The **P**athogen-**H**ost **I**nteractions database: PHI-base

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

Martin Urban and Alayne Cuzick

Kim Hammond-Kosack

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**

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 and animal epidemics
- Emerging fungicide resistance

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 ~ 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 — F
Magnaporthe oryzae — F
Ralstonia solanacearum — B
Ustilago maydis — F
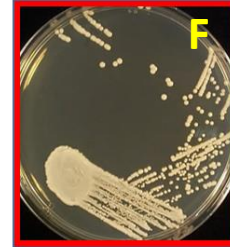*Xanthomonas oryzae* — B
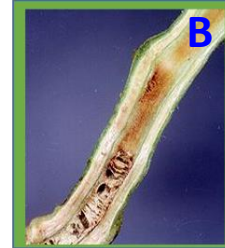
Candida albicans — F
Botrytis cinerea — F
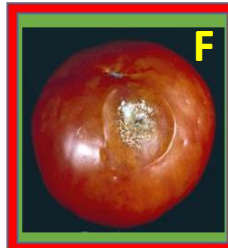*Salmonella enterica* — B
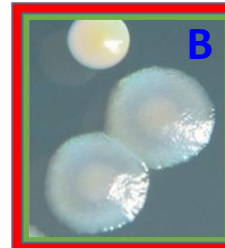*Cryptococcus neoformans* — F
Pseudomonas syringae — B

Aspergillus fumigatus — F
Fusarium oxysporum — F
P. aeruginosa — B
Hyaloperonospora arabidopsidis — P
Staphylococcus aureus — B

Plant pathogen | Animal pathogen | Plant and animal pathogen

Current PHI-base version 4.6 release

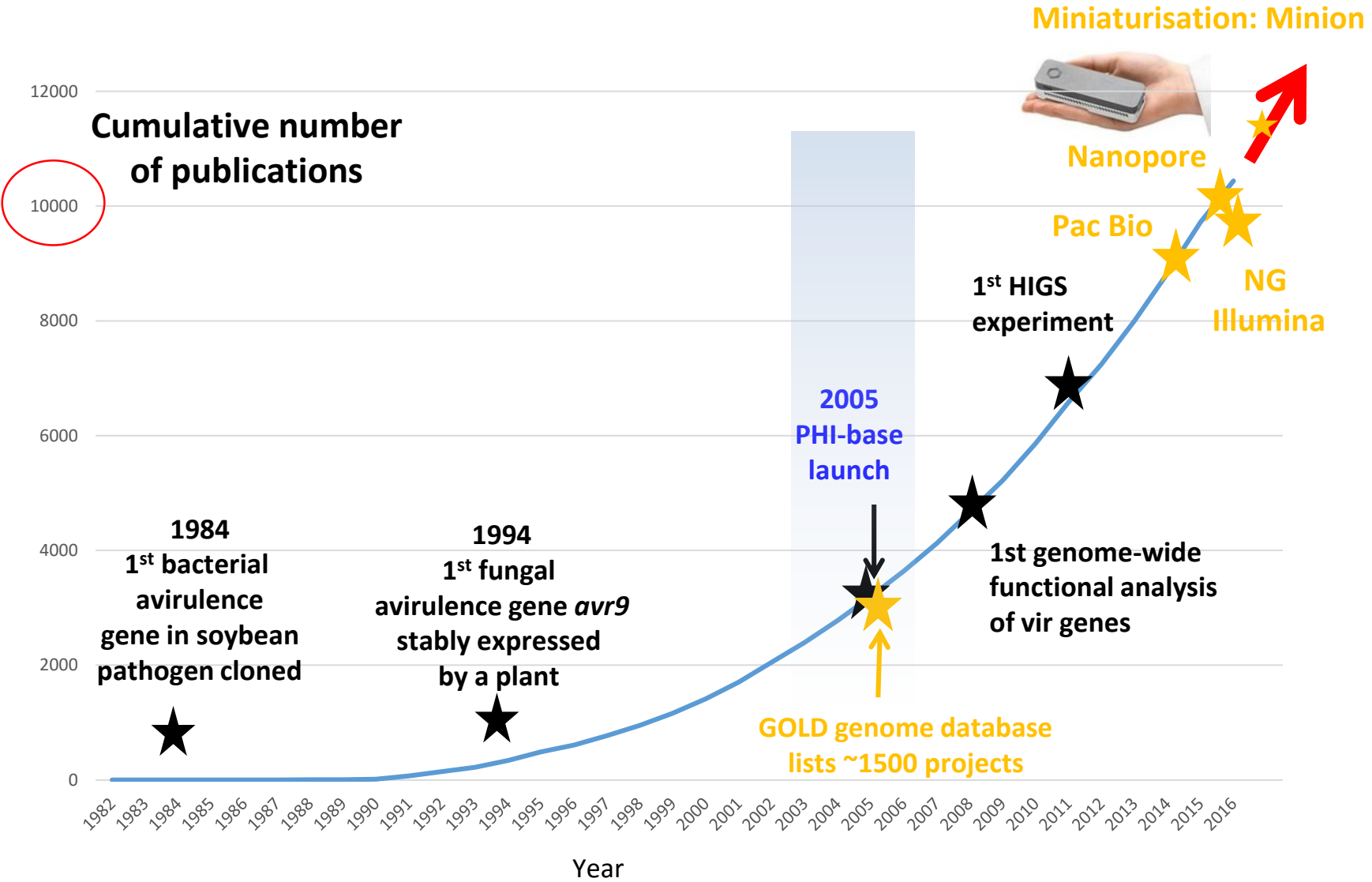|	www.phibase.org
References	**3011**
Genes	**6438**
Interactions	**10382**
Pathogens	263
Hosts	194
Diseases	510
Chemistries	34 fungicides 149 anti-infectives

