

Using a FAIR database and bioinformatics analyses to improve plant, human, animal and ecosystem health

Martin Urban and Alayne Cuzick
Kim Hammond-Kosack



UK node



The Pathogen-Host Interactions database: PHI-base

A multispecies database used to understand the mechanisms underlying pathogenesis (disease formation) and explore new options to protect crop, human, animal and ecosystem health

PHI-
base Pathogen
Host
Interactions

What is PHI-base?

- Web accessible relational **phenotype** database that **catalogues genes whose contribution to the pathogenic process has been tested experimentally**
 - **Manually curated** by domain / species experts
 - **Entries supported by peer reviewed literature** references
(*pathogen gene deletion / gene modification vs WT strain*)
 - Broad scope of pathogen and host species
 - Positive and **negative data** included
- Freely accessible at www.phi-base.org since 2005
- **Regular users based in 130 countries**
- **> 12,000 users / annum**
 - Winnenburg et al. (2006 and 2008) *Nucleic Acids Research (Database issues)*
 - Urban et al. (2015 and 2017) *Nucleic Acids Research (Database issues)***
 - Urban et al. (2015) *Frontiers in Plant Sciences*, doi: 10.3389/fpls.2015.00605
 - Brown et al. (2016) *FEMS Microbiological Reviews* 40, 19-40

Motivation for Pathogen-Host interactions database

- Major crop disease outbreaks, human and animal epidemics
- Emerging fungicide / anti-infective resistance



PHI-base is a multi-species database for fungi, protists and bacteria pathogens (Version 4.8- Sept 2019)

60% plant pathogens, 35% human_animal_fish pathogens, 5% insect_others

Total ~ 250 pathogenic species (fixed)

Top 15 pathogen species in PHI-base

Host Plant Focus

- Agricultural crops
- Horticultural crops
- Commercial trees (timber, fruit)
- Model species
- Emerging crop plant threats (pandemics)

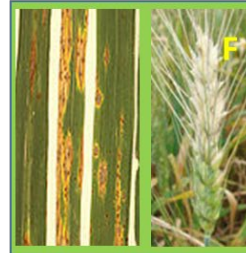
Host Animal Focus

- Key human pathogens
- 3Rs species alternative
- Plant attacking nematodes
- Plant attacking insects

Fusarium graminearum



Magnaporthe oryzae



Ralstonia solanacearum



Salmonella enterica



Xanthomonas oryzae



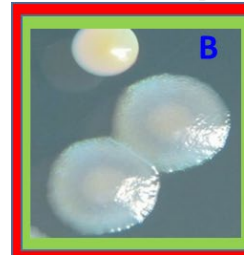
Erwinia amylovora



Candida albicans



Pseudomonas aeruginosa



Botrytis cinerea



Ustilago maydis



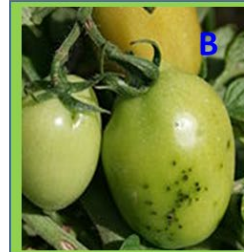
Aspergillus fumigatus



Cryptococcus neoformans



Pseudomonas syringae



Escherichia coli



Staphylococcus aureus



Plant pathogen

Animal pathogen

Plant and animal pathogen

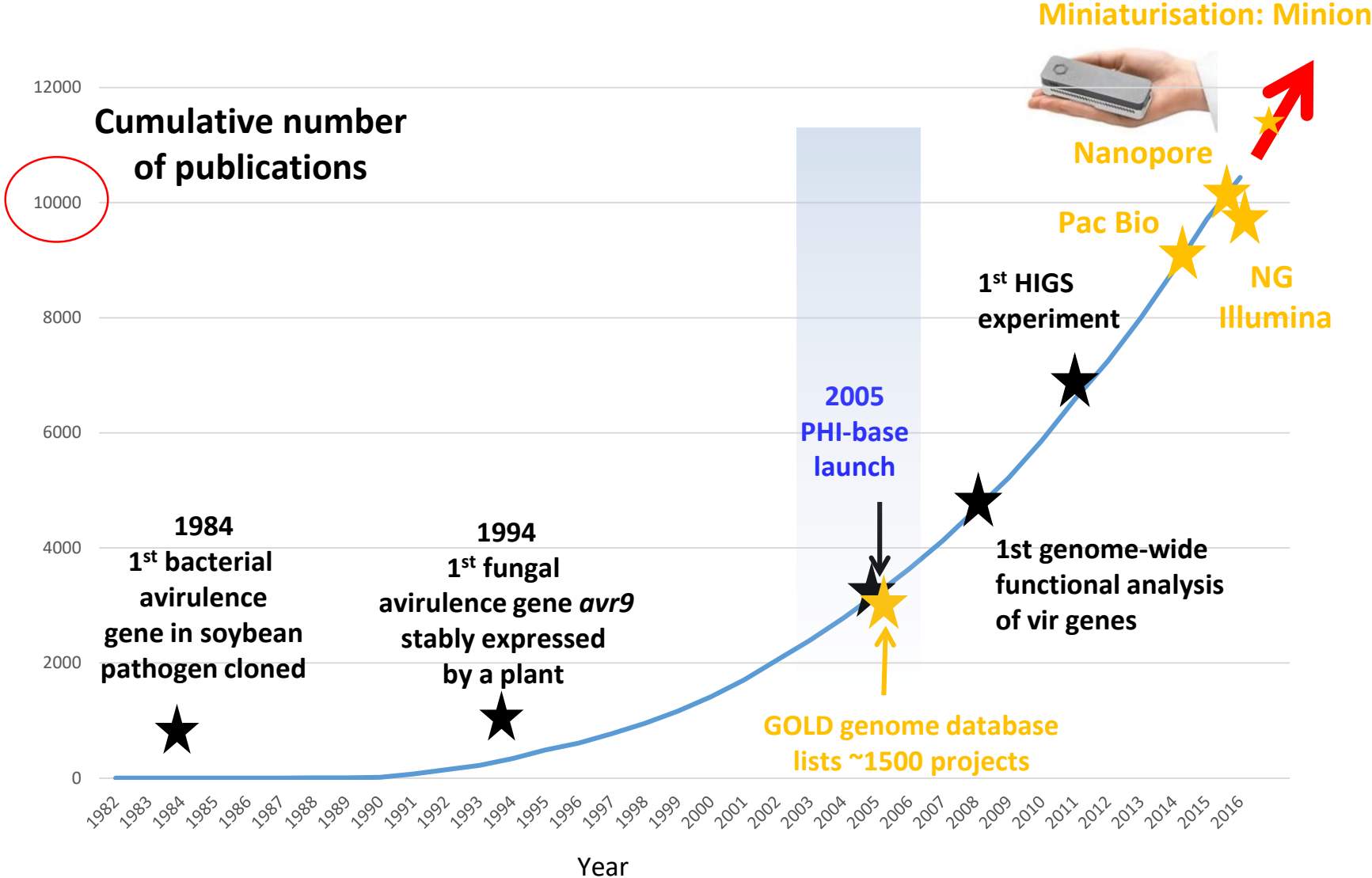
Current PHI-base version 4.8 release (Sept 2019)

	www.phibase.org
References	3454
Genes	8351
Interactions	13801
Pathogens	268
Hosts	210
Diseases	502
Chemistries	34 fungicides 149 anti-infectives

