

**Sharing Mutants and Experimental Information Prepublication using FgMutantDb
(<https://scabusa.org/FgMutantDb>)**

Authors: Thomas T. Baldwin^{1*}, Evelina Basenko², Omar Harb³, Neil A. Brown^{4,5}, Martin Urban⁴, Kim E. Hammond-Kosack⁴, and Phil P. Bregitzer¹

¹ National Small Grains Germplasm Research Facility, USDA-ARS, Aberdeen, Idaho, United States of America

² Institute of Integrative Biology, Functional and Comparative Genomics, University of Liverpool, Liverpool, L69 7ZB UK

³ Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104 USA

⁴ Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ UK

⁵ Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK.

*Corresponding author

Email: tom.baldwin@ars.usda.gov

Keywords: database, mutants, collection, curation, *Fusarium graminearum*, filamentous fungi, FungiDB, Ensembl, PHI-base, Phytopath, crowd-sourcing

Abstract:

There is no comprehensive storage for generated mutants of *Fusarium graminearum* or data associated with these mutants. Instead researchers relied on several independent and non-integrated databases. FgMutantDb was designed as a simple spreadsheet that is accessible globally on the web that will function as a centralized source of information on *F. graminearum* mutants. FgMutantDb aids in the maintenance and sharing of mutants within a research community. It will serve also as a platform for disseminating prepublication results as well as negative results that often go unreported. Additionally, the highly curated information on mutants in FgMutantDb will be shared with other databases such as FungiDB, Ensembl, PhytoPath, and PHI-base, through updating reports. Here we describe the creation and potential usefulness of FgMutantDb to the *F. graminearum* research community, and provide a tutorial on its use. This type of database could be easily emulated for collaborating and tracking research generated mutants in other fungal species.

Introduction:

For many non-model and pathogenic fungal species there is no comprehensive storage for mutant strains or the information of individual scientists' mutant collections (and their associated experimental datasets). Obtaining important information on mutants from publications requires follow up with the corresponding authors. Thus creating a disjointed and inefficient system for collection of material and data that sometimes results in the loss of critical data and/or mutant strains, necessitating the reproduction of mutant strains and/or experiments.

In comparison, the comprehensive and collaborative storage system of model species, including *Neurospora* and *Aspergillus*, maintained by the Fungal Genetic Stock Center (FGSC) offers an extensive collection of strains and mutants useful to the core research community (McCluskey et al., 2010). However, the FGSC has a very limited selection of non-model pathogenic fungal mutant strains developed with genetic engineering, including gene deletion mutants, which require a Biotechnology Regulatory Services (BRS) permit from USDA-APHIS (7 CFR 340). Permit requirements for genetically engineered strains have also impacted the ability to back up collections of such strains at the USDA ARS-NLGRP (K. McCluskey, personal communication).

A partial solution employed by several close-knit scientific communities researching non-model pathogenic fungal species includes circulating spreadsheets which may contain useful mutant data including phenotypes, negative results, lethal observations, or any other relevant information. Shareable data spreadsheets like this

foster collaboration, unite research efforts, minimize experimental duplications, and are a vital resource for early career researchers.

What is FgMutantDb? *FgMutantDb* is a community-driven and community-curated web accessible resource that offers easy to navigate records on the location of mutants and experimental data generated by researchers studying *Fusarium graminearum* (aka *Gibberella zeae*, the causal agent of Fusarium head blight on wheat, barley and other cereal species as well as a range of other disease types on cereal and non-cereal hosts). Prior to the advent of this resource, an informal spread sheet was passed around between investigators to compile information on mutants and their location within the global *F. graminearum* research community. The main effort, led by Dr. Jin-Rong Xu and colleagues (Purdue University, USA) created an extensive reference list which compiled information on thousands of gene deletion mutants along with information on the source, the location of each mutant, and any associated scientific articles. With the permission of Dr. Xu, this information was transformed into a Google Sheets based database for centralized sharing/curation and was presented at the 29th annual Fungal Genetics Conference *Fusarium* workshop in Asilomar CA, USA in March 2017 as an open source resource for researchers to collaborate on. Interest and participation in this initiative from the research community has contributed vital information to this database and was aptly named the *F. graminearum* Mutant Database (FgMutantDb).

