

Pathogen-host interactions phenotype analysis using integrated phenotype and genome resources

Martin Urban



ELIXIR all hands meeting - 9th June 2020



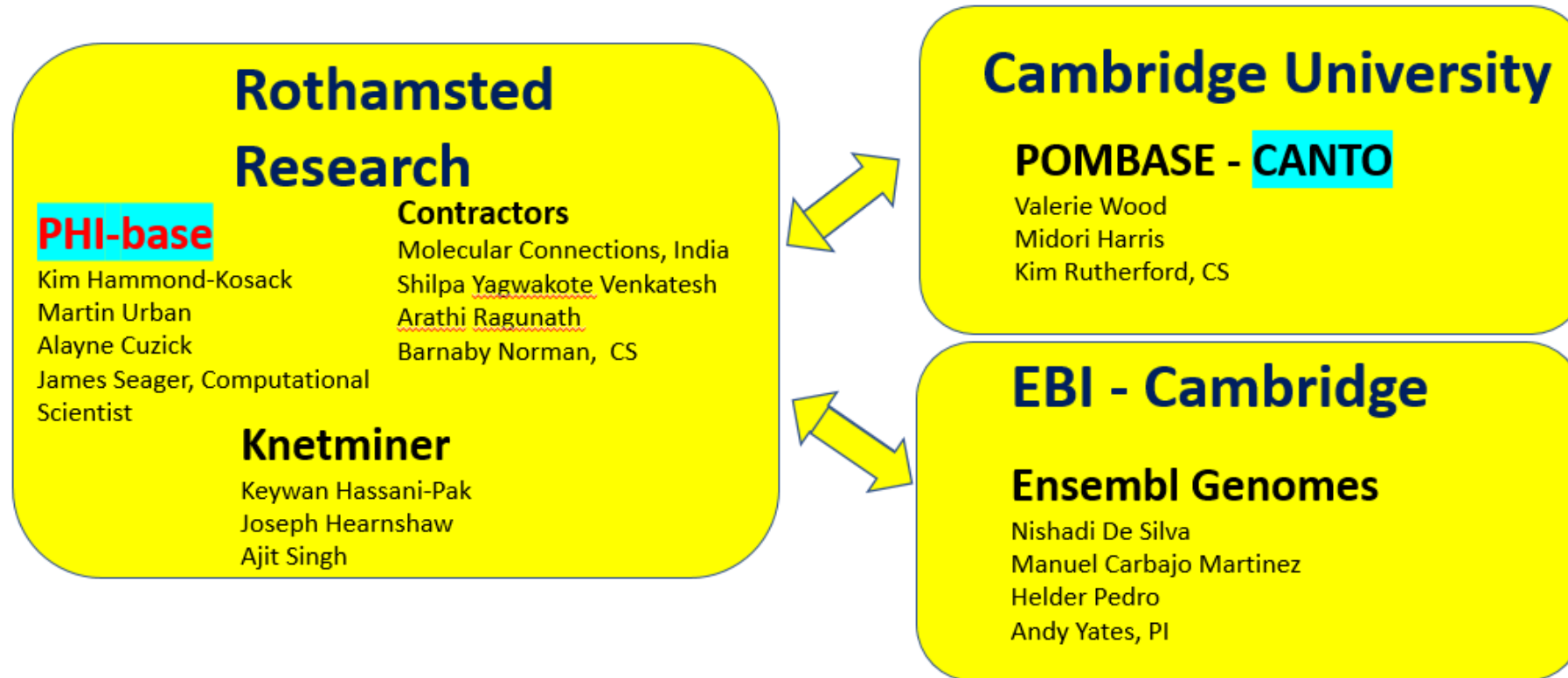
Outline

- Pathogen-Host interactions database
- Database content
- Links to genome resources
- Deep curation of publications

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**
- Part of ELIXIR (European life science project since 2016)
- **Regular users based in 130 countries**