- Pathogens: 75% eukaryotes (fungi_protists), 25% bacteria
- Hosts: 60% plant (50:50 cereal:non-cereal), 40% animal_human

Data and knowledge explosion over the last 40 years

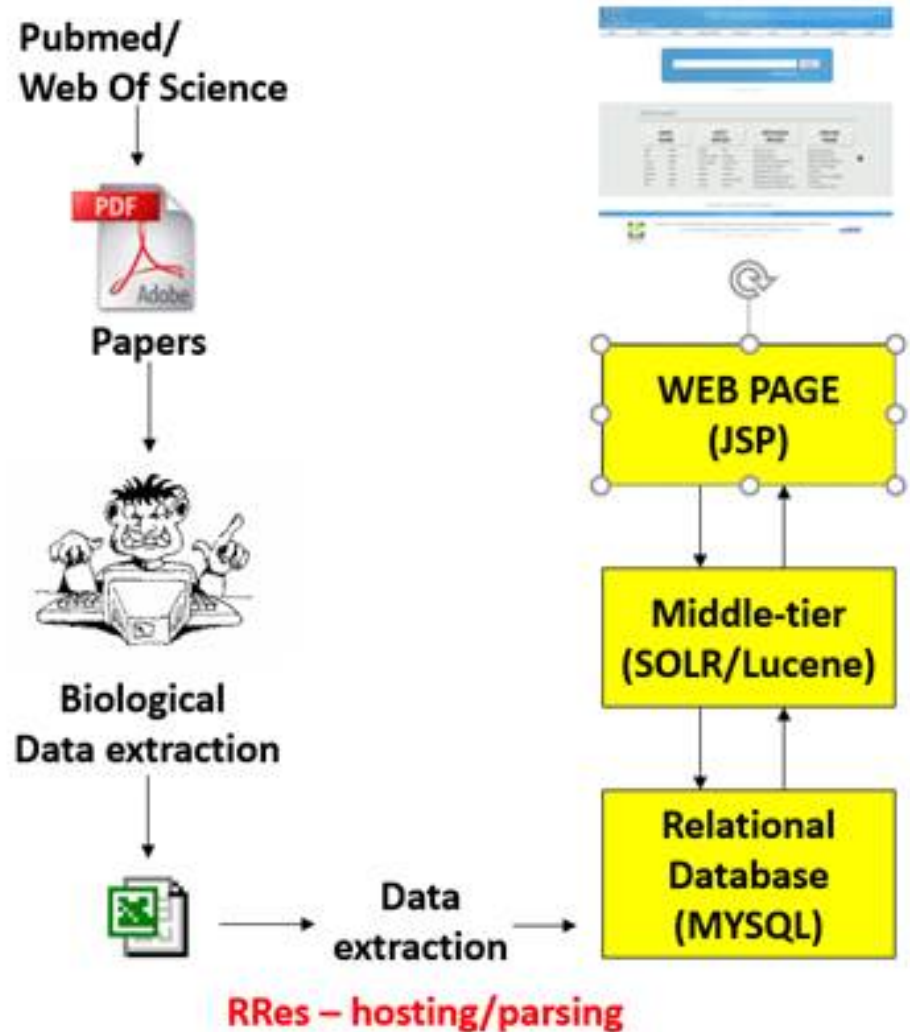
Increasing number of publications on virulence associated genes and their phenotypes in microbial pathogens (Web of Sciences/PubMed)



Database curation workflow

at RRes and Molecular Connections (MC) India

- MC: keyword searches for new papers to generate candidate list (trailing machine learning)
- RRes: review of list and add in submissions from species experts/users, collect PDFs, send to MC
- MC: curation of 10 papers/week; forward data in monthly batches
Up to 81 data entries / gene
- RRes: review of data
2 database releases/year



The PHI-base interface has a faceted views to allow complete visualisation of all curated PHI-base data!

PHI-base Pathogen Host Interactions

Home About Us Search Release notes Download Disclaimer Errors Help Community PHIB-BLAST Anti-infective

Free text entry with autocomplete (1) SEARCH RESET ADVANCED SEARCH (5)

(Use AND, OR for multiple operations eg: Barley AND Loss of pathogenicity)

Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species
PMK1	reduced virulence	Magnaporthe oryzae	Rice blast	Oryza sativa (related: Rice)

RESET REFINER SEARCH

Year

2010(11)

Gene (4)

- CBP1(2)
- CPKA(2)
- EXP5(2)
- Lf pmk1(1)
- MST12(2)

Disease

- Dry bubble disease(1)
- Rice blast(10)

Gene (2)

Gene:PMK1
Phibase accession ID:PHI:2163
Gene ID:EHA52368.1
Protein ID: G4N0Z0
Sequence strain:70-15
Pathway:PMK1 signalling pathway
Essential gene:no

Pathogen (3)

Pathogen species:Magnaporthe oryzae
Pathogen ID: 318829
Pathogen strain:Guy11

Host

Host species:Oryza sativa (related: Rice)
Host classification:Monocots
Host ID: 4530
Host strain:CO39
Tissue:roots

Reference

Pmid:20348434
Ref source:Pubmed
Year:2010
DOI:10.1105/tpc.109.066340
Author reference:Sara L. Tucker

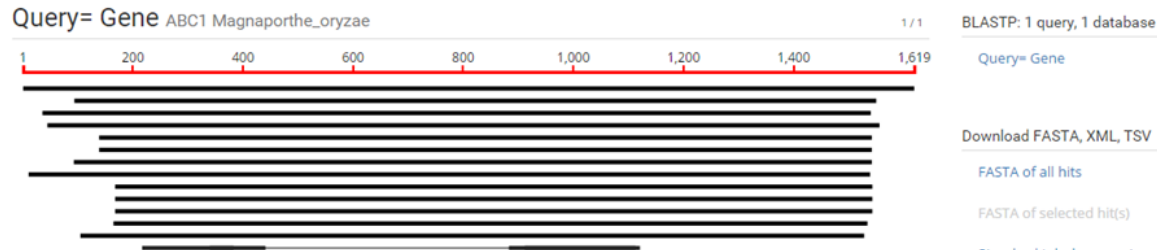
Disease Disease Process Disease Intervention

