

Rothamsted Repository Download

A - Papers appearing in refereed journals

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., De Silva, N., Venkatesh, S. Y, Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K. and Hammond-Kosack, K. E. 2019. PHI-base - the Pathogen–Host Interactions database . *Nucleic Acids Research*. Database issue.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1093/nar/gkz904>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/96z12/phi-base-the-pathogen-host-interactions-database>.

© 16 November 2019, Please contact library@rothamsted.ac.uk for copyright queries.

Some supplementary files may need to be viewed online via your Referee Centre at <http://mc.manuscriptcentral.com/nar>.

PHI-base - the Pathogen-Host Interactions database

Journal:	<i>Nucleic Acids Research</i>
Manuscript ID	NAR-03083-Data-E-2019.R1
Manuscript Type:	6 Database Issue
Key Words:	multiple pathogen species, multiple host species, manually curated literature, fully downloadable content, BLAST tool

SCHOLARONE™
Manuscripts

PHI-base - the Pathogen–Host Interactions database

Martin Urban¹, Alayne Cuzick¹, James Seager¹, Valerie Wood², Kim Rutherford², Shilpa Yagwakote Venkatesh³, Nishadi De Silva⁴, Manuel Carbajo Martinez⁴, Helder Pedro⁴, Andy D Yates⁴, Keywan-Hassani-Pak⁵ and Kim E Hammond-Kosack^{1*}

¹Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden AL5 2JQ, United Kingdom

²Cambridge Systems Biology Centre and Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, United Kingdom

³Molecular Connections, Kandala Mansions, Kariappa Road, Basavanagudi, Bengaluru 560 004, India

⁴European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom

⁵Department of Computational and Analytical Sciences, Rothamsted Research, Harpenden AL5 2JQ, United Kingdom

* Correspondence:

kim-hammond-kosack@rothamsted.ac.uk

Keywords: multiple pathogen species, multiple host species, manually curated literature, fully downloadable content, BLAST tool

ABSTRACT

The pathogen–host interactions database (PHI-base) is available at www.phi-base.org. PHI-base contains expertly curated molecular and biological information on genes proven to affect the outcome of pathogen–host interactions reported in peer reviewed research articles. PHI-base also curates literature describing specific gene alterations that did not affect the disease interaction phenotype, in order to provide complete datasets for comparative purposes. Viruses are not included, due to their extensive coverage in other databases. In this article, we describe the increased data content of PHI-base, plus new database features and further integration with complementary databases. The release of PHI-base version 4.8 (September 2019) contains 3454 manually curated references, and provides information on 6780 genes from 268 pathogens, tested on 210 hosts in 13,801 interactions. Prokaryotic and eukaryotic pathogens are represented in almost equal numbers. Host species consist of approximately 60% plants (split 50:50 between cereal and non-cereal plants), and 40% other species of medical and/or environmental importance. The information available on pathogen effectors has risen by more than a third, and the entries for pathogens that infect crop species of global importance has dramatically increased in this release. We also briefly describe the future

1
2
3 direction of the PHI-base project, and some existing problems with the PHI-base curation
4 process.
5
6

7 8 **THREE KEY POINTS** 9

10 (i) Improved PHI-base 4 resource with an increase in genotypic and phenotypic data.

11
12 (ii) PHI-base annotations displayed in over 1000 genomes in Ensembl, and in 18 genomes in
13 FungiDB.
14

15
16 (iii) Increased curation of pathogen effectors from species that infect model plants, crop plants,
17 and humans.
18

19 20 **INTRODUCTION** 21

22
23 Infectious diseases have a profound influence on every aspect of society. Diseases are a major
24 concern to plant, animal, human, and ecosystem health. Globally infectious diseases threaten
25 food security, human community structures, and the biodiversity of natural ecosystems (1-3).
26 The increasing effects of climate change, human migration, and the globalisation of the trading
27 of fresh goods have resulted in a rise in the incidence and severity of existing disease problems,
28 as well as the emergence of a cohort of novel pathogen species and zoonoses (4). In addition,
29 the (re)acquisition of resistance to anti-infective chemistries – coupled with a rise in legislation
30 banning or restricting existing chemistries – means the burden of microbial infections is of ever
31 growing concern to human, animal, and plant welfare (5,6). In the United Kingdom alone, the
32 total economic burden from infectious diseases is estimated at £30 billion annually, and
33 accounts for 7% of all deaths (7).
34
35

36
37 Infectious diseases are a consequence of complex and dynamic interactions between pathogen
38 virulence factors, and host cell recognition and response systems (8-10). It is increasingly clear
39 that studying these interactions across the tree of life is a fertile ground for uncovering crucial
40 biological principles that control the interaction outcome. In addition, in the post-genomics era
41 – with the ever-decreasing costs for whole genome sequencing, genome assembly, and gene
42 prediction – there is intense scientific and commercial interest in comparative pathogen
43 genomics, as well as whole genome protein–protein interaction predictions and comparisons to
44 identify functionally homologous genes, and to pinpoint species-unique genes and pathways.
45 This increased understanding of the dynamics of a wide range of interactions contributes to the
46 two predominant approaches available for combating infectious disease: namely, stimulating
47 the host immune system to prevent infections, and minimising the use of chemicals to
48 eliminate infectious agents (11-13).
49
50

51
52 The pathogen–host interactions database (PHI-base) was established in 2005 and is freely
53 available at www.phi-base.org. PHI-base contains expertly curated molecular and biological
54
55
56
57
58
59
60

