term *plant egg cell.* These list of annotations include those directly annotated to *plant egg cell* and the terms associated with it as child terms and/or parts (for an example, see **Fig. 6**). (F) Hyperlinks listed in the Name/Symbol Column link the user out to the Data Provider's Page. (G) An additional link often available from annotations page will link out to the gene annotation pages on the Gene Ontology website provided the same annotated object exists in both the PO and GO database.

http://oboformat.org) and Web Ontology Language (OWL; http://www.w3.org/TR/owl2-overview/) format from the links provided on the PO Download webpage (http://plantontology.org/download/download.html). Ontology files and bulk annotation data files are available for download from the SVN repository (http://palea.cgrb.oregonstate.edu:/svn/Poc). The ontology (but currently not the annotations) is also available via web services as described below.

Glossary, translations and subsets. Three additional features have been added to the PO to enhance the ability of users to access the ontology and the associated data. In addition to the ontology browser, another means of accessing terms, synonyms and definitions is by using the glossary feature (http://www.plantontology.org/db/glossary/glossary) on the PO website. Here, the user can browse through *plant anatomical entity* child terms alphabetically or search for a specific term of interest. In order to increase the utility and acceptance of the PO for plant scientists in other countries and non-native English-speaking researchers, Spanish and Japanese translations have been added for the term names in the *plant anatomical entity* branch of the PO and are available on the online ontology browser (**Fig. 4**) as well as in the downloadable ontology files.

Several subsets of PO terms have been created to help make the corresponding terms more easily accessible to specific groups of users (**Supplementary Table S3**). Subsets provide a way for users to search for terms relevant to a particular topic or taxon, and they also provide a means of quality control. For example, a user trying to choose between two related terms can select the term tagged to the most appropriate taxonomic subset. Subsets can also be used to create pared-down versions of the PO—also known as 'slims'–that contain a subset of ontology terms. Existing subsets in the PO have been complemented with new subsets which include: Plant Functional Traits (general terms needed for plant ecology, added at the request of TraitNet; http://traitnet.ecoinformatics.org/); terms used for banana (*Musa*); terms used for potato (*Solanum tuberosum*); and separate subsets for terms used for angiosperms, gymnosperms, pteridophytes and bryophytes. In future releases of the PO, taxonomic subsets may be enhanced with the use of *only_in_taxon* or *never_in_taxon* relations [Deegan (nee Clark) et al. 2010] along the lines described in Walls et al. (2012a).

PO web services. Developers who wish to use the PO in mobile or desktop applications, such as those for annotation and curation tools, can now access terms, synonyms, definitions and comments using web services. The PO has developed its own web services, to complement other existing services. The PO web services (see link below) were built with Hypertext Preprocessor (PHP; http://www.php.net/), a widely used general-purpose scripting language, model aspects of RESTful software architecture (Fielding 2000), and provide PO data encoded in JSON format (http://www.json.org), a widely used standard for providing data over the internet. There are two types of PO services available at this time: (i) the short and quick 'term search web service' (**Fig. 5A**) provides term name and synonym search results, given a partial term name or synonym. For example, a search for '*basal*' will return multiple terms and/or synonyms with '*basal*' in their names, such as *axillary hair basal cell* and *basal flower*; and (ii) the web service providing extensive details on multiple pieces of term data, given a PO accession ID (**Fig. 5B**). A search for 'PO:0000252' will return the term name, aspect, definition, comment and any synonyms for the PO term *endodermis*. These services could be used, for example, in applications that allow users to provide PO terms as keywords for image annotation, gene and phenotype curation, adding mark-ups on scientific literature and help autofill/autocomplete the database query searches, etc. Future development will include a web service delivering PO annotation data in a similar manner. Full documentation is available on the Plant Ontology website documentation page: (http://www.plantontology.org/docs/otherdocs/web_services_guide.html).

BioPortal web services (Whetzel et al. 2011) also offer PO web services as part of a larger set of methods providing access to ontological data, and generally return data in XML, although JSON format was more recently made available for most of their methods. In addition to serving term data, they provide relationship and hierarchy data connecting terms in the ontologies that they host. The iPlant's Simple Semantic Web Architecture and Protocol (SSWAP; http://sswap.info/) (Gessler et al. 2009, Nelson et al. 2010) offers the PO as a complex set of graph-based query services based on the OWL sublanguage (OWL-DL; http://www.w3.org/TR/owl-guide/) and resource description protocols.

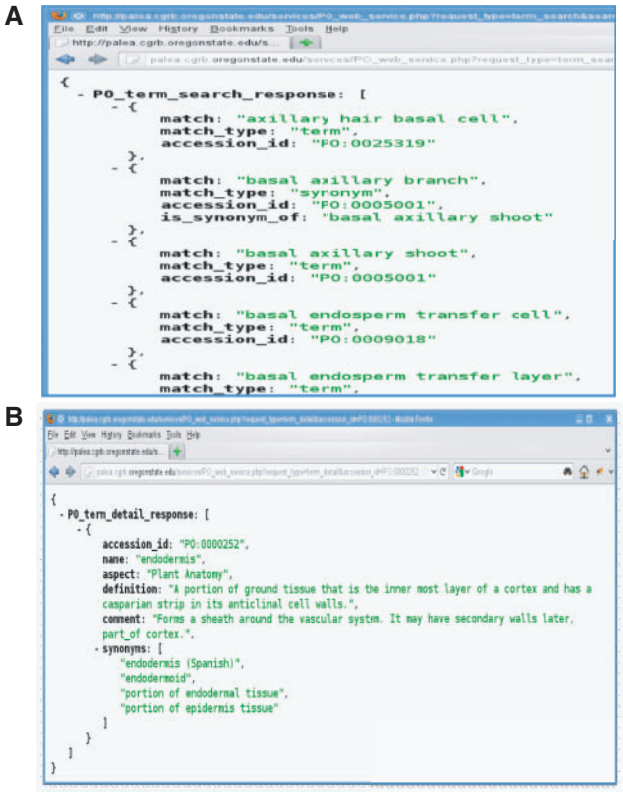

Fig. 5 Two types of PO web services have been developed for mobile or desktop applications to access terms, synonyms, definitions and comments. Built with PHP (http://www.php.net/; http://www.php.net/credits.php) and modeling aspects of RESTful software architecture (Fielding 2000), these services provide PO data encoded in JSON format (http://www.json.org). (A) Example term search request for '*basal*', where the web service returns term name, match type, accession_id and synonym matches. (B) Term detail request for accession ID PO:0000252 provides multiple pieces of term data, given a PO ID. A search for 'PO:0000252' will return the name, aspect, definition, comment and any synonyms for the PO term *endodermis*. Full documentation is available on the Plant Ontology website page (http://www.plantontology.org/docs/otherdocs/web_services_guide.html).

