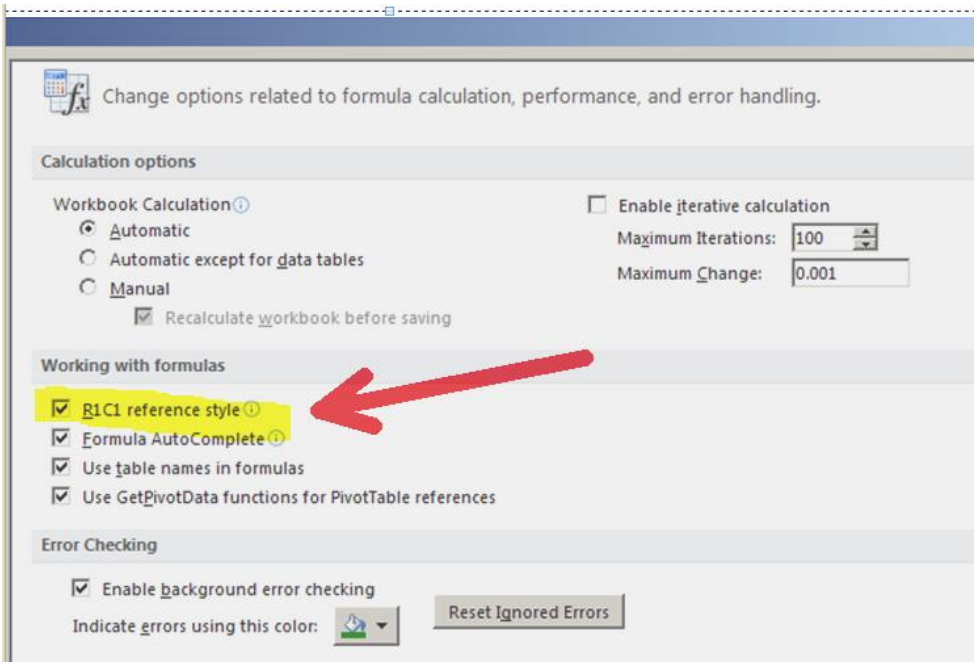
GENERAL

Pay particular attention on using correct text for controlled vocabulary entries –be careful of upper/lower case, exact punctuation, not adding extra spaces etc.
Also be careful of carriage returns etc in free text entries.

The Excel spread sheet uses "R1C1" reference style to allow easy access to numbered columns. Excel default is usually set to the "A1" reference style, but can easily be reset (see beneath for details.)

In Excel: Tab: File > Options >Formulas



SPREADSHEET 1: PHIBASE CURATION

Useful web links:
http://www.uniprot.org/
http://www.uniprot.org/help/text-search
http://www.ebi.ac.uk/chebi/downloadsForward.do
http://www.ncbi.nlm.nih.gov/taxonomy
http://www.ncbi.nlm.nih.gov/pubmed

For fungicides the 2015 FRAC list: http://www.frac.info/docs/default-source/publications/frac-code-list/frac-code-list-2015-finalC2AD7AA36764.pdf?sfvrsn=4

Suppl Text 1: PHI-BASE curator's guidelines for curation of excel spreadsheet

Layout of document

Excel column header

Definition -
Vocabulary -
Mandatory -
Multiple entries -
Example -
Additional notes -
Recent changes -

1 Curation_comments_temporary

Definition - Column for temporary notes during curation
Vocabulary - Free text
Mandatory - No
Multiple entries - Yes
Example -
Additional notes -
Recent changes -

2 _to_do_

Definition - Column for temporary notes during curation
Vocabulary - Free text
Mandatory - No
Multiple entries - Yes
Example -
Additional notes -
Recent changes -

3 Record_ID

Definition - Required at RRes only (for integrity check of spread sheet to avoid row sorting errors)
Vocabulary - Controlled
Mandatory - Yes for RRes
Multiple entries - No
Example - Record XXXX
Additional notes -
Recent changes -

4 PHI-base_accession_ID

Definition - One number corresponding to one gene of one organism (even if there is more than one interaction) from one paper.
Vocabulary - Controlled (PHI: XXXX)
Mandatory - Yes
Multiple entries - No

Example - PHI:3012
Additional notes - There may be conflicting results from different papers so genes from different papers require different PHI:IDs.
If there is the same gene name by more than one organism, every gene of one organism gets a separate tracking number.
In case "alleles" are described, assign a new number for each allele (example: PMID-19454732 describes multiple alleles of one avr gene: pik1 (PHI:2137, pikm (PHI:3498), pikp (PHI:3499)).
Assign the same ID for tests of one gene in different isolates, on different host species, cultivars or tissue type.
Recent changes - Updated for different IDs of genes from different papers (11Mar2015AC)
Note: In case of alleles with diff. PHI:IDs may have same uniprot and EMBL/Genbank ID e.g. PHI:3706 and PHI:3707. (AC 18Mar2015)

5 Identifier_typeOfProteinID

Definition - Source of protein identifier (UniProt)
Vocabulary - Controlled
Mandatory – No (molecular data needs to be entered for a minimum of either Columns 5/6 (Identifier_typeOfProteinID/ ProteinID), 7/8 (IdentifierTypeOfGeneLocusID / GeneLocusID) or 9/10 (AA sequence/ NT sequence)).
Multiple entries - No
Example - 'Uniprot' or 'no data found'
Additional notes - Hint: Look up in Results or Method section or Figure legend. Search for accession at UniprotKB (http://www.uniprot.org/).
Recent changes - Changed to Uniprot only (10Nov2014AC)

6 ProteinID

Definition - UniProt ID <Protein accession number>
Vocabulary – Controlled by UniProtKB
Mandatory - No (molecular data needs to be entered for a minimum of either Columns 5/6 (Identifier_typeOfProteinID/ ProteinID), 7/8 (IdentifierTypeOfGeneLocusID / GeneLocusID) or 9/10 (AA sequence/ NT sequence)).
Multiple entries - No
Example - 'E1WFQ7' or 'no data found'
Additional notes - In rare cases when more than 1 UniProt ID try to choose the most specific entry not the whole genome study. The Swiss-Prot manually annotated and reviewed entries are more reliable than the TrEMBL automatically annotated and not reviewed entries.
Hint: Look up in Results or Method section or Figure legend. Search for accession at UniprotKB (http://www.uniprot.org/).
Recent changes - Changed to Uniprot only (10Nov2014AC)

7 IdentifierTypeOfGeneLocusID

Definition - Source of gene locus ID
Vocabulary – Controlled by databank

Mandatory - No (molecular data needs to be entered for a minimum of either Columns 5/6 (Identifier_typeOfProteinID/ ProteinID), 7/8 (IdentifierTypeOfGeneLocusID / GeneLocusID) or 9/10 (AA sequence/ NT sequence)).
Multiple entries - No
Example - EMBL, Genbank, Broad, JGI etc.
Additional notes -
Hint: Please make an effort to obtain EMBL or Genbank accession, so that an outlink for seq information can be provided. Look up in Results, Method or Figure legend.
Recent changes - Changed to EMBL/Genbank or species specific accessions Broad/JGI (also see Column 71/72 (Additional identifier type of gene locus ID/ Additional gene locus ID)). No uniprot IDs in this column (10Nov2014 AC)