1
2
3 information on genes proven to affect the phenotypic outcome of pathogen–host interactions
4 (14,15). All PHI-base entries are supported by strong experimental evidence from a peer
5 reviewed publication. PHI-base catalogues experimentally verified pathogenicity, virulence, and
6 effector genes from fungal, protist, and bacterial pathogens which infect plant, human, animal,
7 and insect hosts. Genes tested but found not to affect the interaction outcome are also
8 expertly curated. In PHI-base, the term ‘interaction’ is used to describe the observable function
9 of one gene, on one host, on one tissue type (14). Nine high-level phenotypic outcome terms
10 have been developed to permit the comparison of interactions across the entire tree of life
11 (16). These terms are ‘loss of pathogenicity’, ‘reduced virulence’, ‘increased virulence’,
12 ‘unaffected pathogenicity’, ‘effector’, ‘lethal’, ‘increased virulence (hypervirulence)’, ‘resistance
13 to chemical’ and ‘sensitivity to chemical’. These high-level phenotypic outcome terms –
14 although not yet supported by formal controlled vocabulary – are particularly useful for
15 bioinformaticians and biologists unfamiliar with the nuances of multiple pathogen–host
16 interactions, but who wish to include pathogens with different lifestyles and host ranges in
17 their comparative analyses. In addition, a PHIB-BLAST tool has been introduced to permit
18 simple or advanced queries arising from functional genomics, transcriptomics, and proteomics
19 experimentation.

20
21
22 In 2017, PHI-base joined the UK node of ELIXIR’s ‘Data for Life’ project as a gold-standard
23 ‘agricultural omics data’ provider (17). PHI-base follows the FAIR data principles in order to
24 make data findable, accessible, interoperable, and reusable (18). PHI-base also reuses data
25 provided by external resources, including PubMed, NCBI taxonomy, UniProtKB, and the Gene
26 Ontology (GO). A number of complementary multi-species databases on pathogens exist that
27 also provide gene function annotation (recently reviewed by (14,19,20)). PHI-base is unique in
28 describing a broad range of plant and animal pathogen–host interactions using the same
29 controlled vocabulary consistently across >250 species.

30
31
32 In this article, we report on a major increase in PHI-base gene content, new database features,
33 integration with complementary databases, and our immediate plans using new funding.

34 35 36 37 38 39 40 41 42 43 44 45 46 47 **RESULTS AND DISCUSSION**

48 49 50 **Biological data**

51
52 Version 4.8 of PHI-base (released in September 2019 and described in this article), contains
53 data on 6780 genes, 13,801 interactions, 268 pathogens, 210 hosts, and 3454 references. This
54 version includes 71% more interactions, each annotated with a phenotype, compared to PHI-
55 base version 4.2 described in (14). Bacteria and fungal pathogens represent the majority of the
56 interaction data, with a near 50:50 split of entries; whilst protists, nematodes and insects
57 represent 3.6% of the species (**Table 1**). The fungal pathogen interactions are dominated by the
58
59
60

1
2
3 Ascomycetes, which covers 88.5% of annotated fungal interactions (5929 interactions, 100
4 species); this is followed by the Basidiomycetes, which only cover 11.4% of annotated
5 interactions (762 interactions, 11 species). In total, 5755 phenotype interactions describing
6 experimental data on 2320 genes from 1235 newly curated publications are included up to
7 March 2019.
8
9

10
11 The number of pathogenic species in PHI-base was capped at 268, because only a small number
12 of newly emerging pathogens under intense investigation was added to the already included
13 species. Plant infecting pathogens – namely bacteria, fungi, protists, nematodes and insects –
14 represent 60% of the species in PHI-base (**Table 2**). Amongst these, there is an almost equal
15 split between cereal and non-cereal infecting species. Woody tree infecting species provide
16 1004 interaction entries (7.3% of plant pathogen interactions). Amongst the 32 human and
17 animal infecting pathogens, an increasing number are now being tested on non-vertebrate
18 species: for example, various insects, nematodes and crustaceans. These non-vertebrate
19 pathogen interactions now account for 23% of database entries (**Table 2**).
20
21
22
23
24

25 As in previous versions of PHI-base, the highest number of pathogen–host interactions tested
26 in molecular genetic studies and reported in the literature are from the filamentous fungal
27 pathogens *Fusarium graminearum* and *Magnaporthe oryzae*, which cause various diseases on
28 staple crops, such as wheat, rice and maize (**Table 3**). The most highly represented plant-
29 infecting bacteria are *Ralstonia solanacearum*, a pathogen of potato and other Solanaceae
30 crops; and *Xanthomonas oryzae*, a pathogen of rice. For the animal kingdom, the most
31 frequently studied pathogens include the human pathogens *Salmonella enterica*, *Candida*
32 *albicans* and *Pseudomonas aeruginosa* (**Table 3**). Amongst the top 30 species present in PHI-
33 base, phenotypic interaction information – from single, double, and occasionally multiple gene
34 deletions – is provided for each species: from a minimum of 32 genes to a maximum of 1340
35 genes. However, for the cereal infecting fungus *Pyrenophora tritici-repentis*, only five genes
36 have been explored over 142 interactions. Overall, the 30 top species in PHI-base consist of 12
37 fungi, 1 protist and 17 bacteria, and together these covers 71% of total interactions and 88% of
38 total genes.
39
40
41
42
43
44

45 Since 2015, there has been an emphasis on increasing the curation of pathogen gene
46 modifications that result in a hypervirulence phenotype on the host. This has steadily risen
47 from 112 genes (ver. 3.8) to 233 genes (tested in 324 interactions) (ver. 4.2), to 475 genes
48 (tested in 635 interactions) (ver. 4.8). Hypervirulence phenotype interactions now account for
49 4.6% of all database entries and are particularly prevalent amongst bacterial pathogen entries
50 (**Table 1**). This increasing number of hypervirulent interactions indicates that many additional
51 aspects of the negative regulation of key pathogenicity processes – occurring during infection
52 and colonisation of both plant and animal hosts – have been identified. This gene set continues
53 to warrant close monitoring in pathogen populations when attempting to explore, and then
54 mitigate, the emergence and spread of hypervirulent pathogens associated with severe disease
55 outbreaks (21).
56
57
58
59
60