PHI-base interacts closely with Ensembl Genomes and PomBase DB



joint grant by the BBSRC until July 2022

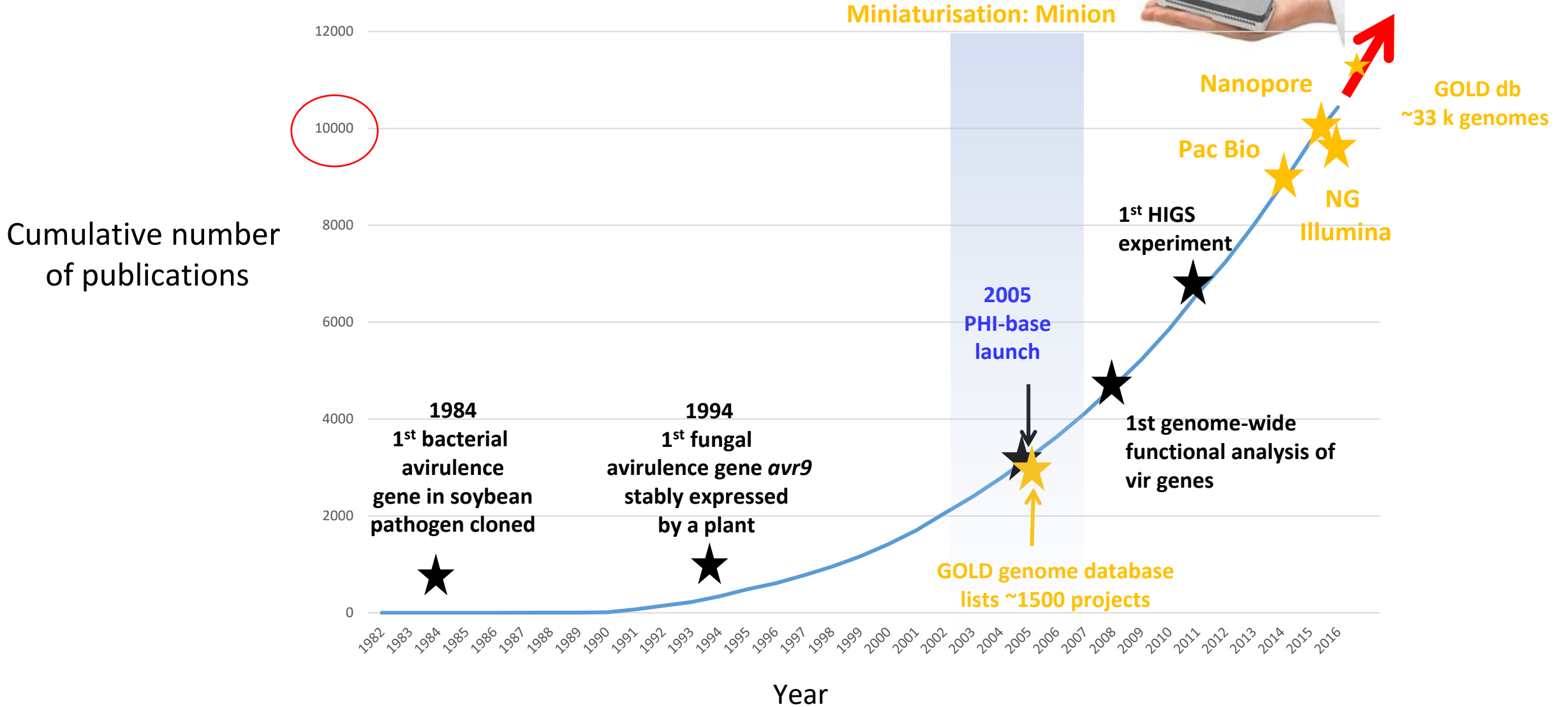
Motivation for Pathogen-Host interactions database

- Major crop disease outbreaks and animal epidemics
- Emerging fungicide resistance



Phenotype 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)



PHI-base is a multi-species database for fungi, protists and bacteria pathogens

65% plant pathogens, 30% human_animal_fish pathogens, 5% insect_others

Total 263 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)

