Tatiana Stead

Development of a central data repository and visualisation framework for *Fusarium* transcriptomic data

Faculty of Engineering and Applied Sciences
Applied Bioinformatics

MSc
Academic Year: 2024 - 2025

Supervisors: Prof. Fady Mohareb, Dr Tom Kurowski
September 2025

Faculty of Engineering and Applied Sciences
Applied Bioinformatics

MSc

Academic Year 2024 - 2025

Tatiana Stead

Development of a central data repository and visualisation framework for *Fusarium* transcriptomic data

Supervisors: Prof. Fady Mohareb, Dr Tom Kurowski
September 2025

This thesis is submitted in partial fulfilment of the requirements for the degree of Applied Bioinformatics

© Cranfield University 2025. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

Academic integrity declaration

I declare that:

- the thesis submitted has been written by me alone.
- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

Background: Contamination of grain crops by mycotoxins is a serious threat to food security, affecting up to 40% of cereals worldwide. One of the main contaminants are members of the *Fusarium* family, which produce trichothecene mycotoxins. Trichothecene biosynthesis is influenced by climatic conditions, and this interaction is essential to study to understand how mycotoxin production will change in the future under more extreme climate change. Transcriptomic studies display gene expression under different conditions, and many datasets are publicly available. However, existing repositories specifically holding fungal transcriptomic data can be tricky to navigate and lack clear annotation of experimental conditions, limiting their effectiveness for exploratory analysis.

Results: To address this issue, the Fungal Transcriptomic Database (FTD) was developed as a *Fusarium* specific repository to hold gene expression data. It was populated with differential expression data from five studies, covering four *Fusarium* species and four experimental conditions. The FTD provided clear visualisation of data through interactive tables and graphs, including volcano plots and heatmaps. The rapid and accurate recreation of a heatmap showing differential expression of trichothecene genes was used as a benchmark to validate the functionality and efficiency of the repository.

Conclusions: The FTD provides a straightforward and interactive tool for exploring and re-analysing fungal transcriptomic datasets. Its modular design supports efficient integration of new features, ensuring the repository remains compatible with emerging technologies and scientific advancements. By highlighting experimental metadata alongside interactive gene expression visualisations, the FTD facilitates data interpretation and reveals new insights regarding *Fusarium* behaviour. These capabilities provide a deeper understanding of mycotoxin production, enabling accurate predictions regarding trichothecene biosynthesis under future climatic conditions.

The source code for the interactive tool is available at: https://github.com/tanyastead/fungal-repository

Keywords: Database, gene expression, mycotoxin, R Shiny, RNA-seq, software, SQLite, trichothecene

Acknowledgements

I would like to express my gratitude to the entire bioinformatics team for the help and support they have provided throughout the year. In particular, I would like to thank my supervisors, Fady and Tom, for all their guidance and for teaching me so much over the course of the year.

I would also like to thank Mariam for all the assistance and advice she has given me throughout the project.

Furthermore, special thanks are given to Green Future Investments Ltd. group, whose award of the Brian Meredith Net Zero Scholarship granted me the opportunity to further my education.

Lastly, I would like to thank my family for continuous support they have always given me. This project could not have been completed without their encouragement and constant motivation.

Table of Contents

Academic integrity declaration i
Abstract ii
Acknowledgements iv
List of Figures vi
List of Tables vii
List of Abbreviations viii
1 Introduction 1
1.1 Current state of food security 1
1.2 Impact of disease on yield 1
1.3 Mycotoxin contamination of cereals 1
1.4 Fungal genomics 2
1.5 *Fusarium* trichothecene biosynthesis 3
1.6 Influence of climate change on mycotoxin contamination 4
1.7 Existing fungal transcriptomic data repositories 5
2 Aims and Objectives 9
3 Materials and Methods 10
3.1 Repository architecture and construction 10
3.2 Input data 17
3.3 Case studies 18
4 Results 19
4.1 Repository overview 19
4.2 Case Study A: reproducing trichothecene gene expression 19
4.3 Case Study B: trichothecene gene expression under variable pH 21
5 Discussion 22
5.1 Usability and utility 22
5.2 FTD and other fungal transcriptomic databases 24
5.3 Strengths and weaknesses 25
5.4 Future development 27
6 Conclusions 29
References 31
Appendices 44
Appendix A Ethical approval letter 44
Appendix B Supplementary Tables 45

List of Figures

Figure 1: Simplified biosynthetic pathway of trichothecene toxins..... 4

Figure 2: Database schema showing tables and relationships between tables.10

Figure 3: Screenshots of the different sections of the frontend. 14

Figure 4: Screenshots showing the different options for data visualisations in the Experiments tab..... 15

Figure 5: Screenshot showing the Upload tab, where users can upload their own datasets to the repository. 16

Figure 6: Schematic overview of repository construction..... 16

Figure 7: Heatmap of TRI gene expression across 27 contrasts..... 20

Figure 8: Screenshot of the repository displaying gene expression of trichothecene-related genes under varying pH. 21

List of Tables

Table 1: Overview of publicly available fungal transcriptomic databases. 8

Table 2: R packages and versions required by the repository. 13

Table 3: Type and description of studies entered into the repository. 18

List of Abbreviations

a_w	Water Activity
DE	Differential Expression
DEG	Differentially Expressed Gene
FTD	Fungal Transcriptomic Database
GO	Gene Ontology
scRNA-seq	Single Cell RNA sequencing

1 Introduction

1.1 Current state of food security

Cereal grains, primarily wheat, maize, and rice, serve as staple crops, feeding the majority of the global population (Wang et al., 2018). However, by 2050, the global population is expected to reach around 10 billion (United Nations, Department of Economic and Social Affairs, 2019). To meet population demands, total food production must increase by 70%. Within this, cereal production will need to increase from 2.1 billion to about 3 billion tonnes to sustain the growing population (Food and Agriculture Organization of the United Nations, 2009). A major challenge to food security is the threat caused by pests and plant pathogens, which cause yield losses averaging 21.5% in wheat, 30.0% in rice, and 22.5% in maize (Savary et al., 2019). Therefore, minimising yield loss caused by microorganisms will be critical to ensure global food production is sufficient to feed the growing population.

1.2 Impact of disease on yield

Plants are at risk of infection from a diverse range of pathogens, including bacteria, fungi, oomycetes, viruses and nematodes (Singh et al., 2023). They can significantly impact crop productivity, with annual crop losses ranging from 20% to 40% (Ayaz et al., 2023). For example, the cassava mosaic virus causes annual yield losses of 13 million tonnes in East Africa, resulting in losses of over $1 billion (Ristaino et al., 2021). Extensive research and strategies are being employed to control crop diseases, including the development of disease-resistant crops, application of chemical fungicides and pesticides, agricultural strategies, and increasing reliance on plant microbiomes and biocontrol agents (Ayaz et al., 2023; Jeger et al., 2021; Singh et al., 2023).

1.3 Mycotoxin contamination of cereals

In addition to causing disease, pathogens pose a risk to food security through the contamination of foods with mycotoxins, resulting in food unsuitable for human or animal consumption. This is a serious threat to global agriculture, with 25%-40% of cereal grains worldwide being contaminated, causing significant yield losses

(Kumar & Kalita, 2017; Luo et al., 2021). Mycotoxins are secondary metabolites produced by certain species of the fungal genera, including *Fusarium, Penicillium, Aspergillus,* and *Alternaria* (Khodaei et al., 2021; Mousavi Khaneghah et al., 2019). Over 300 mycotoxins have been identified, however only a small number are dangerous to humans, namely aflatoxins, ochratoxins, fumonisins, trichothecenes and zearalenones (El-Sayed et al., 2022; Khodaei et al., 2021; Luo et al., 2021). They have diverse molecular structures resulting in various adverse effects on human health, but "generally have an effect of organ toxicity, teratogenicity, carcinogenicity, mutagenicity and genotoxicity" (Luo et al., 2021, p. 1).

The level and type of mycotoxin contamination is highly related to geographic region and climatic conditions, which greatly influence fungal growth. In Europe, *Fusarium* is the most common contaminant, with trichothecenes being the major mycotoxin produced (Luo et al., 2021; McCormick et al., 2011). *Fusarium* is a genus of agriculturally important ascomycete fungi that are common agents of plant diseases such as wilts, blights, rots, and cankers, resulting in major yield losses over $1B (Ma et al., 2013; Rampersad, 2020). Although the majority of *Fusarium* are soil-inhabiting fungi, they exhibit a wide range of behaviours, such as the level of necrosis, infection strategies, and host specificity (Ma et al., 2013; Rampersad, 2020). As plant pathogens, *Fusarium* employs numerous strategies to aid in host colonisation, including the production of trichothecenes as virulence factors to overcome host defences and worsen disease (Laraba et al., 2021; Rauwane et al., 2020).

1.4 Fungal genomics

Understanding of fungal behaviours has progressed greatly since the whole genome sequence of *Saccharomyces cerevisiae* in 1996 (Goffeau et al., 1996). Now over 2000 fungal genomes have been sequenced and annotated, and are available on the JGI Mycocosm database, aiding in the research and understanding of fungal behaviour (Ahrendt et al., 2023; Grigoriev et al., 2014). The Mycocosm database contains fungi from all lifestyles, including mycotoxin-producing species. Regarding the *Fusarium* genera, Sweeney (1998) identified

24 trichothecene producing *Fusarium* strains. Of these, 15 species have had genomes sequenced and are hosted on Mycocosm, including *F. graminearum*, *F. langsethiae, F. culmorum* and *F. sporotrichiodes*. In total, Mycocosm hosts 40 unique *Fusarium* genomes, as well as 20 *Alternaria* genomes, 52 *Penicillium* genomes and 148 *Aspergillus* genomes.