The overall goal was to develop sharable lists in a simple Google Sheets platform to provide easily accessible, curated data that also lends itself to crowd-

sourcing contributed information, while tracking user contributed changes in real time. Since March 1st, 2017, FgMutantDb has been accessed in eight countries by *F. graminearum* researchers a total of 112 times and these users have contributed 14 new deletion mutants (full stats here: https://goo.gl/#analytics/goo.gl/UljcTG/all_time). The research community is encouraged to enter published and unpublished data and curate any previously submitted entries by users to facilitate acquisition of complete records. FgMutantDb records can be quickly searched and sorted for necessary information. For example, free word search for gene locus ID, gene description, or type of mutant is available. As of this publication, there are 1269 entries with information on location of mutants with associated authors and literature. This data is not currently compiled in any other freely accessible place.

The FgMutantDb spreadsheet contains the following five tabs:

Tab 1: *“Mutant strain(PI entry)”* is the core part of the database containing multiple columns such as the target gene identifier (column B), the type of mutant (column E), corresponding researcher (column H), ID mapping between FungiDB, Ensembl, PHlbase, Genbank, and MIMPS (columns J, K, L, N & Q) – note that IDs are hyperlinked to their respected source databases, GenBank accession numbers (column N) and citation (column Q). This is the tab where anyone can enter or edit information about gene modifications.

Tab 2: *“All Genes (ID) cross-ref”* contains a cross referencing of gene IDs between databases including FungiDB, Broad, Ensembl, Phi-Base and GenBank.

Tab 3: “*Top 10*” includes data on the frequency of important *F. graminearum* gene names as they appear in the literature.

Tab 4: “*Fungidb vs. Ensembl MultiBLAST*” contains the parameters and results of blasting translated proteins from the *F. graminearum* genome in FungiDB (PH1) to Ensembl (CS3005).

Tab 5: “*Acknowledgments*” includes information about the funding agencies and curator contact.

Certain features of google spreadsheets add to the utility and maintenance of the database, such as the ability to view and revert to previous versions in case of accidentally data deletion or false data entry. Additionally, the chat and comments feature allows for communication with participants in real time.

How FgMutantDb Contributes to Data Sharing: In the past few years the funding climate for fungal genome and bioinformatics resources has been patchy (Kaiser, 2016). For example, the community has lost the Broad Fungal Genome Initiative and the R01 grants extension for AspGD project, and the Munich Information Center for Protein Sequencing (MIPS) is accessible but underfunded. Fungal research communities relied heavily on these databases and their closures caused significant disruptions including diverting researchers to use different genome assemblies and causing multiple gene IDs to be attributed to the same gene. Despite significant setbacks a number of database initiatives emerged to provide much needed support for fungal genomes. One such resource is FungiDB (<http://fungidb.org>), which is part of the larger Eukaryotic Pathogen Genomics Database Resources (EuPathDB) project

sponsored by NIAID and the Wellcome Trust (Stajich et al., 2011) (PMID: 27903906). Additional resources include, Ensembl Fungi (<http://fungi.ensembl.org/>), part of Ensembl databases (Flicek et al., 2012), with PhytoPath (<http://www.phytopathdb.org>) (Pedro et al., 2016). The multi-species database PHI-base (<http://www.phi-base.org>) provides phenotypic information for mutant genes of fungal pathogens (Urban et al., 2017). The phenotypic information is manually curated from peer-reviewed articles to provide a gold-standard data resource. However, it does not allow researchers to submit and share unpublished data. Also gene IDs are not modified when new genome assemblies become available.

FgMutantDb can help alleviate these shortcomings caused by database transitions, losses, and scope. It provides a community driven data platform that exports and shares data with other fungal bioinformatics resources. FgMutantDb establishes a cross-reference index for available data from FungiDB, Ensembl, and PHI-base. For this *F. graminearum* data sets were imported, sorted, and merged. FgMutantDb has the potential to feed information back to all of these databases and to synchronize data across platforms with input from respective databases and manual researcher entries (Fig. 1). Additionally, starting last year *Fungal Genetics and Biology* publications require genes to be linked out to FungiDB (Momany, 2016). FgMutantDb's complete "All Gene IDs cross-ref" tab will allow researchers using previous gene IDs or different assemblies to quickly find new relevant gene records.

