

## The Pathogen–Host Interactions database (PHI-base): Summary of improvements January 2025

The new version, called PHI-base 5, increases the amount of information that can be curated from peer-reviewed publications about pathogen–host interactions. Information is now recorded in a more formal way, using stable identifiers and standardised terminology, to make it easier for both humans and machine learning systems to use the information.

The new PHI-Base 5 website has several new features:

- Displays all information about a given pathogen or host gene on a single page (one page per gene), which is intended to make it easier for researchers both to use PHI-base as a single place to explore their genes of interest, and to identify findings in the literature that either confirm, or conflict with, existing knowledge about a gene.
- Makes it easier for information from PHI-base to be integrated with other biological databases and knowledge-bases that can organise information around genes, such as UniProtKB and KnetMiner.
- Enables the dissemination of richer information that is collected using a new curation process. This new process is supported by an online curation tool, PHI-Canto (released in 2023), that allows researchers to contribute their latest findings to PHI-base as soon as their findings are published. The availability of the PHI-Canto tool was publicised in 2023 ([New Tools Empower Plant Disease Detectives](#)).
- Allows the curation of information that is either entirely new to PHI-base, or curated in more detail than before. This new information includes:
  - changes to pathogen genes that cause resistance to antibiotics or antifungals;
  - changes to host genes that improve the host's ability to resist infection;
  - phenotypes for interactions that are common in plants, such as gene-for-gene interactions; and
  - comparisons between the normal phenotype causally linked to a gene to what happens when the gene is mutated or deleted.
- Represents a step forward in the use of standardised language for describing phenotypes and experimental variables. Phenotypes are defined using standard terminology provided by the *Pathogen–Host Interaction Phenotype Ontology* (PHIPO), which removes the possibility of different scientists describing a phenotype using different wording. PHIPO was developed by the PHI-base team specifically for curating phenotypes. Conditions of experiments, such as 'high temperature' and 'high pH' are similarly defined by reusing standard terms developed by the *PomBase* team at the University of Cambridge, which were supplemented by terms defined by the PHI-base team.

- Connects to information in many other biological databases through the use of widely-used identifiers for genes, species, host tissue types, and more. These databases include the Gene Ontology, the BRENDA Tissue Ontology, the UniProt Knowledgebase, the Ensembl database, and the NCBI Taxonomy database. This increases the usefulness of PHI-base to other systems that enable knowledge discovery, such as knowledge graphs and machine learning systems.

PHI-base developments have been supported over the project's lifetime by various competitive BBSRC awards. These have included a 5 year National Capability award (BB/J/004383/1), three *Bioinformatics and Biological Resources* awards (BB/I/001077/1, BB/K020056/1, BB/S020020/1) and various PhD projects. PHI-base is currently funded within the *Growing Health and Designing Future Wheat* Institute Strategic Programmes (detailed below).

## References

Martin Urban, Alayne Cuzick, James Seager, Nagashree Nonavinakere, Jahobanta Sahoo, Pallavi Sahu, Vijay Lakshmi Iyer, Lokanath Khamari, Manuel Carbajo Martinez, Kim E.

Hammond-Kosack, PHI-base – the multi-species pathogen–host interaction database in 2025, *Nucleic Acids Research*, 2024;, gkae1084, <https://doi.org/10.1093/nar/gkae1084>

Alayne Cuzick, James Seager, Valerie Wood, Martin Urban, Kim Rutherford, Kim E. Hammond-Kosack (2023) A framework for community curation of interspecies interactions literature, *eLife*, 12:e84658, <https://doi.org/10.7554/eLife.84658>

Thomas K. Baldwin, Rainer Winnenburg, Martin Urban, Chris Rawlings, Jacob Koehler, and Kim E. Hammond-Kosack, The Pathogen-Host Interactions Database (PHI-base) Provides Insights into Generic and Novel Themes of Pathogenicity (2006). *Molecular Plant-Microbe Interactions*, 19:12, 1451–1462, <https://doi.org/10.1094/MPMI-19-1451>

Rainer Winnenburg, Thomas K. Baldwin, Martin Urban, Chris Rawlings, Jacob Köhler, Kim E. Hammond-Kosack, PHI-base: a new database for pathogen host interactions, *Nucleic Acids Research*, Volume 34, Issue suppl\_1, 1 January 2006, Pages D459–D464, <https://doi.org/10.1093/nar/gki047>

### Acknowledgements and funding

Software development work on the website and database for PHI-base 5 was provided by Molecular Connections Pvt Ltd (Bangalore, India). The PHI-Canto curation tool was developed in collaboration with the PomBase team at the University of Cambridge, with Kim Rutherford providing software development and Val Wood providing consultancy and training on the new curation process. Development of PHIPO and other PHI-base ontologies was assisted by the Ontology Development Kit, developed by INCATools, and the Unified Phenotype Ontology, developed by the uPheno project, in consultation with Nico Matentzoglou (Semanticy Ltd, formerly EMBL-EBI). The editors of the Gene Ontology, in particular Pascale Gaudet (Swiss-Prot, Swiss Institute of Bioinformatics), developed new ontology terms to assist curation of pathogen–host processes in PHI-base. Since 2011, PHI-base data has been hosted within each Ensembl release. Development of PHI-Canto and PHIPO was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) (BB/S020020/1). Ongoing development of PHI-base is currently funded by two Institute Strategic Programmes at Rothamsted Research: Growing Health (BB/X010953/1; BBS/E/RH/230003A) and Delivering Sustainable Wheat (BBS/E/RH/230001B).