# Plant-Pathogen Interactions Ontology (PPIO)

Alejandro Rodríguez Iglesias [a,*], Mikel Egaña Aranguren [a] Alejandro Rodríguez González [a]
Mark D. Wilkinson [a]

[a] *Biological Informatics Group, Centre for Plant Biotechnology and Genomics (CBGP), Technical University of Madrid (UPM), Spain*

**Abstract.** Plant-pathogen interactions are an important knowledge domain within plant biology and 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 axiomatization of plant-pathogen interactions knowledge. PPIO encourages consistent annotation and supports both query and inference.

Keywords: Plant pahogenic bacteria, Ontologies, Semantic Web, PPIO

## 1. Introduction

Plants can be susceptible to the attack of many different pathogenic bacterial genera [12], and if the different levels of the host plant defense barriers are overcomed by the pathogen, the infection process can ultimately lead to the death of the plant. This phenomenon affects crop yield, which automatically will translate into considerable economic loss [13]. Therefore, being able to accurately explore and query biological data regarding plant-pathogenic bacteria interactions is a necessary step towards the improvement and protection of worldwide crops. This issue continues being subject of intensive research worldwide. Hundreds of articles that focus on the mechanisms of pathogenic bacteria interactions with their hosts reflect its biological significance[6] [7].

The area of plant-pathogenic bacteria interactions is an illustrious example of biodiversity richness [3]. Traditionally, *Agrobacterium, Erwinia, Pseudomonas and Xanthomonas* were considered the four main plant pathogenic genera some years ago. Nowadays, and thanks to the development of new sequencing techniques, the number of plant pathogenic bacteria genera has increased up to 30. One of the direct consequences from these discoverys is the generation of vast amounts of unexplored data. Semantic technologies have proven extremely effective and have been widely adopted in other branches of the life sciences community. However, they have not been extensively applied to the knowledge domain of plant-pathogenic bacteria to date. Despite the clear need for automation of data annotation, categorization, integration, interpretation, and quality-control, the data and knowledge in this domain remains largely computationally opaque, limiting the utility and usability of these important datasets. Here we present the Plant Pathogen Interactions Ontology (PPIO), an ontology developed to semantically collect data in the context of the plant-pathogenic bacteria interactions. The main goal of PPIO is to ultimately serve as a reference ontology in this area, providing the knowledge necessary to assist in the interpretation of the phenotypic responses that result from the interaction established during the infection of the pathogenic bacteria and the potential host plants.

## 2. Modelling

The initial data necessary to start building the ontology was collected manually by consulting a number of different of web resources. The web page http://pseudomonas-syringae.org/ was the starting point; it contains diverse pieces of state-of-art information re-

*Corresponding author. Email:alejandroriglesias@gmail.com

lated with various *Pseudomonas syringae* strains. This web page was a bridge to other web resources where more datasets were collected, like the (NCPPB, etc). After an initial collection of data was gathered, the modelling of these datasets was performed using the ontology editor Protégé (version).

### 2.1. Desing principles

The main goal pursued during the modelling and designing process was to semantically capture as many biological data existing as possible, and special effort is being done in accurately modelling the *disease triangle* [1] . This term is one of the most essential milestones in the plant pathology ambit, and it principal statement asserts that three factors must be present for a disease to occur: the virulent pathogen, a susceptible host and a propitious environment for disease progression. Two classes have been created to represent these three elements, the Environmental_parameter and the Organism classes. (introducing more text here about the relation of these classes). Physiological state of plants can be inferred visually by observing plant phenotypes. Thus, significant work/endeavour is also being done in the modelling of plant phenotypic representations in PPIO. To this end, a number of classes have been specifically created to represent plant phenotypic traits in a precise and pertinent manner, being especially important the Phenotpyic process and the Phenotype? Classes (figura con un esquema de la jerarquÃ■a de clases). These two classes semantically illustrate the output of the interaction between the host and the bacteria.

## 3. Creation methodology

### 3.1. URI design

The ontology URI (`http://purl.oclc.org/PPIO`) is HTTP resolvable and permanent (the PURL server redirects to our current server at `biordf.org`). The identifiers for entities (classes, individuals and object properties) are alphanumeric, with a URI of the type `http://purl.oclc.org/PPIO#PPIO_NNNNNNN`, and every entity has an informative `rdfs:label` annotation. Currently hash URIs are used due to the small size of the ontology, but since URIs are generated programmatically[2], when the ontology grows into a Knowledge Base or Linked Data dataset (see section **??**) slash URIs can be generated.