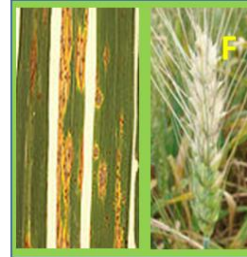
Host Animal Focus

- Key human pathogens
- Animal testing alternatives (3Rs)
- Plant attacking nematodes
- Plant attacking insects

Fusarium graminearum



Magnaporthe oryzae



Ralstonia solanacearum



Salmonella enterica



Xanthomonas oryzae



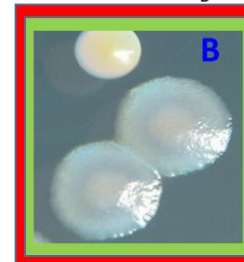
Erwinia amylovora



Candida albicans



Pseudomonas aeruginosa



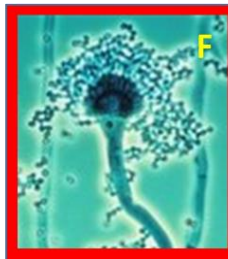
Botrytis cinerea



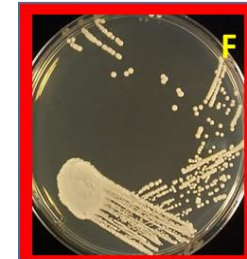
Ustilago maydis



Aspergillus fumigatus



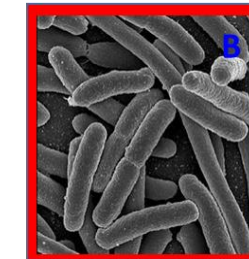
Cryptococcus neoformans



Pseudomonas syringae



Escherichia coli



Staphylococcus aureus



Plant pathogen

Animal pathogen

Plant and animal pathogen

New PHI-base version 4.9 release (May 2020) online!

	www.phibase.org
References	3682
Genes	7126
Interactions	14787
Pathogens	270
Hosts	212
Diseases	530
Chemistries	34 fungicides 149 anti-infectives

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

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. Increased resistance to chemistry
9. Increased sensitivity to chemistry

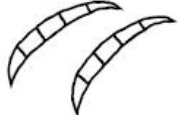





35 fungicides

Terms are routinely published in research articles but mapping to GO terms is not supported. Instead **PHIPO** for pathogen-host interactions is developed.

Interactions are an important concept in PHI-base

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>F. graminearum</i>		1	Reduced virulence
		2	Reduced virulence
		3	Unaffected pathogenicity

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

From mutant genes to phenotypes! The mission of PHI-base is to provide expertly curated molecular and biological information on genes proven to affect the outcome of pathogen-host interactions. Information is also given on the target sites of some anti-infective chemistries.

Pathogens

Hosts

Diseases

Ontologies

Taxonomies

Literature citations

Community resources

External resources

PhytoPath

Wiki page

PHI-base publications

Developer resources (API)

PHI-base data

- [Search PHI-base](#)
- [Download PHI-base data](#)
- [View PHI-base statistics](#)
- [Submit your published phenotype data](#)

Getting Started

- [Text search](#)
Search terms are auto completed when entries are found.
- [PHIB-BLAST](#)
Find homologs and their phenotypes
- [Submit data](#)
Manual curation of the pathogen interaction literature
- [How to cite us:](#)
Urban, M. et al (2019). NAR doi: 10.1093/nar/gkz904

Tweets by @PHI_base



Martin Urban

@phi_base

PHI-base created protein-protein interaction networks for 15 Ascomycete fungal species including rice blast, Fusarium and Botrytis. In a use case study virulence interacting proteins of siRNA targets were identified. Read more at doi.org/10.3389/fmicb....

Version: 4.8 Last revision: 16/09/19

PHI-base interface with faceted views to allow complete visualization of all curated data!