1.5 *Fusarium* trichothecene biosynthesis

The trichothecenes family contains over 200 toxins, which can be split into four groups: types A, B, C and D (McCormick et al., 2011). The most toxic trichothecenes include T-2 toxin, HT-2 toxin, neosolaniol and diacetoxyscirpenol, which are type A trichothecenes, and deoxynivalenol (DON) and nivalenol, which are type B trichothecenes (Placinta et al., 1999). Type A trichothecenes show high levels in oats, whilst type B trichothecenes are more often detected in raw cereals such as barley, maize and rye (Pereira et al., 2014). Trichothecene mechanism of action involves the inhibition of protein synthesis and mitochondrial protein synthesis, as well as the disruption of DNA and RNA synthesis, ultimately resulting in apoptosis (Janik et al., 2021; McCormick et al., 2011; Mousavi Khaneghah et al., 2019).

The genes for trichothecene biosynthesis in *Fusarium* are located in three gene clusters: a 12 core *TRI* cluster containing the genes *TRI8, TRI7, TRI3, TRI4, TRI6, TRI5, TRI10, TRI9, TRI11, TRI12, TRI13, TRI14*, a two-gene *TRI1-TRI16* cluster, and the *TRI101* locus. The functions of these genes includes biosynthesis, regulation of biosynthesis, and trichothecene transport (McCormick et al., 2011). The first step of the biosynthesis pathway is controlled by *TRI5*, which encodes the enzyme trichodiene synthase and cyclises the precursor farnesyl pyrophosphate to trichodiene. Subsequent catalytic reactions controlled by *TRI4, TRI101, TRI11* and *TRI13* catalyse the formation of calonectrin from trichodiene. The specific mycotoxin that is produced from calonectrin depends on the *Fusarium* species and the orthologues of the *TRI* genes present. In *F. graminearum*, *TRI7* and *TRI13* synthesize nivalenol, and in DON-producing strains *TRI1* and *TRI8* control the formation of DON. In *F. sporotrichioides* and other T-2/HT-2 producing strains, *TRI7, TRI1* and *TRI8* catalyse the formation of

T-2, with HT-2 being formed following hydrolysis of T-2's acetyloxy group (Kimura et al., 2007; Meneely et al., 2023). A simplified version of this biosynthetic pathway is portrayed by Figure 1 below.

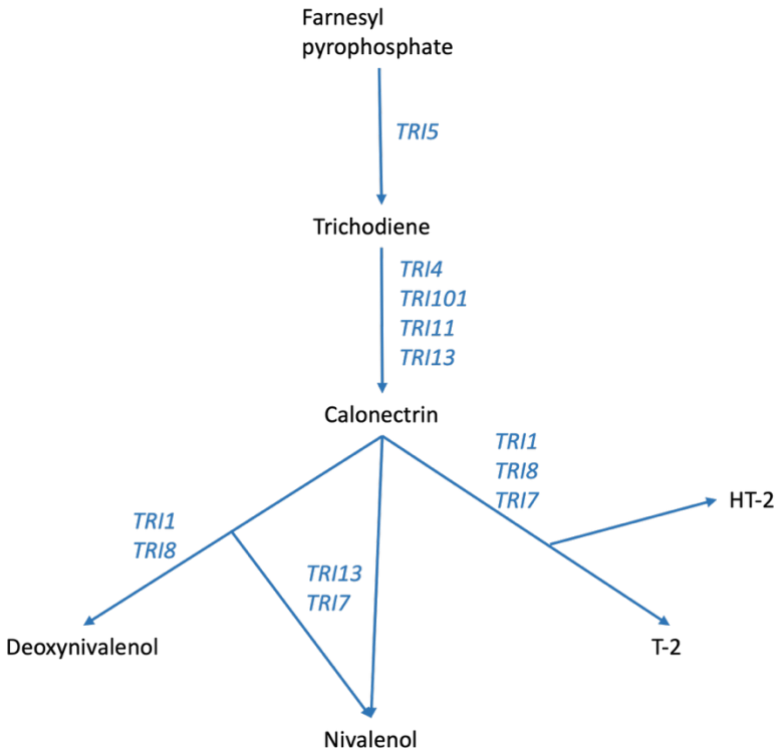

Figure 1: Simplified biosynthetic pathway of trichothecene toxins. The pathway is identical up to the biosynthesis of calonectrin, after which toxin synthesis is species-dependent.

1.6 Influence of climate change on mycotoxin contamination

Atmospheric CO_2 concentrations are projected to be between 550 and 1000 ppm, leading to air temperature increases between 1.0-3.7 °C by the end of the century (Dusenge et al., 2019). Plant metabolism is directly influenced by climatic conditions, and so climate change is predicted to have an enormous impact on crop yield (Dusenge et al., 2019). Moreover, climate change could significantly influence pathogen growth and behaviour, and thus plant-pathogen interactions. Notably, changes in CO_2 concentration, temperature, and water activity have been shown to significantly impact pathogen dynamics (Singh et al., 2023). Regarding *Fusarium*, increased temperature could change species distribution. For example, higher temperature favours the more aggressive *F. graminearum*

over the milder *F. culmorum*, with this shift in prevalence being reported in Norway, UK and Netherlands (Moretti et al., 2019; Parikka et al., 2012). Furthermore, under experimental conditions, elevated CO_2 concentration resulted in greater disease severity in *F. graminearum*, whilst increased water activity increased infection efficiency (Andersen et al., 2015; Váry et al., 2015).

Changes in temperature, CO_2 concentration and water activity have also been shown to influence mycotoxin biosynthesis in *Fusarium* (Bencze et al., 2017; Ding et al., 2024; Verheecke-Vaessen et al., 2019, 2021). Across species and toxins, the influence of climatic conditions on mycotoxin biosynthesis was inconsistent, and highly dependent on species and mycotoxin. For example, Bencze et al. (2017) found that under elevated CO_2 mycotoxin concentration was very high in all wheat grain samples. However, in whole-spike inoculation treatments, concentrations increased, decreased, or did not change depending on the wheat variety. When investigating *F. verticillioides* infection of maize, Ding et al. (2024) determined that greatest mycotoxin production occurred at lower temperatures (15°C) and higher water activity (0.98), whilst Verheecke-Vaessen et al. (2021) determined optimum mycotoxin production in *F. langsethiae* to occur at 0.995-0.98 and 20-30°C. In addition, the combined tripartite effect of interacting abiotic factors, which will more closely mimic future climatic conditions, will likely have a more nuanced effect on mycotoxin production (Medina et al., 2017; Singh et al., 2023; Verheecke-Vaessen et al., 2019). Altogether, these studies highlight the need to study mycotoxin production across a large range of species and under various interacting climatic conditions to more accurately understand mycotoxin contamination under future scenarios.

1.7 Existing fungal transcriptomic data repositories

Predicting mycotoxin production under future climate change conditions will be integral for the development of agricultural practices that safeguard against mycotoxin contamination (Casu et al., 2024). To achieve this, easy access to large amounts of fungal transcriptomics data will be invaluable, necessitating a specialised repository for storing and displaying such data.

Multiple online databases storing plant transcriptomics data exist. These include both more generalised databases holding data for multiple plant species, such as GENEVESTIGATOR or ePlant, as well as specialised databases for single species, such as TomExpress for tomato or AtCAST and ARS for *Arabidopsis* (Hruz et al., 2008; Sasaki et al., 2011; Waese et al., 2017; Zhang et al., 2020; Zouine et al., 2017). These tools consolidate the data from multiple studies into a single platform, enabling users to rapidly access, analyse and compare data across multiple studies and conditions. This ensures researchers can focus all their attention on conducting research instead of spending time navigating across multiple sites to access data (Waese et al., 2017). Furthermore, these platforms offer several visualisation and analysis tools within the same interface, allowing users to easily explore multiple levels of biological data (Waese et al., 2017). For example, enhanced visualisation displaying expression in different tissues or cell compartments, or co-expression networks visualising similarly or oppositely expressed genes (Waese et al., 2017; Zouine et al., 2017). These tools improve the researchers' abilities to analyse and extract information from available data, leading to a deeper understanding of biological processes, more precise modelling of biological systems and more insightful conclusions about biological systems (Hruz et al., 2008).

In contrast, available fungal transcriptomics databases are much more limited. An extensive literature search has only identified three repositories storing transcriptomic data. The first, Fungi.guru, is a tool allowing analysis of co-expression networks and gene expression profiles of 19 fungi. These features are highlighted by the authors, who created a custom co-expression network in *Aspergillus fumigatus,* revealing a high degree of co-expression between metabolite biosynthesis enzymes, including *EAL85149* which was found to be co-expressed with a large variety of genes (Lim et al., 2020). The second tool, FungiDB, is an extensive online resource for data mining and functional genomics analysis. For example, it can be used to conduct *in silico* RNA-Seq experiments, comparing transcriptomes of two strains of *Cryptococcus neoformans*, which identified 281 up-regulated genes in the HC1 strain but not in the G0 strain (Basenko et al., 2018). Lastly, FungiExp, is a platform for exploring fungal gene

expression and alternate splicing. These capabilities were demonstrated in the analysis of the public RNA-seq dataset (SRP188527), which showed that during *Ustilaginoidea vivrens* infection samples were clustered into four groups, reflecting the four infection stages, based on overall gene expression and alternative splicing profiles (Liu et al., 2023).

