

This file is part of the supplemental information provided with the following article.

The KnownLeaf literature curation system captures knowledge about Arabidopsis leaf growth and development and facilitates integrated data mining^{1[W]}

Dóra Szakonyi^{*}, Sofie Van Landeghem^{*}, Katja Baerenfaller, Lieven Baeyens, Jonas Blomme, Rubén Casanova-Sáez, Stefanie De Bodt, David Esteve-Bruna, Fabio Fiorani², Nathalie Gonzalez, Jesper Grønlund, Richard G.H. Immink, Sara Jover-Gil, Asuka Kuwabara, Tamara Muñoz-Nortes, Aalt D.J. van Dijk, David Wilson-Sánchez, Vicky Buchanan-Wollaston, Gerco C. Angenent, Yves Van de Peer, Dirk Inzé, José Luis Micol, Wilhelm Gruissem, Sean Walsh^{*}, Pierre Hilson^{*}

Department of Plant Systems Biology, VIB, B-9052 Ghent, Belgium (D.S., S.V.L., L.B., J.B., S.D.B., F.F., N.G., Y.V.d.P., D.I., P.H.); Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium (D.S., S.V.L., S.D.B., J.B., F.F., N.G., Y.V.d.P., D.I., P.H.); Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland (K.B., A.K., W.G., S.W.); División de Genética and Instituto de Bioingeniería, Universidad Miguel Hernández, 03202 Elche, Alicante, Spain (R.C.S., D.E.B., S.J-G., T.M.N., D.W.S., J.L.M.); Warwick Systems Biology Centre, University of Warwick, Coventry CV4 7AL, United Kingdom (J.G., V.B.-W.); School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom (J.G., V.B.-W.); Plant Research International, Bioscience, 6708 PB Wageningen, The Netherlands (R.G.H.I., A.D.J.v.D., G.C.A.); Genomics Research Institute (GRI), University of Pretoria, Private bag X20, Pretoria, 0028, South Africa (Y.V.d.P.); INRA, UMR1318, Institut Jean-Pierre Bourgin, RD10, F-78000 Versailles, France (P.H.); AgroParisTech, Institut Jean-Pierre Bourgin, RD10, F-78000 Versailles, France (P.H.)



Dóra Szakonyi, Sean Walsh, Katja Baerenfaller and Pierre Hilson

Leaf Knowtator Literature Annotation Scheme



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Table of contents

LEAF KNOWTATOR LITERATURE ANNOTATION SCHEME.....	2
TABLE OF CONTENTS	3
1 ANNOTATION STRUCTURES	4
1.1 INTRODUCTION	4
1.2 ONTOLOGIES UPLOADED.....	4
1.3 AVAILABLE SLOTS	5
1.3.1 <i>Construct Description</i>	5
1.3.2 <i>Developmental stage</i>	5
1.3.3 <i>DNA target ID</i>	5
1.3.4 <i>DNA target sequence</i>	5
1.3.5 <i>Factuality ID</i>	5
1.3.6 <i>Gene expression ID</i>	6
1.3.7 <i>Gene ID</i>	6
1.3.8 <i>Gene studied</i>	6
1.3.9 <i>Gene target</i>	6
1.3.10 <i>Genetic interactor</i>	6
1.3.11 <i>Genotype ID</i>	7
1.3.12 <i>Genotype Zygoty</i>	7
1.3.13 <i>Growth Condition Slot</i>	7
1.3.14 <i>Interaction type</i>	7
1.3.15 <i>Interactor protein</i>	7
1.3.16 <i>Localisation ID</i>	8
1.3.17 <i>Methodology ID</i>	8
1.3.18 <i>Molecular feature ID</i>	8
1.3.19 <i>Mutant LOF_GOF ID</i>	8
1.3.20 <i>Mutant type</i>	8
1.3.21 <i>Plant part ID</i>	9
1.3.22 <i>Process ID</i>	9
1.3.23 <i>Property Slot</i>	9
1.3.24 <i>Protein ID</i>	9
1.3.25 <i>Protein studied</i>	9
1.3.26 <i>Regulation ID</i>	9
1.3.27 <i>Value ID</i>	10
1.4 ANNOTATED RELATIONSHIPS	11
1.4.1 <i>Phenotype</i>	11
1.4.2 <i>Gene expression</i>	11
1.4.3 <i>Feature</i>	12
1.4.4 <i>DNA-protein interaction</i>	12
1.4.5 <i>Genetic interaction</i>	13
1.4.6 <i>Protein-protein interaction</i>	13
1.4.7 <i>Process</i>	13
1.4.8 <i>Regulation of process</i>	14
1.4.9 <i>Regulation of gene expression</i>	14
1.4.10 <i>Regulation of phenotype</i>	15
1.5 HOW TO FIND THE RIGHT ONTOLOGY TERM.....	15