- (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 Pathogen Host Interactions
From genes to mutant phenotypes

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

1 Free text entry with autocomplete

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

Pathogen Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species
<input type="checkbox"/> TRI5	unaffected pathogenicity	Fusarium graminearum	Fusarium Ear Blight	Arabidopsis thaliana (related: thale cress)
<input type="checkbox"/> TRI5	reduced virulence	Fusarium graminearum	Fusarium Ear Blight	Triticum aestivum (related: bread wheat)

4

Gene

TRI5(2)

Pathogen species

Fusarium graminearum(2)

Host species

Arabidopsis thaliana (related: thale cress)(1)
 Triticum aestivum (related: bread wheat)(1)

Pathogen Gene Allele

Gene: TRI5
PHI-base entry: PHI:3518
Gene ID: AAM90953
Protein ID: [Q00909](#)
Sequence strain: A18
Gene function: Trichodiene synthase
Database: GO

Pathogen

Pathogen species: Fusarium graminearum
Pathogen ID: [5518](#)
Pathogen strain: PH-1
Pathogen strain ID: [229533](#)

Host

Host species: Triticum aestivum (related: bread wheat)
Host classification: Monocots
Host ID: [4565](#)
Host strain: Bobwhite
Tissue: spike

Reference

Pmid: [18179606](#)
Ref source: PubMed
Year: 2008
DOI: [10.1111/j.1469-8137.2007.02333.x](#)
Author reference: Cuzick et al.

PHI Phenotype

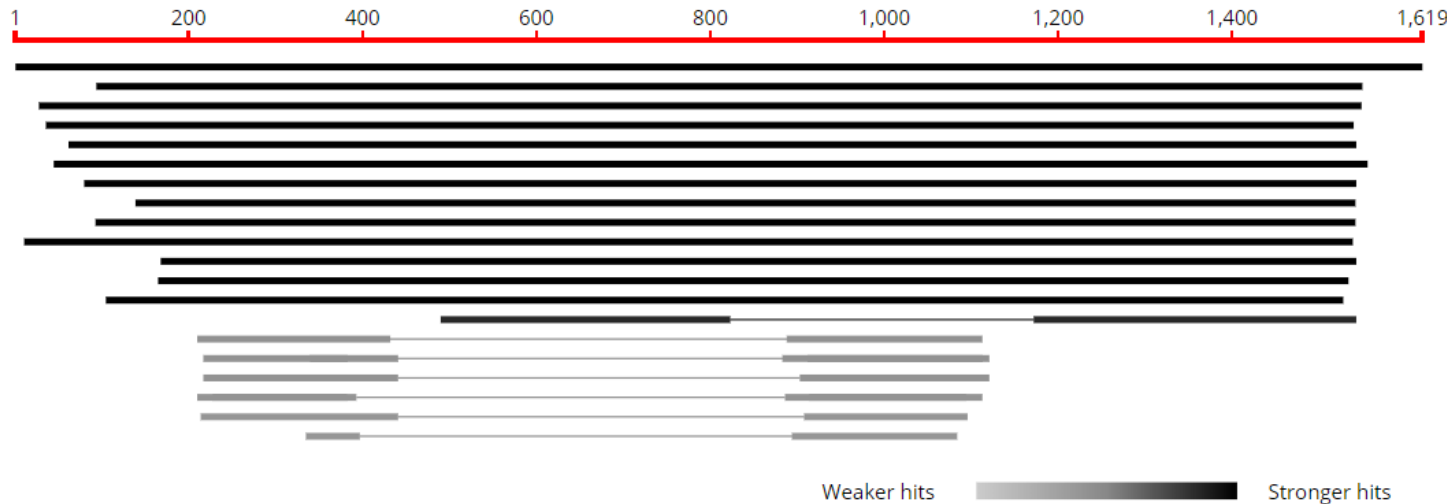
Phenotype: reduced virulence
Disease name: Fusarium Ear Blight
Tissue: spike
Experimental technique: Gene deletion; biochemical analysis

Protein-to-Phenotype BLAST search functionality



Typical BLAST result output provides score list

Query= tr|O13407|O13407_MAGGR ABC1 transporter OS=Magnaporthe grisea OX=148305 1/1
GN=ABC1 PE=3 SV=1