1
2
3 A second major curation effort for PHI-base has been to increase coverage of pathogen
4 effectors (14). An effector is an entity derived from a pathogenic or non-pathogenic species,
5 that either activates or suppresses host defences or other host responses. Interactions
6 involving effectors have risen by 35%: from 1668 (ver. 4.2) to 2261 (ver. 4.8). This category now
7 represents 16% of the dataset, with data derived from 83 species, mostly plant pathogens
8 (**Table 4**). In total, 67% of the effector entries (1511 interactions) are from bacterial species;
9 there is also a considerable number of entries from five obligate fungal rust or powdery mildew
10 species, and one obligate protist species (*Hyaloperonospora arabidopsidis*). Based on data
11 curated in PHI-base, the experimental method of choice for studying effector function is
12 evaluating transient expression in a host or non-host species: transient expression tests
13 account for 573 interactions across 28 pathogen species.
14
15
16
17
18
19

20 In 2015, nine high level phenotypic terms were introduced to the curation process, to permit
21 researchers to explore the database across a wide range of taxonomically diverse species which
22 exhibit varied pathogenic lifestyles (16). The phenotype term 'reduced virulence' is the most
23 highly represented and applies to 44% of database entries. The second most frequent term is
24 'unaffected pathogenicity', at 26%. The majority of the 'unaffected pathogenicity' phenotypes
25 have been reported for plant pathogens (64%), however an increasing number (1004) are from
26 animal pathogens (compared to 80 interactions in version 3.6, and 280 interactions in version
27 4.2). This change appears to have arisen primarily because, within an individual publication, the
28 number of host species tested, or the number of pathogen genes tested has increased; also,
29 comparative results may be included from single, double, and multiple-gene deletion mutants.
30 The number of articles reporting entirely negative data remains small. These negative
31 outcomes are usually presumed (by the respective authors) to indicate that the gene product
32 does not have a functional role in the pathogenic process under investigation, or that gene
33 redundancy exists.
34
35
36
37
38
39

40 The high-level phenotypes for all interactions are summarized in **Table 1** (for pathogen species)
41 and **Table 2** (for host species). A total of 183 PHI-base entries have been assigned the 'lethal'
42 phenotype, consisting of 7 plant-infecting pathogens, 12 animal-infecting pathogens and 1
43 insect-infecting pathogen. The majority of lethal phenotype annotations are for fungal species,
44 in particular *Fusarium graminearum* (94 entries), for which genome-wide single gene
45 replacement studies have been completed for all predicted transcription associated proteins
46 (22), the predicted protein kinases (the kinome) (23), protein phosphatases (the phosphatome)
47 (24), and – most recently – the predicted plasma membrane spanning G-protein coupled
48 receptors (25,26). In these large-scale experiments, no transformants were recovered in repeat
49 experiments, whilst transformants were recovered for many other genes. Thus, the authors
50 considered that the gene's function was 'essential for life'. The human pathogen *Aspergillus*
51 *fumigatus* has also contributed a disproportionately high number of lethal phenotype entries,
52 with 42 of the 207 genes tested (20%) falling into this category where a targeted screen for
53 essential genes has been initiated (27). However, amongst the 30 species with the most
54
55
56
57
58
59
60

1
2
3 interactions in PHI-base (**Table 3**), 17 species have no 'lethal' category entries, whilst a further
4 8 species only provide 1 or 2 lethal entries.
5
6

7 An increasing number of interactions involving human and animal pathogens are now being
8 tested in non-vertebrate species (**Table 2**). In these bioassays, a wide range of insect larvae are
9 used, including: *Galleria mellonella* (greater wax moth), *Plutella xylostella* (diamondback moth),
10 and *Bombyx mori* (domestic silkworm); as well as adult insects, specifically *Drosophila*
11 *melanogaster* (fruit fly). Other studies have used the nematode *Caenorhabditis elegans*
12 (roundworm), the slime mold *Dictyostelium discoideum*, the free-living amoeba *Acanthamoeba*
13 *castellanii*, or various crustaceans: such as shrimp species from the genus *Artemia* and
14 *Penaeus*; and bivalve species, such as oysters from the genus *Crassostrea*. The increasing
15 adoption of the 3Rs principle (replacement, reduction, and refinement) in place of animal
16 models is the main contributing factor to the rising number of non-vertebrate entries (28).
17
18
19
20
21

22 With increasing concerns over global food security, researchers in the international community
23 are being encouraged to investigate host plant-pathogen interactions in crop species, rather
24 than just model pathosystems (29). In addition, the availability of the published completed
25 reference genome for hexaploid bread wheat (*Triticum aestivum*) from the International Wheat
26 Genome Sequencing Consortium (RefSeq v1.0) (<https://www.wheatgenome.org/>) (30) is
27 increasing the pace of discovery for many wheat infecting species. **Table 5** shows the
28 interaction entries involving major food and feed crops: namely wheat, rice, maize, barley,
29 tomato, potato, and *Brassica*. Together, these seven host plant species provide 37% of the data
30 in PHI-base (5096 interactions) and involve 79 pathogenic species (60% of plant pathogen
31 species in PHI-base). In contrast, the three model species *Arabidopsis thaliana*, *Nicotiana*
32 *benthamiana*, and *Nicotiana tabacum* provide only 6.4% of the data (888 interactions). The
33 high number of 48 pathogenic species tested using *N. benthamiana* and *N. tabacum* is
34 predominantly due to the availability of *Agrobacterium*-mediated transient expression assays
35 to test the function of effector proteins.
36
37
38
39
40
41
42

43 Mapping PHI-base phenotypes to Ensembl Genomes and FungiDB

44
45 PHI-base supplies phenotypic annotation for over 100 crop-plant-infecting microbial pathogens
46 into Ensembl Genomes (31). This contribution was initiated as part of the PhytoPath project
47 (32). Recently, the set-up of an improved mapping pipeline developed by Ensembl has
48 contributed to an increase in the number of genomes with PHI-base annotations by a factor of
49 8.7 in the total genomes of bacteria, fungi and protists compared to 2017 (De Silva et al, NAR
50 Database issue 2020, submitted). Also, as a result of extrapolating annotations for conserved
51 genes to closely related species, Ensembl have now applied PHI-base annotations to over
52 14,000 genes in over 1000 genomes. These can provide potential clues for experimental
53 validation in other pathogens. Phenotype annotations are also provided to FungiDB (33).
54 FungiDB release 45 (September 2019) integrates 2633 PHI-base annotations, mapping to 1636
55 genes for 18 FungiDB hosted genomes. In addition to pathogen–host interaction annotations,
56
57
58
59
60