- Pathogens: 75% eukaryotes (fungi_protists), 25% bacteria
- Hosts: 65% plant (50:50 cereal:non-cereal), 30% animal_human

Entries for Key Crop and Model Plant Hosts

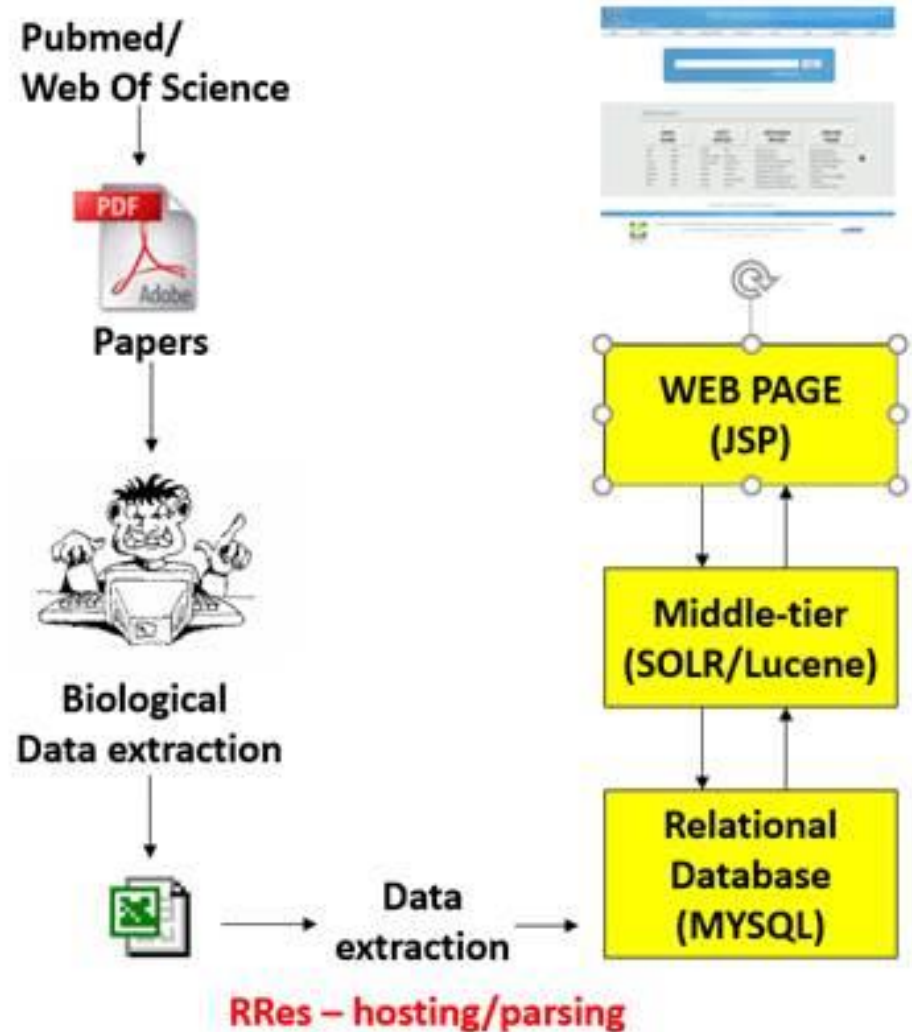| Host plant	Entries / species	Loss of Pathogenicity	Reduced Virulence	Effector	TOTALS
Wheat	1350	65 (5)	382 (13)	6 (3)	453 - 34%
Barley	436	85 (4)	144 (7)	32 (3)	261 - 60%
Potato	128	1 (1)	54 (11)	8 (2)	63 - 49%
Tomato	692	45 (9)	176 (17)	193 (17)	414 - 60%
Brassica	108	16 (6)	53 (10)	15 (3)	84 - 77%
Arabidopsis	347	7 (5)	94 (17)	190 (15)	291 - 84%
Tobacco (benth)	168	2(1)	30 (8)	124 (21)	156 - 93%
TOTALS (7 species)	3229 - 31%	221 - 26%	933 - 17%	568 -24%	1722 - 26%
TOTALS -PHI-base	10382	849	5488	2391	6568

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)

Cumulative number of publications chart (Year vs. Cumulative number of publications), with annotations:

- **Miniaturisation: Minion**
- **Nanopore**
- **Pac Bio**
- **NG Illumina**
- **1st HIGS experiment**
- **2005 PHI-base launch**
- **1984 1st bacterial avirulence gene in soybean pathogen cloned**
- **1994 1st fungal avirulence gene *avr9* stably expressed by a plant**
- **1st genome-wide functional analysis of vir genes**
- **GOLD genome database lists ~1500 projects**

Y-axis: 0, 2000, 4000, 6000, 8000, 10000, 12000

X-axis (Year): 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016

Database curation workflow
at RRes and Molecular Connections (MC) India

- **MC: keyword searches for new papers to generate candidate list**

- **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

Pubmed/ Web Of Science

Papers

Biological Data extraction

Data extraction

WEB PAGE (JSP)

Middle-tier (SOLR/Lucene)

Relational Database (MYSQL)

RRes – hosting/parsing

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

pmk1 | ① Free text entry with autocomplete | SEARCH | RESET | ADVANCED SEARCH

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

RESET | REFINE SEARCH

Year
☑ 2010(11)

Gene
☐ CBP1(2) ④
☐ CPKA(2)
☐ EXP5(2)
☐ Lf pmk1(1)
☐ MST12(2)

Disease
☐ Dry bubble disease(1)
☐ Rice blast(10)

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

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

Ref: http://www.sequenceserver.com/

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/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)