The capabilities and limitations of these databases are summarised in Table 1. Although collectively these tools offer a variety of gene expression analyses and access to transcriptomic data from multiple studies, there have certain limitations. Most notably, there is no option to identify and visualise differentially expressed genes (DEGs) based on experimental conditions or functional annotation, and the user cannot easily compare expression levels between contrasts within the same experiment. As a result, these databases are unsuitable for exploratory analysis of DEGs. Specifically, they do not provide an intuitive platform to investigate expression of mycotoxin-related genes under varying environmental conditions, limiting their usefulness as a foundation for constructing predictive models. To address this issue, it is necessary to develop a dedicated platform focused on gene expression under varying climatic conditions.

Table 1: Overview of publicly available fungal transcriptomic databases.

Database Name	***Nº. Fungal Species***	***Key Features & Analyses***	***Limitations***	***Reference***
Fungi.guru	19	Co-expression networks, gene expression heatmaps, expression specificity	○ No specificity of gene expression based on abiotic factors ○ Limited number of species	(Lim et al., 2020)
FungiDB	303	Transcript expression, genome browser	○ Difficult to navigate and understand what's going on ○ No specificity of gene expression based on abiotic factors	(Basenko et al., 2018)
FungiExp	220	Co-expression, comparative expression, blast search, expression specificity, cross species, co-analysis	○ No specificity of gene expression based on abiotic factors	(Liu et al., 2023)

2 Aims and Objectives

The overarching aim of this project was to create a specialised *Fusarium* database holding transcriptomic data with a clear user interface displaying gene expression under different abiotic conditions. To ensure this database can be used to investigate mycotoxin production under future climatic conditions, the following objectives were established:

- Develop a software platform with an intuitive and interactive user interface.
- Integrate a structured database to store gene expression data.
- Implement interactive plots and tables for clear data visualisation.
- Design the repository following a modular architecture to enable easy integration of additional fungal genera and abiotic conditions.

3 Materials and Methods

3.1 Repository architecture and construction

The Fungal Transcriptomic Database (FTD) was developed as a centralised repository specialised for *Fusarium* differential expression analysis. The general architecture consists of a central SQLite (Hipp, 2023) database which holds differential expression (DE) data generated from different studies. SQLite was chosen over other SQL databases due to its compact size, serverless design, and open-source nature. The database contains nine tables, displayed in Figure 2 below, which hold information on gene expression, functional annotation, and experimental metadata. This design allows rapid extraction of experimental and DE data based on pre-set queries in response to user commands.

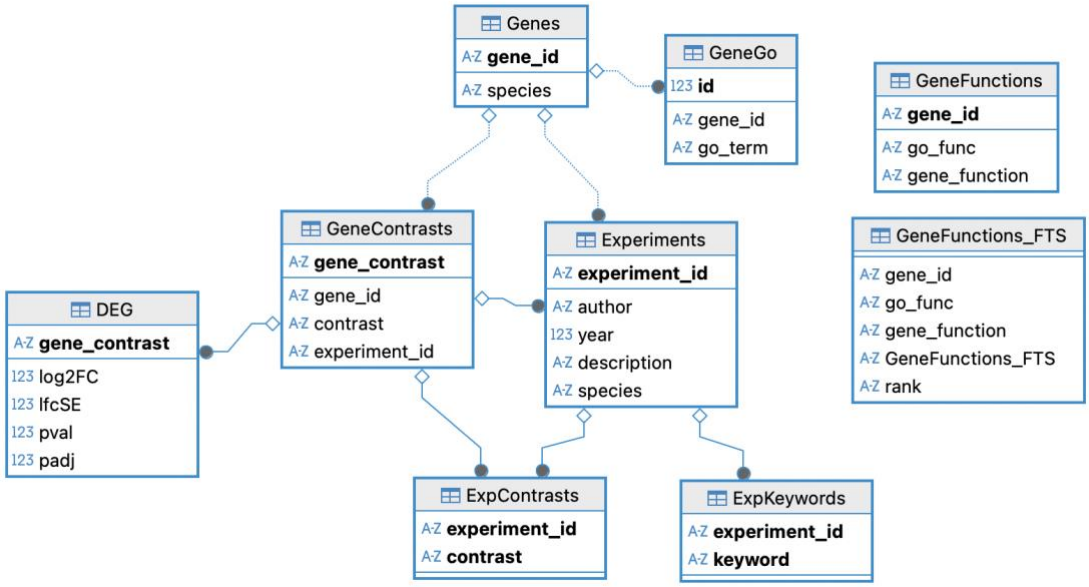

Figure 2: Database schema showing tables and relationships between tables. Bolded names indicate primary keys, whilst connecting lines highlight relationships between tables through foreign keys.

The repository frontend is built using the Shiny package for R (Chang et al., 2024; R Core Team, 2024). This framework was selected due to the extensive documentation available, customisable visualisation, and clear syntax for building an interactive user interface. The R Shiny application connects to the SQLite database to access stored data and displays it in the interface in response to user

inputs. Key packages required by the repository for tasks such as interactivity, data manipulation, layout and plotting are listed in Table 2.

The frontend consists of five tabs. The ‘Search’ tab (Figure 3A), where based on the user input of gene ID, gene function, or experimental condition, the SQLite database is queried and returns a table of genes or experiments matching the user query in the ‘Results’ tab (Figure 3B). This tab offers options to filter the table based on fungal species, experimental condition, or year. Gene functional annotation is set as a hyperlink that brings the user to the gene ontology (GO) web page for that gene description. Experimental contrasts are also set as hyperlinks which navigate the user to the “Experiments” tab. This tab displays a table containing all the differentially expressed genes in the contrast with their associated functional annotation, $\log_2$-fold change, p-value and p-adjusted value (Figure 4A). Options are presented to the user to filter the table based on gene ID, functional annotation, p-value, p-adjusted value, and $\log_2$-fold change. Sub-tabs at the top allow the user to switch between ‘Data Table’, ‘Volcano Plot’, and ‘Expression Heatmap’. The ‘Volcano Plot’ subtab (Figure 4B) displays an interactive volcano plot of $\log_2$-fold change against $-\log_{10}$ p-value, dynamically reflecting changes to p-value and $\log_2$-fold change thresholds set by the user. The ‘Expression Heatmap’ subtab (Figure 4C) displays a heatmap showing $\log_2$-fold change for a maximum of 30 selected genes across all contrasts associated with the experiment. For both the volcano plot and heatmap, hovering over a data point displays the gene ID and functional annotation, and clicking on the data point opens the ‘Gene Info’ tab (Figure 3C). Here, a table is displayed showing the GO terms and functional annotation associated with the selected gene. The GO terms are set as hyperlinks, bringing the user to the respective GO web page. The ‘Upload’ tab (Figure 5) provides an interface for the user to upload their own datasets to the repository. This tab consists of two sections. In the first section, the user can upload DE datasets. It also contains input fields where the user can manually enter author name, year, fungal species, study description, and experimental conditions. In the second section, the user can upload functional annotation data. This section contains radio buttons where the user can specify if the functional annotation dataset contains GO terms.

Two python scripts, `populate_genes.py` and `populate_annot.py`, populate the database with DE datasets and gene functional annotations. The library `argparse` is used to specify arguments containing the following information: author name, study year, fungal species, study description, and experimental conditions. These scripts parse through CSV or TXT files containing DE and gene functional annotation data and enter each unique gene into the database. The files are called from the frontend 'Upload' tab, providing the user with an easy and rapid method to upload their own datasets to the repository.

The modular architecture of the repository, with a separate frontend and backend database, allows specialised programming languages to be used for the different components (Figure 6). This approach enables the storage and rapid access of large amounts of data, whilst simultaneously maintaining high responsiveness and interactivity with the user. Furthermore, the separation between the frontend and the backend allows additional datasets to easily be added to the repository without affecting functionality. The source code for the repository is available at https://github.com/tanyastead/fungal-repository, which includes a detailed user manual and instruction for installation and running of the repository.

Table 2: R packages and versions required by the repository.

Package	Version	Reference
bsicons	0.1.2	Sievert (2023)
bslib	0.8.0	Sievert et al (2024)
DBI	1.2.3	Wickham & Müller (2024)
dplyr	1.1.4	Wickham et al (2023)
DT	0.33	Xie et al (2024)
ggplot2	3.5.1	Wickham (2016)
gridExtra	2.3	Auguie (2017)
jsonlite	1.8.9	Ooms (2014)
purr	1.0.2	Wickham & Henry (2023)
RSQLite	2.3.9	Müller et al (2024)
reticulate	1.40.0	Ushey et al (2024)
shiny	1.10.0	Chang et al (2024)
shinyFeedback	0.4.0	Merlino & Howard (2021)
shinydashboard	0.7.3	Chang & Borges Ribeiro (2025)
shinyjs	2.1.0	Attali (2021)
stringr	1.5.1	Wickham (2023)
tippy	0.1.0	Coene (2021)
xml2	1.3.6	Wickham et al (2023)
plotly	4.10.4	Sievert (2020)
rlist	0.4.6.2	Ren (2021)