Discussion

Applications of the PO in comparative genomics analyses

The power of the PO is its ability to link anatomical and morphological descriptions to genomics and genetic data sets

and to facilitate data mining and inter- and intraspecific comparative genomics analysis. This can be most effective if ontology terms are integrated in metadata annotations of plant structures (spatial aspects) and growth and developmental stages (temporal aspects) in gene expression or phenotype studies. For example, gene expression analysis annotated to *plant anatomical entities* across a wide range of taxa can be combined with taxonomic studies to compare the patterns of expression of gene orthologs.

PO hierarchy and relationships facilitate comparative genomics analyses of the LFY/ZFL homologs. One advantage of an ontology, compared with a simple glossary, is that by making use of the relationships between the terms (**Fig. 6**, **Table 1**), a user (including a computer) may explore up and down the ontology graph to learn more about *plant anatomical entities* and their constituent parts (through *part_of* relations) and/or their ontogenic development, (through the *develops_from* relation). For example, *ear floret* is *part_of ear spikelet* and *flower develops_from flower primordium* (**Fig. 6A**). Additionally, you can query the graph for annotations by entering at any level, because the annotations flow through certain ontology relationships (**Fig. 6B**, **Table 1**). This allows annotations assigned directly to a term to be percolated to the *is_a* or *part_of* parent terms, but not through the *develops_from* relation. For example, the *A. thaliana* gene *AtLFY* was annotated to the *inflorescence* and *flower* (**Fig. 6B**) terms based on mutant phenotype and gene expression studies, and its role in the regulation of flower and inflorescence development (Schultz and Haughn 1991, Weigel et al. 1992, Mandel and Yanofsky 1995, Siriwardana and Lamb 2012, Yamaguchi et al. 2012). Because *inflorescence* and *flower* are child terms (*is_a* children) of *reproductive shoot system*, it can be inferred that *AtLFY* is expressed in a *reproductive shoot system*. Thus, the ontology structure can guide the user to find the *AtLFY* annotation on *reproductive shoot system*, a less granular term in the ontology,

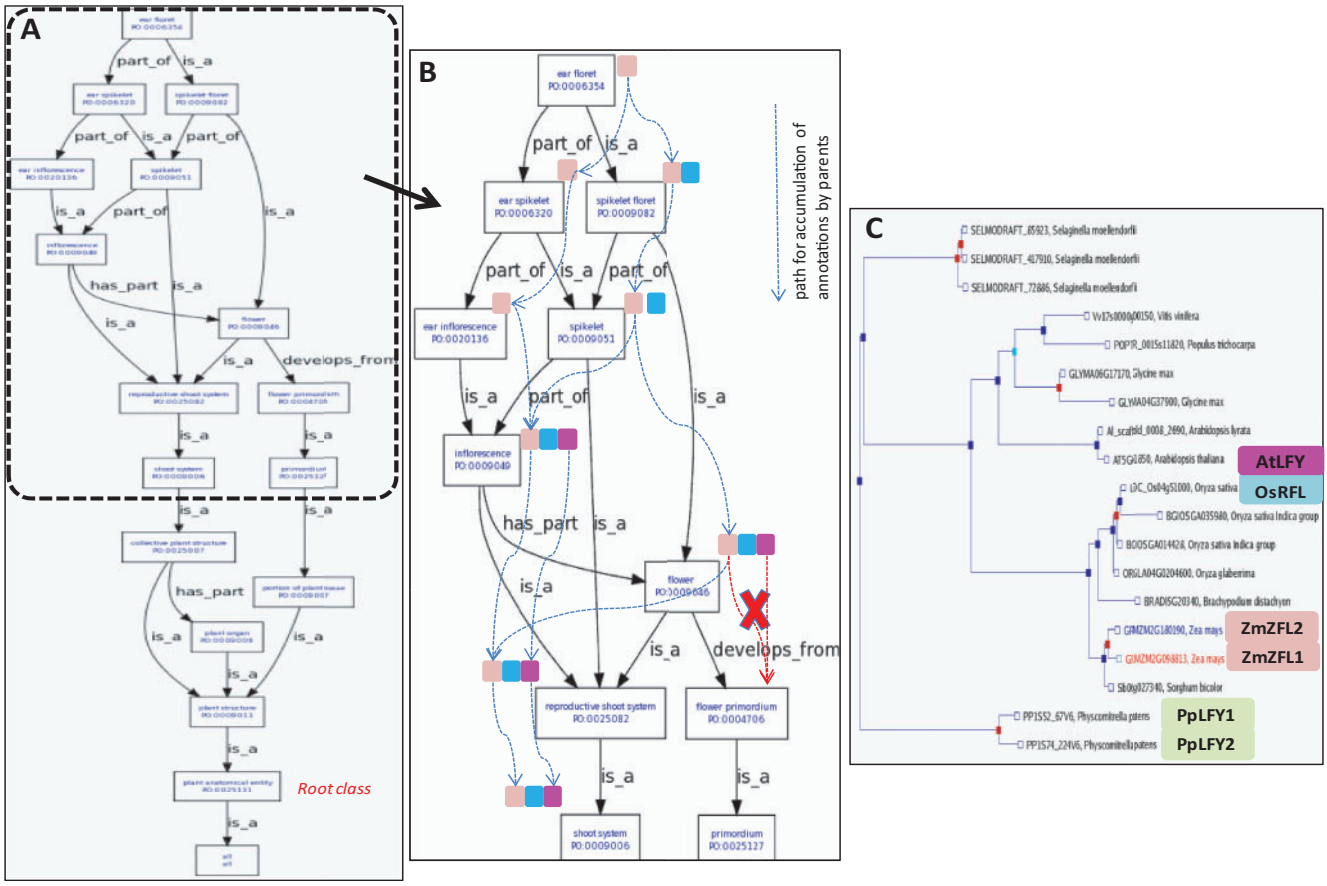

Fig. 6 The PO hierarchy and relationships facilitate comparative genomics analyses using annotated genomics information. (A) Placement of the term *ear floret* and its parent terms in the ontology tree. Terms in the ontology are linked by relations such as *is_a*, *part_of* and *develops_from* (black arrows). (B) A zoomed-in view of the ontology tree showing annotations to *LFY/ZFL* homologs (colored boxes). Annotations flow through a subsumption path (blue dotted arrows), moving to the immediate *is_a* and/or *part_of* parent terms, but not through the *develops_from* relation (red dotted arrows). (C) A phylogenetic gene tree of the *LFY/ZFL* homologs shows that this gene family is widespread across the plant and animal kingdoms. The tree was generated by the Gramene database (http://www.gramene.org/) using the method of Vilella et al. (2009).

to facilitate comparative genomics analysis with species that have a reproductive shoot system but not flowers (such as gymnosperms).