1 Annotation structures

1.1 Introduction

This document provides a general explanation of the structure that was adopted for the annotation of relations via the Knowtator interface. It embodies a compromise between the collection of very detailed information and the necessity to capture this information in a useful, constraint, and machine readable format, based on existing ontologies, and relatively rapidly.

As you read this document then learn to annotate text within the Leaf Knowtator project, you will recognize the annotation structures in the different categories and functions that were created within the Protégé program and Knowtator plug in.

1.2 Ontologies uploaded

We decided to work with a selection of ontology collections that are uploaded into the Leaf Knowtator annotation scheme.

- BTO (BRENDA Tissue Ontology) <http://www.brenda-enzymes.info/>
- EO (Plant Environment Ontology) http://www.gramene.org/plant_ontology/index.html#eo
- GO (Gene Ontology) <http://www.geneontology.org/>
- MGED (Microarray experimental conditions) <http://mged.sourceforge.net/ontologies/MGEDontology.php>
- MI (Molecular Interaction Ontology) <http://www.psidev.info/>
- PATO (Phenotype, Attribute and Trait Ontology) http://obofoundry.org/wiki/index.php/PATO:Main_Page
- PO (Plant Ontology) <http://www.plantontology.org/>
- TAIR10 <http://arabidopsis.org/>
- TO (Cereal Plant Trait Ontology) http://www.gramene.org/plant_ontology/

Files for different ontologies can be found at <http://www.obofoundry.org/>

It can happen that certain ontology terms do exist but are not included in the used collections. In that case, the corresponding ontology term is entered as a single class entry. For example we use Wild type_SO:0000817 (SO: Sequence types and features), loss of function_APO:000001, homozygous diploid _APO:0000229 (APO: Ascomycete phenotype ontology) etc.

The complete ontology collections can be found in Knowtator under the ‘External ontologies’ class. However we work with only a subset of these. The most used terms and additional ontology entries are moved to ‘Internal ontologies’ that is structured according to the Slots that need to be filled.

1.3 Available slots

1.3.1 Construct Description

Used	This can be filled manually in the genotype panel if any information is available about the given transgenic construct.
Slot value	String
Ontology	-

1.3.2 Developmental stage

Used	To describe the developmental stage of the plant part.
Slot value	Instance
Ontology	PO

1.3.3 DNA target ID

Used	It is used in DNA-protein interaction relationships. AGI code of the DNA target is entered.
Slot value	Instance
Ontology	TAIR 10

1.3.4 DNA target sequence

Used	It is used in DNA-protein interaction relationships. The identified DNA target sequence is entered manually.
Slot value	String
Ontology	-

1.3.5 Factuality ID

Used	Negation, Speculation, Modifier.
Slot value	Instance
Ontology	-(Negation, Speculation, Modifier)

Note: In special cases, annotations can be marked with this category. **Negation** was used for negative sentences (is not, do not, lack of etc.) but since we decided to transform this sentences into positive statements. **Speculation** is used for words such as putative, may, might, hypothesis, conclusion etc. that implies the sentence is not a direct observation but derives from analysis of experimental data. **Modifier** can mark any words that slightly change the meaning of the sentence. Examples include ‘partial loss of adaxial identity’, ‘often darker green leaves’, or ‘weak lines had bigger leaves’.

1.3.6 Gene expression ID

Used	It describes how the level of a gene product changed or refers to the described spatial pattern.
Slot value	Instance
Ontology	Few selected PATO and MI terms

Note: If the mRNA or protein level or spatial distribution are unchanged usually **normal_PATO:000046** is used. If they are changed but in an unspecified manner then **abnormal_PATO:0000460** is inserted. If the level is increased or decreased over **expressed level_MI:0506** or **under expressed level_MI:0223** are used. Other possible terms are **absent_PATO:0000462**, **delayed_PATO:0000502**, **mislocalised_PATO:0000628** and **present_PATO:0000467**.