1
2
3 several *in-vitro* phenotypes including growth, sporulation and penetration defects are
4 displayed.
5

6 7 **Migration to reference sequence UniProt IDs**

8
9
10 PHI-base provides links to UniProt IDs when these accessions exist in UniProt Knowledgebase
11 (34). These links can provide further molecular protein annotation including GO terms.
12 However, new genomes are sequenced, and existing genomes are re-annotated. This can
13 generate multiple gene IDs and protein IDs for the same gene causing interoperability issues.
14 We are currently migrating to a system where we consistently use the UniProt identifier from
15 the reference strain as listed by UniProt, rather than IDs from alternative (non-reference)
16 strains. PHI-base has over 15 years of curated literature, and therefore contains ~11% legacy
17 genes with no link to UniProt; here in most cases **GenBank** and EMBL records are referenced.
18 For the genes originally curated with Uniprot IDs, ~10% were in the meantime moved to the
19 UniParc sequence archive. Thus, a challenge exists to frequently review and update PHI-base
20 records, until microbial pathogen proteomes become sufficiently refined and available at
21 UniProtKB. In the meantime, single-species community-based efforts, such as
22 FusariumMutantDB (<https://scabusa.org/FgMutantDb>) (35), can effectively support PHI-base
23 by providing mapping files for legacy gene IDs in several genome assemblies/strains to
24 reference strain IDs available at UniProtKB. BLAST mapping of PHI-base proteins using Blast2GO
25 software (Vers. 5.2.5) (36) using default parameters against the UniProtKB/TrEMBL
26 (release2019_07) identified 937 sequences without GO associations. These sequences include
27 many fungal species-specific effectors, for which currently GO terms are being created.
28
29
30
31
32
33
34
35

36 37 **PHI-base BLAST tool**

38
39 PHI-base has a strong focus on providing curated phenotype data, with less emphasis on
40 providing bioinformatics tools. Excellent tools for genome browsing and sequence
41 investigations are provided for example by Ensembl Genomes, FungiDB and other genomic
42 resource providers (33,37). However, **since 2017** we have provided an online sequence-to-
43 phenotype BLAST search tool, called PHIB-BLAST. This allows users to map their own sequences
44 to PHI-base accessions and the reported phenotypic outcomes are displayed in the BLAST result
45 header, to give immediate comparisons between species. Additionally, this information is also
46 made available for download in FASTA format, where PHI-base information is embedded in the
47 single-line FASTA header for each protein sequence.
48
49
50
51
52

53 54 **PHI-base usage**

55 All of the publications citing PHI-base use are cited in the 'about' section of the database.
56 Currently, 367 articles have cited PHI-base and 60% of these have been published in the past 5
57 years. New research investigations using PHI-base information include the 'hot' topics of gut
58 microbiomes, effector discovery, diagnostic markers for the early 'in host' detection of
59
60

1
2
3 pathogens and finding lethal phenotypes in human pathogens to aid the drug discovery
4 process. For those wishing to query past versions of PHI-base, these have been made available
5 on our 'data' repository on GitHub <https://github.com/PHI-base>. PHI-base is accessed by users
6 in 130 countries over six continents. Over the past 3 years PHI-base usage has remained
7 relatively stable at between 9,000 to 16,000 searches and > 400 to 600 full downloads per
8 annum.
9
10
11

12 **Outreach to inspire the next generation**

13
14 For PHI-base to reach a different audience a STEM (Science, Technology, Engineering and
15 Mathematics) outreach article was recently published highlighting the importance of big data,
16 bioinformatics and plant pathology (<https://futurumcareers.com/saving-plants-from-disease>).
17 This article was aimed at an audience of 11 - 19-year olds to inform and enable them to
18 consider career options within these fields. Example case studies were taken from PHI-base and
19 PhytopathDB. Accompanying worksheets ([https://futurumcareers.com/Kim_Hammond-Kosack-
20 activity-sheet.pdf](https://futurumcareers.com/Kim_Hammond-Kosack-activity-sheet.pdf)) were provided to stimulate discussions and ideas within classrooms and
21 beyond.
22
23
24
25
26
27

28 **Future directions**

29 *PHI-Canto and Ontologies*

30
31 As reported previously (14) we have developed the multi-species web-based curation tool PHI-
32 Canto (canto.phi-base.org). PHI-Canto is an implementation of the Canto community curation
33 tool, developed and used by the fission yeast database PomBase (38). In addition to supporting
34 professional biocurators, researchers will be able to directly contribute annotations from their
35 publications to PHI-base. PHI-Canto supports the annotation of GO, phenotypes, modifications
36 and interactions.
37
38
39
40
41

42 Curation in PHI-Canto involves specifying a publication (using a PubMed ID), entering
43 experimental pathogen and host genes (using UniProtKB IDs), creating genotypes (by listing
44 alleles), annotating genotypes with one or more PHIPO (pathogen host interactions phenotype
45 ontology) terms, and selecting an experimental evidence code. Pathogen–host interaction
46 phenotypes are connected to the underlying genotypes of both the pathogen and the host
47 (multi-species genotypes). Physical protein interactions – such as those identified in yeast two-
48 hybrid or co-immunoprecipitation experiments – can also be curated which will be particularly
49 useful in recording the direct interacting host targets of pathogen effectors.
50
51
52
53

