**Semantic representation of PHI-Base plant-pathogen interactions (PPI) in PPIO**

PHI-base organized their entries around the gene instead of the PPI. As PPIO organizes its knowledge based on the PPI interaction we have to create some additional classes to represent the new knowledge.

PPIO originally assumes a single plant-pathogen interaction. However, PHI base represents several entries with the same interaction but with different genes involved. In order to allow PPIO to represent this situation, and continue with the original idea that the plant-pathogen interaction is the nucleus, we have to modify PPIO to allow represent multiple interactions of the same plant-pathogen pair.

In this context, we have created the “Interaction” class which represents a [plant-pathogen]-gene interaction.

In the following page is summarized the conversion of PHI-base tags to the PPIO semantic representation (Note: OP is Object Property, DP is Data Property).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Phi Base Tags** | **Semantic PPIO Representation** | | | |
| **Domain** | **Property** | **Range** | **Class** |
|  | Interaction | pathogen\_involved [OP] | Pathogen | Interaction |
| Experimental\_host | Interaction | host\_involved [OP] | Host Plant |
|  | Interaction | gene\_involved [OP] | Gene |
| Disease\_Name | Interaction | has\_phenotypic\_process [OP] | Phenotypic Process |
| Host\_Response | Interaction | has\_host\_response [OP] | Host\_Response |
| Host\_Target | Interaction | has\_host\_target [OP] | CCO Ontology - CCO Class |
| Comments | Interaction | comments -> rdfs:comments [DP] | String |
|  |  |  |  |  |
| Function | Gene | has\_function [OP?] | LOD? | Gene |
| Gene name | Gene | name [DP] | String |
| GO annotation | Gene | dbx\_ref [OP] | owl:Thing |
| Locus ID | Gene | locusID [DP] | String |
| AA\_Sequence | Gene | AA\_Sequence [DP] | String |
| NT\_Sequence | Gene | NT\_Sequence [DP] | String |
| Accession | Gene | accession\_ID [DP] | String |
| DB\_Type | Gene | accession\_DB [DP] | String |
|  | Gene | has\_gene\_mutation\_parameters [OP] | GMP |
|  |  |  |  |  |
| Essential\_gene\_Lethal\_knockout | GMP | letal\_knockout [DP] | Boolean | Gene Mutation Parameters (GMP) |
| Experimental\_evidences | GMP | experimetal\_evidences [DP] | String |
| Phenotype\_of\_mutant | GMP | phenotype\_of\_mutant [OP] | Mutant Phenotype |
|  |  |  |  |  |
| *Associated\_strain* | Pathogen | associated\_strain [DP] | String | Pathogen |
| Strain | Pathogen | strain | String |
| In\_vitro\_growth | Pathogen | in\_vitro\_growth\_type [OP] | Growth Type |
| Pathogen\_species | Pathogen | species [DP?] | String |
| Pathogen\_NCBI\_Taxonomy\_ID | Pathogen | ncbi\_taxonomy\_id [DP] | String |
|  |  |  |  |  |
| Host\_NCBI\_Taxonomy\_ID | Host Plant | ncbi\_taxonomy\_id [DP] | String | Host Plant |

In green are marked some doubts about the design of the selected properties.

**has\_function**:

The function tag represents the function of the disrupted gene. A brief look to the PHI-base content shows that function is only represented by a “string” which contains the “name of the function” without any reference to any kind of code or ID. With this situation is difficult to obtain a complex entity which represents the function.

Hence, we have suggested that this property should be a data property where just the text provided by PHI-Base is inserted.

Mikel suggests this property as an object property and the use of LOD and search URIs that can match with the function provided.

Any suggestion?

**Gene name**:

Same situation as before. Given that this is just the name, we consider introducing the gene name through the “label” property.

Mikel suggests this property as an object property and the use of LOD and search URIs that can match with the name provided.

Any suggestion?

**GO annotation**:

PHI-Base provides in some genes a reference to GO. We think that we can use dbx\_ref or seeAlso or any similar property to link our instance of the gene with external databases.

Any suggestion?

**Pathogen species/taxonomy ids (pathogen and host)**:

According to PHI-Base this tag represents the names of the pathogen species (without names of subspecies, strains, isolates, etc.). We thought that this should be represented with a data property which just contains the species name. **Note**: Currently, when we add a new pathogen to the ontology, automatically is generated the taxonomy hierarchy from this concrete pathogen specie to “cellular organism” class.

Mikel suggests this property as an object property and the use of LOD for NCBI taxa or OntoBee to get the taxon species instance.

Any suggestion?