Plant-Pathogen Interactions Ontology (PPIO)

Alejandro Rodríguez Iglesias, Mikel Egaña Aranguren, Alejandro Rodríguez González, and Mark D. Wilkinson

Biological Informatics Group
Centre for Plant Biotechnology and Genomics (CBGP)
U.P.M. – I.N.I.A.
28223 Pozuelo de Alarcón (Madrid)
Spain
`alejandro.rodriguez.iglesias@alumnos.upm.es`
`{mikel.egana.aranguren,alejandro.rodriguezg,mark.wilkinson}@upm.es`
`http://www.wilkinsonlab.info`

Abstract. Plant-pathogen interactions are an important knowledge domain within plant biotechnology, both scientifically and in economic terms. Unlike other knowledge domains within life sciences, however, semantic technologies have not been used extensively to codify it, therefore, there is a lack of axiomatic models amenable to automated integration and inference. We present the Plant-Pathogen Interactions Ontology (PPIO), a first step towards the axiomisation of plant-pathogen interactions that results in a model that encourages consistent annotation and supports both query and inference.

Keywords: Plant-pathogen interaction, Ontology, OWL, RDF, Life Sciences Semantic Web

1 Introduction

Plant-pathogen interactions are the subject of intensive research, since crop losses cause considerable economical loss. Despite its importance, the knowledge domain has not been semantically codified to date, even though robust semantic technologies capable of the task do exist (see Section 2). This makes the data and knowledge of the domain computationally opaque, which hinders inferencing and integration with other information resources. We present the Plant-Pathogen Interactions Ontology[1] (PPIO), an ontology that describes plant-pathogen interactions and offers a scaffold into which important domain data can be embedded in a precise and computationally-transparent manner (see Section 4).

This platform combines previous ontology-related data extracted from other sources such as the Plant Ontology Consortium Database[2], with newly created ontology terms covering this specific knowledge domain. We utilized the infection process of *Pseudomonas syringae pv. tomato (PsPto) DC3000* as a template to

---

[1] `http://code.google.com/p/plant-pathogen-interacions-ontology/`

[2] `http://www.plantontology.org/`

identify new terms that describe the plant-pathogen interactions that take place during disease establishment (see Section 3). The aim is to develop a system that integrates data related with both plant physiology and plant pathology, in order to leverage the combined knowledge of both domains to assist in interpretation of plant-pathogen phenotypic responses and disease processes. This development will make it possible for researchers to pose questions like the following (and more) and obtain a meaningful answer:

- Is *Arabidopsis thaliana* susceptible to the attack by *PsPto*?
- What is the mode-of-entry for *Pseudomonas syringae pv. phaseolicola*?
- Do the pathogens *PsPto DC300* and *Pseudomonas syringae pv. phaseolicola* produce similar symptoms, and why?

2 Life Sciences Semantic Web and Plant Biotechnology

The Life Sciences Semantic Web is the result of the application of semantic technologies (RDF[3], OWL[4], and automated reasoners) to the problem of knowledge management in life sciences [8]. It is a network of data, ontologies and associated services that can be used to infer new knowledge automatically; this knowledge would be very difficult to obtain by other traditional means such as, for example, Web screen scraping.

Certain life sciences projects stand out in this endeavour through their extensive use of semantics to aid in automated data integration from non-collaborating resources: the OBO foundry [18], a collection of biomedical ontologies including Gene Ontology (GO) [7] and Cell Type Ontology [1]; Bio2RDF [2], an RDF resource that represents the most important Bioinformatics Data Bases; the W3C Semantic Web for Health Care and Life Sciences Interest Group[5]; *etc.* In the context of plant biotechnology, however, semantic technologies have only been applied to a limited number of domain resources. For example, the previously mentioned Plant Ontology Consortium developed the Plant Ontology [11] to describe plant anatomy and developmental stages. The Plant Trait Ontology [12] is related to the Gramene project[6] and defines a vocabulary for describing the specific appearance or qualities of various plant anatomical structures. Finally, the contribution most directly related to PPIO is the creation of the Plant Disease Ontology [19] by the extension of the human Infectious Diseases Ontology (IDO) [5] (see Section 4 for a comparison).

3 *Pseudomonas syringae* as an archetype of plant-pathogen interactions

Many bacterial genera are responsible for causing diseases in different plant species. For example, the *Pseudomonas* genus contains different members, some

---

[3] `http://www.w3.org/standards/techs/rdf`

[4] `http://www.w3.org/standards/techs/owl`

[5] `http://www.w3.org/blog/hcls/`

[6] `http://www.gramene.org`

of them lacking pathogenic life cycles, such as *Pseudomonas putida*, used commonly as a bioremediation agent. Other members, however, do exhibit pathogenic activity. *Pseudomonas syringae* is an extensively researched example, and is the species we used for our initial ontological modeling for this reason; the extensive knowledge about its infection process allows us to achieve good coverage of the axioms required to model a variety of plant-pathogen interaction phenomena, while examining only one pathogen.