Tissue:roots
Mutant phenotype:reduced virulence
Multiple mutation:no
Comments:no comments

(1) Search panel for free text
(2) Detail view of gene results
(3) Data on different tabs
(4) Facet view with summary of results with 'click-box' selection for follow up
(5) BLAST search function

PHI-base protein BLASTP search functionality

Typical BLAST result output provides score list



Sequence label format: PHI-base Accession#Gene Name#UniProt Accession#Pathogen Taxon ID#Pathogen Species#Phenotype

Number	Sequences producing significant alignments	Total score	E value	Length
1.	PHI:132#ABC1#AAB86640#318829#Magnaporthe_oryzae#Reduced_virulence	3371.64	0.00	1619
2.	PHI:2693#GcABC-G1#F0XP73#226899#Grosmanella_clavigera#Reduced_virulence	2065.81	0.00	1530
3.	PHI:258#GpABC1#CAC40023#5128#Gibberella_pulicaris#Reduced_virulence	1961.81	0.00	1491
4.	PHI:543#BCATRD#CAC41639#40559#Botrytis_cinerea#Unaffected_pathogenicity	1722.98	0.00	1501
5.	PHI:1159#MgAtr7#ASH456#54734#Mycosphaerella_graminicola#Chemistry_target			
6.	PHI:867#MgAtr7#ABN41482#54734#Mycosphaerella_graminicola#Unaffected_pathogenicity			
7.	PHI:2815#pdr15#Q04182#4932#Saccharomyces_cerevisiae#Unaffected_pathogenicity			
8.	PHI:310#MgAtr4#AAK15314#54734#Mycosphaerella_graminicola#Reduced_virulence			
9.	PHI:2309#BcatrB#Q9UW03#40559#Botrytis_cinerea#Reduced_virulence	891.34	0.00	1439
10.	PHI:1160#BcatrB##40559#Botrytis_cinerea#Chemistry_target	891.34	0.00	1439

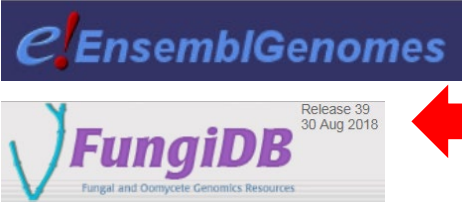
Header includes PHI-base ID and high level phenotype

[Back to PHI-base](#)

Main uses of PHI-base



> **360 publications citing PHI-base**

All papers are cited in the about section of the database

- Quick lookup of gene mutant_ phenotype relationships
(human readable knowledge database/computer readable)
key word search, BLAST tool and a full download function
- **Provide phenotype annotations for genome browsers and knowledge networks**

- **Annotate novel datasets for candidate gene selection:**
 - Newly sequenced pathogen genomes or exploring variomes (variant genomes)
 - RNA-seq and microarray data sets (enrichment of virulence genes)
 - Other candidate gene lists (i.e. from forward genetics screens, GWAS)
- Comparative genomics/phenomics
 - Conserved themes vs species-specific differences
 - *In silico* predictions of candidate virulence using protein-protein interaction networks

Nine high level PHI-base phenotype outcomes





Generic (species neutral) terms to permit wide inter-species comparisons

- Virulence terms
1. Loss of pathogenicity
 2. Reduced virulence
 3. Increased virulence
 4. Unaffected pathogenicity
 5. Effector (transferred 'entity' in an interaction)
 6. Enhanced antagonism (endophytes)
 7. Essential
 8. Resistance to chemistry
 9. Sensitive to chemistry
-  
- 35 fungicides

Terms are routinely published in research articles but mapping to GO terms is not supported due to their high-level nature.

High level phenotypes are assigned to **each** host-pathogen interaction

An **interaction** is defined as the function of one gene, on one host and one tissue type from one publication. **One high-level phenotype term is assigned.**

	Host/tissue	Interaction	Phenotype
 <i>Tri5</i> mutant Pathogen species: <i>Fusarium graminearum</i>		1	Reduced virulence
		2	Reduced virulence
		3	Unaffected pathogenicity

This permits narrow and wide intra- and inter-species comparisons

Summary of phenotypic outcome content in PHI-base (ver 4.8, Sept 2019)

PHENOTYPIC OUTCOME	%
Loss of pathogenicity	6.6%
Reduced virulence	44.4%
Increased virulence	4.6%
Unaffected pathogenicity	26.4%
Effector (transferred 'entity')	16.4%
Enhanced antagonism	0.1%
Essential	1.3%
Resistance to chemistry	0.3%
Sensitive to chemistry	0.1%



Integration of PHI-base data with Ensembl Genomes



ROTHAMSTED
RESEARCH

PHI-
base Pathogen
Host
Interactions



since 2010

EMBL-EBI



e!EnsemblGenomes

Data:

Phenotype

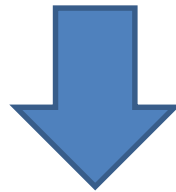
Experimental information

Data:

Reference genomes for pathogens

Variation

Comparative analysis (**Compara Tool to explore closely related sub-taxa**)



Phenotype information mapped onto Ensembl Genomes

(direct hits (**UniProt Ids**) and sequence similarity (**100% identity BlastP within sub-taxa**))

- Ensembl Fungi: 166 genomes
- Ensembl Protists: 44 genomes
- Ensembl Bacteria: 820 genomes