54 Recent developments to PHI-Canto include three main improvements. First, the handling of
55 host genes and genotypes. Second, the ability to capture increasingly complex pathogen host
56 interactions, involving incremental changes to either, or both, the pathogen and the host.
57 Third, mechanisms to capture single species phenotypes for the pathogen and the host. PHIPO
58 was recently registered at the Open Biological and Biomedical Ontology (OBO) Foundry
59
60

1
2
3 (<http://www.obofoundry.org/ontology/phipo.html>) to promote reuse in the pathogen
4 community.
5

6
7 PHI-Canto will enable accurate biocuration of pathogen–host interaction data into PHI-base by
8 the international community. With increased PHI-Canto use this will ensure PHI-base can keep
9 up-to-date with the ever-growing number of publications and newly developed experimental
10 techniques. Researchers interested in trialling PHI-Canto are encouraged to contact us by email
11 (curation@phi-base.org).
12
13

14 15 *Improving strain identification and disease curation*

16
17 Curating strain and disease names are problematic because a wide range of synonyms exist
18 that are inconsistently used and published in different research communities. We have
19 developed a standardized list of strains of importance to PHI-base and currently continue to
20 revise inconsistencies in legacy data. A list of infectious disease names from PHI-base is
21 currently being standardized to a set of external disease ontologies for animals and plants. The
22 revised nomenclature will be used in future PHI-base releases and within PHI-Canto.
23
24
25

26 27 *Curation of the fungicide and anti-infective literature*

28
29 PHI-base is curating publications describing the target sites of some anti-infective chemistries
30 although this has not been a high priority (**Table 1**). We plan to increase the coverage of
31 fungicide and anti-infective literature over the next 2 years. To support this work, we have
32 curated a list of anti-infective agents, including a description of the anti-infectives' function,
33 and cross-references to other databases (FRAC codes, CHEBI IDs, and CAS numbers) where
34 available. This inventory is available from our 'data' repository on Github
35 (<https://github.com/PHI-base>). A pilot text mining study with Molecular Connections (PHI-base
36 curation partner), will test bespoke machine learning algorithms using the anti-infectives list to
37 identify additional papers containing potential fungicide and other anti-infective targets for
38 future curation.
39
40
41
42
43

44 45 *Access to PHI-base annotations in graphical displays of biological networks*

46
47 KnetMiner (<https://knetminer.rothamsted.ac.uk/KnetMiner/>) is a digital research assistant with
48 a Google-like search interface, predictive graph algorithms and interactive features to visualise
49 biological knowledge networks (39,40). KnetMiner mines millions of relations in a genome-
50 scale knowledge network to identify novel clues about genes, gene networks, and diseases
51 (41,42). KnetMiner can search an integrated database of crop and model organism genomes,
52 curated databases such as PHI-base, gene expression, gene interaction information, ontologies
53 and the scientific literature to produce a ranked answer with evidence codes within seconds.
54 The user can then interactively explore the auto generated knowledge network, hiding noisy or
55 untrustworthy relations. So far knowledge networks containing PHI-base data have been
56 developed for the cereal infecting fungal pathogens *Fusarium graminearum* and *Zymoseptoria*
57
58
59
60

1
2
3 *tritici*. For plant and crop species, available networks include Arabidopsis, wheat and rice. In the
4 future, our plans are to link knowledge networks for PHI-genes from Knetminer directly into
5 PHI-base.
6
7

8 *PHI-annotations into additional databases*

9

10
11 PHI-base currently provides phenotype data to a variety of resources, including Ensembl
12 Genomes (Protists, Bacteria and Fungi), FungiDB, Knetminer, FusariumMutantDB and GLOBI
13 (43). Future plans include linking out to Ensembl Plants and to the thousands of fungal
14 genomes and hundreds of protist genomes in the MycoCosm database provided by the Joint
15 Genome Institute (JGI) (44). One of the advantages of MycoCosm is that the genomes of
16 pathogenic and non-pathogenic species are displayed and queried via the use of a navigation
17 tree which assists users with minimal knowledge of taxonomic relationships and groups.
18
19
20

21 We will supply Gene Ontology annotations to GO, from where they will be distributed to other
22 resources, including UniProtKB. PHI-base curators are working closely with the manual curation
23 team of UniProtKB/Swiss-Prot to ensure that gene names and strains are consistent between
24 entries, and we will explore mechanisms to share phenotype annotations with UniProtKB.
25
26
27

28 **TABLE LEGENDS**

29

30
31 Table 1. Summary of pathogen groups, interactions and phenotypes within PHI-base version 4.8
32

33
34 Table 2. Summary of the number of host species and interactions within PHI-base version 4.8
35

36
37 Table 3. Top species and interactions within PHI-base version 4.8
38

39
40 Table 4. Summary of the pathogenic species providing the most information on effectors
41

42
43 Table 5. Crop plant and model plant species contributions to PHI-base version 4.8
44
45

46 **AVAILABILITY**

47

- 48 1. PHI-base 4: www.phi-base.org
- 49 2. PHI-base GitHub page: <https://github.com/PHI-base>
- 50 3. PHIB-BLAST: <http://phi-blast.phi-base.org>
- 51 4. PHI-wiki page: <https://en.wikipedia.org/wiki/PHI-base>
- 52 5. PHI-Canto (multi-species community annotation tool): <https://canto.phi-base.org/>
- 53 6. Linked resource - Ensembl genomes: <http://ensemblgenomes.org>
- 54 7. Linked resource - FungiDB: <https://fungidb.org/fungidb/>
- 55 8. Linked resource - Pombase: <https://www.pombase.org/>
- 56
57
58
59
60

9. Linked resource - KnetMiner: <https://knetminer.org>
10. Linked resource - ELIXIR UK: <https://elixiruknode.org/>

ACKNOWLEDGEMENTS

We thank all the species experts who contributed database annotations from their field of expertise into PHI-base. We are grateful to Dr Paul Kersey at the European Bioinformatics Institute (Cambridge, UK) for helpful discussions. Arathi Raguanath at Molecular Connections is thanked for her contributions to literature curation. We thank Dr Midori Harris (University of Cambridge, UK) for discussing PomBase biocuration, and for advice in phenotype ontology development. We thank Drs Achchuthan Shanmugasundram, Evelina Basenko and the FungiDB consortium for helpful discussions on data sharing. Barnaby Norman is thanked for his contributions to the development of PHI-Canto.