1.3.7 Gene ID

Used	It is used in all genotype annotations to attach AGI codes.
Slot value	Instance
Ontology	TAIR 10

1.3.8 Gene studied

Used	It is used for the gene product in gene expression, process, regulation of gene expression, regulation of process and regulation of phenotype relationships.
Slot value	Instance
Ontology	TAIR 10

1.3.9 Gene target

Used	It is used in regulation of gene expression annotations to include AGI code of the gene that is regulated on mRNA or protein level.
Slot value	Instance
Ontology	TAIR 10

1.3.10 Genetic interactor

Used	It is used in genetic interaction relationships and describes the genotype.
Slot value	Instance
Ontology	TAIR 10

Note: All the genotype annotations are simplified. Any kind of mutant is designated as **mutated gene_MI:0804**. Special constructs (RNAi, amiRNA) are annotated as **knock down_MI:0789**. If a construct results in elevated RNA level the genotype will be **over**

expressed level_MI:0506. **Complex genotype** was introduced to include constructs that do not fit into the above mentioned categories such as heterologous promoter-gene constructs, combination of different genotypes (overexpressor in mutant background) etc. **Wild type_SO:0000817** was introduced manually. You will have to open the Genetic interactor field and attach additional information such as Gene ID, Genotype Zygoty, Mutant LOF GOF ID, Mutant type, Construct description. Please, consult the Knowtator manual for further details.

1.3.11 Genotype ID

Slot features are identical to the one described under 1.2.10 Genetic interactor category.

1.3.12 Genotype Zygoty

Used	It is used to indicate zygoty of the given genotype.
Slot value	Instance
Ontology	homozygous diploid_APO:0000229 heterozygous diploid_APO:0000230

1.3.13 Growth Condition Slot

Used	It describes any environmental conditions and treatments.
Slot value	Instance
Ontology	PEO

1.3.14 Interaction type

Used	It is used in Genetic, Protein-protein and DNA-protein interactions.
Slot value	Instance
Ontology	MI

In case of Genetic interactions, several terms can be used to describe the detected interaction e.g. **negative genetic interaction_MI:0933** (enhancement), **positive genetic interaction_MI:0935**, **suppressive genetic interaction defined by inequality_MI:0796**, **additive genetic interaction defined by inequality_MI:0799**, **epistatic genetic interaction defined by inequality_MI:0797** etc. Protein-protein and DNA-protein interactions can be described mainly by **physical association_MI:0915** or **direct interaction_MI:0407**. In some cases, Protein interactions describe enzymatic reactions between the protein partners, then a general term **enzymatic reaction_MI:0414** is used or it can be further specified if the term exists such as **phosphorylation reaction_MI:0217**, **dephosphorylation reaction_MI:0203**, **methylation reaction_MI:0213** etc.

1.3.15 Interactor protein

Used	It is used in Protein-protein interactions for the identified
-------------	---

	interactor.
Slot value	Instance
Ontology	TAIR10

1.3.16 Localisation ID

Used	This is used for two purposes, first, attaching a second PO term to an annotation that specifies further the plant part slot (e.g. cell type as plant part and organ as localisation). Localisation can also be used to annotate words where no proper ontologies can be found)
Slot value	Instance
Ontology	PO or just annotate as Localisation ID

1.3.17 Methodology ID

Used	Used experimental method entered manually for the given annotation.
Slot value	String
Ontology	-

E.g. RT-PCR, Northern blot, in situ hybridization, confocal microscopy etc.

1.3.18 Molecular feature ID

Used	Any structural elements present in DNA, RNA or protein level.
Slot value	String
Ontology	-

1.3.19 Mutant LOF_GOF ID

Used	It is used to indicate if a mutation is a loss-of-function or a gain-of function allele.
Slot value	Instance
Ontology	loss of function_APO:0000011 gain of function_APO:0000010

1.3.20 Mutant type

Used	It is filled manually if any details are available about the mutant allele.
Slot value	String
Ontology	-

1.3.21 Plant part ID

Used	Plant organs, tissue or cell types. The plant part that is changed or subject of the experiment.
Slot value	Instance
Ontology	PO

1.3.22 Process ID

Used	Used if any biological process, molecular or cellular function can be attached (e.g. cell division, growth, mRNA splicing etc.).
Slot value	Instance
Ontology	GO

1.3.23 Property Slot

Used	Fill this slot with the property of the plant part that has been changed/examined (e.g. width, weight, color etc.)
Slot value	Instance
Ontology	PATO, TO, PO

1.3.24 Protein ID

Used	It is used for the gene product in Feature annotations.
Slot value	Instance
Ontology	TAIR 10

1.3.25 Protein studied

Used	It is used for the gene product in Protein-protein interaction annotations.
Slot value	Instance
Ontology	TAIR 10

1.3.26 Regulation ID

Used	It is used to describe how the gene modifies behaviour of gene expression, process, phenotype etc.
Slot value	Instance
Ontology	Few selected GO terms

Note: major ontologies in use are negative regulation of biological process_GO:0048519, positive regulation of biological process_GO:0048518, positive regulation of gene expression_GO:0010628 and negative regulation of gene expression_GO:0010629.