8 GeneLocusID

Definition - The Gene locus ID corresponding to source in Col. 7 (IdentifierTypeOfGeneLocusID) above.
Vocabulary - Free text
Mandatory - No (molecular data needs to be entered for a minimum of either Columns 5/6 (Identifier_typeOfProteinID/ ProteinID), 7/8 (IdentifierTypeOfGeneLocusID / GeneLocusID) or 9/10 (AA sequence/ NT sequence)).
Multiple entries - No
Example - 'AAC64374' for EMBL ID. 'FGSG_10825, SNOG_16571, VC_0086' for Broad. Otherwise use JGI, MUMdb for species specific accessions.
Additional notes -
Recent changes - Changed to EMBL/Genbank or species specific accessions Broad/JGI (also see Column 71/72 (Additional identifier type of gene locus ID/ Additional gene locus ID)). (10Nov2014 AC)

9 AA sequence

Definition - Amino acid sequence, in case there is no submission to reference database.
Vocabulary - Controlled to amino acid sequence
Mandatory - No
Multiple entries - No (molecular data needs to be entered for a minimum of either Columns 5/6 (Identifier_typeOfProteinID/ ProteinID), 7/8 (IdentifierTypeOfGeneLocusID / GeneLocusID) or 9/10 (AA sequence/ NT sequence)).
Example -
Additional notes -
Recent changes -

10 NT sequence

Definition - Nucleotide sequence, in case there is no submission to reference database.
Vocabulary – Controlled to nucleotide sequence
Mandatory - No

Multiple entries - No (molecular data needs to be entered for a minimum of either Columns 5/6 (Identifier_typeOfProteinID/ ProteinID), 7/8 (IdentifierTypeOfGeneLocusID / GeneLocusID) or 9/10 (AA sequence/ NT sequence)).
Example -
Additional notes -
Recent changes –

11 Genomic sequence providing strain

Definition - Strain, isolate, subspecies etc., of organism used for the submission of the gene/protein sequence.
Vocabulary - Free text
Mandatory – Yes if available or 'no data found'
Multiple entries - No
Example - 'Guy-11' for gene MPG1 from *Magnaporthe oryzae*
Additional notes - This is the strain, isolate, subspecies etc of the organism used for the submission of the gene/protein sequence (the REFERENCE sequence providing strain available in online databases). Find info in Methods, Results or tables.
Recent changes -

12 Gene name

Definition - Name of the gene
Vocabulary - Free text
Mandatory – Yes if available or 'no data found'
Multiple entries - Yes for synonyms
Example - 'Tri5', 'MGG_13324'
Additional notes - If there are synonyms, write them in brackets like 'CaHK1(CHK1)'.
If there is no gene name, use *Locus ID* or record homolog, i.e. '(Sc Atg1)'
Hint: Find in Abstract or Results.
Recent changes -

13 Genome location

Definition - Genome location of the gene
Vocabulary - Free text
Mandatory - No
Multiple entries - No
Example - 'Chromosome 1BS', 'Chromosome 20, contig 214: 152339–1540721', or for (bacterial) pathogenicity islands or plasmids 'SPI1', 'pSLT'.
Additional notes - Hint: Genome location of the (bacterial) gene like pathogenicity islands or plasmids (Example: SPI1) or chromosome location (chromosome 1). Rarely mentioned in paper.
Recent changes -

14 Specific modification/s to the targeted protein or promoter

Definition – Mutation name followed by specific modification/s to the targeted protein or promoter such as a domain swap experiment, amino acid changes introduced or a promoter sequence change.
Vocabulary - Free text

Mandatory - No
Multiple entries - Yes
Example - Use genetic notation given in paper (and describe in free text). 'ΔnasA::Gm (deletion construct containing the gentamicin (Gm) resistance cassette)'
Additional notes – The mutation name may or may not be the same as the 'gene name' in Col.12. If more than one interaction, separate entries using a semicolon.
Recent changes - New Column

15 Accession ID for the modified genetic element

Definition - Accession ID for the modified genetic element within the pathogen.
Vocabulary - Controlled (UniProt: XXXX)
Mandatory - No
Multiple entries - Yes
Example - Give sequence or protein accession for a submitted gene/promoter sequence as: Gene name UniProt: XXXX, Gene name EMBL: XXXX or Gene name Genbank: XXXX; separate multiple entries with semicolon.
Additional notes - Genetic element can be: promoter sequence, gene, protein, RNA.
Recent changes - New Column

16 Known interacting protein(s) or DNA elements in the pathogen

Definition - Known interacting protein(s) or DNA elements within the pathogen
Vocabulary - Free text
Mandatory - No
Multiple entries - Yes
Example - 'PmrA' and 'ssrB' interact. (PHI:4497 and PHI:4498)
Additional notes -
Recent changes - New Column

17 Interacting protein or DNA element– locus ID

Definition - Interacting sequence/protein locus ID.
Vocabulary - Controlled to **(**gene name with multiple entries and) source: ID
Mandatory - No
Multiple entries - Yes
Example – Give sequence or protein accession for a submitted protein or promoter sequence as: UniProt: XXXX, EMBL: XXXX, Genbank: XXXX or PHI: XXXX.
'Uniprot: B0BES1'
Where multiple entries prefix with gene name and separate entries with semicolon.
'HrcJ Uniprot: Q8PQB9; HrcC Uniprot: Q8PQB3'
Additional notes -
If not submitted to these above resources use the GENE LOCUS ID given in the publication. Example: FGSG_06874.3.
Recent changes - New Column