- 5. E9RBG1#PHI:4230#Abcb#746128#Aspergillus_fumigatus#reduced_virulence
- 6. Q96VL9#PHI:543#Bcatrd#40559#Botrytis_cinerea#unaffected_pathogenicity
- 7. Q4X006#PHI:4229#Abca#746128#Aspergillus_fumigatus#unaffected_pathogenicity
- 8. A5H456#PHI:867__PHI:1159#Mgatr7#1047171#Zymoseptoria_tritici
#unaffected_pathogenicity__chemistry_target:_sensitivity_to_chemical

Header includes PHI-base ID and high level phenotype

	Total score	E value	Length
5.	1811.19	0.00	1497
6.	1722.98	0.00	1501
7.	1701.41	0.00	1499
8.	1478.77	0.00	1811

Main uses of PHI-base

>330 publications citing PHI-base

All papers are cited in the about section of the database

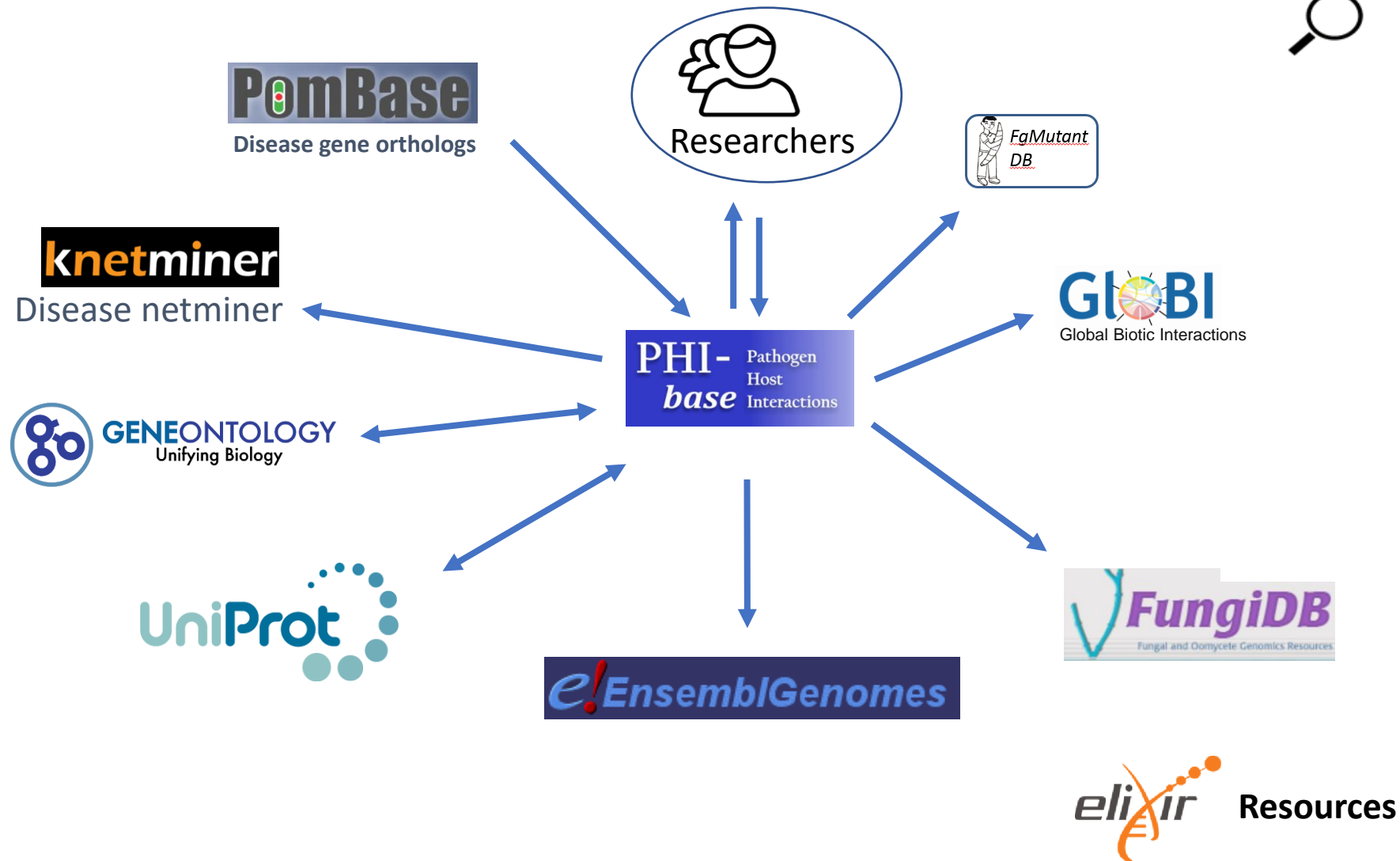
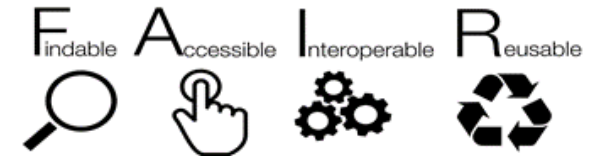
- Quick lookup of gene mutant - phenotype relationships
human readable knowledge database and computer readable!
- Annotate novel datasets for candidate gene selection:
Newly sequenced pathogen genomes or exploring variomes (variant genomes)
RNA-seq and microarray data sets (enrichment of virulence genes)
Other candidate gene lists (i.e. from forward genetics screens)
- Comparative genomics/phenomics
Conserved themes vs species-specific differences
In silico predictions of candidate virulence using protein-protein interaction networks
- Provide phenotype annotations for databases and genome browsers at EMBL-EBI, FungiDB and knowledge networks