In a search for annotations for the *LFY* homologs from maize *ZmZFL1* and *ZmZFL2*, identified in the phylogenetic analysis (Bomblies et al. 2003, Bomblies and Doebley 2006) (**Fig. 6C**), a user could find annotations on the *inflorescence* and *flower* terms, even though in this case these annotations were assigned to a specific *flower* subtype called *ear floret*. The *ZmZFL* genes were annotated to more specific unique terms based on their known roles in regulating the process of floral organ identity and pattern formation, and development of inflorescence architecture. They also regulate flowering time by regulating the transition of the *vegetative shoot apical meristem* to *reproductive shoot apical meristem* (Bomblies et al. 2003, Bomblies and Doebley 2006).

A user, while looking for these *LFY/ZFL* annotations, may also search for known rice (*OsRFL*) (Rao et al. 2008) and moss *Physcomitrella* (*PpLFY1* and *PpLFY2*) (Tanahashi et al. 2005) homologs, based on the gene trees such as those provided by the Gramene database (**Fig. 6C**). The PO database may or may not contain annotations to *OsRFL* and the *PpLFY* genes, but one could hypothesize that *OsRFL* may be associated with *spikelet floret* and *inflorescence* (synonym: panicle in rice), based on the evidence from the homologs. and which we find is true on review of the literature. Though *OsRFL* functions in a manner partially similar to *AtLFY* (Chujo et al. 2003) and the *ZmZFL* genes, it has unique expression patterns and regulates an additional set of interacting genes (Rao et al. 2008). The *PpLFY* genes cannot be compared in this manner because mosses do not have inflorescences like those found in angiosperms, suggesting that the *Physcomitrella* genes may play a different role in moss plant development. Indeed, the *PpLFY* genes are known to control sporophyte development, by regulating the first zygotic cell division (annotations not shown), and *PpLFY1* is expressed in the sporophyte (Tanahashi et al. 2005).

The combination of characterized genes, e.g. the *LFY* homologs and their annotations to PO terms in the ontology tree, allows users to address questions such as: 'Are homologs annotated to the same PO terms describing similar gene expression profiles?' If not, can their annotation tell something about the (dis)similarities between the structures found in the species, such as flowers of monocot grass plants vs. the dicot Arabidopsis? Also, similar to the example mentioned above on C_4 photosynthesis, if the gene products were annotated only with the GO, it would have been difficult to question how homologs with the same or similar function (e.g. *transcription factor activity*; synonym of GO:0000988) regulate the development of taxon-specific plant structures in grasses (rice and maize), Arabidopsis and moss plants. Therefore, by adding the spatial and temporal annotations from PO to the existing GO annotations, it is possible to find answers to such questions.

Comparative analysis of the terpene synthase gene family with PO annotations. Often plant genomes contain sets of related genes as members of a gene family. The terpene synthase (*TPS*) gene family is well studied and characterized (Aubourg et al. 2002, Chen et al. 2011, Tholl and Lee 2011). These families can be identified as arising due to ancient or recent genome duplications and characterized by synteny across phylogenetically distant homologs. Many such homologs may have similar functions, such as enzymatic activities, but have clearly diverged in different lineages (Chen et al. 2011). Tholl and Lee (2011) characterized the genomic organization of the 32 Arabidopsis enzymes of the core biosynthetic pathways producing the 5-carbon building blocks of terpenes. The PO terms and annotation database allows us to ask questions such as: do all the homologs and *TPS* gene family members have similar *plant anatomical entity* annotations or do they differ based on *TPS* subgene family and how do the annotations differ between the same or different species?

In order to address these questions, we first resolved a gene family tree of some the known *TPS* gene family members from five species (*A. thaliana*, *Z. mays*, *O. sativa*, *Selaginella moellendorfii* and *P. patens*) (**Fig. 7**; **Supplementary Table S4**). The tree includes 33 *A. thaliana TPS* gene family members (Tholl and Lee 2011), to ensure the gene families are classified according to the known nomenclature. Based on the classification of TPS genes provided for *A. thaliana* (Tholl and Lee 2011), five major groups (TPS-a, b c, e/f and g) of *TPS* genes were identified in this set (indicated on the tree, **Fig. 7**).

The TPS-a family had a clear subdivision with the dicot (*A. thaliana*) in the TPS-a1 subgroup and the monocots (*Z. mays* and *O. sativa*) in the TPS-a2 subgroup (**Fig. 7**). The moss, *P. patens*, was limited to the TPS-e/f subgroup, along with three *S. moellendorfii* genes, while the majority of the *S. moellendorfii* genes are in the TPS-h group (not shown in **Fig. 7**). TPS-g had representation from *A. thaliana*, *Z. mays* and *O. sativa*. These results agree with the groupings of the *TPS* gene family found by Tholl and Lee (2011). The tree was then probed by overlaying the *plant anatomical entity* annotations hosted currently in the PO database (**Fig. 7**). The PO database currently includes a large number of annotations to the members of the *Z. mays* and Arabidopsis *TPS* families, but lacks extensive data linking *TPS* homologs in *O. sativa* and *S. moellendorfii*.

Based on the current set of annotations, we found that *A. thaliana TPS* genes for each of the subgroups indicate a widespread divergence of tissue- and cell type-specific expression profiles, while the *Z. mays* genes in the subgroups TPS-c and a2 indicate consistency in expression among the paralogs. The *A. thaliana* TPS-g gene AT1G61680 is preferentially annotated to reproductive plant structures compared with the TPS-g homologs from *Z. mays* that are preferentially expressed in vegetative structures. Also evident from this analysis was that the *Z. mays* TPS-a2 genes are expressed in the vegetative structures *leaves* and *primary root* and in the reproductive structures *floret* and *anther*, while the *A. thaliana* TPS-a1 family is more commonly expressed in the parts of the *flower* and *inflorescence*. From these results, guided by the placement