18 Multiple mutation

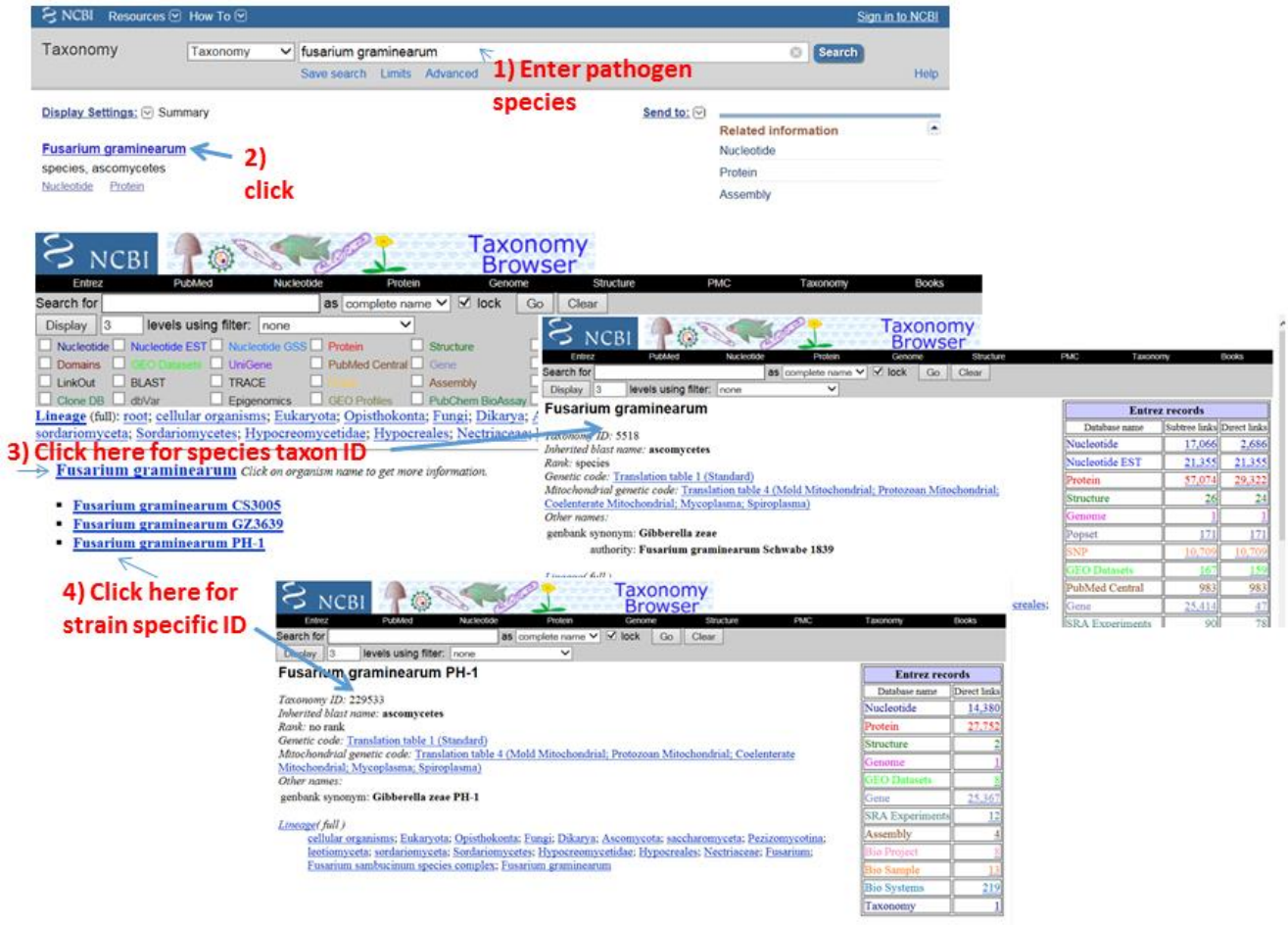
Definition - More than one gene is disrupted
Vocabulary - Controlled ('PHI:XXXX' or 'no')
Mandatory - Yes
Multiple entries - Yes
Example - 'PHI:XXXX' or 'no' **See below**
Additional notes - Hint: If there is more than one gene disrupted, make sure that every gene appears in the column "Gene name" with the corresponding PHI- base accession ID. The additional genes are listed in the column "Multiple Mutation" by their PHI- base accession number (separated by a semicolon). See XYL1 (PHI:567) and XYL2 (PHI:568) beneath as an example.
Recent changes –

EXAMPLE

PHI_MolConn_ID	IdentifierTypeOfProteinID	ProteinID	IdentifierTypeOfGeneLocusID	GeneLocusID	Gene_name	Multiple_mutation	Pathogen_species	Experimental_host_species
PHI:567	Uniprot	P0CT48	EMBL	AAC41683	XYL1 (XYN22)	PHI:568	Magnaporthe oryzae	Oryza sativa (related: Rice)
PHI:567	Uniprot	P0CT48	EMBL	AAC41683	XYL1 (XYN22)		Magnaporthe oryzae	Hordeum vulgare subsp. vulgare (related: Domesticated barley)
PHI:567	Uniprot	P0CT48	EMBL	AAC41683	XYL1 (XYN22)	PHI:568	Magnaporthe oryzae	Hordeum vulgare subsp. vulgare (related: Domesticated barley)
PHI:567	Uniprot	P0CT48	EMBL	AAC41683	XYL1 (XYN22)		Magnaporthe oryzae	Eragrostis curvula (related: Weeping love grass)
PHI:567	Uniprot	P0CT48	EMBL	AAC41683	XYL1 (XYN22)		Magnaporthe oryzae	Oryza sativa (related: Rice)
PHI:567	Uniprot	P0CT48	EMBL	AAC41683	XYL1 (XYN22)	PHI:568	Magnaporthe oryzae	Eragrostis curvula (related: Weeping love grass)
PHI:568	Uniprot	Q01176	EMBL	AAC41684	XYL2 (XYN33)		Magnaporthe oryzae	Eragrostis curvula (related: Weeping love grass)
PHI:568	Uniprot	Q01176	EMBL	AAC41684	XYL2 (XYN33)		Magnaporthe oryzae	Oryza sativa (related: Rice)
PHI:568	Uniprot	Q01176	EMBL	AAC41684	XYL2 (XYN33)		Magnaporthe oryzae	Hordeum vulgare subsp. vulgare (related: Domesticated barley)
PHI:568	Uniprot	Q01176	EMBL	AAC41684	XYL2 (XYN33)	PHI:567	Magnaporthe oryzae	Hordeum vulgare subsp. vulgare (related: Domesticated barley)
PHI:568	Uniprot	Q01176	EMBL	AAC41684	XYL2 (XYN33)	PHI:567	Magnaporthe oryzae	Oryza sativa (related: Rice)
PHI:568	Uniprot	Q01176	EMBL	AAC41684	XYL2 (XYN33)	PHI:567	Magnaporthe oryzae	Eragrostis curvula (related: Weeping love grass)

19 Pathogen species NCBI Taxonomy ID

Definition - NCBI taxonomy ID of the pathogen species
Vocabulary - Controlled to NCBI ID
Mandatory - Yes
Multiple entries - No
Example - '5518' for *Fusarium graminearum*
Additional notes - lookup NCBI taxonomy IDs at http://www.ncbi.nlm.nih.gov/taxonomy
Use Species ID, i.e. *Fusarium graminearum* (5518) instead of strain specific ID for PH-1 (229533). i.e. *Pseudomonas syringae* (317) instead of pathovar and strain specific ID for pv. tomato DC3000 (223283)
Recent changes –



20 Pathogen species name

Definition - Names of pathogen species (without names of subspecies, strains, isolates, pathovars etc.)
Vocabulary - Controlled to NCBI taxonomy species name
Mandatory - Yes
Multiple entries - No
Example - Species names 'Phaeosphaeria nodorum (related: Stagonospora nodorum)'
Additional notes - Hint: Find in Abstract.
Recent changes -

21 Pathogen strain NCBI Taxonomy ID

Definition - Strain specific NCBI taxonomy ID of the pathogen species
Vocabulary - Controlled to NCBI ID
Mandatory - No
Multiple entries - No

Example - '229533' for strain PH-1 of *Fusarium graminearum*
Additional notes - lookup NCBI taxonomy ID of the STRAIN (if available) http://www.ncbi.nlm.nih.gov/taxonomy Use STRAIN specific taxonomy ID here. Example: for *Fusarium graminearum* strain PH-1 the ID is 229533 instead of 5518 (species taxonomy ID used in Col.19). For *Pseudomonas syringae* pv. *tomato* DC3000 the ID is 223283 instead of 317.