1.3.27 Value ID

Used	This describes the way how the plant part changed (e.g. increased size, decreased width, dark green etc.).
Slot value	Instance
Ontology	PATO

1.4 Annotated relationships

1.4.1 Phenotype

It is used for any phenotypic description about wild-type, mutant or transgenic plants.

Example: The reduced leaf area in the hub1-1 mutant was confirmed by morphological measurements of the fully expanded leaves 1 and 2.

Information type/Slots	Original text	Ontologies attached
Developmental stage	fully expanded leaves	3 leaf fully expanded_PO:0001053
Factuality ID		
Genotype ID	hub1-1	mutated gene_MI:0804
Gene ID		RDO4 HUB1_AT2G44950
Mutant LOF_GOF ID		loss of function_APO:0000011
Genotype_Zygoty		homozygous diploid _APO:0000229
Growth Condition		juvenile leaf_PO:0006339
Slot		
Localisation ID		
Methodology ID		
Plant part ID	leaf	leaf_PO:0025034
Process ID		
Property Slot	area	area_PATO:0001323
Value ID	reduced	decreased area_PATO:0002058

1.4.2 Gene expression

It is used to annotate either observations about mRNA or protein levels of wild type, mutant or transgenic plants in experiments including (q)RT-PCR, Northern or Western blot, microarray) or description of gene expression pattern of the gene studied.

Example: some overexpressed genes in hub1-1 were related to meristem development, such as the homeodomain genes KNOTTED1-LIKE2 (KNAT2).

Information type/Slots	Original text	String entry	Ontologies attached
Developmental stage			
Factuality ID			
Gene expression ID	overexpressed		over expressed level_MI:0506
Gene studied	KNAT2		KNAT2 ATK1_AT1G70510
Genotype ID	hub1-1		mutated gene_MI:0804
Gene ID			RDO4 HUB1_AT2G44950
Mutant LOF_GOF ID			loss of function_APO:0000011
ID			homozygous diploid _APO:0000229
Genotype_Zygoty			
Growth Condition			
Slot			

Localisation ID	
Methodology ID	microarray
Plant part ID	shoot apex_PO:0000037
Property Slot	transcript expression level evidence_ECO:0000048

1.4.3 Feature

This category includes any structural elements on DNA, mRNA or protein level (e.g. promoter binding elements, miRNA target sites, protein domains).

Example: HUB1 contains a RING finger domain (PF00097) at position 826 to 864 and is classified as an Hca-RING-type protein.

Information type/Slots	Original text	String entry	Ontologies attached
Factuality ID			
Molecular feature ID		a RING finger domain (PF00097) at position 826 to 864 and is classified as an Hca-RING-type protein	
Protein	HUB1		RDO4 HUB1_AT2G44950

1.4.4 DNA-protein interaction

This relationship is used when direct interaction was proved between the protein studied and the DNA molecule experimentally (mobility shift assay, yeast one-hybrid assay, chromatin IP).

Example: The putative ANT binding sites in the FIL and YAB3 promoters match the in vitro-determined ANT consensus binding site in 10 of 14 conserved positions (Fig. 1A). Gel mobility shifts revealed that ANT binds in vitro to fragments of both promoters that contain these sequences.

Information type/Slots	Original text	String entry	Ontologies attached
DNA target ID	FIL		AFO FIL YAB1_AT2G45190
DNA target sequence			
Factuality ID			
Methodology ID		mobility shift assay	
Plant part ID			
Protein	ANT		ANT DRG CKC CKC1_AT4G37750
Regulation ID	binds		direct interaction_MI:0407

1.4.5 Genetic interaction

It is used when a genetic interaction is observed between different genotypes.

Example: An enhancer-trap insertion (ET5446) into AGO7/ZIP (ago7-m5446, also known as ago7-jason) was also found to enhance as1.