 e!EnsemblGenomes

 FungiDB
Fungal and Oomycete Genomics Resources
Release 39
30 Aug 2018

 UniProt

PHI-base – part of a connected network of FAIR data resources

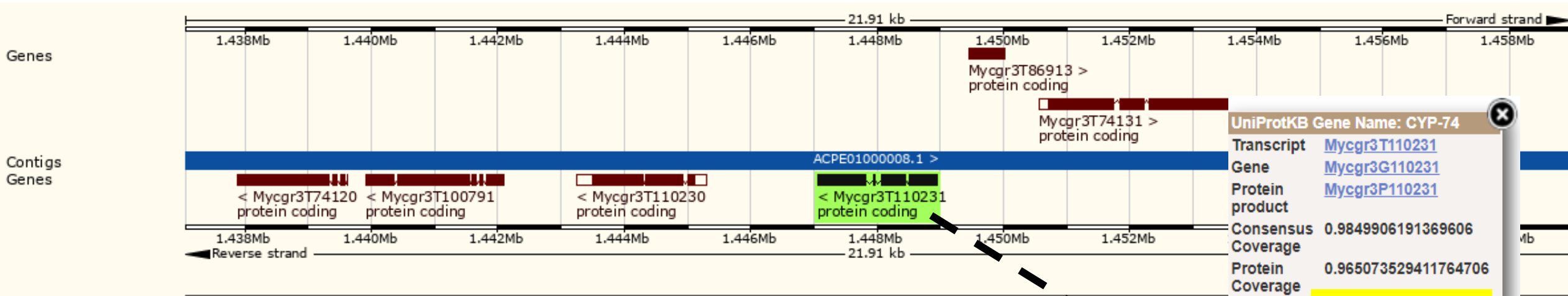


Display phenotypes directly in genome browsers since 2011

Ensembl Fungi



Zymoseptoria tritici (MG2)



UniProtKB Gene Name:	CYP-74
Transcript	Mycgr3T110231
Gene	Mycgr3G110231
Protein product	Mycgr3P110231
Consensus Coverage	0.9849906191369606
Protein Coverage	0.965073529411764706
PHI-base mutant	increased resistance to chemical
PHI	PHI:1149
Gene GC	54.43
Location	Chromosome 7: 1,447,054-1,448,960
Gene type	Protein coding
Transcript type	Protein coding
Strand	Reverse
Base pairs	1,635
Amino acids	544
Analysis	JGI
Annotation method	Protein coding genes annotated by the JGI