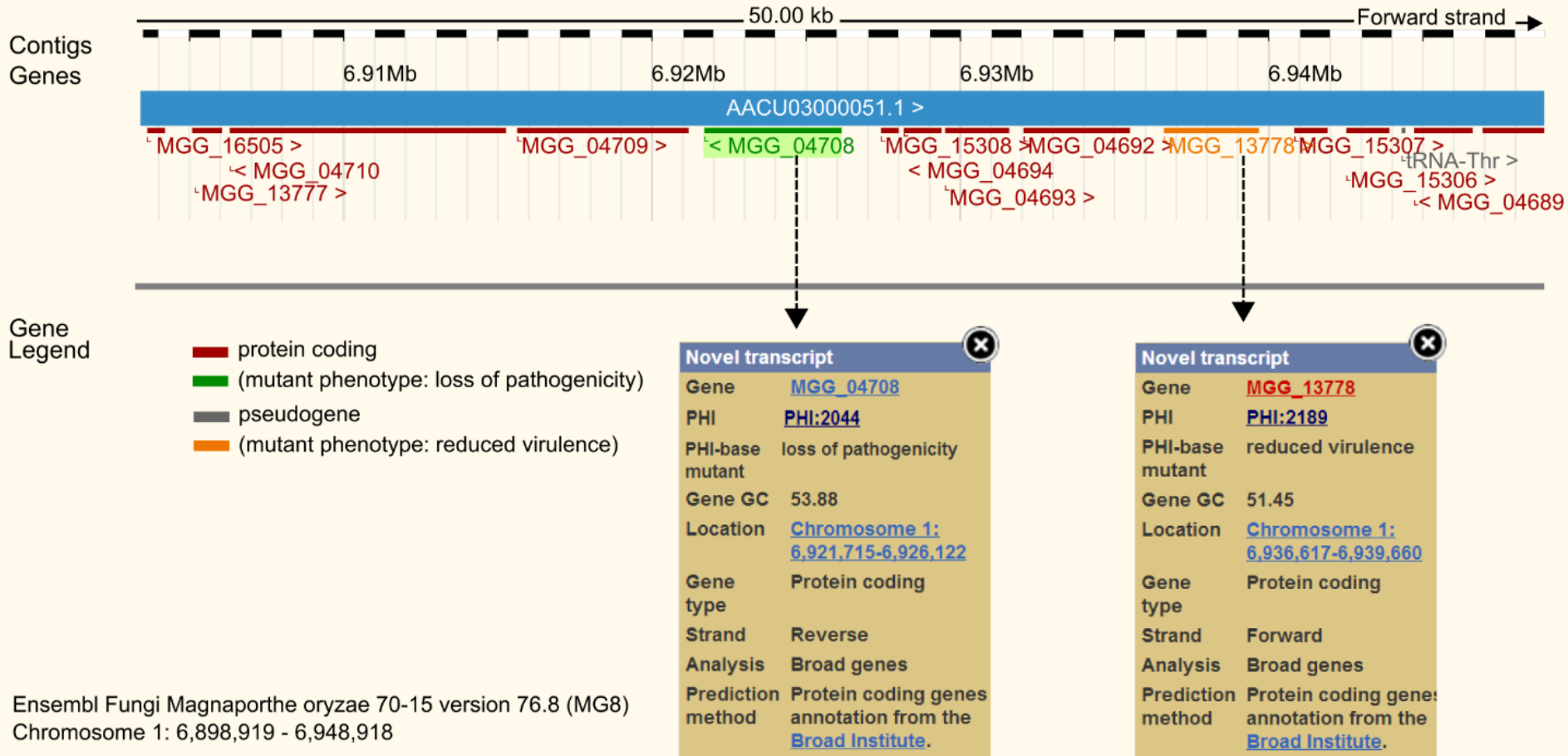
BioMART search for PHI-base annotated pathogen genes

Display phenotypes directly in Ensembl Genomes browser

Ensembl Fungi



Magnaporthe oryzae
Magnaporthe oryzae 70-15



Search *Zymoseptoria tritici* genes by phenotype using the BioMart tool

Gene stable ID	Transcript stable ID
Mycgr3G98311	Mycgr3T98311
Mycgr3G70181	Mycgr3T70181
Mycgr3G70181	Mycgr3T70181
Mycgr3G46840	Mycgr3T46840
Mycgr3G46840	Mycgr3T46840
Mycgr3G46840	Mycgr3T46840
Mycgr3G46840	Mycgr3T46840
Mycgr3G46840	Mycgr3T46840
Mycgr3G80707	Mycgr3T80707
Mycgr3G80707	Mycgr3T80707
Mycgr3G69942	Mycgr3T69942

Results of search for genes with a PHI-base annotation of 'reduced virulence'

e!EnsemblFungi

New | Count | Results

Dataset
Zymoseptoria tritici genes (MG2)

Filters
[None selected]

Attributes
Gene stable ID
Transcript stable ID

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:

GENE:

PATHOGEN PHENOTYPES (PHI-BASE):

Host (source: PHI-base)

Pathogenic phenotype (source: PHI-base)

Experimental condition (source: PHI-base)

GENE ONTOLOGY:

MULTI SPECIES COMPARISONS:

PROTEIN DOMAINS AND FAMILIES:

VARIANT:

Wheat

chemistry target: resistance to chemical
chemistry target: sensitivity to chemical
effector (plant avirulence determinant)
increased virulence (hypervirulence)
loss of pathogenicity

gene complementation
gene mutation: characterised
biochemical analysis
gene deletion
gene deletion: full

Full BioMart results output and export options

Dataset
Zymoseptoria tritici genes (MG2)

Filters
Pathogenic phenotype (source: PHI-base): reduced virulence

Attributes
Gene stable ID
Transcript stable ID
Pathogenic phenotype
PHI-base ID
Host
Experimental condition

Dataset
[None Selected]