### 3.2. Ontology production

The development of PPIO is automated as much as possible. Once the main structure is set, most of the remaining parts are produced programmatically using the Galaxy platform, a bioinformatics-oriented workflow environment [8]. By using Galaxy, the specific workflow we need is defined once and executed for each release; also, we can plug PPIO directly with other Bioinformatics tools.

The workflow adds the necessary entities and axioms[3] (Figure **??**):

1. The organism taxa hierarchy is produced by the tool NCBITaxonomy2OWL[4]: it gets the user-defined taxa from the NCBI taxononomy through a BioPortal Web Service [**?** ] and injects them in PPIO, reproducing the original taxonomical hierarchy (representing each rank-subrank as a simple subsumption relation [**?** ]) and adding each taxon with a resolvable OntoBee[5] URI.

2. Since pathogens in PPIO are modelled as individuals, they cannot be directly related with class hierarchies like the NCBI taxonomy and the symptoms hierarchy. Therefore, PPIO exploits OWL punning[6] and an individual with the same URI as each type class is generated programmatically (see bellow) for those hierarchies: the linking of pathogens to those hierarchies (*e.g.* `NCBITaxon_552 types Erwinia amylovora`, `NCBITaxon_552 causes Canker`, `NCBITaxon_552 causes Blight`) is done manually. This is achieved by defining two Ontology Pre Procesor Language (OPPL)[7] scripts and executing them via OPPL-Galaxy [1][8]:

---

[2] `https://github.com/wilkinsonlab/OWLNumericIDGenerator`

[3] The workflow can be reproduced at `http://biordf.org:8090/u/alejandroriglesias/w/ppio-taxa-punning`

[4] `https://github.com/wilkinsonlab/NCBITaxonomy2OWL`

[5] `http://www.ontobee.org/`

[6] `http://www.w3.org/TR/owl2-new-features/Punning`

[7] `http://oppl.sf.net`

[8] Since we are using OPPL, any complex axiomisation -not only puning- can be defined once and automatically applied -expanded to different parts of the ontology- every time the workflow is executed.

---

[1] `http://www.apsnet.org/edcenter/instcomm/TeachingArticles/Pages/DiseaseTriangle.aspx`

```
?x:CLASS,
?y:INDIVIDUAL = create(?x.RENDERING)
SELECT ?x SubClassOf NCBITaxon_1
WHERE ?x != Nothing, ?x != Thing
BEGIN
ADD ?y Type ?x
END;

?x:CLASS,
?y:INDIVIDUAL = create(?x.RENDERING)
SELECT ?x SubClassOf PPIO_0000069
WHERE ?x != Nothing, ?x != Thing
BEGIN
ADD ?y Type ?x
END;
```

## 4. Discussion

Certain semantics-oriented projects are particularly noteworthy as a result of their success in using semantics to aid in automated data integration from non-collaborating resources, like the OBO foundry [14],which includes Gene Ontology (GO) [4], Bio2RDF [2] and the W3C Semantic Web for Health Care and Life Sciences Interest Group[9]. In the context of plant biotechnology and phytopathology, however, semantic technologies have been applied to only a limited number of domain resources. For example, the Plant Ontology Consortium developed the Plant Ontology[10] platform to describe the anatomy, morphology and developmental stages of plants [9]. The Plant Trait Ontology [10], related to the Gramene project[11], defines a vocabulary for describing the specific appearance or qualities of various plant anatomical structures. The Plant Disease Ontology (IDOPlant) [15], was modeled after the human Infectious Diseases Ontology (IDO) [5] to serve as a reference ontology that contained and covered any plant infectious disease. Apart from IDO-Plant, there is another contribution directly related to the plant-pathogenic bacteria area, and is the GO extension for description of the Type III Effectors [11], which is part of the Plant-Associated Microbe Gene Ontology project (PAMGO).