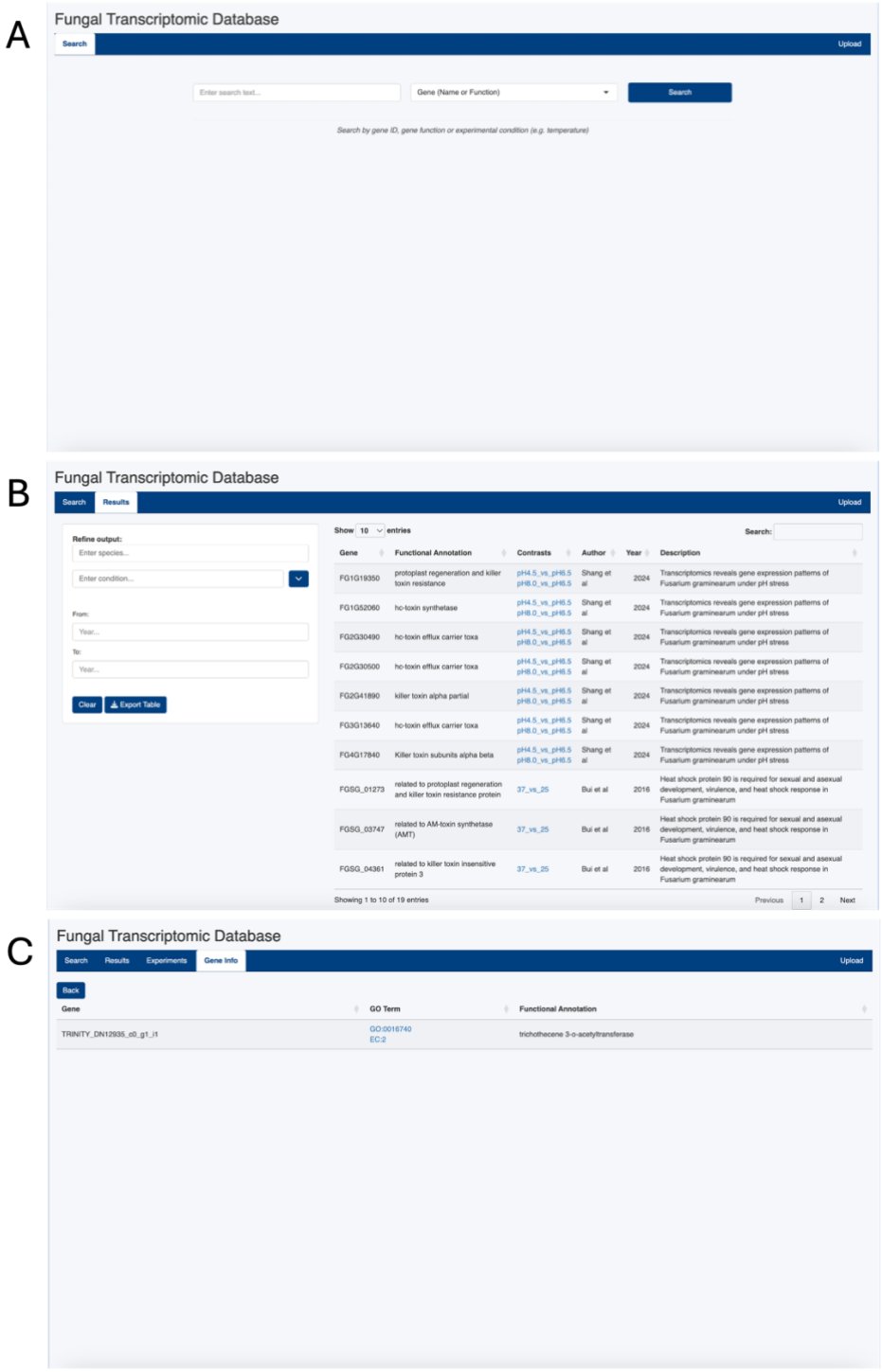

Figure 3: Screenshots of the different sections of the frontend. These include the initial 'Search' tab (A) where the user can search based on gene ID, gene function, or experimental condition; the 'Results' tab (B), which displays a table with the results from the initial search; and the 'Gene Info' tab which displays gene ontology terms and functional annotation for selected genes.

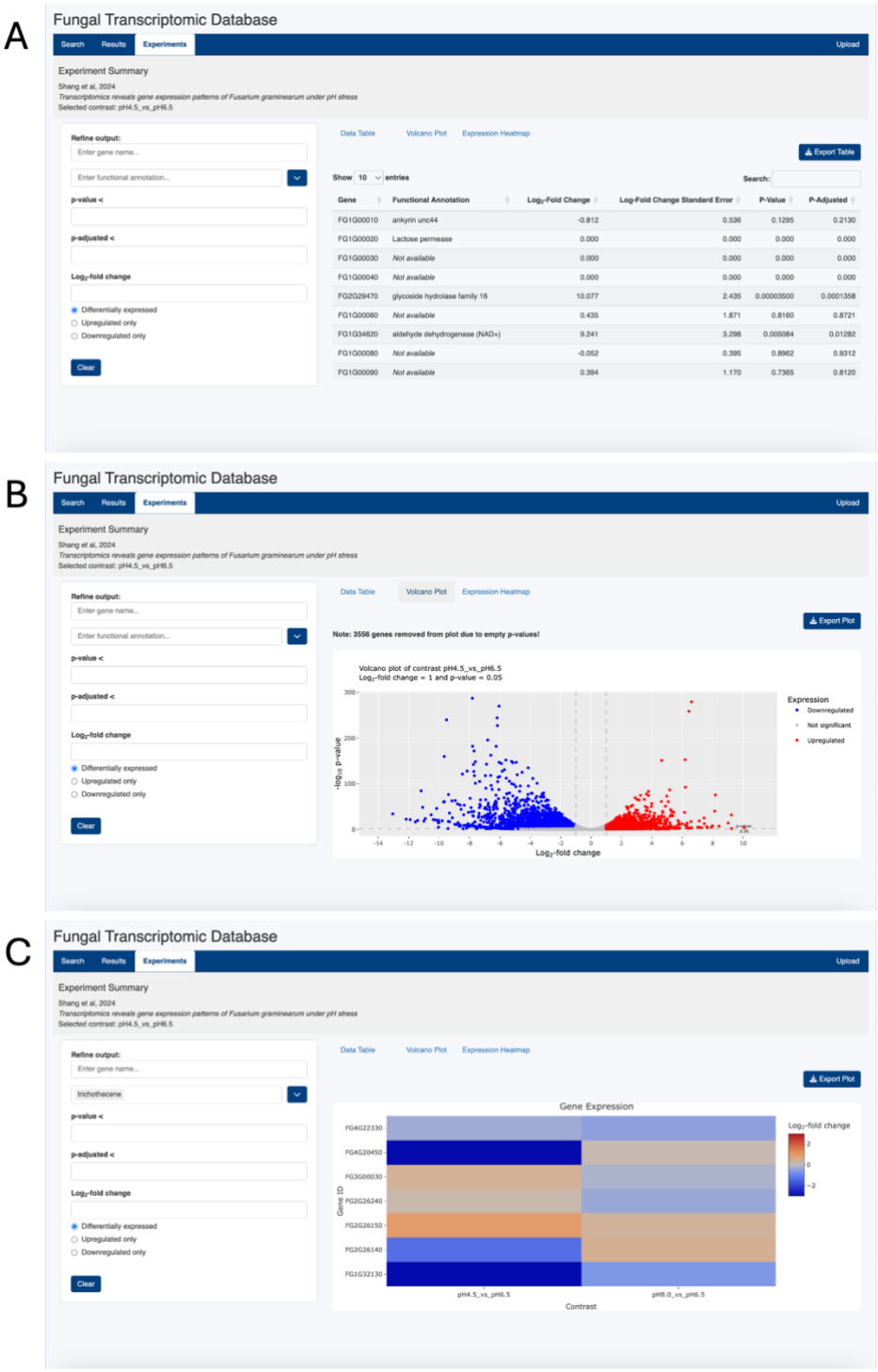

Figure 4: Screenshots showing the different options for data visualisations in the Experiments tab. These include a table displaying all differentially expressed genes for the selected contrast (A), a volcano plot of log_2-fold change against -log_{10} p-value (B), and a heatmap showing differential expression for selected genes against all contrasts investigated in the study (C).

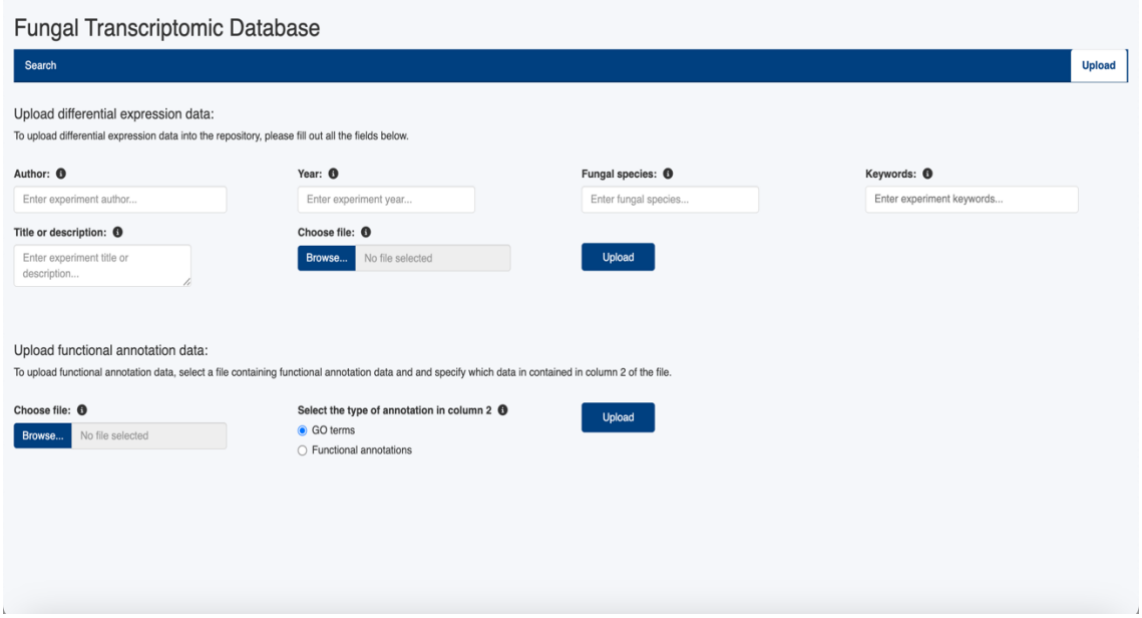

Figure 5: Screenshot showing the Upload tab, where users can upload their own datasets to the repository.