Export all results to Unique results only

Email notification to

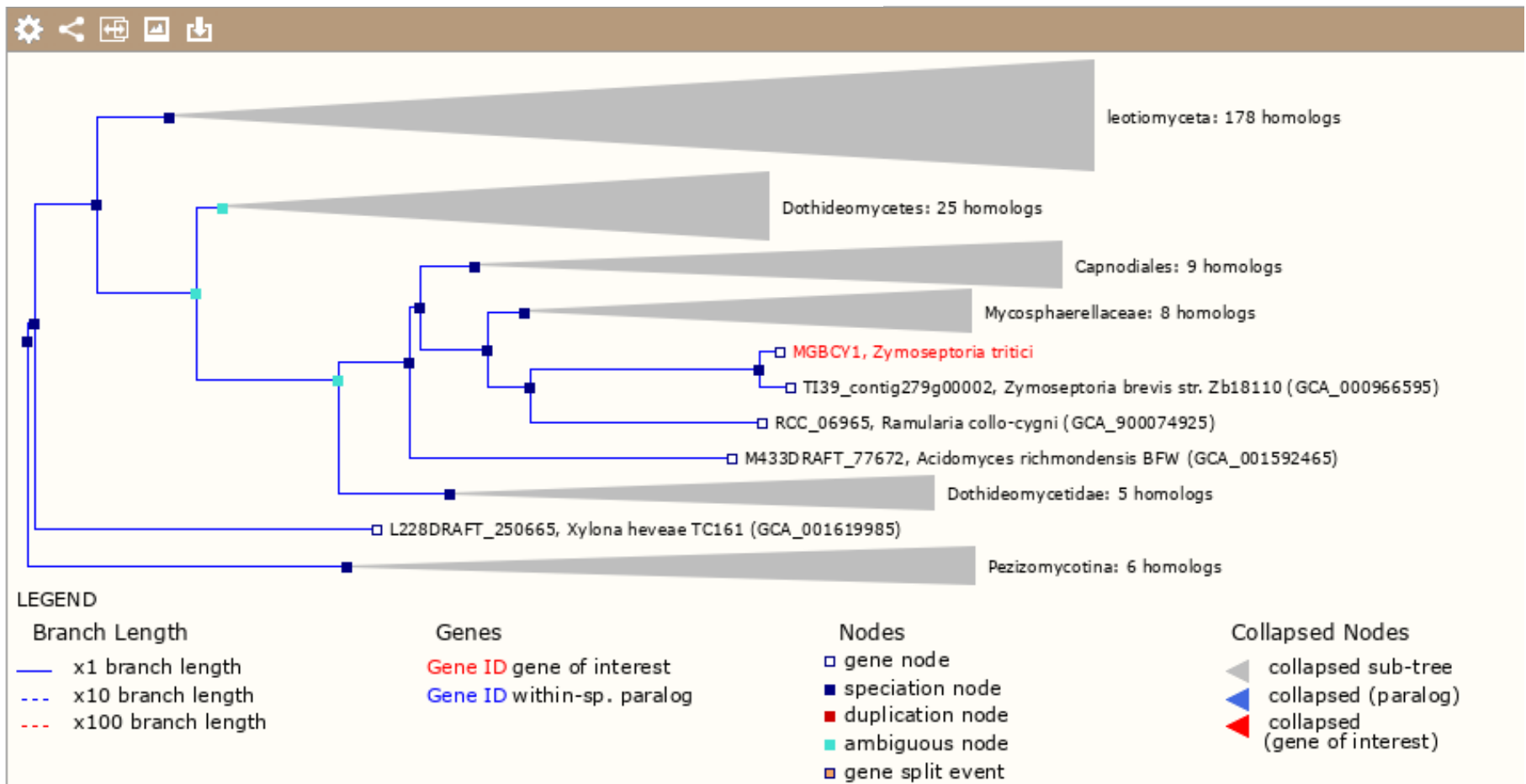
View rows as Unique results only

Gene stable ID	Transcript stable ID	Pathogenic phenotype	PHI-base ID	Host	Experimental condition
Mycgr3G98311	Mycgr3T98311	reduced virulence	PHI:1045	Wheat	gene disruption
Mycgr3G70181	Mycgr3T70181	reduced virulence	PHI:7445	Wheat	gene deletion: full
Mycgr3G70181	Mycgr3T70181	reduced virulence	PHI:7445	Wheat	gene complementation
Mycgr3G46840	Mycgr3T46840	reduced virulence	PHI:1075	Wheat	gene deletion
Mycgr3G46840	Mycgr3T46840	reduced virulence	PHI:1075	Wheat	gene complementation
Mycgr3G46840	Mycgr3T46840	reduced virulence	PHI:2126	Wheat	gene deletion: full
Mycgr3G46840	Mycgr3T46840	reduced virulence	PHI:2126	Wheat	gene complementation
Mycgr3G80707	Mycgr3T80707	reduced virulence	PHI:6546	Wheat	gene complementation
Mycgr3G80707	Mycgr3T80707	reduced virulence	PHI:6546	Wheat	gene mutation
Mycgr3G69942	Mycgr3T69942	reduced virulence	PHI:4914	Wheat	gene disruption

Using the COMPARA tool to explore homologues in other taxa

Show/hide columns (1 hidden)		Filter				
Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
Novel	Mycgr3T98311	1514	460aa	Protein coding	F9WW98	

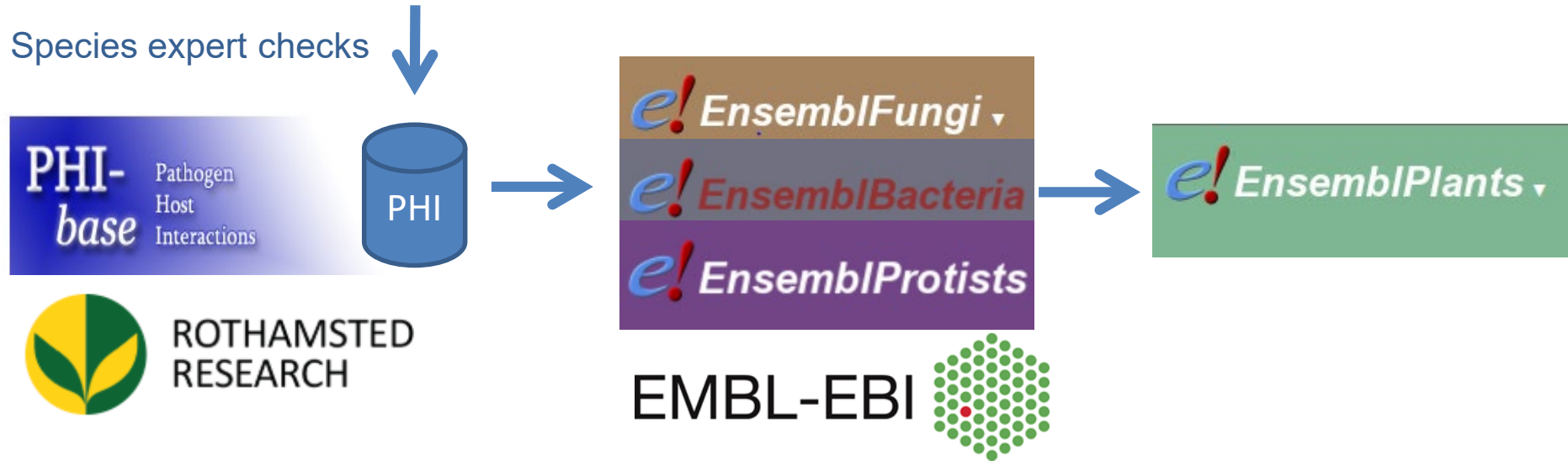
GeneTree [ENSGT00930000004479](#)



Two new features for PHI-base since 2017

1. Improved author curation tool: **PHI-Canto**

2. Curation of the **first host targets** in plants of pathogen effectors



Further integration of PHI-base data with Ensembl Genomes and the BioMart tools

Community multi-species curation tool for authors PHI-Canto

<http://curation.phi-base.org> (beta version)



Curation application for the PHI-base database

[Questions?](#) [Contact curators ...](#)

Curate genetic and molecular data from your publications

Curate a paper

Start curating using a PubMed ID:

20569395

Find ...

Once you have found your paper, you can curate GO, phenotypes, interactions and protein modifications -- or reassign the paper to a colleague for curation. See the [help pages](#) for more information.

Your curation sessions

- PMID:17506324

Documentation

Find out about community curation and the types of data that can be curating by viewing our help pages:

Visit ...

Or try the demo version of Canto:

Demo ...