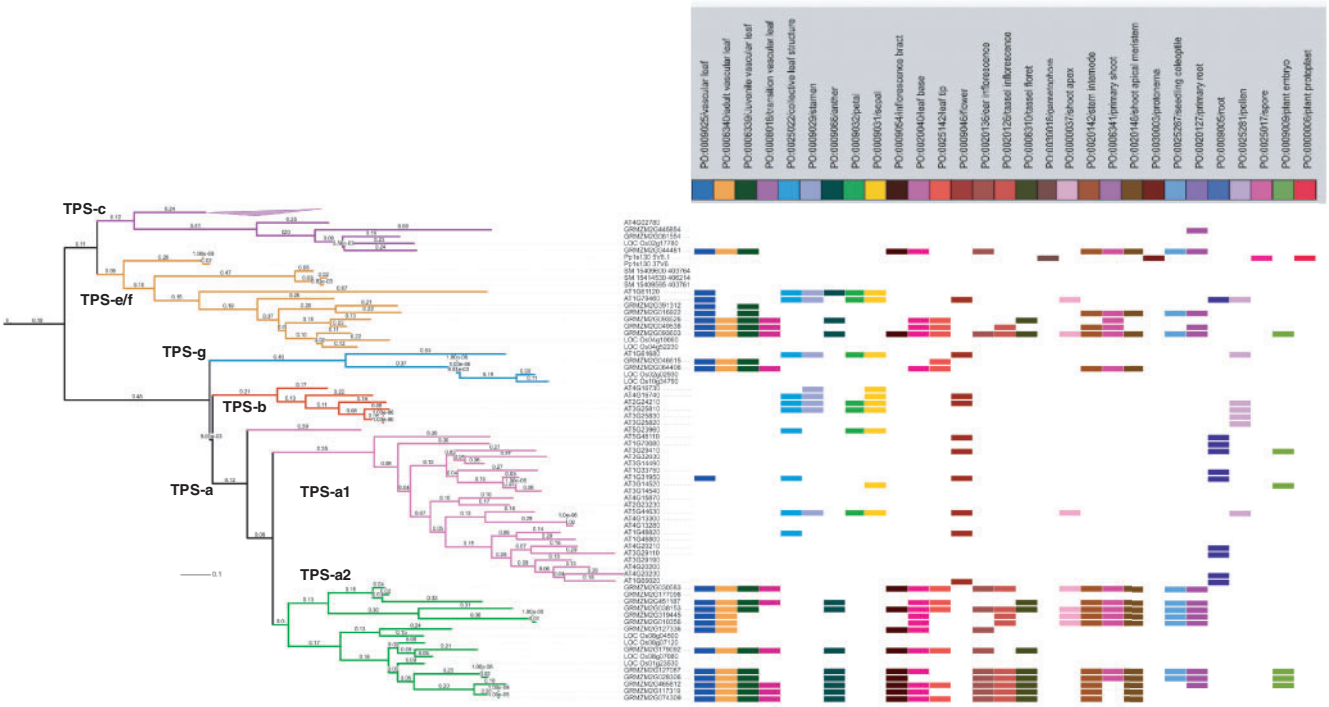

Fig. 7 Expression profiles of *TPS* orthologs based on annotations to plant structures in the PO. Using Arabidopsis *TPS* gene sequences, we identified the *TPS* homologs in four other species (*Zea mays*, *Oryza sativa*, *Physcomitrella patens* and *Selaginella moellendorfii*) and resolved their expression on a *TPS* gene family tree. Bioinformatics analysis of the expression of *TPS* genes was performed by aligning the genes annotated in the PO database to *plant anatomical entity* terms. Groups of the *TPS* gene family members are indicated on the gene family tree. Some branches were collapsed to avoid empty blocks due to unavailability of annotations for those genes. Branch lengths are shown on the gene tree. The iTOL (http://itol.embl.de/index.shtml) online tool was used to make this figure (Letunic and Bork 2007, Letunic and Bork 2011).

of *TPS* homologs in the gene family tree, one can hypothesize about gene expression in other closely related plant species, such as *O. sativa* and other monocots, and *S. moellendorfii*. For example, a user might expect to find the expression of the *S. moellendorfii TPS* genes in the *non-vascular leaves*. A recent study by Li and co-workers (2012) characterized the *TPS* genes in the above-ground portions of plants after treatment with a fungal elicitor, but, to our knowledge, no one has yet examined the tissue-specific expression of *TPS* genes in *Selaginella*.

The *Physcomitrella TPS* homolog Pp1s130_5V6.1 is annotated in the PO database to four plant structures: *gametophore*, *protenema*, *plant protoplast* and *plant spore* (**Fig. 7**). This gene was characterized as encoding an *ent*-kaurene synthase, *PpCPS/KS* (Hayashi et al. 2010). The gametophore is a shoot that bears *non-vascular leaves* (phyllids) and ultimately the *megagametophyte* and *microgametophyte*. Thus, by using the PO annotations, users can compare not only across taxa, but also across plant life cycles.

Integration of the PO in online plant genomics portals and databases

The PO is widely adopted among plant genomics databases and websites (**Table 4**). There are too many to describe them all in detail, but we present a few representative examples here.

The Arabidopsis Information Resource (TAIR). As a founding member of the Plant Ontology Consortium, TAIR (http://Arabidopsis.org/) has contributed to the development and use of the PO from its inception (Berardini et al. 2004, Jaiswal et al. 2005). TAIR's current participation in the PO consortium is through the large-scale contribution of PO annotations and new term requests. PO terms are used within TAIR to annotate Arabidopsis gene expression patterns reported in published research articles, along with the evidence supporting the annotations. A notable example of such a large-scale submission is the gene expression data from the multinational Arabidopsis expression atlas project (AtGenExpress) (Schmid et al. 2005), which resulted in 480,444 PO annotations. As of June 21, 2012, the combined efforts of TAIR curators and community data submitters have produced a total of 532,336 PO annotations for 20,007 Arabidopsis genes. A total of 397 distinct PO terms (326 *plant anatomical entities* and 71 *plant structure development stages*) have been used to capture Arabidopsis gene expression patterns. These annotations are based on experimental data from 2,123 research articles as well as from personal communications. TAIR's PO annotations are updated in the TAIR curation database and the TAIR website, and submitted to the PO SVN repository (http://palea.cgrb.oregonstate.edu/viewsvn/Poc/trunk/associations/) on a weekly basis. These new data are integrated into the PO database with

each PO release (roughly quarterly). Although the current TAIR annotation files may be accessed through the PO SVN repository site, they are not displayed on the PO browser until the next release.

The Sol Genomics Network (SGN). The SGN (http://solgenomics.net) database hosts genomic, phenotypic and taxonomic information on Solanaceae and related species, mostly from the asterid clade. As a clade-oriented database, SGN's main focus is to exploit the high level of genome conservation in the Solanaceae family for comparative querying of phenotype and genotype data. For this purpose, PO is extensively used for annotating functional genes, gene models and phenotyped germplasm, such as mutants and mapping populations. SGN also utilizes PO for scoring plant traits, thereby assisting quantitative and qualitative phenotyping in breeding programs. The two predominant species in SGN are *Solanum lycopersicum* (tomato) and *S. tuberosum* (potato), both having high-quality sequenced genomes (Potato Genome Sequencing Consortium 2011, Tomato Genome Consortium 2012). These species are important food crops and serve as models for studying developmental processes such as fruit ripening and tuberization. By including the required vocabulary for describing the plant anatomical entities and plant structure development stages in tomato and potato, the PO provides the resources to represent their counterparts in other Solanaceae species, such as *Solanum melongena* (eggplant), *Nicotiana tabacum* (tobacco) and *Capsicum annuum* (pepper). Overall, SGN has contributed more than 20,000 manually curated gene and phenotype annotations for 14 Solanaceae species, and plans to develop PO annotations for expression data for each published Solanaceae transcriptome in the near future.