Different strains of *Pseudomonas syringae* differ mainly in their host specificity [9]. Tomato crops, for instance, may be susceptible to *Pseudomonas syringae pv. tomato (PsPto) DC3000*. This bacterium causes the plant disease known as bacterial speck of tomato, characterized by the presence of necrotic lesions in the affected tissue, often surrounded by chlorotic halos. The disease this bacterium causes is known to result in extensive economic loss, and is a worldwide extended significant disease. Fortunately, this bacterium is also a pathogen for the model plant *Arabidopsis thaliana* [20], and as a result there is an unusual amount of high-quality data about all stages of infection and disease-progression. The infection process begins after the bacteria encounters a susceptible host plant (Figure 1). At this point, it grows epiphytically on the leaf surface and the entry through the leaf stomata takes place. Once PsPto has entered the plant and colonization of the apoplastic space has occurred, necrotic symptoms can be visible. Non-host and resistant plants trigger a series of molecular responses leading to the production of the hypersesintive response, therefore activating the plant's resistance pathways against this pathogen.

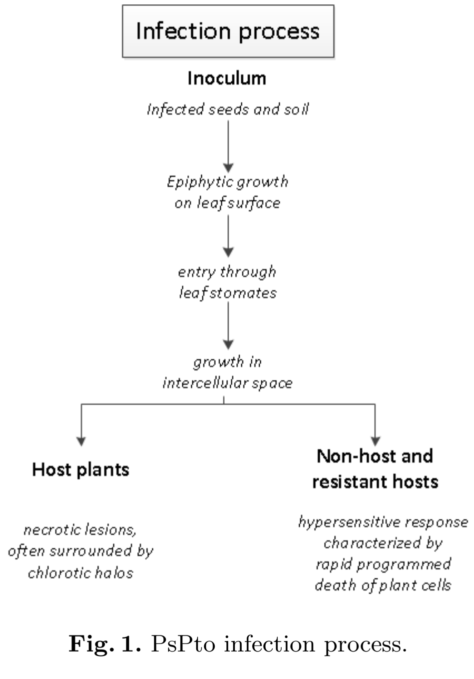

Fig. 1. PsPto infection process.

4 Plant-Pathogen Interactions Ontology

The justification for building PPIO lies in its coverage of a domain not represented by other ontologies. PPIO has been constructed, in part, through integrating terms and axioms from other ontologies, including the Plant Ontology and Plant Trait Ontology, and adding new axioms and terms that describe pathogenic effects. The Plant Ontology is a Knowledge Base (KB) focused on the description of plant anatomy and the developmental/morphological stages of plants. The Plant Trait Ontology has been developed to describe plant phenotypic traits with an ontology-based vocabulary. To complement these two KBs, PPIO is being developed from plant-pathogen interaction point of view, and intends to describe all biological features related to this process.

The basic structure of PPIO is shown in Figures 2 and 3. PPIO follows a "minimal axiomatic commitment" strategy: it sets the ground for further modeling, offering a "skeleton", expressive enough for modeling what a plant pathologist needs to describe but not overly expressive such as to be demanding in terms of axiomatic commitment (the commitment needed by a prospective ontologist is minimal). Therefore it differs from the ontology described in [19], which is based in the Basic Formal Ontology [17] and hence demands from the modeler to comply with a concrete "view of the world".

For modeling taxa PPIO follows the simplest pattern described in [16], thus each taxon is modeled as an OWL class reproducing the Linnean taxonomy in a simple class-subclass hierarchy. Taxonomic information is directly retrieved from the NCBI taxonomy [15] and with each release of the PPIO is added to the ontology through the OWL API [10] at design-time, automating the prospective addition of more taxa as needed, in particular, as additional disease-causing bacteria are added to the KB.

5 Conclusion

Plant diseases provoked by microorganisms have emerged as an important problem worldwide, causing substantial losses in many susceptible crops. Plant-pathogen interaction studies have discovered a wide range of molecular and physiological aspects of this process [6]; however there is currently no formal model for capturing this information. The PPIO is designed to re-use existing plant-related ontologies, but is the first ontological platform that encompasses and integrates terms related to plant pathogenicity and its phenotypic consequences.