**Start with Manuscript ID,
enter details including gene IDs (UniProt),
phenotypes, submit**

Anticipated curation time per article: 15-30 min

New BBSRC-BBR grant (July 2019 for 3 years)

Focus: Further PHI-Canto, metagenotype and ontology developments

PHI-base uses that are relevant to EPPO / Euphresco mission in the era of metagenomics

Emerging infectious diseases (EIDs)

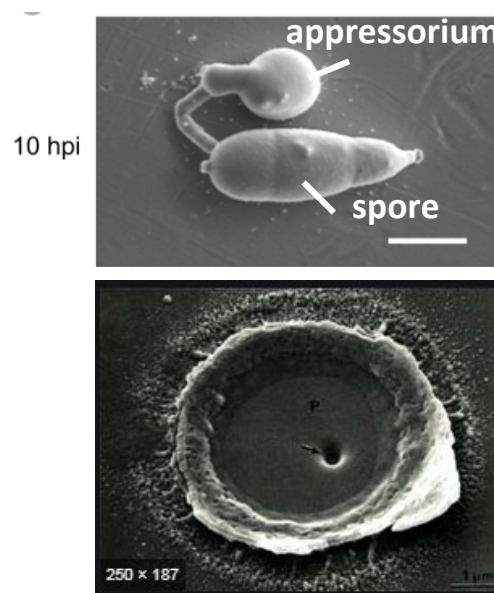
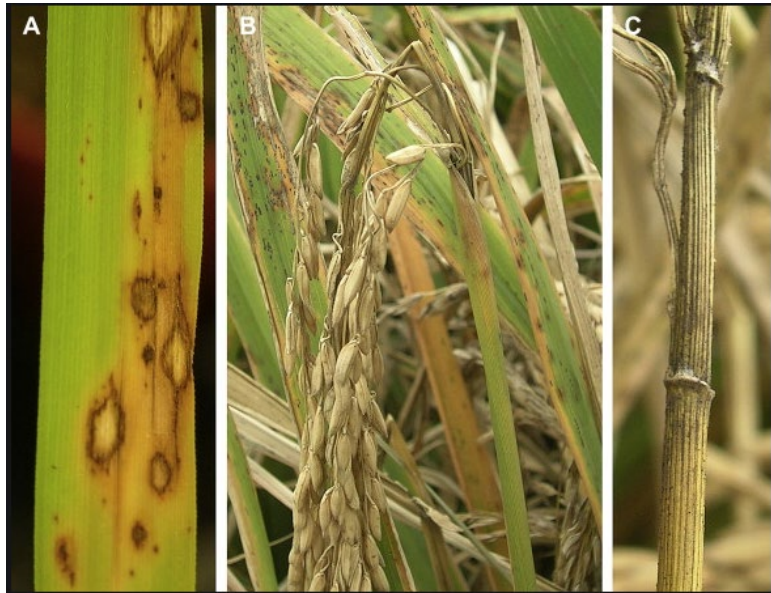
Typical workflow

- ➔ Identify the problematic pathogen species taxa
 - ➔ Full genome sequencing, assembly and ORF predictions
 - ➔ **Annotation**
 - Limited gene set
 - Full genome annotation

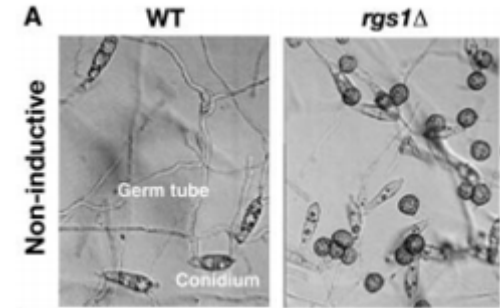
PHI-base - hypervirulence category (4.6%) - ability to cause more disease
635 interactions (529 genes) - 108 pathogenic species
244 interactions (198 genes) - 49 plant pathogenic species
29 fungi, 15 bacteria
4 protists, 1 nematode

Rice and grape vine hosts have the highest number of entries

Magnaporthe oryzae (rice blast pathogen) – 10 hypervirulence entries



**Negative regulator – RGS1
(intracellular signalling)**



**more
appressorium**

***rgs1* mutant produces more
asexual spore**

Many fungal pathogens need to produce an appressorium to penetrate plant cell surfaces including:

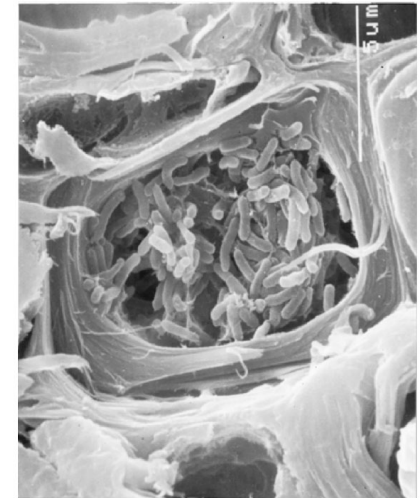
Powdery mildews, the rust and *Colletotrichum* species

Exploiting knowledge on a generic ‘lifestyle’ theme

Xylella fastidiosa (Grape vine infecting bacteria) - 13 hypervirulence entries



2013 1st recorded in Europe



Gram-negative bacteria
in xylem vessels

Outer membrane vesicles (OMVs) produced by *Xylella*, block its interaction with various surfaces including xylem vessel walls. **OMVs stops bacterial adhesion**

Low OMVs (bacteria are static)
- insect vector acquisition



High OMVs – systemic bacteria spread in host

A *X. fastidiosa* Δ *rpfF* mutant in which quorum signalling is disrupted, has increased OMV release (5 fold), and as a result adheres less to plant cell surfaces, becomes more systemic within the olive plant and causes more disease.

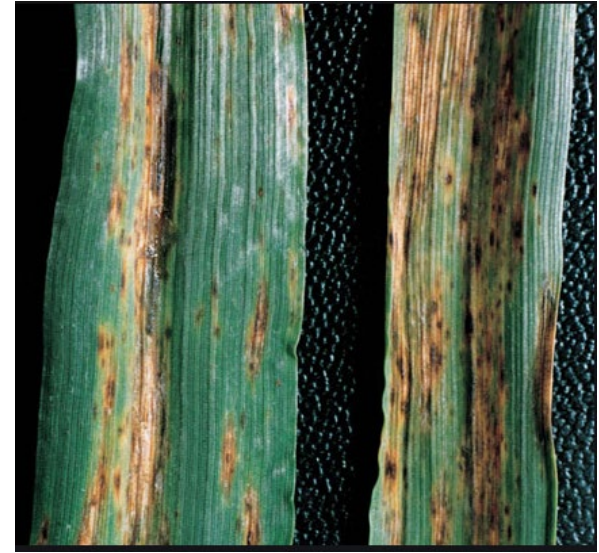
Exploiting knowledge on a generic 'lifestyle' theme