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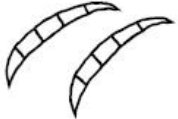
Generic (species neutral) terms to permit wide inter-species comparisons

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

FRAC
FUNGICIDE RESISTANCE
ACTION COMMITTEE

ChEMBL

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
Wheat ear	1	Reduced virulence
Wheat seedling	2	Reduced virulence
Maize cob	3	Unaffected pathogenicity

Tri5 mutant

Pathogen species:
Fusarium graminearum

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

Summary of phenotypic outcome content in PHI-base

PHENOTYPIC OUTCOME	%
Loss of pathogenicity	8.6%
Reduced virulence	39.4%
Increased virulence	4.0%
Unaffected pathogenicity	24.9%
Effector (transferred 'entity')	20.6%
Enhanced antagonism	0.1%
Essential	1.9%
Resistance to chemistry	0.4%
Sensitive to chemistry	0.1%

Integration of PHI-base data with Ensembl Genomes

Annotation:
phenotype and
experimental information

pathogen genomes

since 2010

Ensembl fungi: 83 species, 111 genomes

Ensembl protists: 16 species, 24 genomes

Ensembl bacteria: 16 species, 119 genomes

115 species, 254 genomes
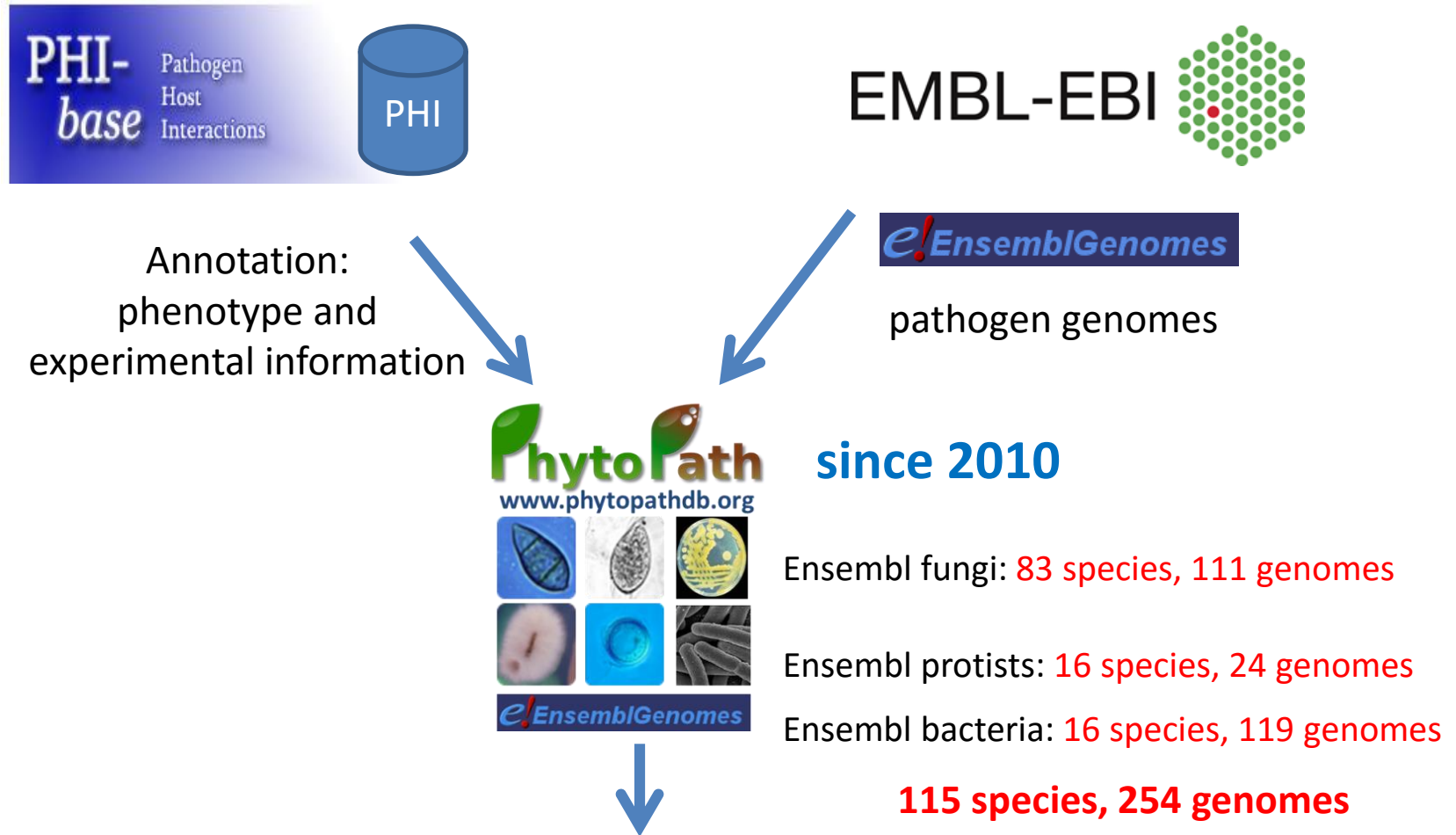
❖ Phenotype information directly accessible in *Ensembl* Genomes
❖ BioMART search of PHI-base annotated pathogen genes