Information type/Slots	Original text	Ontologies attached
Factuality ID		
Genetic interactor	ago7-jason	mutated gene_MI:0804
Gene ID		AGO7 ZIP_AT1G69440
Mutant LOF_GOF ID		loss of function_APO:0000011
Genotype_Zygoty		homozygous diploid _APO:0000229
Genotype ID	as1	mutated gene_MI:0804
Gene ID		ATPHAN AS1 ATMYB91 MYB91_AT2G37630
Mutant LOF_GOF ID		loss of function_APO:0000011
Genotype_Zygoty		homozygous diploid _APO:0000229
Growth Condition		
Slot		
Interaction type	enhance	negative genetic interaction_MI:0933
Methodology ID		
Plant part ID		leaf_PO:0025034
Property Slot		shape_PATO:0000052
Process ID		
Value ID		

1.4.6 Protein-protein interaction

This relationship is used when a direct interaction was shown between two protein molecules (yeast two-hybrid assay, pull-down experiment, coimmunoprecipitation, FRET, split molecular tag studies).

Example: our interaction assay showed that RHL1 can bind to AtSPO11-3/RHL2.

Information type/Slots	Original text	String entry	Ontologies attached
Factuality ID			
Interaction type	bind		direct interaction_MI:0407
Interactor protein	RHL2		RHL2 BIN5_AT5G02820
Methodology ID		yeast two-hybrid assay	
Protein studied	RHL1		RHL1 HYP7_AT1G48380

1.4.7 Process

Process is used when the sentence is giving information about the gene itself or the corresponding gene product (RNA, protein) to which a GO term can be attached.

Example: TCH4 gene presumably plays a role in the cell expansion process.

Information type/Slots	Original text	Ontologies attached
Developmental stage		
Factuality ID		
Gene studied	TCH4	TCH4 XTH22_AT5G57560
Growth Condition Slot		
Localisation ID		
Methodology ID		
Plant part ID		
Process ID	cell expansion	cell growth_GO:0016049
Property Slot		
Value ID		

1.4.8 Regulation of process

Similar to process only a regulation word is included (promote, enhance, repress, negatively regulate etc.)

Example: KAN promotes abaxial identity.

Information type/Slots	Original text	Ontologies attached
Factuality ID		
Gene studied	KAN	KAN KAN1_AT5G16560
Growth Condition Slot		
Localisation ID		
Methodology ID		
Plant part ID		
Process ID	abaxial identity	determination of dorsal identity_GO:0048263
Property Slot		
Regulation ID	promotes	positive regulation of biological process_GO:0048518
Value ID		

1.4.9 Regulation of gene expression

It is a wider category than DNA-protein interaction. The regulator does not have to be involved in direct interaction between the protein and the target. Not only regulation on DNA level but also mRNA and protein level are included (miRNA, tasiRNA cleavage, proteasome dependent degradation etc.).

Example: AtCPL1 ... negative regulators of RD29A expression

Information type/Slots	Original text	Ontologies attached
Developmental stage		
Factuality ID		
Gene studied	AtCPL1	CPL1 FRY2 ATCPL1_AT4G21670
Gene target	RD29A	COR78 LTI78 RD29A LTI140_AT5G52310
Methodology ID		
Plant part ID		
Regulation ID	negative regulators	negative regulation of gene expression_GO:0010629

1.4.10 Regulation of phenotype

If a gene product is described to modulate plant phenotype.

Example: PHABULOSA ... influences leaf shape

Information type/Slots	Original text	Ontologies attached
Factuality ID		
Gene studied	PHABULOSA	
Growth Condition Slot		
Localisation ID		
Methodology ID		
Plant part ID	leaf	leaf_PO:0025034
Property Slot	shape	shape_PATO:0000052
Regulation ID	influence	regulation of developmental process_GO:0050793

1.5 How to find the right ontology term

The most difficult part of the project is to get familiar with the existing ontologies and to pick the most appropriate terms when building relationships. To facilitate this process, we have created Supplemental Table S4. It lists all the terms included in the KnownLeaf annotations, based on 283 primary research papers. We foresee that the vast majority of additional relationships can be built on the basis of this focused vocabulary list.

If an appropriate term cannot be found in this list, we recommend using the following site to search for the right formulation in <http://www.ebi.ac.uk/ontology-lookup/>. Other useful resources are <http://www.obofoundry.org/> and <http://bioportal.bioontology.org/>.