Although the existence of ontologies related with plants and plant pathology has already been reported [10] [11] [15], the justification for building this ontology lies in its coverage of a domain not fully represented by other resources. In a comparison between the IDOPlant and PPIO, a more generalistic approach of data modelling can be seen in the case of the first platform, which semantically describes plant infectious diseases caused by either biotic or abiotic agents . The ontology reported here pursues a knowledge capture strategy specifically focused on data that concerns plant-pathogenic bacteria interactions. On the other hand, although the GO extension for the type III effectors is built for capturing processes in the host-pathogen level, effector proteins data capture is emphasized. Of course, PPIO has been developed in order to be used with these ontological resources. In this aspect, PPIO complements these previous ontologies by introducing accurate and concrete biological information concerning plant pathogenic bacteria and the interactions with plants. The final goal of this initiative is to use this platform, combined with others, as a diagnosis/prevention/alert system. PPIO will make it possible for users to pose, and answer, questions like the following ones and obtain a meaningful answer:

1. Is tomato plant susceptible to the attack of *Pseudomonas syringae* pv. *tomato* DC3000?
2. Does a high humidity and and low temperature rates favour the development of *Pectobacterium carotovorum* subsp. *carotovorum*?
3. What is the phenotype of the disease produced by *Dickeya dadantii* in *Solanum tuberosum*?

Knowledge acquisition is based on the process of transcripting the knowledge from unstructured sources into a format that is machine-readable and useful. This approach can report a great benefit if big amounts of datasets are required in order to populate ontologies. One of the main ideas behind this that the participation of field experts should ensure the fiability of the data content captured. A knowledge capture project is now being developed in our laboratory, and it is our thought that it will aid in the process of collecting data that will ultimately populate PPIO.

---

[9]http://www.w3.org/blog/hcls/
[10]tp://www.plantontology.org/
[11]http://www.gramene.org

Fig. 1. Galaxy workflow for producing a release of PPIO. In the first step, NCBITaxonomy2OWL is executed; it gets the ontology and a flat file containing the NCBI taxonomy IDs, and it adds them to the ontology. Then two OPPL scripts are executed against the resulting ontology, adding axioms and entities to create

## References

[1] M. E. Aranguren, J. T. F. Breis, E. Antezana, C. Mungall, A. R. González, and M. Wilkinson. OPPL-Galaxy, a Galaxy tool for enhancing ontology exploitation as part of bioinformatics workflows. *Journal of Biomedical Semantics*, 4(1):2+, 2013.

[2] F. Belleau, M.-A. 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.

[3] C. T. Bull, S. H. D. Boer, T. P. Denny, G. Firrao, M. F.-l. Saux, G. S. Saddler, M. Scortichini, D. E. Stead, and Y. Takikawa. LETTER TO THE EDITOR COMPREHENSIVE LIST OF NAMES OF PLANT PATHOGENIC BACTERIA , 1980-2007. 92:551–592, 2010.

[4] G. O. Consortium. The Gene Ontology in 2010: extensions and refinements. *Nucleic acids research*, 38(Database issue):D331–5, 2010.

[5] L. G. Cowell and B. Smith. Infectious disease ontology. *Infectious Disease Informatics*, pages 373–395, 2010.

[6] P. J. G. M. de Wit. How plants recognize pathogens and defend themselves. *Cellular and molecular life sciences : CMLS*, 64(21):2726–32, 2007.

[7] P. Dodds and J. Rathjen. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Review Genetics*, 11:539–548, 2010.

[8] J. Goecks, A. Nekrutenko, J. Taylor, and Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*, 11(8):R86+, 2010.

[9] P. Jaiswal, S. Avraham, K. Ilic, E. A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S. Y. Rhee, M. M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, D. Ware, and F. Zapata. Plant ontology (po): a controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics*, 6(7-8):388–397, 2005.

[10] P. Jaiswal, D. Ware, J. Ni, K. Chang, W. Zhao, S. Schmidt, X. Pan, K. Clark, L. Teytelman, S. Cartinhour, L. Stein, and S. McCouch. Gramene: development and integration of trait and gene ontologies for rice. *Comparative and Functional Genomics*, 3(2):132–136, 2002.

[11] M. Lindeberg and A. Collmer. Gene Ontology for type III effectors: capturing processes at the host-pathogen interface. *Trends in microbiology*, 17(7):304–11, 2009.

[12] J. Mansfield, S. Genin, S. Magori, V. Citovsky, M. Sriariyanum, P. Ronald, and et al. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13(6):614–629, 2012.

[13] E. Montesinos. Pathogenic plant-microbe interactions. What we know and how we benefit. *International microbiology : the official journal of the Spanish Society for Microbiology*, 3(2):69–70, 2000.

[14] B. Smith, M. Ashburner, C. Rosse, J. Bard, and W. B. et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, 25(11):1251–1255, 2007.

[15] R. Walls, B. Smith, J. Elser, A. Goldfain, D. W. Stevenson, and P. Jaiswal. A plant disease extension of the Infectious Disease Ontology. *In Ronald Cornet and Robert Stevens, editors, ICBO*, pages 1–5.