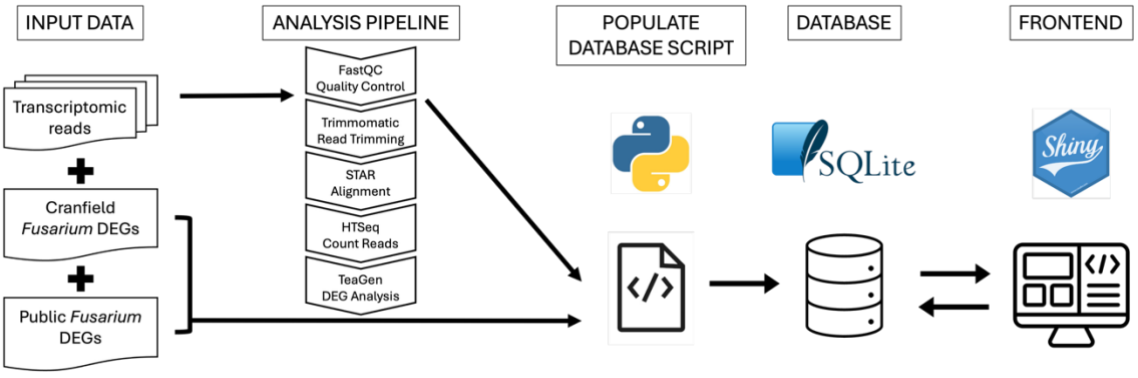

Figure 6: Schematic overview of repository construction. Input data consisting of DEGs, is inputted into the SQLite database via a python script. Input data consisting of raw transcriptomic reads are put through an analysis pipeline to generate DE data, after which they are inputted into the SQLite database via a python script. The user interface is built as an R Shiny application, which calls the SQLite database to access DE data and present it to the user, based on user inputs.

3.2 Input data

The SQLite database requires DE data as input in the form of DEGs. These are calculated from count data using software packages such as DESeq2 and edgeR (Chen et al., 2025; Love et al., 2014). The main metrics in DE datasets include log_2-fold change, p-value, and p-adjusted value, which are required by the repository to generate volcano plots and heatmaps. Count data for DE calculations is generated from aligning raw transcriptomic reads to a reference genome and calculating coverage, using tools such as HTSeq (Anders et al., 2015).

Input data was either generated at Cranfield University, downloaded directly from public repositories such as the Gene Expression Omnibus (Edgar, 2002), or calculated from raw transcriptomic reads downloaded from NCBI (Sayers et al., 2025). DE data included studies by Goodisman (2023) which analysed the effect of temperature, water activity and CO_2 concentration on *Fusarium langsethiae* gene expression; by Shang et al (2024), which investigated the effect of pH on gene expression in *Fusarium graminearum*; and DeIulio et al (2018), which examined kinomes and the effect of heat shock in *Fusarium oxysporum*, and was supplemented with gene functional annotations from Ma et al (2010).

Datasets consisting of count data included the study by Bui et al (2016), which investigated heat shock response in *Fusarium graminearum*. To calculate DEGs, the count data was passed into TeaGen (Tolic et al., 2024), which outputted DE data.

Raw transcriptomic reads included the study by Ding et al (2024), which analysed the transcriptome of *Fusarium verticilliodes* in response to temperature and water activity. To calculate DE for this dataset, reads underwent quality control and trimming using FastQC (Andrews, n.d.) and TRIMMOMATIC (Bolger et al., 2014). Reads were aligned using STAR (Dobin et al., 2013) against the *F. verticillioides* strain 7600 reference genome (assembly ASM14955v1,Syngenta Biotechnology, Inc., 2011) and counted using HTSeq (Anders et al., 2015). Finally, counts were inputted into TeaGen (Tolic et al., 2024) to generate DE data, and subsequently

entered into the repository. The input data used to populate the repository is summarised in Table 3.

Table 3: Type and description of studies entered into the repository.

Data Type	***Fungal Species***	***Experimental Conditions***	***Nº of Contrasts***	***Reference***
DEGs	*F. langsethiae*	Temperature CO_2 Water activity	27	Goodisman (2023)
DEGs	*F. graminearum*	pH	2	Shang et al (2024)
DEGs	*F. oxysporum* f. sp *lycopersici* 4287	Temperature	1	DeIulio et al (2018)
Count data	*F. graminearum* Z-3639	Temperature	1	Bui et al (2016)
Transcriptomic reads	*F. verticillioides*	Temperature Water activity	6	Ding et al (2024)

3.3 Case studies

The accuracy of the repository was demonstrated by two case studies, which recreated gene expression findings identified in the original paper and generated novel visualisations of DE in mycotoxin-related genes. Case Study A reproduced the expression heatmap displaying expression of trichothecene biosynthesis genes at elevated temperature, CO_2 and water activity conditions as reported by Goodisman (2023). In contrast, Case Study B re-analysed the dataset generated by Shang et al (2024) to identify and visualise DE of trichothecene-related genes at varying pH.

4 Results

4.1 Repository overview

The Fungal Transcriptomic Database is a repository holding fungal gene expression data. In this study, it was configured to store data from five studies, encompassing four fungal species, *F. graminearum, F. langsethiae, F. oxysporum, F. verticillioides*, across four experimental conditions: temperature, CO_2 concentration, water activity, and pH. The repository offers a simple and intuitive interface for the user to search through datasets and identify experiments and DEGs depending on gene ID, gene function, or experimental keyword. The database also offers several visual representations of the data, including tables displaying DE, volcano plots of log_2-fold change against p-value, and heatmaps of DEGs across different contrasts. In addition, gene functional annotation is clearly listed and functions as hyperlinks to their respective GO web pages. Lastly, the repository provides an easy and rapid way for users to upload their own data via the frontend interface.

4.2 Case Study A: reproducing trichothecene gene expression

Figure 7A shows a screenshot of the 'Experiments' page displaying an expression heatmap for 14 TRI genes (Table B-1) identified in Goodisman (2023), whilst Figure 7B shows the downloaded heatmap modified to show changing factors above each contrast group. When only temperature was changed, TRI15 (TRINITY_DN25107_c0_g1_i1) was consistently upregulated across all temperatures, water activities (a_w) and CO_2 conditions, whilst the other genes showed a general upregulation at 0.995 a_w and downregulation at 0.980 a_w. In the a_w alone condition, at low a_w there was an overall trend of upregulation at 25 ^oC and downregulation at 30 ^oC. When solely changing CO_2 concentration, gene expression was generally upregulated across all conditions, except for at 25 ^oC and 0.9980 a_w.

When temperature and a_w were both changed, TRI15 (TRINITY_DN25107_c0_g1_i1), TRI6 (TRINITY_DN43694_c0_g1_i1) and TRI10 (TRINITY_DN17003_c3_g1_i1) stood out as showing high log_2-fold

change at high temperature (30 ºC) and low a_w (0.980 a_w), with an average value of 1.706. This is in contrast to other TRI genes which showed a general trend of downregulation under these conditions. Changing both temperature and CO_2 resulted in overall upregulation when a_w was at 0.995, whilst varying both CO_2 and a_w resulted in a general gene upregulation at low aw and high CO_2. The exception to this was TRI15 (TRINITY_DN25107_c0_g1_i1), which was downregulated at 30 ºC, with a log_2-fold change of -1.066. When all three factors were changed between contrasts, no overall trend across TRI genes could be identified.

Figure 7: Heatmap of TRI gene expression across 27 contrasts. (A) Screenshot of the heatmap displayed in the repository, with the sidepanel displaying how the dataset has been refined by gene ID. (B) Enlarged view of the heatmap, modified to display changing factors above each group, with T denoting temperature, a_w denoting water activity, and CO_2 denoting carbon dioxide concentration. Displayed datasets were generated by Goodisman (2023).

4.3 Case Study B: trichothecene gene expression under variable pH

Figure 8 shows a gene expression heatmap for seven trichothecene-related *F. graminearum* genes across varying pH levels, from the dataset generated by Shang et al (2024). The genes were identified by searching for the keyword 'trichothecene', which refined the dataset to seven genes involved in trichothecene biosynthesis and regulation (Table B-2).

At both low and high pH, there was a general trend of slight gene downregulation, with the exceptions of FG4G20450 and FG1G32130, which were significantly downregulated at low pH, showing log_2-fold change values of -9.118 and -6.753 respectively. Overall, gene expression showed little variation at pH 8.0 compared to pH 6.5, with an average absolute log_2-fold change of 0.388. In contrast, at pH 4.5 four genes had an absolute log_2-fold change of over 0.5.