Display phenotypes directly in genome browser

http://fungi.ensembl.org
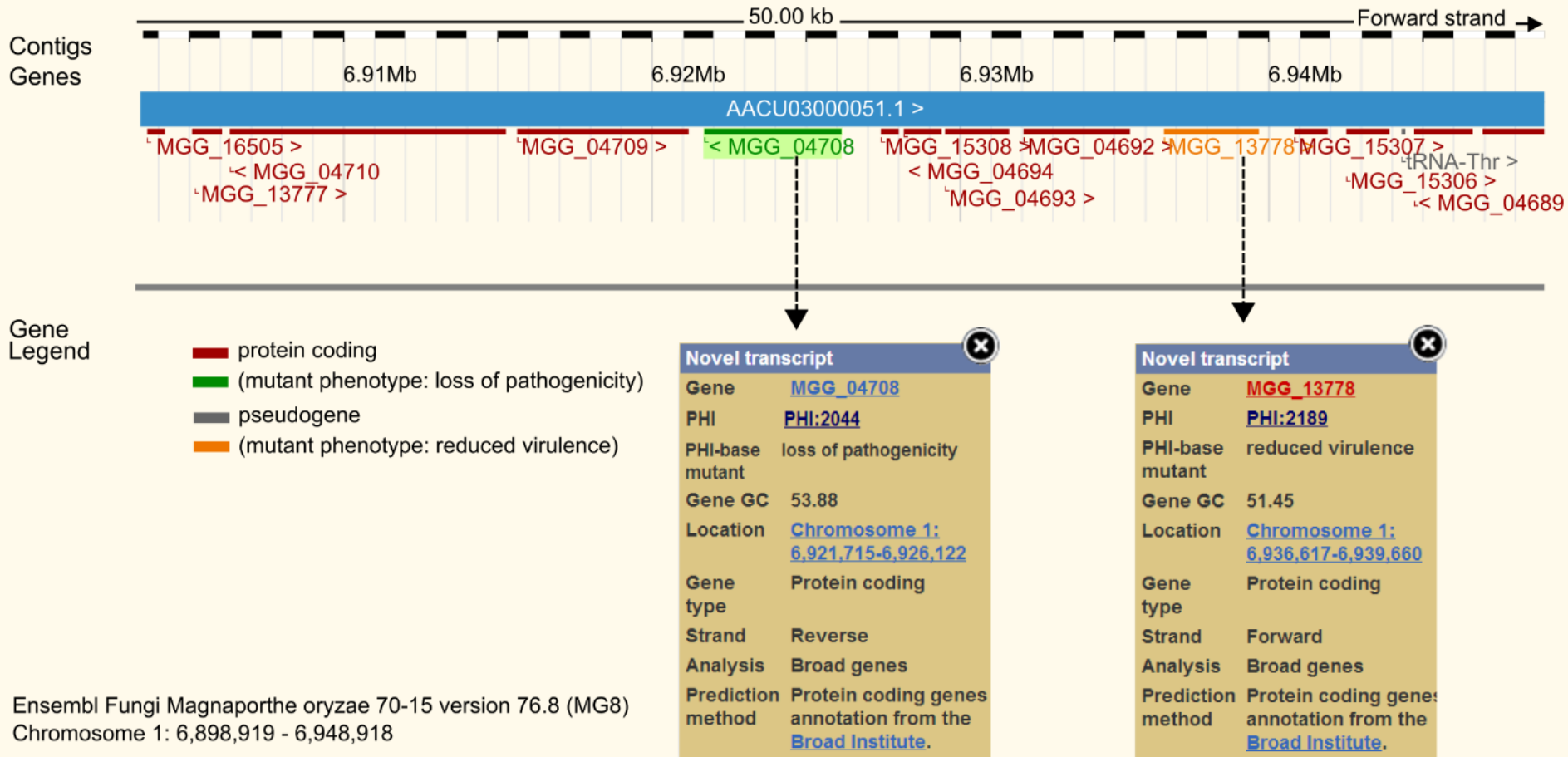
Search *Fusarium graminearum* genomes by phenotype