Emergence of a new disease as a result of interspecific virulence gene transfer

Parastagonospora nodorum causes
Septoria nodorum blotch (SNB) on wheat



Pyrenophora tritici-repentis causes
Tan Spot on wheat



ca ~ 1940



~11kb DNA
Tox A gene



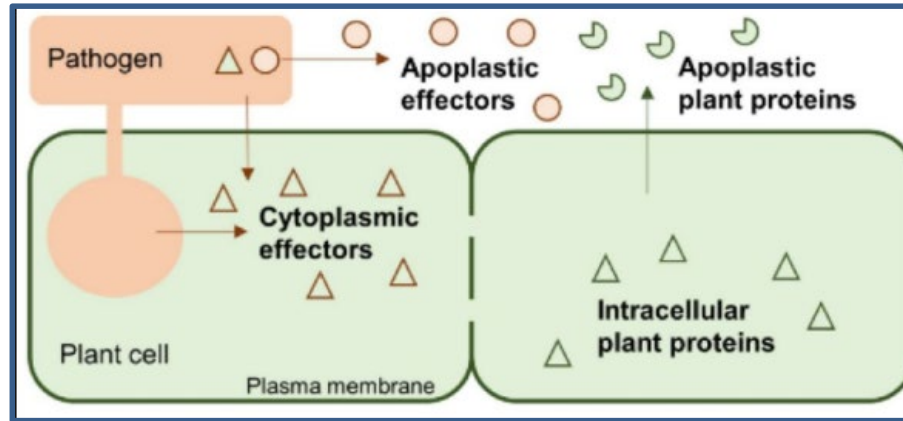
Wheat target is Tsn1 (susceptibility locus)

**Emergence of new diseases as a result of acquisition or modification
of a secondary metabolite cluster**

Tracking pathogen effectors and their 1st host targets

An effector is an entity produced by a pathogenic species to facilitate the infection process

- activate or suppress host defences or other host responses



PHI-base data – Version 4.8

2,261 effector entries (16%) coming from 83 species, mostly plant pathogens

PHI-base started to curate 1st host targets - soon to be displayed in ENSEMBL

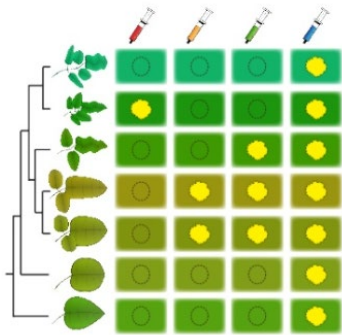
Most effectors are species / strain specific

– maximum discovery by exploring the variome

**But a few are generic and essential for pathogenicity - chitin binding
Ecp6, 3LysM**

PHI-base entries can contribute to plant breeding via effectoromics

Effectoromics, a high-throughput functional genomics approach that uses effectors to probe plant germplasm to detect disease resistance (*R*) genes.



Effectoromics

- accelerates *R* gene identification
- distinguishes functional redundancy
- pinpoints recognition specificities
- assists *R* gene deployment

Application particularly successful

Phytophthora infestans (late blight) – potato / *Solanum tuberosum*
model / *Nicotiana benthamiana*

Experimental techniques - agroinfiltration and agroinfection

Breeders' aim to maintain the widest possible repertoire of disease resistance genes in elite germplasm

Discovery of novel anti-infective targets

Lethals 183 entries in PHI-base (1.3%) - 101 are plant pathogen entries
4 fungi, 3 bacteria

Experimental evidence - no transformants recovered when deleting a single gene

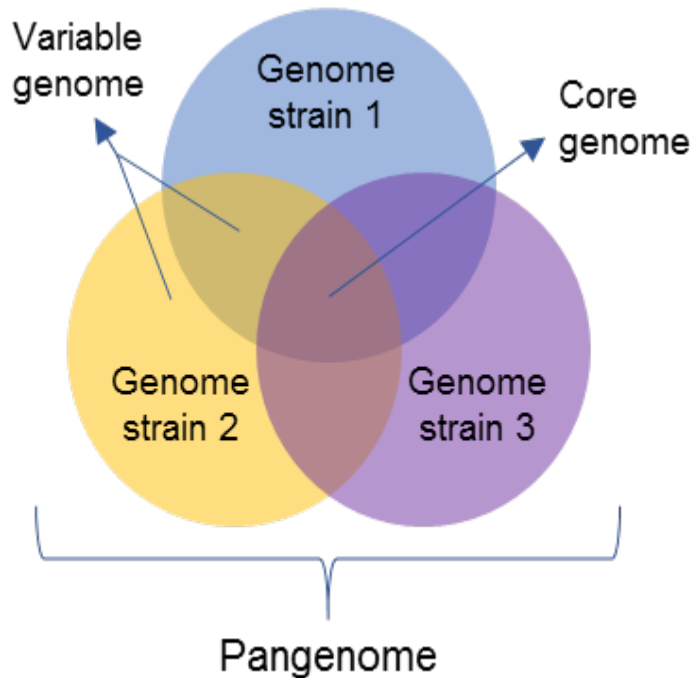
Use this new information about 'essential for life' genes to nominate novel intervention targets to be explored by bioscientists and the AgChem /AgBiotech industries

'Essential for life' proteins to be included in chemical screens – natural products and / or synthetic chemistries

High likelihood that 'essential for life' genes are conserved between species

Pan-genome analyses

Core genomes vis variable genome



Historic isolates compared to **newer problematic isolates**

Previous control practices compared to **existing / new control practices**

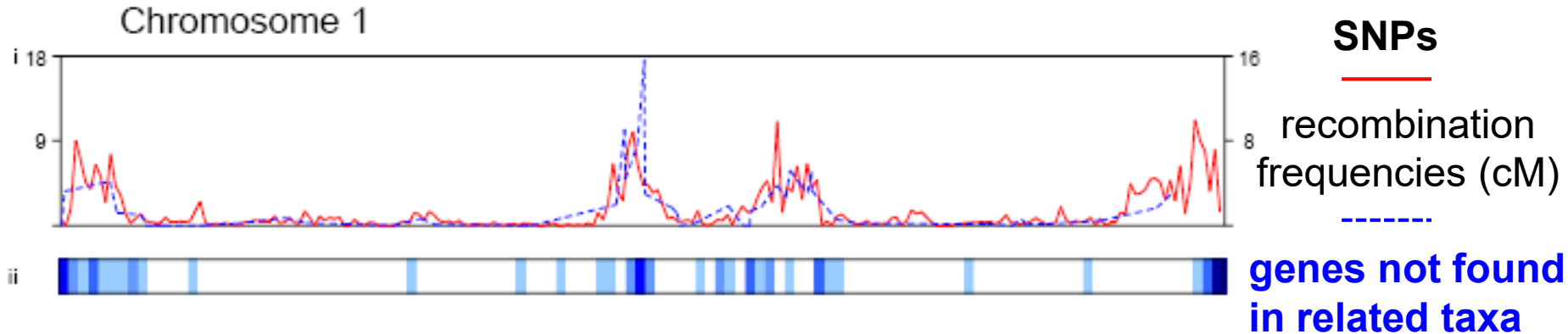
Fungicide use → **Biocontrol options plus *R* gene use**