The Maize Genetics and Genomics Database (MaizeGDB). The PO grew out of its third founding member MaizeGDB's (http://www.maizegdb.org/) contribution to maize-specific controlled vocabulary (Vincent et al. 2003). Currently, the maize data hosted in the PO database include genes, genetic stocks and gene models. Associations with 7,067 stocks and 11,436 alleles, representing 1,157 genes, are inferred from more than 800 phenotypes that are annotated with *plant anatomical entity* and/or *plant structure development stage* terms. The phenotype curation efforts have been mostly supplied by the Maize Genetics Cooperation Stock Center (Neuffer et al. 1997, Sachs 2009), with annotations to PO terms under the purview of MaizeGDB staff. A recent collaborative project involved associating PO terms to gene models from a comprehensive atlas of global transcription profiles across 60 combinations of plant structures and developmental stages of the maize inbred line B73 (Sekhon et al. 2011). In this project, each tissue sampled was annotated with both PO terms and the corresponding MaizeGDB-specific synonym. For example, the MaizeGDB record labeled 'tassel meiotic V18 B73' (http://www.maizegdb.org/cgi-bin/termrefs.cgi?id=2366346) is annotated in the PO to the *plant anatomical entity* term *tassel inflorescence*, as well as the *plant structure development stage* terms *D pollen mother cell meiosis stage* and *LP.18 eighteen leaves visible*. To make the gene expression data more interactive with genome data about other plants, MaizeGDB provides enhanced access to the PO. A stable reference page is provided for each expression experiment, which lists the PO terms and plant sample images. The PO database hosts about 1.5 million MaizeGDB annotations to 35,323 gene models. A new tool for phenotype query that leverages the PO is being developed at MaizeGDB. It will be similar to the tools described by Green et al. (2011) and Harnsomburana et al. (2011), which use parent and child terms, along with synonyms, to search both annotations and full text descriptions for any ontology supplied. Currently, you can search the prototype, VPhenoDBS:Maize (http://www.phenomicsworld.org) for associations to the GO, PO, and TO, returning both text data and any images associated with a phenotype.

Oryzabase Database. Oryzabase (http://www.shigen.nig.ac.jp/rice/oryzabaseV4) is an integrated database of rice science in Japan (Yamazaki et al. 2010) that has been continuously providing information such as traits, genes, mutants, wild rice collections and organ-specific developmental stages for more than 15 years. Most contents are available in both English and Japanese. The phenotype information of genes, mutants and wild rice as well as anatomical terms and developmental stages are annotated using the PO (Yamazaki and Jaiswal 2005). While the features of DNA sequences and enzyme names/reactions are mostly described in a common language of English, the phenotypes and anatomical names in Japanese have been used historically. Even though today all scientists publish their articles in English, it is still difficult for non-native English speakers to describe the exact meaning of each term of the PO in English. To overcome these difficulties and enable Japanese scientists to contribute more to the development of the PO, a newly introduced 'Japanese version of the PO browser' is available at www.shigen.nig.ac.jp/plantontology/ja/go.cgi and provides term names and keyword search of *plant anatomical entities* in both English and Japanese, allowing Japanese users to grasp the hierarchy of the PO intuitively.

Gramene database. The Gramene database (http://www.gramene.org) is a curated online resource for plant comparative plant genomics and genetics analysis (Liang et al. 2008, Jaiswal, 2011, Youens-Clark et al. 2011). As a founding member of the PO, Gramene has integrated PO in their spatial and temporal aspects of annotation of plant gene products and QTL phenotypes to describe the spatial and temporal associations. Gramene contributes by sharing their PO annotations for about 1,700 rice genes and about 8,500 QTLs in addition to requesting new terms required for annotating cereal crop genomes. The Gramene project team, in collaboration with Plant Ensembl (http://plants.ensembl.org), mirrors PO annotation in gene pages of Plant Ensembl. In a new collaboration with the European Bioinformatics Institute's ATLAS and Array projects,

the PO will be integrated into annotations of the source plant samples used in developing the microarray and RNA-seq transcriptome data sets submitted to these database archives for analysis. The Gramene project is also the primary developer of the TO for plants (Jaiswal et al. 2002, Yamazaki and Jaiswal 2005). The curators at Gramene and the PO are working on aligning the TO and PO as described earlier.

VirtualPlant. VirtualPlant (http://virtualplant.org) is a software platform designed to allow scientists to mine lists of genes, microarray experiments and gene networks from *A. thaliana* and to visualize, integrate and analyze genomic data from a systems biology perspective (Katari et al. 2010). The project's data browser provides access to the annotations and functional categories in the VirtualPlant database, including all of the *A. thaliana* annotations associated with PO terms.

The Plant Expression Database. The Plant Expression Database (PLEXdb; http://www.plexdb.org/) is a gene expression resource for plants and plant pathogens that leverages highly parallel expression data with portals to related genetic, physical and pathway data (Wise et al. 2008). PLEXdb provides access to whole-genome transcriptome expression data sets contributed by authors for barley, maize, rice, sugarcane, wheat, Arabidopsis, citrus, cotton, grape, *Medicago*, poplar, soybean and tomato. The PO is used by the resource for consistent annotation of the plant samples used as RNA library source in the experiments.

Conclusions and Future Directions

The standard names and definitions used in the *plant anatomical entity* branch of the PO constitute a controlled vocabulary that is designed to foster consistency in annotation and querying of genomics data sets such as gene expression profiles and phenotypes pertaining to plant anatomy. The consistent use of PO terms in annotations and publications will allow plant biologists and breeders to make meaningful cross-database and cross-species queries, in order to discover patterns of similarity and dissimilarity. This, in turn, will facilitate determination of the functions of genes and their genetic interactions associated with plant development processes, and thus their contribution to the agronomic and commercially significant traits, such as improved disease resistance and yield. Textual definitions provided for each term in the ontology serve to assist researchers in understanding the precise meaning of the term in question, while the logical definitions based on ontology relationships allow for different types of computer processing of the associated data (e.g. for purposes of cross-species integration or for data quality assurance).

Future versions of the ontology will probably include additional high-level terms that will allow some unique structures to be better classified and circumscribed. For example, *plant structure* includes three direct child terms (*plant ovary*, *trichome* and *rhizoid*) that cannot be categorized as child terms of any other *plant structures* in the current version of the PO, because they consist of more than one type of plant structure. We will be looking at closer integration of PO and GO by cross-referencing each with the other to suggest which plant-specific GO *biological processes*, *molecular functions* and *cellular components* are associated with respective *plant anatomical entity* terms from PO. One example is the *C4 photosynthesis* mentioned previously, and which is specific to *mesophyll cell* and cells of the *bundle sheath*. Future enhancements to the database would include an integrated tool to query gene homologs and their annotations, links from *plant structure* term pages to the images in image archives annotated with the PO terms and enrichment of annotations by adding gene and gene product annotations for existing and new species.