UniProtKB provides links to PHI-base phenotypes

UniProt

[BLAST](#) [Align](#) [Retrieve/ID mapping](#) [Peptide search](#) [SPARQL](#) [Help](#) [Contact](#)

UniProtKB - F9XMT4 (F9XMT4_ZYMTI)

Display


[BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#) [Help video](#) [Add a publication](#) [Feedback](#)

-
- [Publications](#)
- [Feature viewer](#)
- [Feature table](#)

Protein Submitted name: **Uncharacterized protein**

Gene **MYCGRDRAFT_111221**

Organism *Zymoseptoria tritici* (strain CBS 115943 / IPO323) (Speckled leaf blotch fungus) (*Septoria tritici*)

Status  Unreviewed - Annotation score: ●○○○○○ - Protein predictedⁱ

- [None](#)
- Function
 - Names & Taxonomy

Pathology & Biotechⁱ

Miscellaneous databases

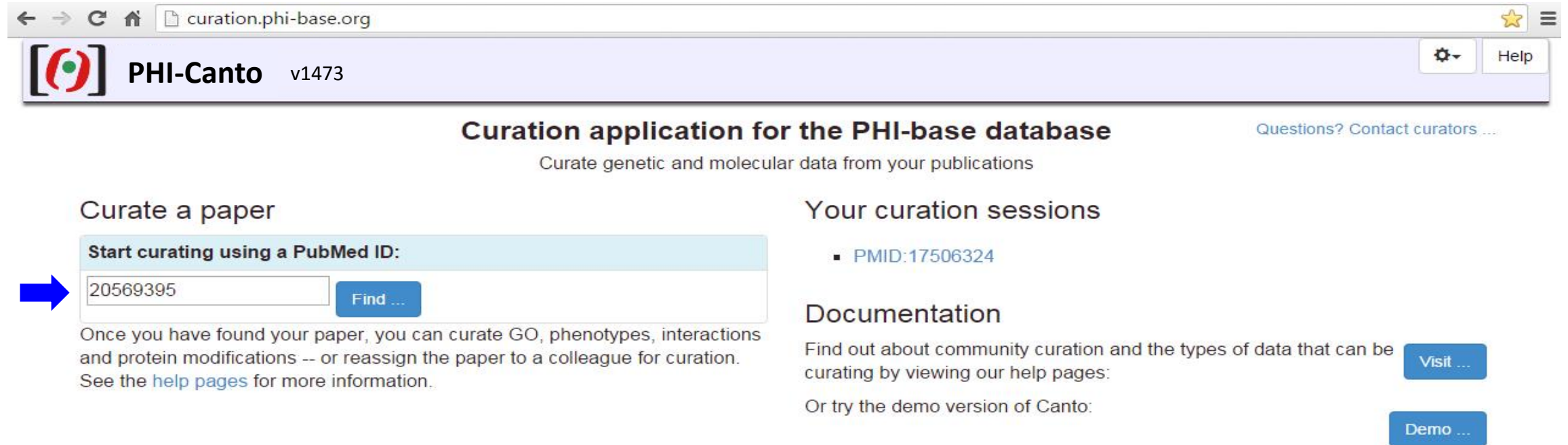
PHI-base ⁱ	PHI:6494
-----------------------	----------

Switching to **Deep** curation of publications

- complex phenotypes using annotation extensions
- meaningful analyses and comparisons within and between organisms
- using well-established [author-curation tool](#) used by PomBase.org (fission yeast)



PHI-Canto is an author curation tool for phenotypes of pathogen-host interactions, anti-infectives and hosts