22 Experimental strain

Definition - The name of the strain, isolate, subspecies, pathovars, serotypes etc. of the organism used in the experiment.
Vocabulary - Free text
Mandatory – Yes if available if not 'no data found'
Multiple entries - Yes
Example - For *Pseudomonas syringae* pv. *tomato* DC3000 only enter 'pv. tomato DC3000' here
Additional notes - Hint: In Methods. Tables and Results also possible. Please add this information here and NOT in Col. 20 (Pathogen species name).

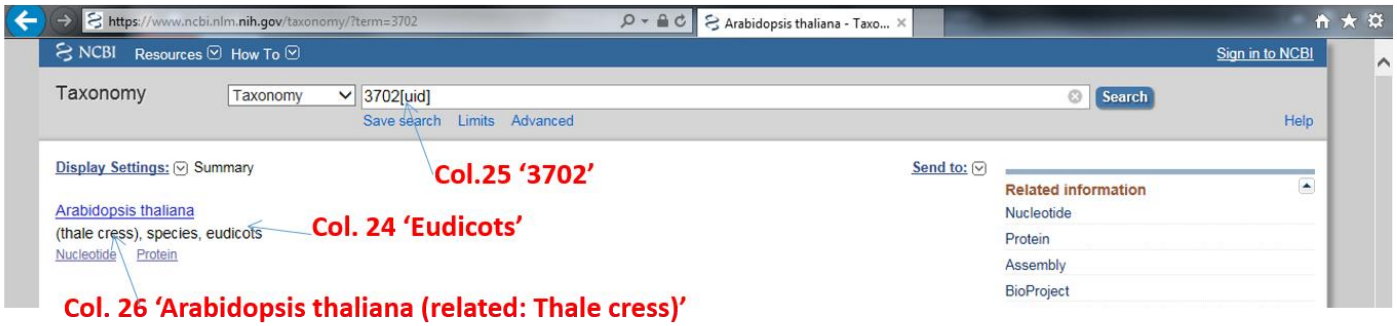
23 Disease name

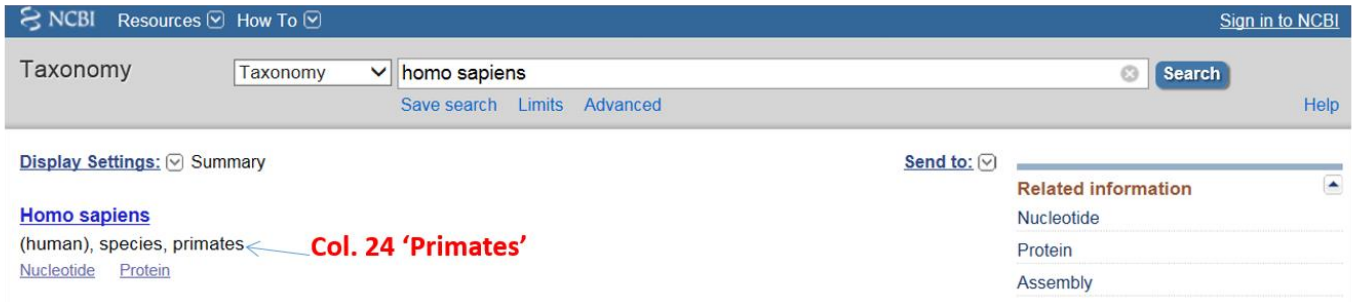
Definition - Common name of disease
Vocabulary - Free text
Mandatory - No
Multiple entries – Yes separated by semicolon
Example - 'Fusarium ear blight; Fusarium head scab'
Additional notes - Names can vary from paper to paper. For curation use the name given in the paper. Post-curation tidy ups of synonyms can be implemented. (Related names can be recorded in brackets.) Hint: Find in earlier entries of spreadsheet.
Recent changes -

24 Host_description

Definition - NCBI taxonomy description of host
Vocabulary - Controlled to NCBI taxonomy
Mandatory - Yes
Multiple entries - No
Example - 'Eudicots', 'Primates' see image below
'Lethal pathogen phenotype' –therefore no host bioassay reported
'No host tests done' –for chemistry only entries
Additional notes - Enter host NCBI taxonomy ID (required for Col. 25 (Host NCBI Taxonomy ID) see below) or experimental host species (required for Col. 26 (Experimental host species) see below) http://www.ncbi.nlm.nih.gov/taxonomy
See image below and select correct host description labelled Col. 24 (Host description) with arrow e.g. 'Eudicots', 'Primates'
Recent changes - updated AC 18Mar2015

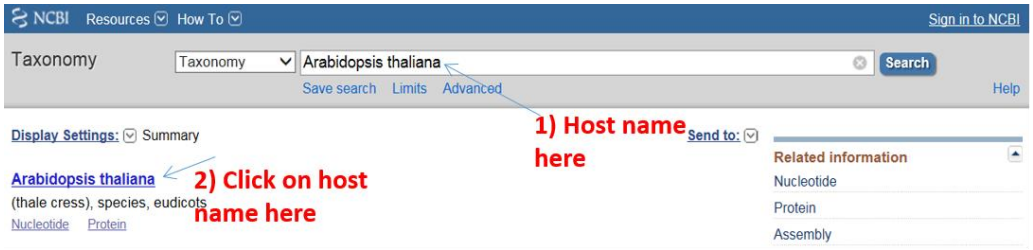
How to look up NCBI info. For Columns 24/25/26