www.phytopathdb.org

Two NEW functionality for PHI-base since 2017

Curation of the 1st HOST TARGETS in plants of pathogen effectors
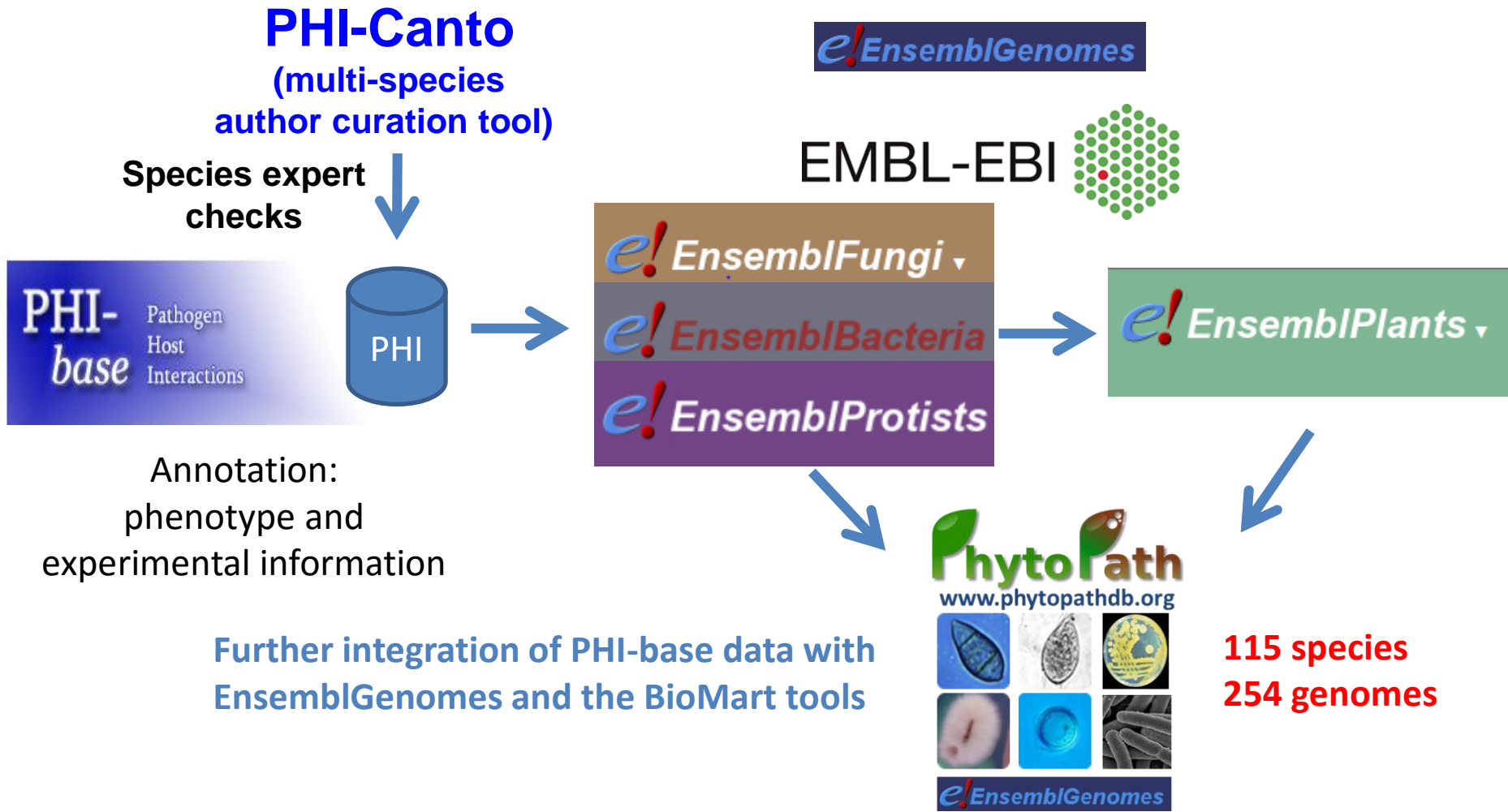
Community multi-species curation tool for authors PHI-Canto

http:// demo-canto.phi-base.org (beta version)

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

From genes to mutant phenotypes

Summary

Multi-pathogen

- PHI-base is a multi-species microbial mutant database with wide range of crop and non-crop plants and **an increasing number of 1st host targets**

- **PHI-base is a FAIR database** and part of the Elixir 'Data for Life' project since 2016

- Use for the researcher:
  access to published phenotypes and genotypes on pathogens and hosts