In summary, these examples demonstrate how the PO can serve as a reference ontology for all plants. The structure of *plant anatomical entity* and its child terms in the ontology will continue to be developed to describe and annotate plants from all taxa. This will set the stage for widening the scope of the genomic data annotated using terms from the PO and other ontologies.

Materials and Methods

Analysis of terpene synthase gene families using annotations to *plant anatomical entity* terms

Sequences in FASTA format of the 33 *A. thaliana TPS* gene family members (Tholl and Lee 2011) and the *TPS* homologs in four other species (*Z. mays*, *O. sativa*, *P. patens* and *S. moellendorfii*) were obtained from Gramene (http://www.gramene.org) and Phytozome (www.phytozome.org). The homologs retrieved from the two sources were further refined, by querying the homolog gene clusters generated in a large-scale analysis (done previously) by using a modified version of the InParanoid (Östlund et al. 2010) program (Shulaev et al. 2011). For this analysis, the primary homolog hits (score 1.0) were listed, plus any additional matches with a homology score $\geq$0.25, restricting the results to the canonical form of the gene model (longest transcript/peptide). The homolog list was compiled (see **Supplementary Table S4**) along with their protein sequences in FASTA format. In a further analysis, the *TPS* homolog sequences were analyzed using ClustalW (http://www.ebi.ac.uk/) and MUSCLE (Edgar 2004) at http://www.phylogeny.fr/ to create the best alignments. Branch lengths are shown on the gene tree and the tree is rooted between the higher plants *A. thaliana*, *Z. mays* and *O. sativa*, and the lower plants *P. patens* and *S. moellendorfii*. These alignments were then used to generate the *TPS* gene family tree by using the PhyML 3.0 tool (http://www.phylogeny.fr/). A series of MySQL searches were performed on the PO database using the *TPS* orthologs from Arabidopsis, maize and moss. The list of PO annotations for *TPS* orthologs from Arabidopsis, maize and moss was overlaid on the gene family at the Interactive

Tree of Life site (http://itol.embl.de/index.shtml (Letunic and Bork 2007, Letunic and Bork 2011) to create an integrated image (**Fig. 7**).

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was supported by the US National Science Foundation [grant No. IOS: 0822201].

Acknowledgments

The authors thank Chris Sullivan (Center for Genome Research and Biocomputing at Oregon State University) for hosting and maintenance of the Plant Ontology project web servers; David Cove, Tomoaki Nishiyama and Peter Szovenyi for contributing to the Moss Ontology (https://www.cosmoss.org/physcome_project/wiki/Moss_Ontology); the German Research Foundation/Deutsche Forschungsgemeinschaft (DFG; cosmoss.org; RE 837/10-2); Hong Cui and James Macklin (Flora of North America; http://floranorthamerica.org/), and Kevin C. Nixon (Cornell University) for access to and development of the Plantsystematics.org and Cornell University Plant Anatomy Collection (CU-PAC) databases and image servers; Rex Nelson (SoyBase), Fiona McCarthy and the curators at AgBase (http://www.agbase.msstate.edu/) for their help with term enrichment of and providing PO annotations; and numerous volunteers and reviewers of the Plant Ontology. We also thank the Gene Ontology Consortium (www.geneontology.org) for useful discussion and sharing the GO database package and the AmiGO ontology browser software (software developer Seth Carbon), that were customized to run the PO database and the ontology browser.

References

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408: 796–815.

Arp, R. and Smith, B. (2008) Function, role, and disposition in Basic Formal Ontology. *Nature Precedings* 1941:1, 1–4.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. et al. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25: 25–29.

Aubourg, S., Lecharny, A. and Bohlmann, J. (2002) Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* 267: 730–745.

Avraham, S., Tung, C.-W., Ilic, K., Jaiswal, P., Kellogg, E.A., McCouch, S. et al. (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res.* 36: D449–D454.

Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P. et al. (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol.* 135: 745–755.

Blake, J.A. and Bult, C.J. (2006) Beyond the data deluge: data integration and bio-ontologies. *J. Biomed. Inform.* 39: 314–320.

Bombarely, A., Menda, N., Tecle, I.Y., Buels, R.M., Strickler, S., Fischer-York, T. et al. (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res.* 39: D1149–D1155.

Bomblies, K. and Doebley, J.F. (2006) Pleiotropic effects of the duplicate maize *FLORICAULA/LEAFY* genes *zfl1* and *zfl2* on traits under selection during maize domestication. *Genetics* 172: 519–531.

Bomblies, K., Wang, R.-L., Ambrose, B.A., Schmidt, R.J., Meeley, R.B. and Doebley, J. (2003) Duplicate *FLORICAULA/LEAFY* homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development* 130: 2385–2395.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S. et al. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* 25: 288–289.

Chen, F., Tholl, D., Bohlmann, J. and Pichersky, E. (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66: 212–229.

Chujo, A., Zhang, Z., Kishino, H., Shimamoto, K. and Kyozuka, J. (2003) Partial conservation of *LFY* function between rice and *Arabidopsis*. *Plant Cell Physiol.* 44: 1311–1319.

Dahdul, W.M., Lundberg, J.G., Midford, P.E., Balhoff, J.P., Lapp, H., Vision, T.J. et al. (2010) The teleost anatomy ontology: anatomical representation for the genomics age. *Syst. Biol.* 59: 369–383.

Day-Richter, J., Harris, M.A., Haendel, M. The Gene Ontology OBO-Edit Working Group and Lewis, S. (2007) OBO-Edit an ontology editor for biologists. *Bioinformatics* 23: 2198–2200.

Deegan (nee Clark), J., Dimmer, E. and Mungall, C.J. (2010) Formalization of taxon-based constraints to detect inconsistencies in annotation and ontology development. *BMC Bioinformatics* 11: 530.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792–1797.

Fielding, R.T. (2000) Architectural Styles and the Design of Network-Based Software Architectures. University of California.

Garcia-Hernandez, M., Berardini, T.Z., Chen, G., Crist, D., Doyle, A., Huala, E. et al. (2002) TAIR: a resource for integrated *Arabidopsis* data. *Funct. Integr. Genomics* 2: 239–253.

Gene Ontology Consortium. (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.* 40: D559–D564.

Gessler, D., Schiltz, G., May, G., Avraham, S., Town, C., Grant, D. et al. (2009) SSWAP: a Simple Semantic Web Architecture and Protocol for semantic web services. *BMC Bioinformatics* 10: 309.

Gkoutos, G., Green, E., Mallon, A.-M., Hancock, J. and Davidson, D. (2004) Using ontologies to describe mouse phenotypes. *Genome Biol.* 6: R8.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296: 92–100.

Green, J.M., Harnsomburana, J., Schaeffer, M.L., Lawrence, C.J. and Shyu, C.-R. (2011) Multi-source and ontology-based retrieval engine for maize mutant phenotypes. *Database* 2011: bar012.

