ARTICLE IN PRESS

Fungal Genetics and Biology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Resource

Sharing mutants and experimental information prepublication using FgMutantDb (https://scabusa.org/FgMutantDb)

Thomas T. Baldwin[a,*], Evelina Basenko[b], Omar Harb[c], Neil A. Brown[d,e], Martin Urban[d], Kim E. Hammond-Kosack[d], Phil P. Bregitzer[a]

[a] *National Small Grains Germplasm Research Facility, USDA-ARS, Aberdeen, ID, United States*
[b] *Institute of Integrative Biology, Functional and Comparative Genomics, University of Liverpool, Liverpool L69 7ZB, UK*
[c] *Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, United States*
[d] *Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*
[e] *Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK*

ARTICLE INFO

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 (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 other fungal species.

1. 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 (Mcclusky 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, locations, 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. However, they are generally not widely accessible.

2. What is FgMutantDb?

FgMutantDb[1] 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

---

* Corresponding author.
E-mail address: tom.baldwin@ars.usda.gov (T.T. Baldwin).

[1] https://scabusa.org/FgMutantDb.

https://doi.org/10.1016/j.fgb.2018.01.002
Received 29 September 2017; Received in revised form 20 December 2017; Accepted 3 January 2018
1087-1845/ Published by Elsevier Inc.

Please cite this article as: Baldwin, T.T., Fungal Genetics and Biology (2018), https://doi.org/10.1016/j.fgb.2018.01.002

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 that can foster communication and collaboration among researchers. 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. These data are not 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, PHI-base, Genbank, and MIPS (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 information.

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 accidental data deletion or false data entry. Additionally, the chat and comments feature allows for communication with participants in real time.

3. 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 causing researchers to use different genome assemblies and 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., 2011), 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 and serves as 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.

4. 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 data was for further sorted in FgMutantDb entries using vlookup functions to align entries to 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., 2011). 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 went missing due to the initial database transitions, as discussed above. By cross-referencing the FungiDB RR1 assembly (King et al., 2015) and the 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