Details for Data Import and Export: The metadata for gene records from FungiDB was imported into FgMutantDb by the following path on the FungiDB website (initiate a

Fusarium graminearum search for genes strategy via the Taxonomy menu option) Results were exported from FungiDB as a “Tab delimited (Excel) – custom table” and imported to FgMutantDb where it was for further sorted in FgMutantDb entries using “vlookup” functions to align entries FGRAMPH1 gene IDs in the RR1 assembly (King et al., 2015) and the previous FGSG gene IDs in the MIPS assembly v3.2 (Wong et al., 2001). FungiDB offers crowd-sourced input for their gene record pages under the User Comments section. Contributing to the goal of FungiDB to accumulate the best input from individual public commenters, FgMutantDb supplies comments to FungiDB using the unique user ID “FgMutantDb”. Information is shared with FungiDB with periodic reports. The first report was generated on July, 22nd, 2017 and provided missing previous Gene IDs from the current gene records. These previous IDs were missing due to the initial database transitions, as discussed above. By cross-referencing FungiDB RR1 assembly (King et al., 2015) and Ensembl CS3005v1 assembly (Gardiner et al., 2014) with FGSG entries in FgMutantDb, 253 previously-assigned Gene IDs were identified that had been reported in the literature but had no phenotypic annotation in FungiDB (Fig. S1). Additionally, 1248 associated articles were reported to FungiDB and are now available under the FgMutantDb comments on FungiDB. Other types of missing annotations and genes located at NCBI that could not be found in FungiDB and Ensembl were noted (full report to FungiDB here - <https://goo.gl/tuoZAg>).

Data acquisition from Ensembl was achieved by bulk download of CS3005v1 assembly gene IDs and sequence data in FASTA format was downloaded from the Ensembl website (ftp://ftp.ensemblgenomes.org/pub/fungi/release-37/fasta/fungi_ascomycota1_collection/fusarium_graminearum_gca_000599445). A

multiBLAST analysis was run with FGRAMPH1_ genes as query for the FG05_ using CLC Genomics Workbench 10 software (v.10.0.1). BLAST parameters are located on the “FungiDB vs Ensembl MultiBLAST” tab on the database along with alignment results. Data acquisition from PHI-Base was achieved through downloading an excel sheet with bulk meta-data (<http://www.phi-base.org/downloadLink.htm>) and was sorted for comparisons to meta-data from other databases.

The Importance of Unpublished Data and How to Record Entries: Fungal mutants developed to test a hypothesis often do not contribute immediately to publishable data. For instance, while investigating potential gene targets for RNAi downregulation, it was hypothesized that the production of oxalate presumably by a putative oxaloacetate acetylhydrolase (FGSG_11137) might account for a pH change preceding induction of mycotoxin activation. To test this hypothesis, a gene deletion mutant was generated. Both the pH of the media (i.e. the drop in pH to ~4.0 from ~7.0) and pathogenicity on wheat floral spikes remained unchanged, therefore further experimentation was abandoned. This information was posted on FgMutantDb to give other researchers interested in the same gene(s) the knowledge to proceed with additional hypothesis and the ability to request mutants or experimental material (Fig. S2). Furthermore, should others wish to pursue other potential applications of this mutant then the location of the mutant is known and can be requested. For posting of a detailed prepublication data and/or negative results, it's helpful to utilize the Google forms (<https://goo.gl/yZ77tY>) which gives directions on data submission. A unique address will be generated and placed in slot for this data.

Discussion: In this article, we have described the FgMutantDb database which solves two underlying problems for the global *Fusarium graminearum* community: (1) Compiling information on locations, phenotypes, and any useful (unpublished) data on mutants in the community not stored at FGSC, and (2) ensuring researchers are informationally on the same page across multiple databases. Furthermore, FgMutantDb is actively working with FungiDB, Ensembl Fungi, and PHI-base to create stable links to individual genes and gene pages, tracking changes in the gene models across genome updates, and providing a feedback loop to multiple databases (Fig. 1). The feedback loop will be completed when FgMutantDb records are, on a regular basis, exported to outside databases to ensure that reliable data is available among participating databases. This model of a database can be quickly and conveniently copied for any fungal genetics community that needs access to data within and linking to multiple databases. The corresponding author can be contacted for help in generating a single species tool with similar functionalities as FgMutantDb.