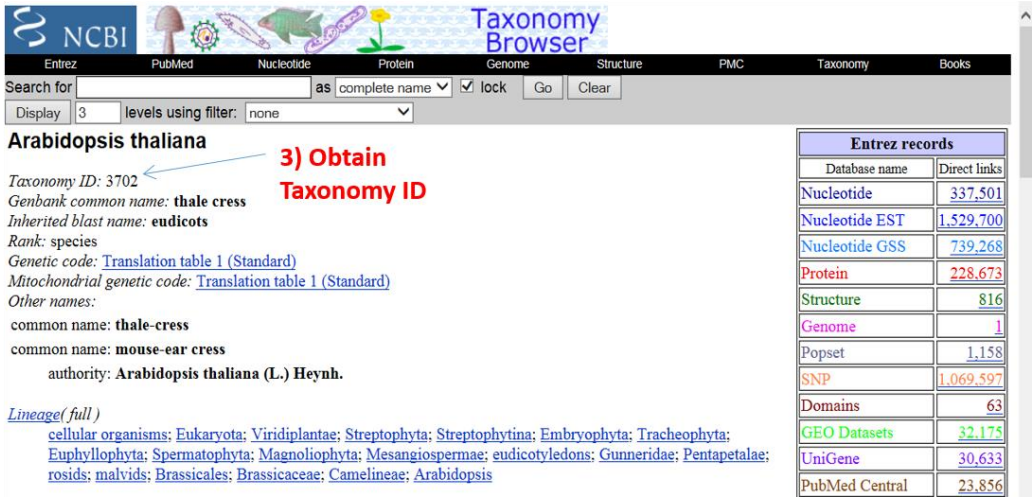




25 Host NCBI Taxonomy ID

Definition - The NCBI taxonomy ID of the host species
Vocabulary - Controlled to NCBI taxonomy ID
Mandatory - Yes
Multiple entries - No
Example - '3702' for Arabidopsis thaliana, see image below
'Lethal pathogen phenotype' –therefore no host bioassay reported
'No host tests done' –for chemistry only entries
Additional notes - lookup NCBI taxonomy ID of the host organism at http://www.ncbi.nlm.nih.gov/taxonomy see diagram for direction
Recent changes - updated AC 18Mar2015
Hint: Find in earlier entries of spreadsheet or at NCBI taxonomy.





26 Experimental host species

Definition - The Latin species name as given by NCBI taxonomy ID followed by NCBI common name of experimental host.
Vocabulary - Controlled by NCBI Taxonomy
Mandatory - Yes
Multiple entries - No
Example - 'Arabidopsis thaliana (related: Thale cress)' (see image above)
'Solanum lycopersicum (related: Tomato)'
'Mus musculus (related: House mouse)'
'Acanthamoeba castellanii (no common name found)' –where NCBI gives no common name
'Lethal pathogen phenotype' –therefore no host bioassay reported
'No host tests done' –for chemistry only entries
'Species name (related: XXXX) (non-host bioassay)'-where pathogen is tested on a non-host
Additional notes - Hint: In Introduction or earlier entries of spreadsheet.
Recent changes - updated AC 18Mar2015

27 Host strain/genotype /or cultivar/ taxonomy ID NCBI

Definition - The name of the strain, genotype, cultivar, ecotype etc. of the host species used in the experiment.
Vocabulary - Free text
Mandatory - No

Multiple entries - Yes
Example - for rice cultivar 'Sariceltik', for tomato cultivar 'Moneymaker', for 'NCBI: XXXX'
'Lethal pathogen phenotype' –therefore no host bioassay reported
'No host tests done' –for chemistry only entries
Additional notes - (Use only one cultivar to indicate the different interactions.) Hint: In Introduction or earlier entries of spreadsheet.
Recent changes -

28 HostGenotype_definedGene of interest

Definition - The host gene/s of interest which is/are not necessarily tested, but referred to in the publication with regard to the pathogen gene of interest.
Vocabulary - Free text
Mandatory - No
Multiple entries - Yes
Example - 'Snn1', 'MyoB1' , 'ANXA1', 'GmNDR1a; GmNDR1b'
Additional notes -
Recent changes - New Column

29 HostGenotype_definedGene of interest _AccessionID

Definition – Accession ID of host genotype defined gene of interest
Vocabulary – Controlled to **(**gene name with multiple entries and) source: ID
Mandatory - No
Multiple entries - Yes
Example – Give sequence or protein accession for a submitted protein or promoter sequence as: UniProt: XXXX, EMBL: XXXX, Genbank: XXXX.
'Uniprot: W5I2V6'
'Uniprot: F4HXQ7; Genbank: AC003981_23'
Where multiple entries prefix with gene name and separate entries with semicolon.
'GmNDR1a Genbank: GU132856; GmNDR1b Genbank: GU132854'
Additional notes - If not submitted to these above resources use the GENE LOCUS ID given in the publication.
Note: Genbank is 'NCBI Genbank'.
Often multiple uniprot IDs for host genes-not always clear why therfore choose Swiss-prot reviewed entry over TrEMBL unreviewed for greater accuracy. Also match to strain where possible.
Recent changes - New Column

30 tissue type

Definition – The type of host tissue where the experiment was initiated (and resulting diseased tissues given in brackets)
Vocabulary – Free text
Mandatory – Yes, if no tissue tested use
'Lethal pathogen phenotype' –therefore no host bioassay reported
'No host tests done' –for chemistry only entries
Multiple entries – Yes, separate with semicolon
Example – initial site of inoculation (systemic spread to other tissues where disease phenotype is seen)
'roots (stem, leaves)'

Additional notes -
Recent changes – inclusion of description of inoculation point and resulting diseased tissues in brackets (August 2015)

31 Function

Definition - Function of the disrupted pathogen gene
Vocabulary – Free text
Mandatory - No
Multiple entries - No
Example – Cysteine protease
Additional notes - Hint: Take information as described in the article. This is with regard to the gene of interest and the function is to be captured as assigned by the author.
Recent changes -

32 GO annotation

Definition - List of all appropriate Gene Ontology (GO) terms with their corresponding evidence codes and / or Enzyme Commission (EC) numbers for pathogen gene of interest.
Vocabulary – Controlled to annotation
Mandatory - No
Multiple entries - Yes
Example –
'no data found'
GO:0031176, NAS; EC:3.2.1.8
GO:0009405, IMP; GO:0044409, IMP
Additional notes - GO terms separate from evidence code by comma, GO terms separate from GO terms or EC numbers by semicolon
Recent changes -
Hint: This field will be computationally completed by automatic retrieval via Uniprot giving all general GO terms related to gene and EC number.

33 Database

Definition - Name of the Database generating data in GO annotation Column.
Vocabulary - Controlled
Mandatory - No
Multiple entries - Yes
Example - GO, EC or 'no data found'
Additional notes -
Recent changes -

34 Pathway/secretion systems

Definition - Name of the pathway the disrupted pathogen gene is involved in.
Vocabulary – Free text
Mandatory - No
Multiple entries -

Example – 'trichothecene biosynthesis'
'Chitin biosynthesis'
'MAPK signalling pathway'
Additional notes - Hint: Designated by authors.
Recent changes –

35 Phenotype of mutant

Definition The phenotype that results from a pathogen-host interaction.
Vocabulary – Controlled to underlined lower case text below:

TERMS (NINE):
•loss of pathogenicity
the transgenic strain fails to cause disease
•reduced virulence
the transgenic strain still causes some disease formation but this is less than the wild-type strain (ie. a quantitative effect). Synonymous with the term reduced aggressiveness.
•unaffected pathogenicity
the transgenic strain which expresses no or reduced levels of a specific gene product(s) has wild-type disease causing ability
•increased virulence (hypervirulence)
the transgenic strain causes higher levels of disease than the wild-type strain
•effector (plant avirulence determinant)
currently a plant pathogen specific term which was previously known as an avirulence gene. An effector gene is required for the direct or indirect recognition of a pathogen only in resistant host genotypes which possess the corresponding disease resistance gene. Positive recognition leads to activation of plant defences and the pathogen fails to cause disease. Note some effector genes are required to cause disease on susceptible hosts but most are not.
•lethal
the transgenic strain which expresses no or reduced levels of a specific gene product(s) is not viable. The gene product is essential for life.
•enhanced antagonism
the transgenic strain of endophytes which shows no asymptomatic colonisation, but is gaining the upperhand of the pathogen-host-interaction.
•wild-type mutualism
the transgenetic strain of endophytes which shows no difference in the pathogen-host-interaction comparing to the wild-type.
•chemistry target: resistance to chemical
chemistry target: sensitivity to chemical

Mandatory – Yes
Multiple entries - No
Example - above
Additional notes - Hint: Find in results section, in text but also tables and figures. ALWAYS FILL THIS IN.
Recent changes –

36 Mating defect prior to penetration

Definition – Does the experimental pathogen have a mating defect prior to host penetration?
Vocabulary – Controlled
Use one of the following terms:
'yes', 'no', 'no data found'
Mandatory - No
Multiple entries - Yes separate with semicolon and space
Example – 'yes'
Additional notes - Hint: Find in results section, in text but also tables and figures.
Recent changes -

37 Pre-penetration defect

Definition - Does the experimental pathogen have a pre-penetration defect?
Vocabulary – Controlled
Use one of the following terms:
'yes', 'no', 'no data found'
Mandatory - No
Multiple entries - Yes separate with semicolon and space
Example - 'yes'
Additional notes - Hint: Find in results section, in text but also tables and figures.
Recent changes -

38 Penetration defect

Definition - Does the experimental pathogen have a penetration defect?
Vocabulary – Controlled
Use one of the following terms:
'yes', 'no', 'no data found'
Mandatory - No
Multiple entries - Yes separate with semicolon and space
Example - 'yes'
Additional notes - Hint: Find in results section, in text but also tables and figures.
Recent changes -

39 Post-penetration defect

Definition - Does the experimental pathogen have a post-penetration defect?
Vocabulary – Controlled
Use one of the following terms:
'yes', 'no', 'no data found'
Mandatory - No
Multiple entries - Yes separate with semicolon and space
Example – 'yes'
Additional notes - Hint: Find in results section, in text but also tables and figures.
Recent changes -

40 Disease development (macroscopically visible)

Definition – Description of macroscopically visible disease development.
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example - for M. oryzae:
'no data found; low disease severity; few and isolated lesions; typical blast lesions'
Additional notes -
Recent changes -

41 Vegetative spores

Definition - Description of the experimental pathogen's vegetative spores.
Vocabulary - Controlled
Use the following terms:
'increased', 'wild type', 'reduced', 'defective', 'aberrant', 'no data found'
Mandatory - No
Multiple entries - Yes
Example -
'increased'
'aberrant; reduced'
Additional notes - Hint: Find in results section, in text but also tables and figures. Also consider supplementary materials.
Recent changes -

42 Sexual spores

Definition - Description of the experimental pathogen's sexual spores.
Vocabulary - Controlled
Use the following terms:
'increased', 'wild type', 'reduced', 'defective', 'aberrant', 'no data found'
Mandatory - No
Multiple entries - Yes
Example -
'increased'
'aberrant; reduced'
Additional notes -
Recent changes -

43 In vitro growth

Definition - Description of how the experimental pathogen grows *in vitro*.
Vocabulary - Controlled
Use the following terms:
'increased', 'wild type', 'reduced', 'defective', 'aberrant', 'no data found'
Mandatory - No
Multiple entries - Yes
Example -
'increased'

'aberrant; reduced'
Additional notes -
Recent changes -

44 Spore germination

Definition - Description of the experimental pathogen's spore germination.
Vocabulary - Controlled
Use the following terms:
'increased', 'wild type', 'reduced', 'defective', 'aberrant', 'no data found'
Mandatory - No
Multiple entries - Yes
Example -
'increased'
'aberrant; reduced'
Additional notes -
Recent changes -

45 Essential gene (knockout is lethal)

Definition - Definition of "essential gene": a gene critical for growth in vitro and/or in host.
No: growth in vitro observed; Yes: Conclusion of study by authors (if yes can be different to Col.35 'phenotype of mutant').
Vocabulary - Controlled
Use one of the following terms:
'yes', 'no', 'no data found'
Mandatory - Yes
Multiple entries - No
Example -
Additional notes - Hint: Find in results section, in text but also tables and figures.
Also consider supplementary materials.
Note: Relationship of this column to Col.35 'phenotype of mutant' Example PHI:2534 and 2535 double mutant ERG11A and B are lethal (lethal in col.35 'phenotype of mutant' as entry is double mutation), however not essential on own ('no' in this col regarding individual mutation).
Recent changes -further definition given (15Feb2015AC and MU)

46 Inducer

Definition – Substance that has the capability of changing metabolic reactions within a cell, tissue or intact host
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'wounding' or application of named chemistry **'**menadione (Bion, source Syngenta)'
Additional notes - Hint: Find in Materials and Methods.
Recent changes -

47 ChemicalAccession (Chebi/CAS)

Definition – Chemical accession number from ChEBI or CAS for inducers (Col. 46) or anti-infectives (Col.73).
Vocabulary – Controlled by database
Mandatory - No
Multiple entries - Yes
Example - prefix with 'inducer' or 'anti-infective'. Eg 'inducer: H2O2: Chebi: 7722-84-1' copper sulphate: Chebi: 7758-98-7; anti-infective: fluconazole: Chebi: 46081
Additional notes –
Look up chemical identifier online at http://www.ebi.ac.uk/chebi/downloadsForward.do
(tick the 'All in ChEBI' option) For example, the ChEBI id for 'salicylic acid' is CHEBI: 16914.
If a chemical cannot be found in ChEBI, only then use CAS accession instead. CAS accession numbers are also used to link genes/fungicide targets to their respective fungicide in Excel worksheet 2. The fungicide information originates from the "Fungicide resistance action committee (FRAC)" PDF document the 2015 FRAC list: http://www.frac.info/docs/default-source/publications/frac-code-list/frac-code-list-2015-finalC2AD7AA36764.pdf?sfvrsn=4

Hint: Look up in databases
CAS weblink: https://www.cas.org/
Chebi weblink: https://www.ebi.ac.uk/chebi/

Recent changes -
prefix with 'inducer' or 'anti-infective'

48 Tested Host target

Definition – The first host target protein which interacts with the corresponding pathogen protein and/or the host cellular location in which the (pathogen) experiments have been observed.
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'GmNDR1a', 'GmNDR1b', 'GmNDR1a; GmNDR1b'
'cell wall', 'nucleus', 'no data found'
'cytoplasm; nucleus'
'cell periphery; vacuoles'
Separate multiple entries with semi colon
Additional notes – Need evidence from yeast two hybrid, co-immunoprecipitation experiments etc for interaction with first host target.
Recent changes -