FUNDING

PHI-base is funded from two UK Biotechnology and Biological Sciences Research Council grants BB/K020056/1 and BB/S020020/1. Rothamsted authors MU and KHK receive additional BBSRC grant-aided support as part of the Institute Strategic Programme Designing Future Wheat (BB/P016855/1). Rothamsted author AC receives additional support for the PHI-base project from Smart Crop Protection (SCP) strategic programme (BBS/OS/CP/000001) funded through the BBSRC's Industrial Strategy Challenge Fund. University of Cambridge authors VW and KR are supported by the Wellcome Trust (grant no. 104967/Z/14/Z). EMBL-EBI authors NDS, MCM, HP and AY are supported by funding from the UK Biotechnology and Biological Sciences Research Council Grants BB/K020102/1, BB/I001077/1 and BB/S02011X/1 and The European Molecular Biology Laboratory. This work is integrated into the Ensembl browser, supported in part by Wellcome Trust grant 108749/Z/15/Z and The European Molecular Biology Laboratory.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

1
2
3
4
5
6
7
8
9

TABLES

Table 1. Summary of pathogen groups, interactions and phenotypes within PHI-base version 4.8

Phenotype / pathogen	Bacterium	Fungus	Protist	Nematode	Insect	TOTALS
Number of pathogens	131	112	17	6	2	268
Interactions in total	6608	6696	463	24	10	13801
Loss of pathogenicity	204	696	7	1	0	908
Reduced virulence	3054	2960	96	13	0	6123
Unaffected pathogenicity	1375	2202	60	0	0	3637
Effector (plant avirulence determinant)	1511	468	263	9	10	2261
Increased virulence (hypervirulence)	433	173	28	1	0	635
Lethal	18	156	9	0	0	183
Chemical target: resistance to chemical	7	29	0	0	0	36
Chemical target: sensitivity to chemical	6	8	0	0	0	14
Enhanced antagonism	0	4	0	0	0	4

Table 2. Summary of the number of host species and interactions within PHI-base version 4.8

Phenotype	Plant	Vertebrate	Insect	Nematode	Others
Host species	131	32	24	1	22
Interactions in total	8248	4439	696	258	81
Loss of pathogenicity	650	233	15	9	1
Reduced virulence	2885	2655	386	137	60
Unaffected pathogenicity	2326	1004	193	98	16
Effector (plant avirulence determinant)	2001	233	23	1	3
Increased virulence (hypervirulence)	244	300	77	13	1
Lethal	101	80	2	0	0
Chemical target: resistance to chemical	27	3	0	0	0
Chemical target: sensitivity to chemical	13	1	0	0	0
Enhanced antagonism	4	0	0	0	0

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3. Top species and interactions within PHI-base version 4.8

Pathogen	Interactions	Genes	Loss of pathogenicity	Reduced virulence	Increased virulence	Effector	Unaffected pathogenicity	Lethal	No of tested host species
<i>Fusarium graminearum</i>	1571	1340	36	516	8	0	917	94	13
<i>Magnaporthe oryzae</i>	1273	738	279	501	10	84	398	1	7
<i>Ralstonia solanacearum</i>	666	132	16	43	0	597	9	1	9
<i>Salmonella enterica</i>	664	412	8	381	45	108	122	0	11
<i>Xanthomonas oryzae</i>	512	224	3	96	24	306	83	0	3
<i>Erwinia amylovora</i>	450	135	34	165	55	15	181	0	5
<i>Candida albicans</i>	448	343	48	305	11	0	80	4	12
<i>Pseudomonas aeruginosa*</i>	440	220	19	218	34	4	165	0	16
<i>Botrytis cinerea</i>	368	147	24	205	10	4	123	0	26
<i>Ustilago maydis</i>	360	264	48	187	8	17	100	0	3
<i>Aspergillus fumigatus</i>	309	207	30	128	14	0	93	42	4
<i>Cryptococcus neoformans</i>	305	203	44	184	17	0	50	10	8
<i>Pseudomonas syringae</i>	293	170	1	53	9	191	38	1	13
<i>Escherichia coli</i>	264	169	1	167	15	11	69	1	13
<i>Staphylococcus aureus</i>	212	139	12	137	22	2	38	1	10
<i>Fusarium oxysporum*</i>	209	131	24	90	8	27	60	0	17
<i>Xanthomonas campestris</i>	180	110	11	95	4	0	39	2	8
<i>Klebsiella pneumoniae</i>	173	134	4	72	4	0	93	0	4
<i>Streptococcus pneumoniae</i>	152	106	2	110	4	0	30	6	5
<i>Mycobacterium tuberculosis</i>	150	112	3	64	36	1	46	0	4
<i>Candida glabrata</i>	148	43	0	89	6	0	52	1	3
<i>Verticillium dahliae</i>	145	60	14	64	9	24	34	0	16
<i>Listeria monocytogenes</i>	142	69	2	102	17	3	18	0	10
<i>Pyrenophora tritici-repentis</i>	142	5	0	3	1	138	0	0	3
<i>Enterococcus faecalis</i>	132	32	1	82	3	0	46	0	5
<i>Hyaloperonospora arabidopsidis</i>	127	70	0	1	3	123	0	0	5
<i>Streptococcus pyogenes</i>	121	68	0	67	19	0	33	2	7
<i>Xanthomonas citri</i>	119	33	6	18	3	88	4	0	6
<i>Vibrio cholerae</i>	117	71	1	83	2	0	31	0	6
<i>Beauveria bassiana</i>	108	84	4	70	8	2	24	0	11
TOTALS	9835	5971	675	4296	409	1745	2976	166	

*Pathogen species able to infect both plant and animal hosts.