The screenshot shows the PHI-Canto web application interface. The browser address bar displays 'curation.phi-base.org'. The page header includes the PHI-Canto logo (a red 'C' in a circle) and the text 'PHI-Canto v1473'. A navigation bar contains a settings gear icon and a 'Help' button. The main content area is titled 'Curation application for the PHI-base database' with the subtitle 'Curate genetic and molecular data from your publications'. On the right, there is a link for 'Questions? Contact curators ...'. The interface is divided into two columns. The left column, 'Curate a paper', features a light blue box with the heading 'Start curating using a PubMed ID:'. Below this is a text input field containing '20569395' and a blue 'Find ...' button. A blue arrow points to the input field. Below the input field, text explains that users can curate GO, phenotypes, interactions, and protein modifications, or reassign the paper to a colleague. The right column, 'Your curation sessions', lists a session for 'PMID:17506324'. Below this is a 'Documentation' section with text about community curation and help pages, accompanied by a 'Visit ...' button, and a link to a demo version with a 'Demo ...' button.

**Start with Manuscript ID,
enter details including protein IDs, phenotypes,
GO annotation and submit**

Authors will be invited to curate their articles after manuscript acceptance.

PHIPO: Pathogen Host interaction Phenotype Ontology on OBO Foundry

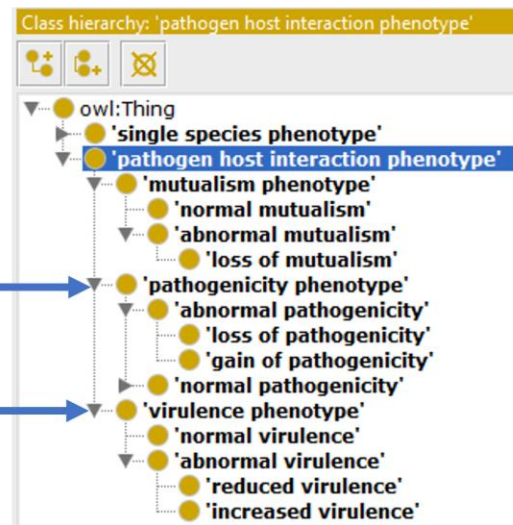
Phenotypes

Migrating the PHI-base controlled vocabulary phenotypes to a formal ontology for use in PHI-Canto and other resources

The nine PHI-base high level phenotypes

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

PHIPO



+ use of 7 external ontologies including GO, PSI-MOD

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

Terms use formal definitions referencing external ontologies such as PATO, CHEBI and GO



www.phi-base.org

PHI-base
Pathogen
Host
Interactions

PHI-base phase 3 is funded by the BBSRC until July 2022



Acknowledgements



Rothamsted Research

- Kim Hammond-Kosack
- Martin Urban
- Alayne Cuzick
- James Seager
- Barnaby Norman
- Dan Smith
- Chris Rawlings
- William Brown
- Keywan Hassani-Pak

Collaborators

Individual species experts globally (~50)

Phytopath team @ EBI

- | | |
|----------------|------------------|
| Paul Kersey | Nishadi De Silva |
| Helder Pedro | Manuel Carbajo |
| Uma Mareshwari | |



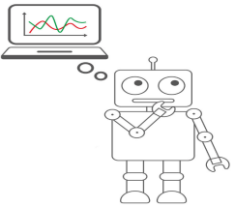
- Valerie Wood
- Kim Rutherford
- Midori Harris



Summary & Questions?

- **PHI-base - multi-species microbial mutant database with wide range of crop plants**
- **Main uses for researchers:**
 - access & download published phenotypes
 - provides annotation to genomes
 - enables searches across species and taxonomies, eg by BLAST-to-phenotype analyses
- **UniProtKB, FungiDB, and ENSEMBL integrate PHI annotation into their databases**
- **Community author-curation tool, PHI-Canto:**
 - allows deep curation of phenotypes ; authors curate themselves

Multi-pathogen/hosts



PHI- Pathogen
base Host
Interactions

From genes to mutant phenotypes

www.phi-base.org

Thank you

martin.urban@rothamsted.ac.uk

PHI-base at Rothamsted Research, UK, near London

PomBase Curation progress

PomBase staff focus their curation efforts on papers published before 2014, community curation is an indispensable route to database inclusion and subsequent dissemination for new publications.