Grenon, P., Smith, B. and Goldberg, L. (2004) Biodynamic Ontology: applying BFO in the biomedical domain. *In* Ontologies in Medicine. Edited by Pisanelli, D.M. pp. 20–38. IOS Press, Amsterdam.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K. Jr, Hannick, L.I. et al. (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31: 5654–5666.

Haendel, M., Neuhaus, F., Osumi-Sutherland, D., Mabee, P.M., Mejino, J.L.V.J., Mungall, C.J. et al. (2007) CARO—the Common Anatomy Reference Ontology. *In* Anatomy Ontologies for Bioinformatics: Principles and Practice. Edited by Burger, A., Davidson, D. and Baldock, R. pp. 327–349. Springer, New York.

Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R. et al. (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant Cell Physiol.* 52: 220–229.

Harnsomburana, J., Green, J., Barb, A., Schaeffer, M., Vincent, L. and Shyu, C.-R. (2011) Computable visually observed phenotype ontological framework for plants. *BMC Bioinformatics* 12: 260.

Hayashi, K., Horie, K., Hiwatashi, Y., Kawaide, H., Yamaguchi, S., Hanada, A. et al. (2010) Endogenous diterpenes derived from ent-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss *Physcomitrella patens*. *Plant Physiol.* 153: 1085–1097.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L. et al. (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics* 2008: 1–5.

Iida, K., Kawaguchi, S., Kobayashi, N., Yoshida, Y., Ishii, M., Harada, E. et al. (2011) ARTADE2DB: improved statistical inferences for Arabidopsis gene functions and structure predictions by dynamic structure-based dynamic expression (DSDE) analyses. *Plant Cell Physiol.* 52: 254–264.

Ilic, K., Kellogg, E.A., Jaiswal, P., Zapata, F., Stevens, P.F., Vincent, L. 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., Rhee, S.Y., Kellogg, E.A. and Stevens, P.F. (2008) Plant Structure Ontology (PSO)—a morphological and anatomical ontology of flowering plants. *In* Anatomy Ontologies for Bioinformatics: Principles and Practice. Edited by Burger, A., Davidson, D. and Baldock, R. pp. 27–42. Springer, New York.

Jaillon, O., Aury, J.-M. and Noel, B. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.

Jaiswal, P. (2011) Gramene database: a hub for comparative plant genomics. *Methods Mol. Biol.* 678: 247–275.

Jaiswal, P., Avraham, S., Ilic, K., Kellogg, E.A., McCouch, S., Pujar, A. et al. (2005) Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comp. Funct. Genomics* 6: 388–397.

Jaiswal, P., Ware, D., Ni, J., Chang, K., Zhao, W., Schmidt, S. et al. (2002) Gramene: development and integration of trait and gene ontologies for rice. *Comp. Funct. Genomics* 3: 132–136.

Jensen, L.J. and Bork, P. (2010) Ontologies in quantitative biology: a basis for comparison, integration, and discovery. *PLoS Biol.* 8: e1000374.

Kakumanu, A., Ambavaram, M.M.R., Klumas, C., Krishnan, A., Batlang, U., Myers, E. et al. (2012) Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.* 160: 846–867.

Katari, M.S., Nowicki, S.D., Aceituno, F.F., Nero, D., Kelfer, J., Thompson, L.P. et al. (2010) VirtualPlant: a software platform to support systems biology research. *Plant Physiol.* 152: 500–515.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R. et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40: D1202–D1210.

Lang, D., Eisinger, J., Reski, R. and Rensing, S.A. (2005) Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism in mosses. *Plant Biol.* 7: 238–250.

Lawrence, C.J., Schaeffer, M.L., Seigfried, T.E., Campbell, D.A. and Harper, L.C. (2007) MaizeGDB's new data types, resources and activities. *Nucleic Acids Res.* 35: D895–D900.

Lens, F., Cooper, L., Gandolfo, M.A., Groover, A., Jaiswal, P., Lachenbruch, B. et al. (2012) An extension of the Plant Ontology project supporting wood anatomy and development research. *IAWA J.* 33: 113–117.

Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23: 127–128.

Letunic, I. and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39: W475–W478.

Li, G., Köllner, T.G., Yin, Y., Jiang, Y., Chen, H., Xu, Y. et al. (2012) Nonseed plant *Selaginella moellendorfii* has both seed plant and microbial types of terpene synthases. *Proc. Natl Acad. Sci. USA* 109: 14711–14715.

Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T. et al. (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.* 36: D947–D953.

Makita, Y., Kobayashi, N., Mochizuki, Y., Yoshida, Y., Asano, S., Heida, N. et al. (2009) PosMed-plus: an intelligent search engine that inferentially integrates cross-species information resources for molecular breeding of plants. *Plant Cell Physiol.* 50: 1249–1259.

Mandel, M.A. and Yanofsky, M.F. (1995) A gene triggering flower formation in *Arabidopsis*. *Nature* 377: 522–524.

McCarthy, F.M., Gresham, C.R., Buza, T.J., Chouvarine, P., Pillai, L.R., Kumar, R. et al. (2011) AgBase: supporting functional modeling in agricultural organisms. *Nucleic Acids Res.* 39: D497–D506.

Meehan, T., Masci, A., Abdulla, A., Cowell, L., Blake, J., Mungall, C.J. et al. (2011) Logical development of the Cell Ontology. *BMC Bioinformatics* 12: 6.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B. et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318: 245–250.

Mochida, K., Uehara-Yamaguchi, Y., Yoshida, T., Sakurai, T. and Shinozaki, K. (2011) Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. *Plant Cell Physiol.* 52: 785–803.

Mungall, C., Torniai, C., Gkoutos, G., Lewis, S. and Haendel, M. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* 13: R5.

Mungall, C.J., Bada, M., Berardini, T.Z., Deegan, J., Ireland, A., Harris, M.A. et al. (2011) Cross-product extensions of the Gene Ontology. *J. Biomed. Inform.* 44: 80–86.

Nagamura, Y., Antonio, B.A., Sato, Y., Miyao, A., Namiki, N., Yonemaru, J. et al. (2011) Rice TOGO Browser: a platform to retrieve integrated information on rice functional and applied genomics. *Plant Cell Physiol.* 52: 230–237.

Nelson, R., Avraham, S., Shoemaker, R., May, G., Ware, D. and Gessler, D. (2010) Applications and methods utilizing the Simple Semantic Web Architecture and Protocol (SSWAP) for bioinformatics resource discovery and disparate data and service integration. *BioData Mining* 3: 3.

Neuffer, M., Coe, E. and Wessler, S.R. (1997) Mutants of Maize. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Östlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D.N., Roopra, S. et al. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* 38: D196–D203.

Potato Genome Sequencing Consortium. (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I. et al. (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329: 223–226.