Table 4. Summary of the pathogenic species providing the most information on effectors

Pathogen - 83 species	Interactions
Bacteria - 40 species	1511
<i>Ralstonia solanacearum</i>	597
<i>Xanthamonas oryzae</i>	306
<i>Pseudomonas syringae</i>	191
<i>Salmonella enterica</i>	122
<i>Xanthamonas citri</i>	88
Fungus - 25 species	471
<i>Pyrenophora tritici-repentis</i>	138
<i>Magnaporthe oryzae</i>	84
<i>Passalora fulva</i>	57
<i>Fusarium oxysporum</i>	27
<i>Ustilago maydis</i>	17
Obligate fungal biotrophs - 5 species	65
<i>Melampsora</i> species	33
<i>Puccinia</i> species	27
<i>Blumeria</i> species	5
Protist / 10 species	263
<i>Hyaloperonospora arabidopsidis</i>	123
<i>Phytophthora sojae</i>	51
<i>Phytophthora capsici</i>	38
<i>Phytophthora infestans</i>	29
Nematodes and insects - 3 species	10

Table 5. Crop plant and model plant species contributions to PHI-base version 4.8

Host plant	Interaction entries	No of pathogen species	Loss of pathogenicity	Reduced virulence	Increased virulence	Effector	Unaffected pathogenicity
Crop species							
Wheat	1790	18	71	513	15	149	923
Rice	1371	9	172	464	33	366	324
Maize	661	13	66	381	19	18	176
Barley	463	9	95	163	2	33	169
Tomato	590	30	56	190	9	195	139
Potato	112	15	1	59	8	20	24
<i>Brassica</i>	109	12	16	55	3	15	20
Model species							
Arabidopsis	359	28	7	97	13	198	44
Tobacco (<i>N. benthamiana</i> and <i>N. tabacum</i>)	329	47	7	71	21	202	28
TOTALS (8 crop species)	5784	102 (different species)	491	1993	123	1196	1847

References

1. Brown, G.D., Denning, D.W., Gow, N.A.R., Levitz, S.M., Netea, M.G. and White, T.C. (2012) Hidden Killers: Human fungal infections. *Science Translational Medicine*, **4**, 1-9.
2. Fisher, M.C., Hawkins, N.J., Sanglard, D. and Gurr, S.J. (2018) Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science*, **360**, 739-742.
3. Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L. and Gurr, S.J. (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, **484**, 186-194.
4. Myers, S.S., Smith, M.R., Guth, S., Golden, C.D., Vaitla, B., Mueller, N.D., Dangour, A.D. and Huybers, P. (2017) Climate change and global food systems: potential impacts on food security and undernutrition. *Annu Rev Publ Health*, **38**, 259-277.
5. Bloom, D.E. and Cadarette, D. (2019) Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response. *Frontiers in Immunology*, **10**.
6. Hahn, M. (2014) The rising threat of fungicide resistance in plant pathogenic fungi: *Botrytis* as a case study. *J Chem Biol*, **7**, 133-141.
7. UK Parliament. (2017), *POSTNOTE*. UK Parliament, Vol. 545, pp. 1-5.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
8. Brown, A.J.P., Gow, N.A.R., Warris, A. and Brown, G.D. (2019) Memory in fungal pathogens promotes immune evasion, colonisation, and infection. *Trends Microbiol.*, **27**, 219-230.
9. Hall, R.A. and Noverr, M.C. (2017) Fungal interactions with the human host: exploring the spectrum of symbiosis. *Curr. Opin. Microbiol.*, **40**, 58-64.
10. Jones, J.D. and Dangl, J.L. (2006) The plant immune system. *Nature*, **444**, 323-329.
11. Dangl, J.L., Horvath, D.M. and Staskawicz, B.J. (2013) Pivoting the plant immune system from dissection to deployment. *Science*, **341**, 746-751.
12. Fairlamb, A.H., Gow, N.A., Matthews, K.R. and Waters, A.P. (2016) Drug resistance in eukaryotic microorganisms. *Nat Microbiol*, **1**, 16092.
13. Spampinato, C. and Leonardi, D. (2013) *Candida* infections, causes, targets, and resistance mechanisms: traditional and alternative antifungal agents. *Biomed Res Int*, **2013**, 204237.
14. Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., Sadanadan, V., Khamari, L., Billal, S., Mohanty, S. *et al.* (2017) PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database. *Nucleic Acids Res.*, **45**, D604-D610.
15. Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Kohler, J. and Hammond-Kosack, K.E. (2006) PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.*, **34**, D459-464.
16. Urban, M., Pant, R., Raghunath, A., Irvine, A.G., Pedro, H. and Hammond-Kosack, K.E. (2015) The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.*, **43**, D645-655.
17. Hancock, J.M., Game, A., Ponting, C.P. and Goble, C.A. (2016) An open and transparent process to select ELIXIR Node Services as implemented by ELIXIR-UK. *F1000Research*, **5**.
18. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, **3**, 160018.
19. Cairns, T.C., Studholme, D.J., Talbot, N.J. and Haynes, K. (2016) New and improved techniques for the study of pathogenic fungi. *Trends Microbiol.*, **24**, 35-50.
20. Urban, M., Irvine, A.G., Cuzick, A. and Hammond-Kosack, K.E. (2015) Using the pathogen-host interactions database (PHI-base) to investigate plant pathogen genomes and genes implicated in virulence. *Front Plant Sci*, **6**, 605.
21. Brown, N.A., Urban, M. and Hammond-Kosack, K.E. (2016) The trans-kingdom identification of negative regulators of pathogen hypervirulence. *FEMS Microbiol. Rev.*, **40**, 19-40.
22. Son, H., Seo, Y.S., Min, K., Park, A.R., Lee, J., Jin, J.M., Lin, Y., Cao, P., Hong, S.Y., Kim, E.K. *et al.* (2011) A phenome-based functional analysis of transcription factors in the cereal head blight fungus, *Fusarium graminearum*. *PLoS Pathog.*, **7**, e1002310.
23. Wang, C., Zhang, S., Hou, R., Zhao, Z., Zheng, Q., Xu, Q., Zheng, D., Wang, G., Liu, H., Gao, X. *et al.* (2011) Functional analysis of the kinome of the wheat scab fungus *Fusarium graminearum*. *PLoS Pathog.*, **7**, e1002460.