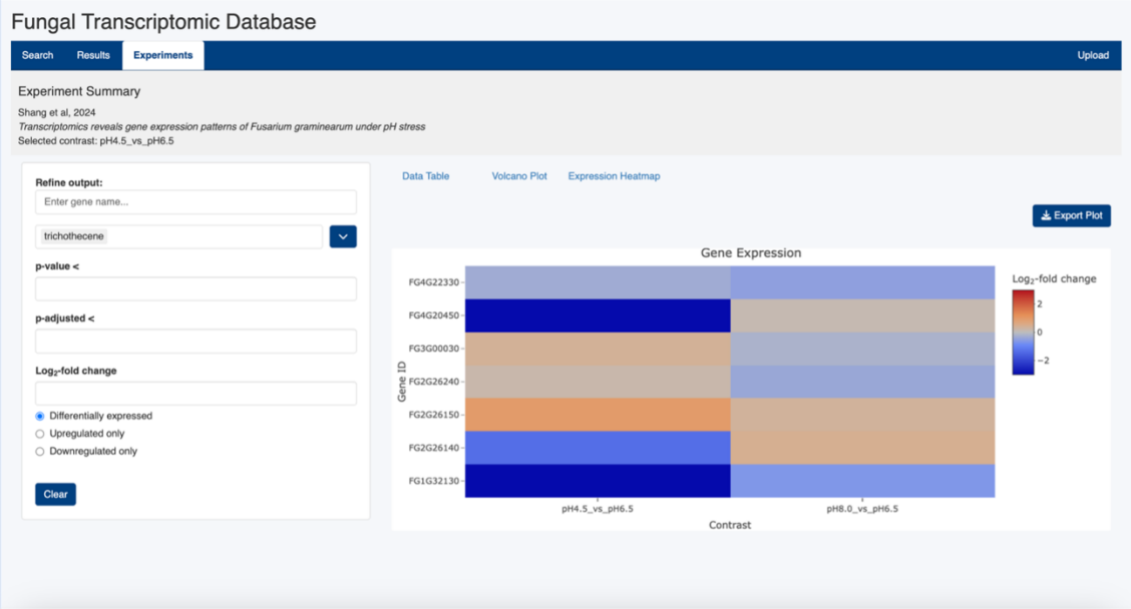

Figure 8: Screenshot of the repository displaying gene expression of trichothecene-related genes under varying pH. Genes were identified by searching for 'trichothecene' functional annotation. Displayed datasets were generated by Shang et al (2024).

5 Discussion

5.1 Usability and utility

The structure of the Fungal Transcriptomic Database is organised into five main tabs: Search, Results, Experiments, Gene Info, and Upload. Each tab consists of a clear and simplistic interface and presents the user with several options to explore and refine the view. This straightforward design provides a user-friendly experience and highlights experimental conditions and contrasts. The FTD takes as input DE data, which is both compact and accessible. DE data is often publicly available in online databases, or easy to generate from transcriptomic reads, ensuring the repository can rapidly and efficiently be populated with multiple studies. Furthermore, since the time-consuming and computationally heavy analyses are conducted in the steps prior to the generation of DE data, the FTD does not require much computing power, resulting in a lightweight application that is highly responsive to user commands. These features create an easy-to-navigate exploratory tool to investigate gene expression across conditions of interest. The FTD can therefore be easily applied to the study of mycotoxin biosynthesis under different climatic conditions, helping visualise changes in gene expression in response to complex and variable external stimuli and contributing to the generation of new hypotheses regarding fungal behaviour.

The construction of the FTD followed a modular design, with the repository split into two main components: the SQLite database backend storing DEGs and the R Shiny frontend interface. This approach enables each component to act independently, allowing for specialised programming languages to optimise speed and efficiency. Furthermore, this architecture ensures each component can be modified or upgraded without affecting the entire system. For example, the frontend can easily be upgraded from an R Shiny application to a more sophisticated web development platform such as Angular without rewriting the backend. Lastly, the repository does not contain hardcoded variable names, such as fungal species, supporting future additions of other mycotoxin-producing fungal genera.

Storing DE data in an SQLite database improved overall performance by enabling rapid retrieval and visualisation of data. This results in high responsiveness and interactivity, as data is accessed and presented within seconds. However, large datasets containing many contrasts strained the plot generating system in the Experiments tab, introducing several seconds of lag when users interacted with the data or changed view. Although inconvenient, this delay does not prevent investigation of gene expression. While repository performance was not formally evaluated through stress testing, the current observed speed and responsiveness are sufficient for exploratory analysis. Nevertheless, these observations indicate future optimisations may be required to ensure the repository can support the addition of larger datasets.

Case Study A validated the repository's functionality by accurately recreating the DE analysis conducted by Goodisman (2023). This was determined by refining the dataset to display only the 14 trichothecene genes and navigating to the expression heatmap. The displayed heatmap was identical to the one generated in the original study, with the same patterns of gene expression clearly visible. From this, it can be determined that the repository accurately displays data stored in the database, and that the data is not altered or transformed during generation of the graphs. Furthermore, data visualisation via graphs, such as heatmaps or volcano plots, is rapid and done automatically, meaning the researcher does not need to spend time or have the skills to analyse large datasets manually. Overall, this demonstrates that the FTD presents an easy and accurate method for re-analysing datasets.

The advantage of the repository as an exploratory tool was demonstrated in Case Study B. This study initially investigated metabolic responses to pH stress, however by using the repository to search for trichothecene-related genes, the dataset was instantly specialised for mycotoxin research. Although the generated heatmap displayed an overall pattern of little differential gene expression, three genes were strongly downregulated at pH 4.5 compared to pH 6.5. The functional annotation for these genes includes trichothecene c-15 hydroxylase, trichothecene c-15 esterase and tri7-trichothecene biosynthesis, indicating that

all three genes are directly involved in the trichothecene biosynthesis pathway. Therefore, their strong downregulation at low pH is surprising and contradictory to previous studies which show that acidic media induce *TRI* gene expression and increases trichothecene production (Gardiner et al., 2009; Merhej et al., 2010, 2011). This result is unlikely to be a repository artefact, since gene expression at alkaline conditions correlates with literature (Merhej et al., 2010). Instead, it could arise from experimental factors such as incubation time or fungal media. Nevertheless, this case study highlights the ease and speed at which the FTD can analyse mycotoxin gene expression in unrelated studies.

5.2 FTD and other fungal transcriptomic databases

While existing repositories, such as Fungi.guru, FungiExp and FungiDB, also provide access to fungal transcriptomic data, they are limited in their search capabilities, restricting searches to only gene ID or GO term. In contrast, the FTD has a more generalised search feature, allowing the user to search by gene function or experiment keyword in addition to gene ID. This broader search capability enhances FTD's flexibility and exploratory power by allowing users to investigate datasets without prior knowledge of specific gene IDs or GO terms. This exploratory functionality is further supported by the clear visualisation of experimental contrasts and gene functional annotation alongside gene IDs, features that are either absent or obscured in existing databases. By clearly displaying these features in FTD, gene expression can be easily and rapidly linked to function or condition.

A unique feature of the FTD is the visualisation of data through interactive heatmaps and volcano plots. In contrast, existing repositories display static plots, with data primarily displayed as bar plots instead of heatmaps. Bar plots limit the volume of data that can be simultaneously displayed, requiring users to use more specific search terminology. Furthermore, the volcano plot is a distinctive feature of the FTD, providing a means for users to rapidly understand the quality and spread of DE in the dataset.

Lastly, a key distinctive functionality of the Fungal Transcriptomic Database is its ability for users to easily upload their own datasets to the repository. This feature

extends the utility of FTD, enabling user to explore gene expression within their own datasets, and compare results with previously published data. Novel findings can therefore be easily and instantaneously analysed in the context of other studies, supporting the generation and validation of new hypotheses based on gene expression from previous experiments.

Despite containing several unique and advantageous features, the FTD lacks important functionalities necessary for in-depth analysis of gene expression. In particular, FTD does not offer co-expression or cluster enrichment analyses, features that are preset in FungiExp and Fungi.guru. These analyses can link metabolic processes based on patterns in differential expression, resulting in a deeper understanding of overall biological system function. In its current state, genes can only be linked by searching for gene description keywords. Furthermore, all three existing repositories contain BLAST functionality, allowing the user to identify related genes or search for gene expression based on sequence similarity. This feature is useful for researchers who are investigating a specific gene or protein. However, since the FTD is designed as an exploratory tool aimed towards researchers who aren't necessarily investigating a specific gene, the BLAST feature is not essential for the FTD.

5.3 Strengths and weaknesses

The main advantage of the repository is the clear and minimalistic interface. By displaying only essential information and features, such as gene descriptions or options to refine the data tables, the interface is understandable at first glance. This straightforward design avoids unnecessary clutter and ensures no data is obscured, aiding interpretation and navigation.

Another strength of the repository is the high interactivity, facilitated by functionalities that allow users to customise displayed tables. For example, users can select which fungal species, experimental conditions, or genes are displayed. The interactivity is further enhanced by the high responsiveness of the repository, achieved by storing DEG data in the SQLite database. The efficient data storage and retrieval offered by SQLite means data is displayed almost instantaneously in response to user commands, leading to a highly interactive interface. Together,

the high responsiveness and customisation of the repository creates the unique exploratory prowess of the tool, making the FTD stand out compared to other databases.

Although the FTD displays numerous advantages, there are certain weaknesses inherent to the repository, primarily relating to the input data. In the current setup, the FTD requires a highly specific format for the input data. As different studies will generate DE data using different tools and software, the format of the DE data is unlikely to match the required format set by the FTD. Therefore, data must be transformed to the correct format prior to upload, a step that limits automation and reduces efficiency. The second weakness relating to the input data is the quality of the functional annotation. For optimal exploratory data analysis, the FTD requires both GO terms and gene descriptions in the functional annotation file. However, many basic analyses only provide gene descriptions, thus inactivating the GO term and gene description hyperlinks. Furthermore, the FTD does not verify functional annotations. Therefore, if the input file contains incorrect annotations, they will be mistakenly represented in the repository, potentially leading to incorrect hypotheses.

The FTD supports visualisation of DE across contrasts within an experiment, enabling rapid comparison of DE across different conditions. However, this feature is restricted to within-experiment comparisons and does not extend across multiple studies, limiting the repository's usefulness for cross-study analysis of gene expression. Nevertheless, such comparisons are challenging due to external factors, such as normalisation techniques or fungal growth media, influencing gene expression. Therefore, this type of visualisation is not a critical analysis for the repository and could thus be considered a limitation rather than a weakness.