Summary: The current challenge is to ensure researchers are united with shared resources and information across platforms to facilitate effective research. Sharable information on centralized community-curated spreadsheet-based databases, similar to FgMutantDb, will help tackle this challenge.

Acknowledgements: FgMutantDb would not have been possible without the contributions made by Dr. Jin-Rong Xu (Department of Botany and Plant Pathology,

Purdue University, West Lafayette, IN, USA). Thanks to Dr. Linda Harris for naming the database FgMutantDb and to all the *Fusarium graminearum* community researchers for supplying mutant information and expanding database records. The generation of FgMutantDb was partially funded by USDA-ARS Project 2050-21000-031-00 and by the US Wheat and Barley Scab Initiative project *Down with DON: Stable Expression of Proven Genes in a Marker-Free Background*. The USDA-ARS is an equal opportunity employer. Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC), UK, as part of the Institute Strategic Programme grants 20:20 wheat [BB/J/00426X/1] and Designing Future Wheat [BB/P016855/1]. Neil Brown was supported by the BBSRC Future Leader Fellowship [BB/N011686/1]. PHI-base receives support from the BBSRC as a National Capability [BB/J/004383/1] and the PhytoPath1 and Phytopath2 projects [BB/I000488/1, BB/K020056/1].

References:

[Flicek et al., 2011](#) P. Flicek, *et al.*

Ensembl

Nucleic acids research, 40(20), pp.D84-D90.

[Gardiner et al., 2014](#) D.M. Gardiner, *et al.*

Genome sequence of *Fusarium graminearum* isolate CS3005.

Genome announcements, 2(2014), pp.e00227-14.

[Kaiser, 2016](#) Kaiser J.

BIOMEDICAL RESOURCES. Funding for key data resources in jeopardy.

Science 351 (2016), pp. 14

[King et al., 2015](#) R. King, *et al.*,

The completed genome sequence of the pathogenic ascomycete fungus

***Fusarium graminearum*.**

BMC genomics, 16(2015), p.544.

[McCluskey et al., 2010](#) K. McClusky, *et al.*, 2010

**The Fungal Genetics Stock Center: a repository for 50 years of fungal genetics
research**

J Biosci. 35(1):119-26.

[Momany, 2016](#) M. Momany

Database whiplash, crowdsourcing, and FungiDB

Fungal Genet Biol. 92 (2016), p. iii.

[Pedro et al., 2016](#) H. Pedro, et al.

PhytoPath: an integrative resource for plant pathogen genomics. Nucleic Acids

Research Database, 44, (2016) D688-D693

[Stajich et al., 2011](#) J.E. Stajich *et al.*

FungiDB: an integrated functional genomics database for fungi.

Nucleic acids research, 40(2011), pp.D675-D681.

[Urban et al., 2017](#) M. Urban, *et al.*

PHI-base: A new interface and further additions for the multi-species pathogen-host interactions database.

Nucleic Acids Res, 45 (2017) D604-D610

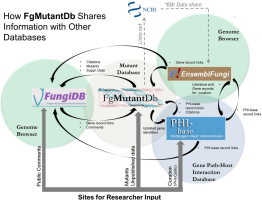
[Wong et al., 2011](#) P. Wong, et al.

FGDB: Revisiting the genome annotation of the plant pathogen *Fusarium graminearum*

Nucleic Acids Res. 39 (2011) D637-D639

Figure 1: How FgMutantDb shares information by importing gene record links, public comments, PHI-base record links, and citations from each of the unique and specialized databases (FungiDB, Ensembl, and PHI-base). Imported data is aligned to curated entries of mutants in FgMutantDb and data is shared back with FungiDB through public comments on gene record pages, including citations, mutants available, and supplementary data. Similarly, FungiDB links out to Ensembl and PHI-base gene records. FgMutantDb can contribute to PHI-base by providing updated gene identifiers to link older accessions PHI-base reports from the literature. Researchers can directly input data at three sites in the sharing scheme, include the addition of public comments on FungiDB gene records, reporting mutants and unpublished data to FgMutantDb, and curating data in PHI-base through PHI-Canto. FgMutantDb conveniently links out to NCBI Genbank records that can be edited to update EnsemblFungi records. Closing the data sharing loop ensures databases remain relevant and allows unique aspects of each to remain useful.

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 publicly available data are released in the next version of Ensembl. Authors can edit Genbank entries to be available in the next Ensembl release.