- 1
- 2
- 3
- 4 24. Yun, Y., Liu, Z., Yin, Y., Jiang, J., Chen, Y., Xu, J.R. and Ma, Z. (2015) Functional analysis of the
- 5 Fusarium graminearum phosphatome. *New Phytol.*, **207**, 119-134.
- 6
- 7 25. Dilks, T., Halsey, K., De Vos, R.P., Hammond-Kosack, K.E. and Brown, N.A. (2019) Non-canonical
- 8 fungal G-protein coupled receptors promote Fusarium head blight on wheat. *PLoS Pathog.*, **15**,
- 9 e1007666.
- 10
- 11 26. Jiang, C., Cao, S., Wang, Z., Xu, H., Liang, J., Liu, H., Wang, G., Ding, M., Wang, Q., Gong, C. *et al.*
- 12 (2019) An expanded subfamily of G-protein-coupled receptor genes in *Fusarium graminearum*
- 13 required for wheat infection. *Nat Microbiol*, **4**, 1582-1591.
- 14
- 15 27. Hu, W., Sillaots, S., Lemieux, S., Davison, J., Kauffman, S., Breton, A., Linteau, A., Xin, C.,
- 16 Bowman, J., Becker, J. *et al.* (2007) Essential gene identification and drug target prioritization in
- 17 *Aspergillus fumigatus*. *PLoS Pathog.*, **3**, e24.
- 18
- 19 28. Tornqvist, E., Annas, A., Granath, B., Jalkestén, E., Cotgreave, I. and Oberg, M. (2014) Strategic
- 20 focus on 3R principles reveals major reductions in the use of animals in pharmaceutical toxicity
- 21 testing. *PLoS ONE*, **9**, e101638.
- 22
- 23 29. Manners, R. and van Etten, J. (2018) Are agricultural researchers working on the right crops to
- 24 enable food and nutrition security under future climates? *Global Environmental Change*, **53**,
- 25 182-194.
- 26
- 27 30. Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Stein, N.,
- 28 Choulet, F., Distelfeld, A. *et al.* (2018) Shifting the limits in wheat research and breeding using a
- 29 fully annotated reference genome. *Science*, **361**.
- 30
- 31 31. Kersey, P.J., Allen, J.E., Allot, A., Barba, M., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen,
- 32 M., Davis, P., Grabmueller, C. *et al.* (2018) Ensembl Genomes 2018: an integrated omics
- 33 infrastructure for non-vertebrate species. *Nucleic Acids Res.*, **46**, D802-D808.
- 34
- 35 32. Pedro, H., Maheswari, U., Urban, M., Irvine, A.G., Cuzick, A., McDowall, M.D., Staines, D.M.,
- 36 Kulesha, E., Hammond-Kosack, K.E. and Kersey, P.J. (2016) PhytoPath: an integrative resource
- 37 for plant pathogen genomics. *Nucleic Acids Res.*, **44**, D688-D693.
- 38
- 39 33. Basenko, E.Y., Pulman, J.A., Shanmugasundram, A., Harb, O.S., Crouch, K., Starns, D.,
- 40 Warrenfeltz, S., Aurrecochea, C., Stoeckert, C.J., Jr., Kissinger, J.C. *et al.* (2018) FungiDB: An
- 41 integrated bioinformatic resource for fungi and oomycetes. *J Fungi (Basel)*, **4**.
- 42
- 43 34. UniProt, C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506-
- 44 D515.
- 45
- 46 35. Baldwin, T.T., Basenko, E., Harb, O., Brown, N.A., Urban, M., Hammond-Kosack, K.E. and
- 47 Bregitzer, P.P. (2018) Sharing mutants and experimental information prepublication using
- 48 FgMutantDb. *Fungal Genet. Biol.*, **115**, 90-93.
- 49
- 50 36. Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M.,
- 51 Talon, M., Dopazo, J. and Conesa, A. (2008) High-throughput functional annotation and data
- 52 mining with the Blast2GO suite. *Nucleic Acids Res.*, **36**, 3420-3435.
- 53
- 54 37. Karp, P.D., Ivanova, N., Krummenacker, M., Kyrpides, N., Latendresse, M., Midford, P., Ong,
- 55 W.K., Paley, S. and Seshadri, R. (2019) A Comparison of Microbial Genome Web Portals. *Front*
- 56 *Microbiol*, **10**, 208.
- 57
- 58
- 59
- 60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
38. Rutherford, K.M., Harris, M.A., Lock, A., Oliver, S.G. and Wood, V. (2014) Canto: an online tool for community literature curation. *Bioinformatics*, **30**, 1791-1792.
 39. Hassani-Pak, K. (2017) *KnetMiner - An integrated data platform for gene mining and biological knowledge discovery. Dissertation*. PhD, Bielefeld University.
 40. Singh, A., Rawlings, C.J. and Hassani-Pak, K. (2018) KnetMaps: a BioJS component to visualize biological knowledge networks. *F1000Research*, **7**, 1651.
 41. Brandizi, M., Singh, A., Rawlings, C. and Hassani-Pak, K. (2018) Towards FAIRer biological knowledge networks using a hybrid linked data and graph database approach. *J Integr Bioinformat*, **15**.
 42. Hassani-Pak, K., Castellote, M., Esch, M., Hindle, M., Lysenko, A., Taubert, J. and Rawlings, C. (2016) Developing integrated crop knowledge networks to advance candidate gene discovery. *Appl Transl Genom*, **11**, 18-26.
 43. Poelen, J.H., Simons, J.D. and Mungall, C.J. (2014) Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, **24**, 148-159.
 44. Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F. *et al.* (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.*, **42**, D699-704.