Finally, the FTD is currently configured as a local application, requiring the user to download and start the R Shiny application prior to running. This process is time consuming and requires technical knowledge, thus reducing overall efficiency and ease of access. In contrasts, online-hosted repositories offer easy and instant access from any machine. On the other hand, online hosting requires

continuous funding, which could be difficult to justify for such a small database. Furthermore, a local setup has its own benefits. Notably, it provides greater security, ensuring unpublished datasets that have been uploaded to the repository remain private and inaccessible to other users.

5.4 Future development

The FTD can be improved by integrating additional data visualisations. Firstly, the repository can be extended to support visualisation of single cell RNA sequencing (scRNA-seq) data. As technologies improve, the number of studies using scRNA-seq is increasing (Jovic et al., 2022). Therefore, adding this functionality is important to ensure the repository is up-to-date and can support the latest technologies. To achieve this, instead of the heatmap the repository could display an interactive dimensionality reduction scatter plot, such as a uniform manifold approximation and projection (UMAP) to visualise organisation and relationships between cell clusters, when displaying scRNA-seq data (Becht et al., 2019). Although scRNA-seq technology has had limited application in fungal studies, it could nevertheless be useful in understanding mycotoxin production during different fungal states, and so supporting this technology would be beneficial for the repository (Hovhannisyan & Gabaldón, 2018).

Another valuable data visualisation technique that could enhance the repository is co-expression analysis. As discussed earlier, this feature is present in existing fungal databases and so implementing it in the FTD would eliminate this weakness in the repository. Co-expression analyses are powerful techniques that can identify patterns and associations between genes and target traits, and thus are invaluable during exploratory analysis (Langfelder & Horvath, 2008; Wu et al., 2022). Tools such as weighted gene co-expression network analysis (WGCNA) could easily be added to the repository to construct networks on the backend, and plots such as module cluster dendrograms and module-trait heatmaps could be displayed in the Experiments tab alongside the heatmap and volcano plot. These graphs group genes into modules, or clusters, of highly correlated genes, and displays the strength of the associations between module and experimental condition (Langfelder & Horvath, 2008; Wu et al., 2022). In

short, the addition of this analysis would enable users to explore genes most strongly associated with experimental contrasts, providing a deeper insight into fungal biology and responses to external stimuli.

Minor changes could also be made to the repository to improve current data visualisation. Currently, the DE heatmap displays all contrasts investigated in the study. This does not pose a problem when few contrasts are displayed, however when large numbers of contrasts were studied the heatmap became cluttered, reducing readability and interpretability. Therefore, adding functionality for the user to choose which contrasts are displayed would greatly improve clarity.

Finally, speed and performance could be optimised to improve overall user experience. Currently, large datasets with many contrasts cause the repository to lag when generating the interactive plots, due to the large volume of data being queried. To improve query speed, techniques such as indexing and parallel execution could be implemented to increase search efficiency. These optimisations would likely reduce lag and improve responsiveness, resulting in enhanced user experience when exploring large datasets.

6 Conclusions

Although several publicly available databases that display fungal transcriptomic data exist, these are insufficient for the exploratory analysis of gene expression under different climatic conditions. To address this problem, the Fungal Transcriptomic Database was built. It improves upon existing repositories by providing a clear and simple interface, highlighting experimental conditions and gene functions, and creating a user-friendly repository for analysing gene expression. The usefulness of the FTD was demonstrated in the field of mycotoxin research by the re-analysis of trichothecene gene expression from Goodisman (2023). The FTD clearly displayed the gene expression data in tables and graphs, while also rapidly and accurately re-creating the heatmap showing trichothecene gene expression across all experimental contrasts. This highlights the advantages of the FTD as an exploratory tool for re-analysing published datasets, resulting in a greater understanding of gene expression across different conditions.

Despite lacking certain functionalities present in other databases, most notably co-expression analysis, the modular design of the FTD ensures new features can easily be added to the repository. For example, the FTD could be updated to support new technologies, such as visualisation of scRNA-seq data, or extended to host transcriptomic data on additional fungal genera, ensuring a broad and versatile tool. Thus, the FTD can easily be kept up-to-date and compatible with the newest scientific advances.

Overall, the FTD's clear and straightforward design creates a reliable and accessible platform for exploratory analysis of differential expression. The repository provides researchers with the tools to visualise, compare and interpret gene expression data across multiple studies, leading to deeper insights and enabling the development of new hypotheses regarding mycotoxin biosynthesis under future climatic conditions.

References

Ahrendt, S. R., Mondo, S. J., Haridas, S., & Grigoriev, I. V. (2023). MycoCosm, the JGI's Fungal Genome Portal for Comparative Genomic and Multiomics Data Analyses. In F. Martin & S. Uroz (Eds), *Microbial Environmental Genomics (MEG)* (Vol. 2605, pp. 271–291). Springer US. https://doi.org/10.1007/978-1-0716-2871-3_14

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andersen, K. F., Madden, L. V., & Paul, P. A. (2015). Fusarium Head Blight Development and Deoxynivalenol Accumulation in Wheat as Influenced by Post-Anthesis Moisture Patterns. *Phytopathology®*, *105*(2), 210–219. https://doi.org/10.1094/PHYTO-04-14-0104-R

Andrews, S. (n.d.). *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Computer software]. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Attali, D. (2021). *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds* (Version 2.1.0) [R]. https://CRAN.R-project.org/package=shinyjs

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for 'Grid' Graphics* (Version 2.3) [R]. https://CRAN.R-project.org/package=gridExtra

Ayaz, M., Li, C.-H., Ali, Q., Zhao, W., Chi, Y.-K., Shafiq, M., Ali, F., Yu, X.-Y., Yu, Q., Zhao, J.-T., Yu, J.-W., Qi, R.-D., & Huang, W.-K. (2023). Bacterial and Fungal Biocontrol Agents for Plant Disease Protection: Journey from Lab to Field, Current Status, Challenges, and Global Perspectives. *Molecules*, *28*(18), 6735. https://doi.org/10.3390/molecules28186735

Basenko, E. Y., Pulman, J. A., Shanmugasundram, A., Harb, O. S., Crouch, K., Starns, D., Warrenfeltz, S., Aurrecoechea, C., Stoeckert, C. J., Kissinger, J. C., Roos, D. S., & Hertz-Fowler, C. (2018). FungiDB: An Integrated

Bioinformatic Resource for Fungi and Oomycetes. *Journal of Fungi*, *4*(1), 39. https://doi.org/10.3390/jof4010039

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, *37*(1), 38–44. https://doi.org/10.1038/nbt.4314

Bencze, S., Puskás, K., Vida, G., Karsai, I., Balla, K., Komáromi, J., & Veisz, O. (2017). Rising atmospheric CO2 concentration may imply higher risk of Fusarium mycotoxin contamination of wheat grains. *Mycotoxin Research*, *33*(3), 229–236. https://doi.org/10.1007/s12550-017-0281-2

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bui, D.-C., Lee, Y., Lim, J. Y., Fu, M., Kim, J.-C., Choi, G. J., Son, H., & Lee, Y.-W. (2016). Heat shock protein 90 is required for sexual and asexual development, virulence, and heat shock response in Fusarium graminearum. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep28154

Casu, A., Camardo Leggieri, M., Toscano, P., & Battilani, P. (2024). Changing climate, shifting mycotoxins: A comprehensive review of climate change impact on mycotoxin contamination. *Comprehensive Reviews in Food Science and Food Safety*, *23*(2), e13323. https://doi.org/10.1111/1541-4337.13323

Chang, W., & Borges Ribeiro, B. (2025). *shinydashboard: Create Dashboards with 'Shiny'* (Version 0.7.3) [R]. https://CRAN.R-project.org/package=shinydashboard

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *shiny: Web Application Framework for R* (Version 1.10.0) [R]. https://CRAN.R-project.org/package=shiny

Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L., & Smyth, G. K. (2025). edgeR v4: Powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Research*, *53*(2), gkaf018. https://doi.org/10.1093/nar/gkaf018

Coene, J. (2021). *tippy: Add Tooltips to 'R markdown' Documents or 'Shiny' Apps* (Version 0.1.0) [R]. https://CRAN.R-project.org/package=tippy

DeIulio, G. A., Guo, L., Zhang, Y., Goldberg, J. M., Kistler, H. C., & Ma, L.-J. (2018). Kinome Expansion in the Fusarium oxysporum Species Complex Driven by Accessory Chromosomes. *mSphere*, *3*(3). https://doi.org/10.1128/msphere.00231-18

Ding, Y., Ma, N., Haseeb, H. A., Dai, Z., Zhang, J., & Guo, W. (2024). Genome-wide transcriptome analysis of toxigenic Fusarium verticillioides in response to variation of temperature and water activity on maize kernels. *International Journal of Food Microbiology*, *410*, 110494. https://doi.org/10.1016/j.ijfoodmicro.2023.110494

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dusenge, M. E., Duarte, A. G., & Way, D. A. (2019). Plant carbon metabolism and climate change: Elevated CO_2 and temperature impacts on photosynthesis, photorespiration and respiration. *New Phytologist*, *221*(1), 32–49. https://doi.org/10.1111/nph.15283

Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, *30*(1), 207–210. https://doi.org/10.1093/nar/30.1.207

El-Sayed, R. A., Jebur, A. B., Kang, W., & El-Demerdash, F. M. (2022). An overview on the major mycotoxins in food products: Characteristics,

toxicity, and analysis. *Journal of Future Foods*, *2*(2), 91–102. https://doi.org/10.1016/j.jfutfo.2022.03.002