  provides instant phenotype annotation to genomes, transcriptomes
  enables searches across any species and any taxonomies

- Computational applications:
  use of controlled vocabularies (PHI-PO ontologies) allows computers to analyse a model phenotype data sets

- Community annotation tool:
  - will allow authors to curate their own data
  - advantage: high visibility of own research to increase citation number

Canto v1251
Community curation
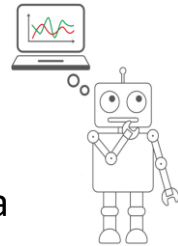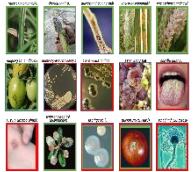
Acknowledgements

www.phi-base.org

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)
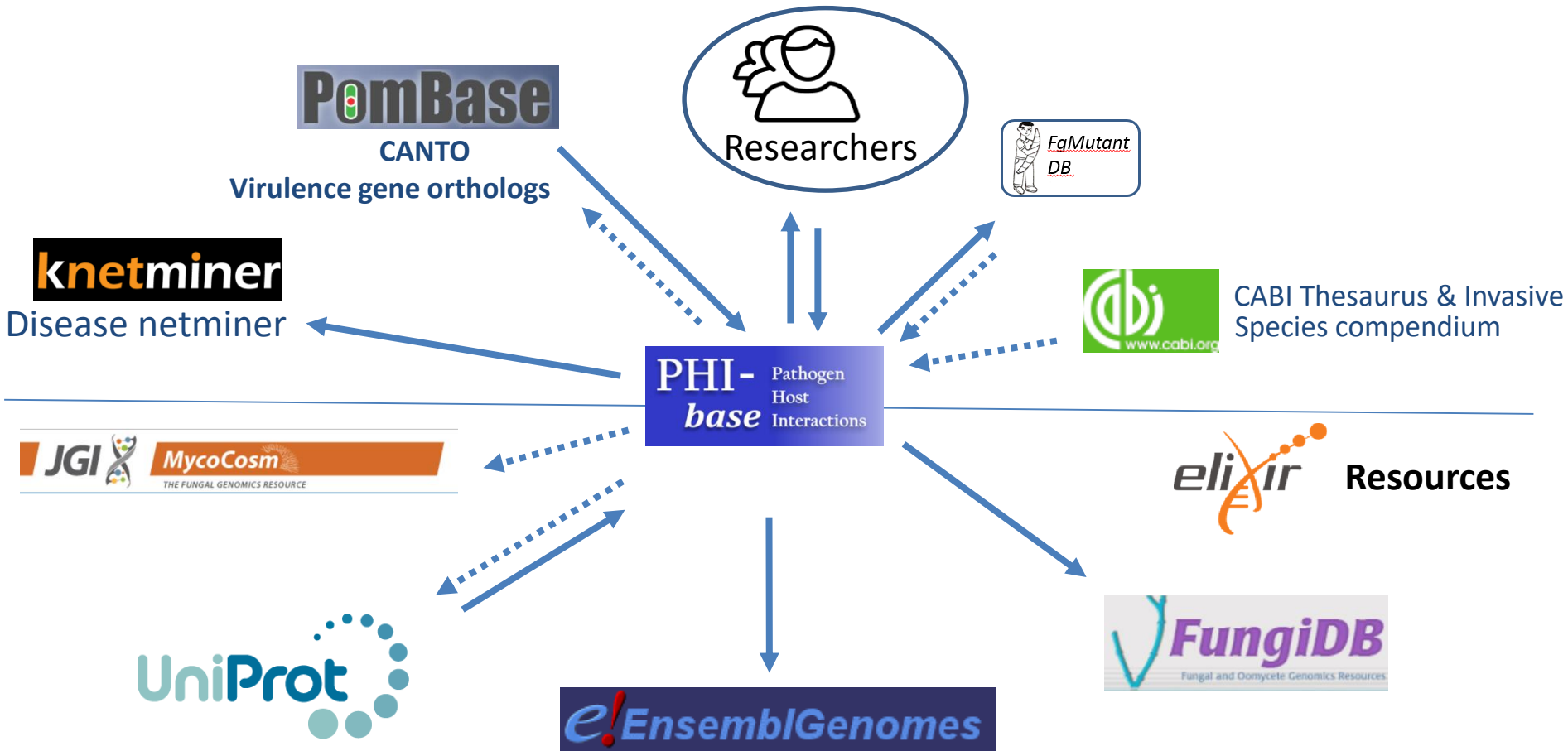
Phytopath team @ EBI