Pinpoint whole genome evolutionary changes

- Identify which **genome locations** and **gene sets** need to be monitored



Monitoring specific genome locations for changes

Pathogen Species A – minimal repetitive DNA or transposons



Pathogen Species B - rich in repetitive DNA and transposons



-  = blocks of simple sequence repeats and / or transposons
-  = **species specific gene and / or small effector gene**

PROBLEM : Repeat masker software obscures these important sequences

Pathogenicity and virulence networks

Using well explored pathogens to extrapolate knowledge onto the under-explored species

Taxonomic and / or lifestyle relatedness

Cereal infecting

Fusarium graminearum – 1571 interactions

Fusarium verticillioides – 55 interactions

5 other *Fusarium* species – 21 interactions

Non-Cereal infecting

Fusarium oxysporum – 209 interactions

Fusarium solani – 6 interactions

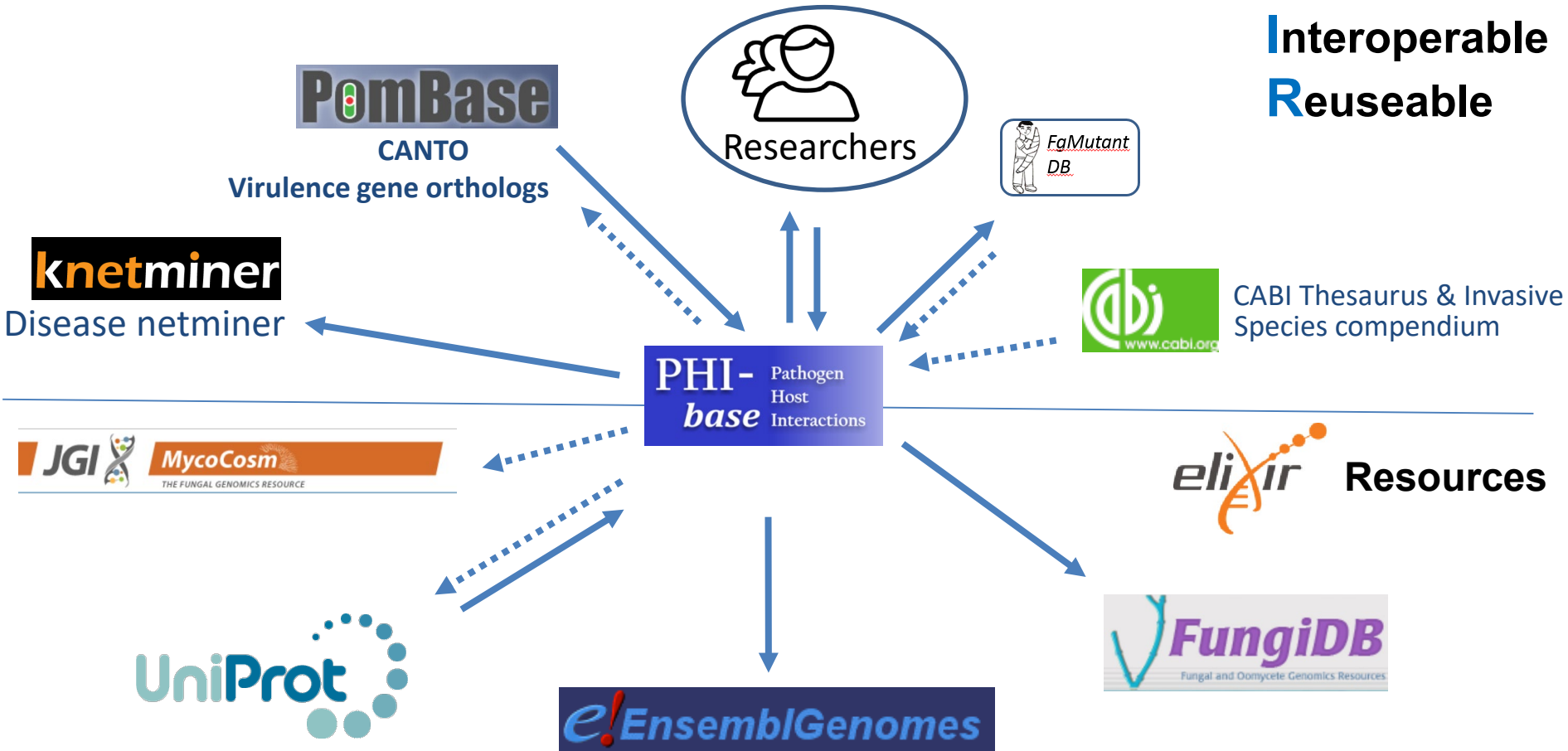
2 other *Fusarium* species – 5 interactions

New types of interspecies comparative analyses that also take advantage of lifestyle similarities or differences

PHI-base Universe - FAIR Data Principles

PHI-base users/associated databases

Findable
Accessible
Interoperable
Reuseable



Acknowledgements

PHI- Pathogen
base Host
Interactions

From genes to mutant
phenotypes

www.phi-base.org



ROTHAMSTED
RESEARCH



DFW



Rothamsted Research

Kim Hammond-Kosack

Martin Urban

Alayne Cuzick

James Seager

Barnaby Norman

Dan Smith

Chris Rawlings

William Brown

Elzbieta Janowska-Sejda – PHI-Nets

Artem Lysenko –PHI-Nets

Keywan Hassani-Pak - Knetminer

 **Canto**

Valerie Wood

Kim Rutherford

Midori Harris

Collaborators

Individual species experts
globally (~50)



Phytopath team @ EBI

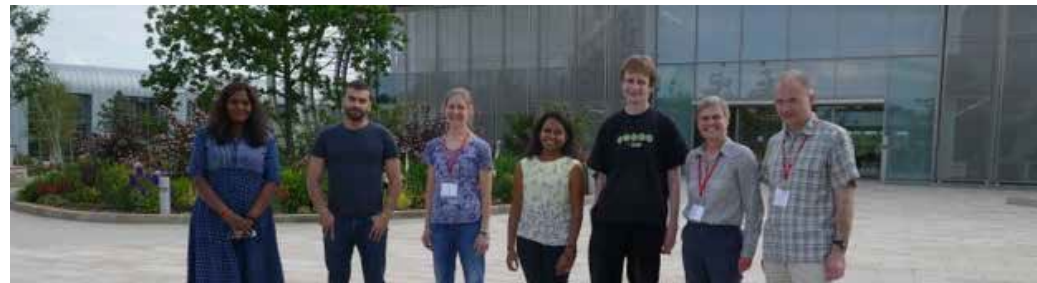
Paul Kersey

Helder Pedro

Uma Mareshwari

Nishadi De Silva

Manuel Carbajo





Thank you