49 TestedHostTarget_AccessionID

Definition – Accession ID of the tested host target gene(s) of interest (from Col.48 Tested host target).
Vocabulary - Controlled to **(**gene/protein name with multiple entries and) source: ID
Mandatory - No
Multiple entries – Yes

Example – Give sequence or protein accession for a submitted protein or promoter sequence as: UniProt: XXXX, EMBL: XXXX, Genbank: XXXX.
'Genbank: GU132856'
Where multiple entries prefix with gene name and separate entries with semicolon.
'GmNDR1a Genbank: GU132856; GmNDR1b Genbank: GU132854'
Additional notes – If not submitted to these above resources use the GENE LOCUS ID given in the publication.
Note: Genbank is 'NCBI Genbank'.
Recent changes - New Column

50 Interaction phenotype

Definition – Description of the interaction phenotype between pathogen and host.
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes, separate with semicolon
Example – 'resistance', 'wild-type susceptibility', 'binding', 'no binding', 'positive YFP fluorescence', 'no data found'
Additional notes -
Recent changes -New Column

51 Host response

Definition – Description of the host response observed upon challenge with the altered pathogen gene/protein of interest.
Vocabulary – Free text
Mandatory - No
Multiple entries – Yes, separate with semicolon
Example - stronger defence response; weaker defence response; wild-type defence response; enhanced callose deposition; reduced callose deposition; wild-type callose deposition; wild-type necrosis; more necrosis; less necrosis; wild-type chlorosis; more chlorosis; less chlorosis; no host response; no data found
Additional notes - Hint: Find in results section, in text but also tables and figures.
Recent changes -

52 Experimental evidence (stable)-

Definition – Description of the experimental procedure.
Vocabulary - Controlled using one or a combination of the following
lower case underlined terms:
promoter mutation
Putative promoter region upstream of a gene was disrupted due to transposon insertion or by the insertion of a resistance gene cassette.
gene disruption
The ORF or intergenic sequences within the gene was disrupted due to transposon insertion or by the insertion of a resistance gene cassette.
gene mutation
The gene was mutated using radiation (UV) or a chemical agent. The mutation itself is unknown.

gene mutation: characterised
The UV or chemical gene mutation was characterised by sequencing or biochemistry.
gene deletion: full
The entire gene was replaced with a resistance gene cassette.
gene deletion: partial
A part of the gene was replaced with a resistance gene cassette.
gene deletion: cluster
A small chromosomal region containing more than one gene was replaced using a resistance gene cassette.
gene complementation
The mutant phenotype was restored to wild-type by introducing a copy of the wild-type gene. This evidence is given together with other evidence such as: "gene deletion: full; gene complementation".
altered gene expression / gene regulation: overexpression
Gene expression was altered by changing the promoter/regulatory sequences. Gene expression is increased as a consequence.
altered gene expression / gene regulation: downregulation
Gene expression was altered by changing the promoter/regulatory sequences. Gene expression is reduced as a consequence.
altered gene expression / gene regulation: silencing
Gene expression is reduced by "gene silencing" using an RNAi approach.
altered gene expression / gene regulation
Gene expression was altered by changing the promoter/regulatory sequences. It is not reported if the gene is down- or up-regulated as a consequence.
biochemical analysis
The function of the gene was investigated by studying the protein in enzymatic or other assays.
functional test in host
The function of the gene was investigated by applying the protein to the host using a biological delivery system (i.e. expression of a stable transgene from the host, transient delivery with a virus, or transient delivery with Agrobacterium). Put details of strains used into the free text comment section and also state if peptide uses a signal peptide sequence or not. (This information will be turned into controlled vocabulary at a later stage.)
functional test in host: direct injection
The protein was injected into the host using a device (i.e. syringe needle).
functional test in host: transient expression
The protein was transiently expressed in the host.
sequence analysis of sensitive and resistant strains
Sequence information of both strains was analysed to infer the gene function.
sexual cross, sequencing of resistance conferring allele
Mendelian crossing was used to delineate the function of an allele.
heterologous expression
i.e. in Pasteuria pastoris
chemical complementation
i.e. addition of Ca2+ rescues MID1 Ca2+ channel gene deletion
other evidence
Format of entry: "Other evidence: <free text>"
Free text field for additional experimental evidence given by the authors if none of the above applies. Precede this text with "Other evidence:"
natural sequence variation this is NEW
Mandatory - Yes

Multiple entries - Yes
Example – 'Gene mutation; gene complementation'
Additional notes – separate multiple entries with a semicolon
Recent changes -

53 Transient Assay Experimental Evidence

Definition - Fill this in when Col.52 'Experimental evidence' is 'functional test in host: transient expression' or when 'other evidence' is used to describe a transient assay
Vocabulary – Free text to describe transient system
Mandatory – Yes when Col.52 'Experimental evidence' is 'functional test in host: transient expression' or when 'other evidence' is used to describe a transient assay
Multiple entries - Yes
Example – 'HIGS', 'Agrobacterium infiltration', 'PVX vector', 'Pseudomonas vector', 'co-bombardment', others
Additional notes -
Recent changes -New Column

54 Species expert

Definition - Abbreviation (two initials) of species expert outside Rothamsted listed in spreadsheet "curators".
Vocabulary - Free text
Mandatory - No
Multiple entries - Yes
Example -
Additional notes - Hint: Do not fill in. Will be filled in by RRes.
Recent changes -

55 Entered by

Definition - The curator's initials.
Vocabulary - Free text
Mandatory - Yes
Multiple entries - Yes
Example - 'RP' for Rashmi Pant
Additional notes - If more than one curator, separate using a semicolon. Example: "RP; AR"
Recent changes -

56 Literature ID

Definition - The PUBMED Accession number for the published article. Website http://www.ncbi.nlm.nih.gov/pubmed
Vocabulary - Accession number or 'data not found'
Mandatory - Yes
Multiple entries - No
Example - 22346755
Additional notes -

Recent changes - In past multiple entries allowed but not now. 1 reference per PHI:ID.

57 Literature source

Definition - Name of library or information resource containing the publication.
Vocabulary – Controlled to source name or 'data not found'
Mandatory - Yes
Multiple entries - No
Example – 'Pubmed', 'not in Pubmed', 'no data found',' WoS'
Additional notes - Do not use 'Web of Science' (WoS) ISSN numbers, which reflect many different papers in different journals.
Pubmed preferred. For old articles this has to be extracted from Pubmed database. For newer articles list the DOI (see Column 58) or article ID.

58 DOI

Definition - The Digital object identifier number which gives direct access to the publication.
Vocabulary – controlled to DOI
Mandatory - No
Multiple entries - No
Example - 10.1105/tpc.5.11.1575
'data not found' or 'blank'
Additional notes -
Recent changes - In past multiple entries allowed but not now. 1 reference per PHI:ID.