Pujar, A., Jaiswal, P., Kellogg, E.A., Ilic, K., Vincent, L., Avraham, S. et al. (2006) Whole-plant growth stage ontology for Angiosperms and its application in plant biology. *Plant Physiol.* 142: 414–428.

Rao, N.N., Prasad, K., Kumar, P.R. and Vijayraghavan, U. (2008) Distinct regulatory role for *RFL*, the rice *LFY* homolog, in determining flowering time and plant architecture. *Proc. Natl Acad. Sci. USA* 105: 3646–3651.

Rensing, S., Ick, J., Fawcett, J., Lang, D., Zimmer, A., Van de Peer, Y. et al. (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol. Biol.* 7: 130.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H. et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.

Rosse, C. and Mejino, J.L.V. Jr (2003) A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J. Biomed. Inform.* 36: 478–500.

Sachs, M.M. (2009) Cereal germplasm resources. *Plant Physiol.* 149: 148–151.

Saito, T., Ariizumi, T., Okabe, Y., Asamizu, E., Hiwasa-Tanase, K., Fukuda, N. et al. (2011) TOMATOMA: a novel tomato mutant database distributing Micro-Tom mutant collections. *Plant Cell Physiol.* 52: 283–296.

Sakurai, T., Kondou, Y., Akiyama, K., Kurotani, A., Higuchi, M., Ichikawa, T. et al. (2011) RiceFOX: a database of *Arabidopsis* mutant lines overexpressing rice full-length cDNA that contains a wide range of trait information to facilitate analysis of gene function. *Plant Cell Physiol.* 52: 265–273.

Schaeffer, M.L., Harper, L.C., Gardiner, J.M., Andorf, C.M., Campbell, D.A., Cannon, E.K.S. et al. (2011) MaizeGDB: curation and outreach go hand-in-hand. *Database* 2011: Article ID bar022.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M. et al. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37: 501–506.

Schultz, E.A. and Haughn, G.W. (1991) *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell* 3: 771–781.

Schulz, S., Daumke, P., Smith, B. and Hahn, U. (2005) How to distinguish parthood from location in bio-ontologies. *AMIA Annu. Symp. Proc.* 2005: 669–673.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Robin Buell, C., De Leon, N. et al. (2011) Genome-wide atlas of transcription through maize development. *Plant J.* 66: 553–563.

Shrestha, R., Arnaud, E., Mauleon, R., Senger, M., Davenport, G.F., Hancock, D. et al. (2010) Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature. *AoB Plants* 2010: plq008.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L. et al. (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43: 109–116.

Siriwardana, N.S. and Lamb, R.S. (2012) The poetry of reproduction: the role of *LEAFY* in *Arabidopsis thaliana* flower formation. *Int. J. Dev. Biol.* 56: 207–221.

Smith, B. (2012) On classifying material entities in Basic Formal Ontology. *In* Interdisciplinary Ontology. Proceedings of the Third Interdisciplinary Ontology Meeting pp. 1–13. Keio University Press, Tokyo.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W. et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25: 1251–1255.

Smith, B., Ceusters, W., Klagges, B., Kohler, J., Kumar, A., Lomax, J. et al. (2005) Relations in biomedical ontologies. *Genome Biol.* 6: R46.

Sohngen, C., Chang, A. and Schomburg, D. (2011) Development of a classification scheme for disease-related enzyme information. *BMC Bioinformatics* 12: 329.

Tanahashi, T., Sumikawa, N., Kato, M. and Hasebe, M. (2005) Diversification of gene function: homologs of the floral regulator *FLO/LFY* control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* 132: 1727–1736.

Tholl, D. and Lee, S. (2011) Terpene specialized metabolism in *Arabidopsis thaliana*. *In* The Arabidopsis Book 9. E0143 American Society of Plant Biologists, Rockville, MD.

Timmerhaus, G., Hanke, S.T., Buchta, K. and Rensing, S.A. (2011) Prediction and validation of promoters involved in the abscisic acid response in *Physcomitrella patens*. *Mol. Plant* 4: 713–729.

Tomato Genome Consortium. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635–641.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U. et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R. and Birney, E. (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19: 327–335.

Vincent, L., Coe, E.H. and Polacco, M.L. (2003) *Zea mays* ontology—a database of international terms. *Trends Plant Sci.* 8: 517–520.

Walls, R.L., Athreya, B., Cooper, L., Elser, J., Gandolfo, M.A., Jaiswal, P. et al. (2012a) Ontologies as integrative tools for plant science. *Amer. J. Bot.* 99: 1263–1275.

Walls, R.L., Cui, H., Macklin, J.A., Mungall, C., Cooper, L.D., Stevenson, D.W. et al. (2012b) Mapping of glossary terms from the Flora of North America to the Plant Ontology enhances both resources. *In* Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012). Edited by Cornet, R. and Steven, R.in . Graz, Austria.

Ware, D.H., Jaiswal, P., Ni, J., Yap, I.V., Pan, X., Youens-Clark, K. et al. (2002) Gramene, a tool for grass genomics. *Plant Physiol.* 130: 1606–1613.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992) *LEAFY* controls floral meristem identity in Arabidopsis. *Cell* 69: 843–859.

Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T. et al. (2011) BioPortal: enhanced functionality via new web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* 39: W541–W545.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M. et al. (2009) SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 37: D380–D386.

Wise, R.P., Caldo, R.A., Hong, L., Shen, L., Cannon, E. and Dickerson, J.A. (2008) BarleyBase/PLEXdb. *In* Plant Bioinformatics. Edited by Edwards, D. pp. 347–363. Humana Press, Totowa, NJ.

Wolf, L., Rizzini, L., Stracke, R., Ulm, R. and Rensing, S.A. (2010) The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-B radiation. *Plant Physiol.* 153: 1123–1134.

Yamaguchi, N., Yamaguchi, A., Abe, M., Wagner, D. and Komeda, Y. (2012) *LEAFY* controls *Arabidopsis* pedicel length and orientation by affecting adaxial–abaxial cell fate. *Plant J.* 69: 844–856.

Yamazaki, Y. and Jaiswal, P. (2005) Biological ontologies in rice databases. an introduction to the activities in Gramene and Oryzabase. *Plant Cell Physiol.* 46: 63–68.

Yamazaki, Y., Sakaniwa, S., Tsuchiya, R., Nonomura, K.-I. and Kurata, N. (2010) Oryzabase: an integrated information resource for rice science. *Breed. Sci.* 60: 544–548.

Yoder, M.J., Mikó, I., Seltmann, K.C., Bertone, M.A. and Deans, A.R. (2010) A gross anatomy ontology for Hymenoptera. *PLoS One* 5: e15991.

Youens-Clark, K., Buckler, E., Casstevens, T., Chen, C., DeClerck, G., Derwent, P. et al. (2011) Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 39: D1085–D1094.

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296: 79–92.