PPIO has been developed using PsPto infection as its initial modeling problem due to the extensive knowledge about this pathogenic process, but will now be applied to other well-studied plant pathogens to demonstrate its wider utility as a generic PPI knowledge model.

In the future, PPIO will be used to build a Plant-Pathogen Interactions KB (PPIKB), along the lines of, for example, KUPKB [13], allowing for querying, visualization, knowledge management, *etc.* We also plan to add GO annotations

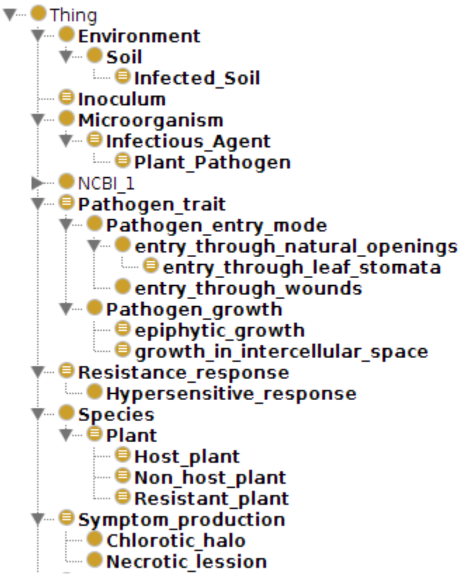

Fig. 2. OWL Class hierarchy of PPIO, showing its main components: `Environment`, `Pathogen`, `Species`, *etc.*

[4] and novel annotations produced in-house by the CBGP wet labs, in a KB like GOAL [14]. PPIKB will be made part of the Linked Open Data (LOD) cloud [3], since it describes a novel information domain. Finally, we also plan to apply inference to discover new knowledge implicit in the KB.

With this tool, we believe pathologists will be able to gain an insight into new molecular and physiological processes, enabled by the ability to easily integrate data from a wide range of pathogenic processes and species.

Acknowledgments

Alejandro Rodríguez Iglesias and Alejandro Rodríguez González are funded by the Isaac Peral Programme. Mikel Egaña Aranguren and Mark D. Wilkinson are funded by the Marie Curie Cofund programme (FP7).

References

1. Jonathan Bard, Seung Y. Rhee, and Michael Ashburner. An ontology for cell types. *Genome biology*, 6(2):R21+, 2005.
2. F. Belleau, M. Nolin, N. Tourigny, P. Rigault, and J. Morissette. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, 41(5):706–716, 2008.

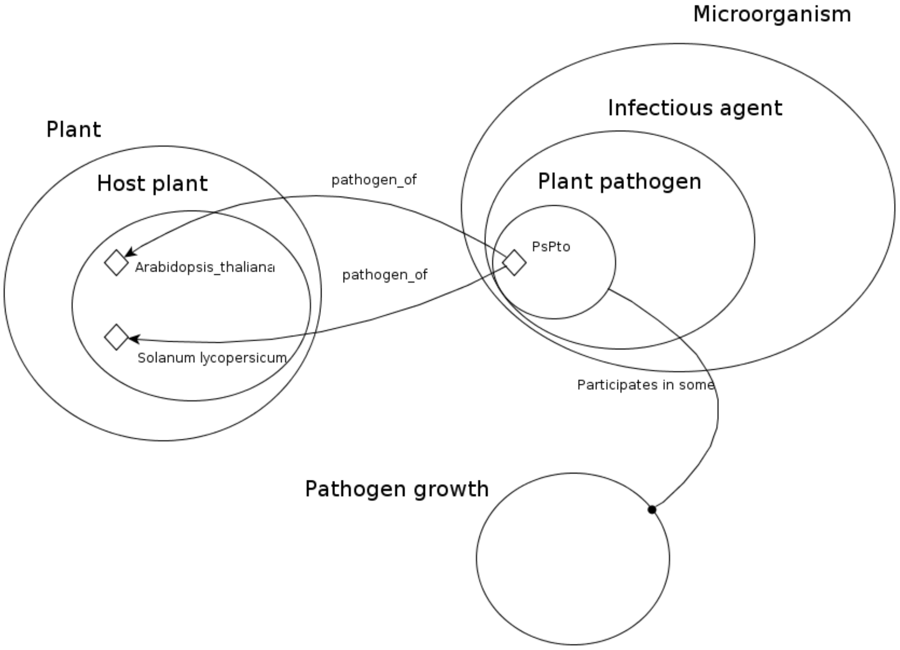

Fig. 3. Detailed depiction of the OWL axioms related to the class `Plant pathogen` and the individual `PsPto`. Circle: class; Diamond: instance; Arrow: triple; Line with circle at the end: restriction (Anonymous OWL class not shown for clarity).

3. Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.
4. Evelyn Camon, Michele Magrane, Daniel Barrell, Vivian Lee, Emily Dimmer, John Maslen, David Binns, Nicola Harte, Rodrigo Lopez, and Rolf Apweiler. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res*, 32:D262–D266, Jan 2004.
5. Lindsay Grey Cowell and Barry Smith. Infectious disease ontology. In Vitali Sintchenko, editor, *Infectious Disease Informatics*, pages 373–395. Springer New York, 2010.
6. Peter N. Dodds and John P. Rathjen. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature reviews. Genetics*, 11(8):539–548, 2010.
7. Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 23(May):25–29, 2000.
8. Benjamin M. Good and Mark D. Wilkinson. The life sciences semantic web is full of creeps! *Briefings in Bioinformatics*, 7(3):275–286, 2006.
9. S. S. Hirano and C. D. Upper. Bacteria in the leaf ecosystem with emphasis on pseudomonas syringae-a pathogen, ice nucleus, and epiphyte. *Microbiology and molecular biology reviews : MMBR*, 64(3):624–653, 2000.
10. Matthew Horridge and Sean Bechhofer. The OWL API: A java API for OWL ontologies. *Semantic Web Journal*, 2(1):11–21, 2011.
11. Pankaj Jaiswal, Shulamit Avraham, Katica Ilic, Elizabeth A. Kellogg, Susan McCouch, Anuradha Pujar, Leonore Reiser, Seung Y. Rhee, Martin M. Sachs, Mary Schaeffer, Lincoln Stein, Peter Stevens, Leszek Vincent, Doreen Ware, and Felipe Zapata. Plant ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comp Funct Genom*, 6(7-8):388–397, 2005.
12. Pankaj Jaiswal, Doreen Ware, Junjian Ni, Kuan Chang, Wei Zhao, Steven Schmidt, Xiaokang Pan, Kenneth Clark, Leonid Teytelman, Samuel Cartinhour, Lincoln Stein, and Susan McCouch. Gramene: development and integration of trait and gene ontologies for rice. *Comp Funct Genom*, 3(2):132–136, 2002.
13. Simon Jupp, Julie Klein, Joost Schanstra, and Robert Stevens. Developing a kidney and urinary pathway knowledge base. *Journal of Biomedical Semantics*, 2(Suppl 2):S7+, 2011.
14. Simon Jupp, Robert Stevens, and Robert Hoehndorf. Logical gene ontology annotations (GOAL): exploring gene ontology annotations with OWL. *Journal of biomedical semantics*, 3 Suppl 1, 2012.
15. E.W. Sayers, T. Barrett, D.A. Benson, E. Bolton, S.H. Stephen, K. Canese, V. Chetvernin, D.M. Church, M. Deanna, M. DiCuccio, S. Federhen, M. Feolo, Lewis Y. L.Y.Geer, Wolfgang Helmberg, Yuri Kapustin, David Landsman, David J. Lipman, Zhiyong Lu, Thomas L. Madden, Tom Madej, Donna R. Maglott, Aron Marchler-Bauer, Vadim Miller, Ilene Mizrachi, James Ostell, Anna Panchenko, Kim D. Pruitt, Gregory D. Schuler, Edwin Sequeira, Stephen T. Sherry, Martin Shumway, Karl Sirotkin, Douglas Slotta, Alexandre Souvorov, Grigory Starchenko, Tatiana A. Tatusova, Lukas Wagner, Yanli Wang, W. John Wilbur, Eugene Yaschenko, and Jian Ye. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 38(suppl 1):D5–D16, 2010.
16. S. Schulz, H. Stenzhorn, and M. Boeker. The ontology of biological taxa. *Bioinformatics (Oxford, England)*, 24(13), 2008.
17. Barry Smith. The basic tools of formal ontology. In *Formal Ontology in Information Systems*, 1998.

18. Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis Goldberg, Karen Eilbeck, Amelia Ireland, Christopher Mungall, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard Scheuermann, Nigam Shah, Patricia Whetzel, and Suzanna Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255, November 2007.
19. Ramona Walls, Barry Smith, Justin Elser, Albert Goldfain, Dennis W. Stevenson, and Pankaj Jaiswal. A plant disease extension of the infectious disease ontology. In Ronald Cornet and Robert Stevens, editors, *ICBO*, 2012.
20. M. C. Whalen, R. W. Innes, A. F. Bent, and B. J. Staskawicz. Identification of pseudomonas syringae pathogens of arabidopsis and a bacterial locus determining avirulence on both arabidopsis and soybean. *The Plant cell*, 3(1):49–59, 1991.