59 Full citation

Definition – Full citation of paper where there is no Pubmed ID or DOI
Vocabulary – Controlled to citation
Mandatory – Yes when no Pubmed ID or DOI
Multiple entries - No
Example – 'Shi et al. (1995) MOLECULAR PLANT-MICROBE INTERACTIONS 8 (6): 949-959'
Additional notes -
Recent changes -

60 Author email

Definition – The corresponding author's email address
Vocabulary – Controlled to email address
Mandatory - No
Multiple entries - No
Example – 'rouxel@versailles.inra.fr'
Additional notes - Hint: Mostly given at bottom of first page.
Recent changes -

61 Comments

Definition - Field provided for any further free text information which is only found in the abstract of the paper and is not reported in any other column.
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'Mutants more sensitive to ROS'
'Mutants showed reduced biofilm formation'
Additional notes -
Recent changes -

62 Reference

Definition –Brief publication reference
Vocabulary – Controlled to reference
Mandatory - Yes
Multiple entries - No
Example - Give name of the first author only. Use: initials, surname and year, i.e. "BN Taylor 2007". Do not use dots.).
Additional notes -
Recent changes -

63 Year published

Definition – Year reference published
Vocabulary – Controlled to year
Mandatory - Yes
Multiple entries - No
Example – '2010'
Additional notes -
Recent changes -

64 Curation details

Definition – used by curation company Molecular Connections (MC) regarding number of genes curated per paper
Vocabulary - Free text
Mandatory -
Multiple entries -
Example – '1 gene', '2 Genes'
Additional notes -
Recent changes -

65 File_name_pdf_files_provided

Definition - used by curation company Molecular Connections (MC) regarding PDF file name
Vocabulary - Free text
Mandatory -

Multiple entries -
Example -
Additional notes -
Recent changes -

66 Batch_number

Definition - RRes/MC curation batch number to allow tracking
Vocabulary – Controlled to batch number
Mandatory - Yes
Multiple entries -
Example – 'Batch022'
Additional notes -
Recent changes -

67 CurationDate

Definition - Time when the article was curated. Please use: Month/Year format, i.e. "01/2013" Allows tracking over the years
Vocabulary – Controlled to Month/Year format
Mandatory -
Multiple entries -
Example -
Additional notes -
Recent changes -

68 CuratorOrganisation

Definition - Please use: "MC" for Molecular Connection. Others are: RRes, etc.
Allows tracking
Vocabulary -
Mandatory -
Multiple entries -
Example -
Additional notes -
Recent changes -

69 __lab__

Definition - Please enter LAST AUTHOR of publication
Allows contacting the lab.
Vocabulary -
Mandatory -
Multiple entries -
Example -
Additional notes -
Recent changes -

70 __FG_mycotoxin__

Definition - Special case: Please enter if study in species Fusarium graminearum (related : Gibberella zeae) and the mycotoxin level is affected. Research interest at RRes.
Vocabulary – Controlled to one of the following: increased, wildtype, reduced, absent.
Mandatory -
Multiple entries -
Example -
Additional notes -
Recent changes -

71 Additional identifier type of gene locus ID

Definition – To prevent loss of information where Uniprot and EMBL/Genbank have already been entered into columns 6-8, the source of species specific accessions Broad/JGI etc. information can be entered here.
Vocabulary -
Mandatory - No
Multiple entries -
Example – 'JGI', 'Broad', 'MUMDB'
Additional notes -
Recent changes - New Column (10Nov2014 AC)

72 Additional gene locus ID

Definition - To prevent loss of information where Uniprot and EMBL/Genbank have already been entered into columns 6-8, the Gene Locus ID of species specific accessions Broad/JGI etc. assigned by the genome sequencing centre can be entered here.
Vocabulary -
Mandatory - No
Multiple entries -
Example – 'FGSG_08713', 'SNOG_02041,' VC_1459'
Additional notes -
Recent changes - New Column (10Nov2014 AC)

73 Anti-infective (Chemical) (alternate header: Anti infective agent)

Definition – The chemical used to prevent infection
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'Methyl 1-(butylcarbamoyl)-2-benzimidazolecarbamate'
when more than one entry prefix with compound name and colon, separate multiple entries with semicolon. 'benomyl: Methyl 1-(butylcarbamoyl)-2-benzimidazolecarbamate; diethofencarb: 1-methylethyl (3,4-diethoxyphenyl)carbamate'
Additional notes -
Recent changes -

74 Compound (alternate header: Anti infective compound)

Definition – Common name of anti-infective chemical
Vocabulary – Free text
Mandatory - No
Multiple entries – Yes
Example – 'benomyl'
'benomyl; diethofencarb'
Additional notes -
Recent changes -

75 Target site (alternate header: Anti infective-target site)

Definition – Site where compound binds (protein or nucleic acid)
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'b-tubulin assembly in mitosis'
'benomyl: b-tubulin asssembly in mitosis; diethofencarb: b-tubulin asssembly in mitosis'
Additional notes -
Recent changes -

76 Group name (alternate header: Anti infective group name)

Definition – Based on chemical relatedness of structures as given in the literature (e.g. the Pesticide Manual)
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'MBC-fungicides (Methyl Benzimidazole Carbamates)'
'benomyl: MBC-fungicides (Methyl Benzimidazole Carbamates); diethofencarb: N-phenyl carbamates'
Additional notes -
Recent changes -

77 Chemical group (alternate header: Anti infective Chemical group)

Definition – Based on chemical considerations. Nomenclature is according to IUPAC and Chemical abstract name.
Vocabulary – Free text
Mandatory - No
Multiple entries – Yes if multiple chemical used
Example – 'benzimidazoles'
'benomyl: benzimidazoles; diethofencarb: N-phenyl carbamates'
Additional notes -
Recent changes -

78 Mode in planta (alternate header: Anti infective-Mode in planta)

Definition – Distribution of anti-infective in the plant.
Vocabulary – Controlled
'Contact'
'Translaminar'
'Systemic'
'Locally systemic' (ie within leaf, not into stem)
Mandatory - No
Multiple entries - Yes
Example – 'systemic'
benomyl: systemic; diethofencarb: systemic
Additional notes -
Recent changes -

79 Mode of action

Definition – Chemical pathway of pathogen affected by anti-infective.
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example – 'Mitosis and cell division'
'benomyl: Mitosis and cell division; diethofencarb: Mitosis and cell division'
Additional notes -
Recent changes -

80 FRAC CODE (alternate header: FRAC CODE)

Definition – Code given to anti-infective by FRAC dependent on their cross-resistance behaviour. See FRAC code list 2015 website
Vocabulary – Controlled by FRAC
Mandatory - No
Multiple entries - Yes
Example – '1'
'benomyl: 1; diethofencarb: 10'
Additional notes -
Recent changes -

81 Additional comments on anti-infectives (alternate header: Anti infective-comments)

Definition - Additional comments on anti-infectives
Vocabulary – Free text
Mandatory - No
Multiple entries - Yes
Example -
Additional notes -
Recent changes -