Paul Kersey   Nishadi De Silva
Helder Pedro   Manuel Carbajo
Uma Mareshwari

Valerie Wood
Kim Rutherford
Midori Harris

Thank you

Extra Slides

PHI-base Universe

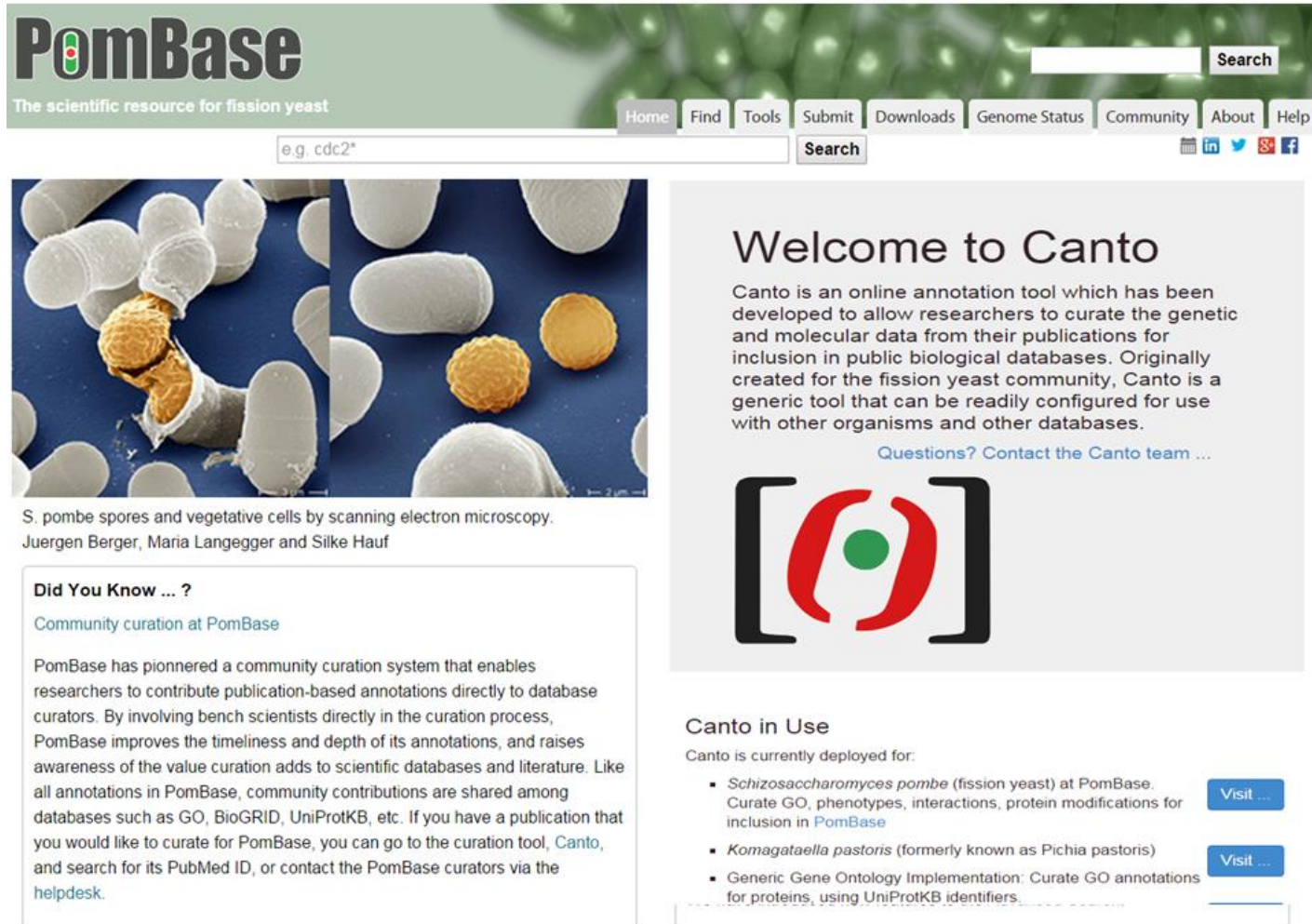PHI-base users/associated databases

How **FgMutantDb** Shares Information with Other Databases
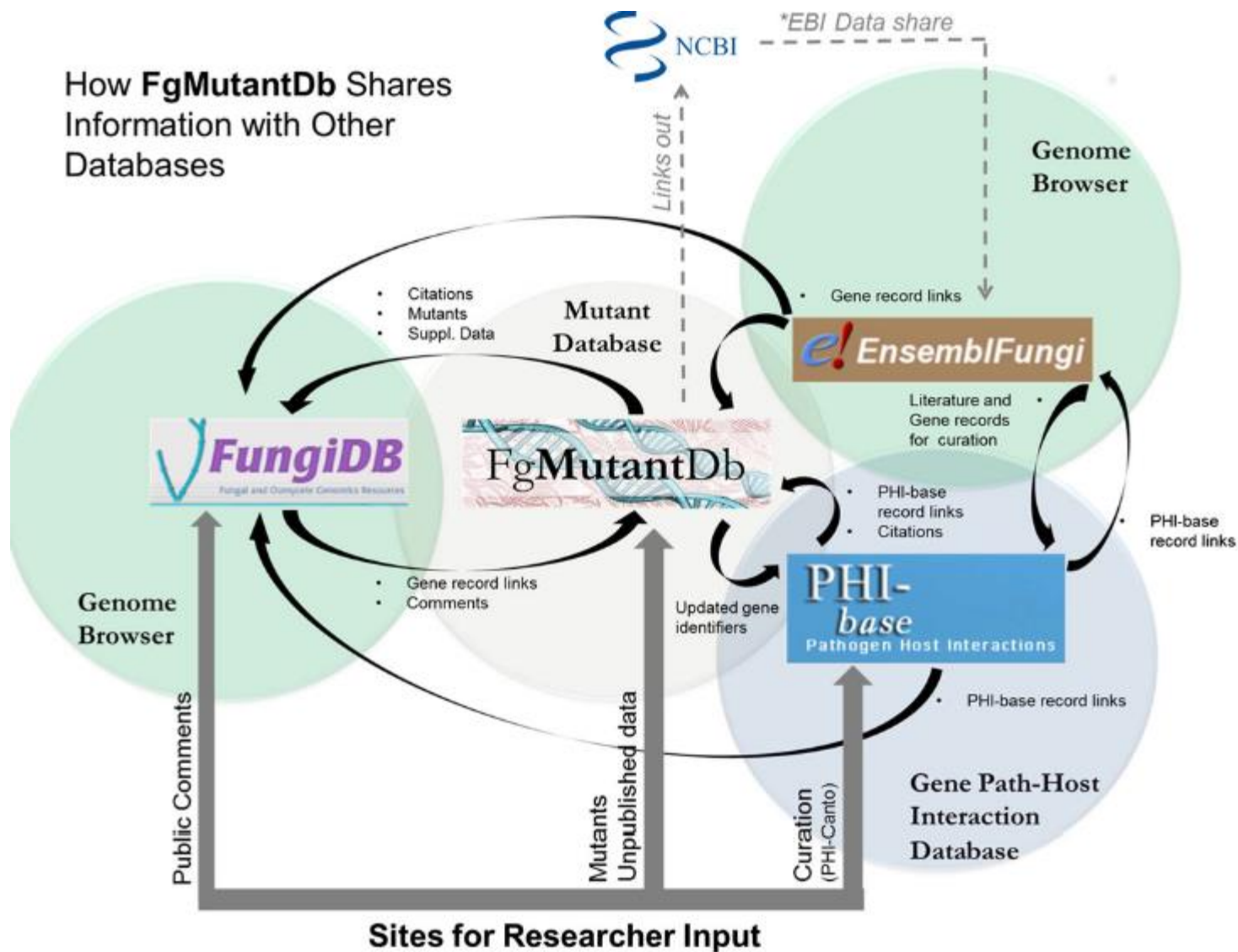
*New entries into NCBI, Genbank, or another public repository are shared by EBI every 3 months. All new publically available data are released in the next version of Ensembl. Authors can edit Genbank entries to be available in the next Ensembl release.

Developing a community curation tool for authors

Multi-species chemogenomics approach
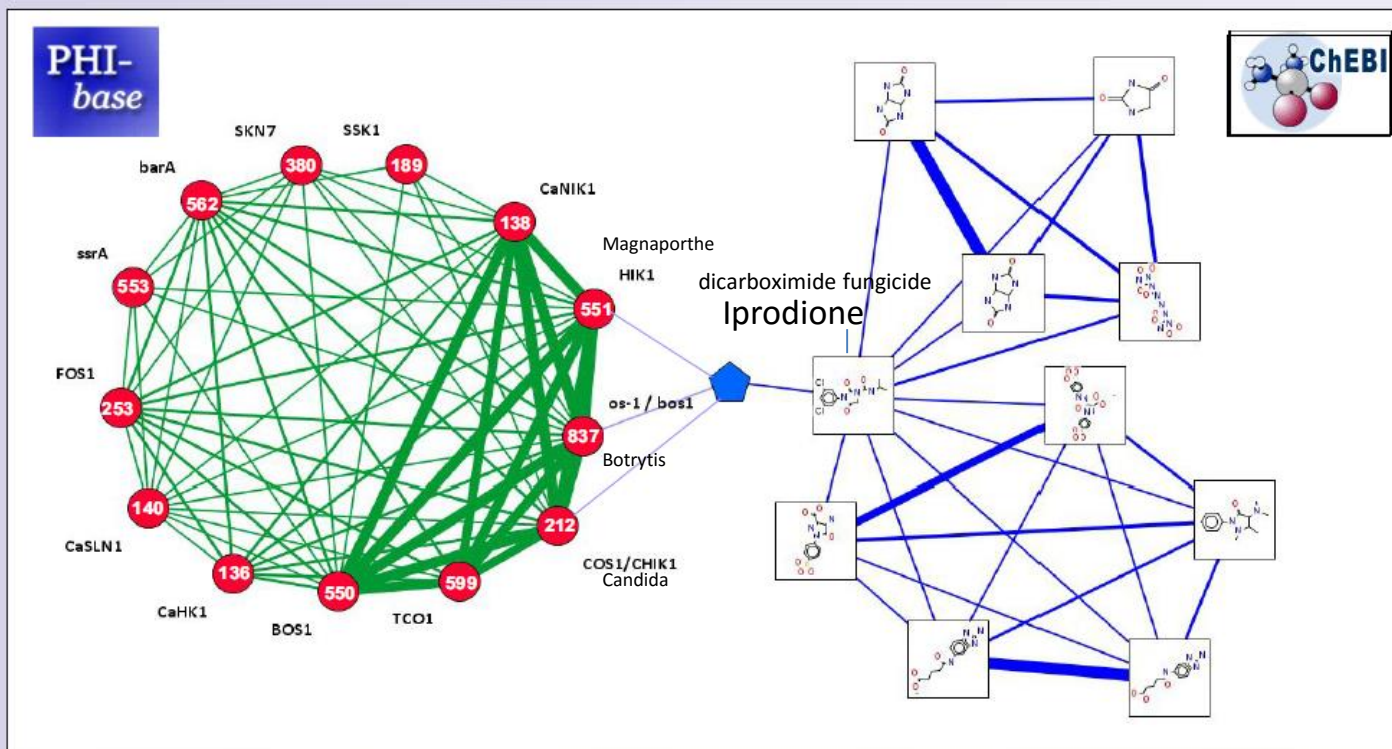
Data integration of PHI-base and ChEBI databases
The blue pentagon indicates where a gene mutation confers resistance to iprodione. Green edges (left) show the sequence similarity derived by reciprocal BLASTP to other PHI-base accessions (white number in red circle). Blue edges (right) show the Tanimoto similarity to other chemicals stored in ChEBI (de Matos et al, 2012).

Collaboration with Mark Wilkinson (Spain) –
A FAIR version of PHI-base for plant pathogens

Conforms to each of the FAIR Principles

2016 International proposition - all scientific data should be "FAIR"—
Findable, Accessible, Interoperable, and Reusable.

May 2016

frontiers
in Plant Science

Publishing FAIR Data: An Exemplar Methodology Utilizing PHI-Base

Alejandro Rodríguez-Iglesias [1†], Alejandro Rodríguez-González [2†], Alistair G. Irvine [3],
Ane Sesma [1], Martin Urban [4], Kim E. Hammond-Kosack [4] and Mark D. Wilkinson [1*]

[1] Center for Plant Biotechnology and Genomics, Universidad Politécnica de Madrid, Madrid, Spain, [2] ETS de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid, Spain, [3] Department of Computational and Systems Biology, Rothamsted Research, Harpenden, UK, [4] Department of Plant Biology and Crop Science, Rothamsted Research, Harpenden, UK

User contribution invited to

PHI-base: contact@phi-base.org

➢ suggest articles for curation by email

➢   inform us on annotation mistakes/paper retractions

➢ Authors asked to submit novel isolate specific gene sequences to Genbank/EMBL to help link published phenotypes to new alleles

➢ publish negative data sets on genes not involved in virulence

➢ Try out the new author curation tool PHI-CANTO

PhD students, Post-Docs and PIs

Table 1. Multispecies databases and websites involving plant, human and/or animal infecting pathogens which contain information complementary to the data in PHI-base

Name and ref[a]	URL (http://)	Comments
Broad-Fungal Genome Initiative	www.broadinstitute.org/scientific-community/ science/projects/fungal-genome-initiative	Genome browsing and comparative analysis for several plant pathogen division
DFVF (12)	sysbio.unl.edu/DFVF	Fungal virulence factor database generated using text-mining of the PubMed database and Internet
e-Fungi (13)	www.cs.man.ac.uk/~cornell/eFungi	Rich source of ESTs obtained by Sanger sequencing
Ensembl Genomes (14)	www.ensemblgenomes.org	Non-vertebrate species genomes portal with links to bacteria, fungi, metazoa, plants and protists
Ensembl Bacteria	bacteria.ensembl.org	Genomes of bacterial and archea
Ensembl Fungi	fungi.ensembl.org	Genomes of fungal species including fungal pathogens
Ensembl Protists	protists.ensembl.org	Genomes of protist species including Phytophthora
Oomycetes Transcriptomics Database (15)	www.eumicrobedb.org/transcripts	Oomycete genomes and transcriptomics
EuPathDB (16)	eupathdb.org	Human pathogens
FRAC	www.frac.info	All known chemical target sites used commercially for the control of pathogens
FungiDB (17)	fungidb.org	Fungal genomics database providing graphical tools for data mining
HPIDB (18)	agbase.msstate.edu	Fifteen human virus pathogens–protein-protein interaction data
JGI-MycoCosm (19)	genome.jgi.doe.gov/programs/fungi	A genome portal for 100s of pathogenic and non-pathogenic fungal species
Pathogen Portal	www.pathogenportal.org	Emerging or re-emerging pathogens, potential biowarfare or bioterrorism pathogens
PHIDIAS (20)	www.phidias.us	Medical fungal and bacterial pathogens
PhytoPath	www.phytopathdb.org	PhytoPath–32 Fungi, 14 Protists, 12 bacterial species linked to PHI-base
PLEXdb (21)	www.plexdb.org	Transcriptomics data only on plants, pathogens and during interactions
USDA	nt.ars-grin.gov/fungaldatabases	Description of all the known hosts of fungi which infect plants
VFDB (22)	www.mgc.ac.cn/VFs	Virulence factors of human and animal bacterial pathogens

[a]Reference provided where available.