Food and Agriculture Organization of the United Nations. (2009). *How to Feed the World in 2050*. Food and Agriculture Organization of the United Nations. https://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Gardiner, D. M., Osborne, S., Kazan, K., & Manners, J. M. (2009). Low pH regulates the production of deoxynivalenol by Fusarium graminearum. *Microbiology*, *155*(9), 3149–3156. https://doi.org/10.1099/mic.0.029546-0

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 Genes. *Science*, *274*(5287), 546–567. https://doi.org/10.1126/science.274.5287.546

Goodisman, T. (2023). *Unravelling Fusarium langsethiae secondary metabolites potential: From transcriptomics to genomics* [MSc Thesis]. Cranfield University.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., & Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*(D1), D699–D704. https://doi.org/10.1093/nar/gkt1183

Hipp, D. R. (2023). *SQLite* (Version 3.43.2) [SQL]. SQLite Consortium. https://www.sqlite.org/index.html

Hovhannisyan, H., & Gabaldón, T. (2018). Transcriptome Sequencing Approaches to Elucidate Host–Microbe Interactions in Opportunistic Human Fungal Pathogens. In M. L. Rodrigues (Ed.), *Fungal Physiology*

and Immunopathogenesis (Vol. 422, pp. 193–235). Springer International Publishing. https://doi.org/10.1007/82_2018_122

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., & Zimmermann, P. (2008). Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Advances in Bioinformatics*, *2008*(1), 420747. https://doi.org/10.1155/2008/420747

Janik, E., Niemcewicz, M., Podogrocki, M., Ceremuga, M., Stela, M., & Bijak, M. (2021). T-2 Toxin—The Most Toxic Trichothecene Mycotoxin: Metabolism, Toxicity, and Decontamination Strategies. *Molecules*, *26*(22), 6868. https://doi.org/10.3390/molecules26226868

Jeger, M., Beresford, R., Bock, C., Brown, N., Fox, A., Newton, A., Vicent, A., Xu, X., & Yuen, J. (2021). Global challenges facing plant pathology: Multidisciplinary approaches to meet the food security and environmental challenges in the mid-twenty-first century. *CABI Agriculture and Bioscience*, *2*(1), 20. https://doi.org/10.1186/s43170-021-00042-x

Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., & Luo, Y. (2022). Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine*, *12*(3), e694. https://doi.org/10.1002/ctm2.694

Khodaei, D., Javanmardi, F., & Khaneghah, A. M. (2021). The global overview of the occurrence of mycotoxins in cereals: A three-year survey. *Current Opinion in Food Science*, *39*, 36–42. https://doi.org/10.1016/j.cofs.2020.12.012

Kimura, M., Tokai, T., Takahashi-Ando, N., Ohsato, S., & Fujimura, M. (2007). Molecular and Genetic Studies of *Fusarium* Trichothecene Biosynthesis: Pathways, Genes, and Evolution. *Bioscience, Biotechnology, and Biochemistry*, *71*(9), 2105–2123. https://doi.org/10.1271/bbb.70183

Kumar, D., & Kalita, P. (2017). Reducing Postharvest Losses during Storage of Grain Crops to Strengthen Food Security in Developing Countries. *Foods*, *6*(1), 8. https://doi.org/10.3390/foods6010008

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Laraba, I., McCormick, S. P., Vaughan, M. M., Geiser, D. M., & O'Donnell, K. (2021). Phylogenetic diversity, trichothecene potential, and pathogenicity within Fusarium sambucinum species complex. *PLOS ONE*, *16*(1), e0245037. https://doi.org/10.1371/journal.pone.0245037

Lim, J. J. J., Koh, J., Moo, J. R., Villanueva, E. M. F., Putri, D. A., Lim, Y. S., Seetoh, W. S., Mulupuri, S., Ng, J. W. Z., Nguyen, N. L. U., Reji, R., Foo, H., Zhao, M. X., Chan, T. L., Rodrigues, E. E., Kairon, R. S., Hee, K. M., Chee, N. C., Low, A. D., … Mutwil, M. (2020). Fungi.guru: Comparative genomic and transcriptomic resource for the fungi kingdom. *Computational and Structural Biotechnology Journal*, *18*, 3788–3795. https://doi.org/10.1016/j.csbj.2020.11.019

Liu, J., Zhang, Y., Shi, Y., Zheng, Y., Zhu, Y., Guan, Z., Shen, D., & Dou, D. (2023). FungiExp: A user-friendly database and analysis platform for exploring fungal gene expression and alternative splicing. *Bioinformatics*, *39*(1), btad042. https://doi.org/10.1093/bioinformatics/btad042

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luo, S., Du, H., Kebede, H., Liu, Y., & Xing, F. (2021). Contamination status of major mycotoxins in agricultural product and food stuff in Europe. *Food Control*, *127*, 108120. https://doi.org/10.1016/j.foodcont.2021.108120

Ma, L.-J., Geiser, D. M., Proctor, R. H., Rooney, A. P., O'Donnell, K., Trail, F., Gardiner, D. M., Manners, J. M., & Kazan, K. (2013). *Fusarium*

Pathogenomics. *Annual Review of Microbiology*, *67*(1), 399–416. https://doi.org/10.1146/annurev-micro-092412-155650

Ma, L.-J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P. M., Kang, S., Shim, W.-B., Woloshuk, C., Xie, X., Xu, J.-R., Antoniw, J., Baker, S. E., Bluhm, B. H., … Rep, M. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature*, *464*(7287), 367–373. https://doi.org/10.1038/nature08850

McCormick, S. P., Stanley, A. M., Stover, N. A., & Alexander, N. J. (2011). Trichothecenes: From Simple to Complex Mycotoxins. *Toxins*, *3*(7), 802–814. https://doi.org/10.3390/toxins3070802

Medina, A., Gilbert, M. K., Mack, B. M., OBrian, G. R., Rodríguez, A., Bhatnagar, D., Payne, G., & Magan, N. (2017). Interactions between water activity and temperature on the Aspergillus flavus transcriptome and aflatoxin B 1 production. *International Journal of Food Microbiology*, *256*, 36–44. https://doi.org/10.1016/j.ijfoodmicro.2017.05.020

Meneely, J., Greer, B., Kolawole, O., & Elliott, C. (2023). T-2 and HT-2 Toxins: Toxicity, Occurrence and Analysis: A Review. *Toxins*, *15*(8), 481. https://doi.org/10.3390/toxins15080481

Merhej, J., Boutigny, A. L., Pinson-Gadais, L., Richard-Forget, F., & Barreau, C. (2010). Acidic pH as a determinant of *TRI* gene expression and trichothecene B biosynthesis in *Fusarium graminearum*. *Food Additives & Contaminants: Part A*, *27*(5), 710–717. https://doi.org/10.1080/19440040903514531

Merhej, J., Richard-Forget, F., & Barreau, C. (2011). The pH regulatory factor Pac1 regulates Tri gene expression and trichothecene production in Fusarium graminearum. *Fungal Genetics and Biology*, *48*(3), 275–284. https://doi.org/10.1016/j.fgb.2010.11.008

Merlino, A., & Howard, P. (2021). *shinyFeedback: Display User Feedback in Shiny Apps* (Version 0.4.0) [R]. https://CRAN.R-project.org/package=shinyFeedback

Moretti, A., Pascale, M., & Logrieco, A. F. (2019). Mycotoxin risks under a climate change scenario in Europe. *Trends in Food Science & Technology*, *84*, 38–40. https://doi.org/10.1016/j.tifs.2018.03.008

Mousavi Khaneghah, A., Fakhri, Y., Gahruie, H. H., Niakousari, M., & Sant'Ana, A. S. (2019). Mycotoxins in cereal-based products during 24 years (1983–2017): A global systematic review. *Trends in Food Science & Technology*, *91*, 95–105. https://doi.org/10.1016/j.tifs.2019.06.007

Müller, K., Wickham, H., James, D., & Falcon, S. (2024). *RSQLite: SQLite Interface for R* (Version 2.3.9) [R]. https://CRAN.R-project.org/package=RSQLite

Ooms, J. (2014). *The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects.* [R]. https://arxiv.org/abs/1403.2805

Parikka, P., Hakala, K., & Tiilikkala, K. (2012). Expected shifts in *Fusarium* species' composition on cereal grain in Northern Europe due to climatic change. *Food Additives & Contaminants: Part A*, *29*(10), 1543–1555. https://doi.org/10.1080/19440049.2012.680613

Pereira, V. L., Fernandes, J. O., & Cunha, S. C. (2014). Mycotoxins in cereals and related foodstuffs: A review on occurrence and recent methods of analysis. *Trends in Food Science & Technology*, *36*(2), 96–136. https://doi.org/10.1016/j.tifs.2014.01.005

Placinta, C. M., D'Mello, J. P. F., & Macdonald, A. M. C. (1999). A review of worldwide contamination of cereal grains and animal feed with Fusarium mycotoxins. *Animal Feed Science and Technology*, *78*(1–2), 21–37. https://doi.org/10.1016/S0377-8401(98)00278-8

R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (Version 4.4.1) [R]. R Foundation for Statistical Computing. https://www.R-project.org/

Rampersad, S. (2020). Pathogenomics and Management of Fusarium Diseases in Plants. *Pathogens*, *9*(5), 340. https://doi.org/10.3390/pathogens9050340

Rauwane, M. E., Ogugua, U. V., Kalu, C. M., Ledwaba, L. K., Woldesemayat, A. A., & Ntushelo, K. (2020). Pathogenicity and Virulence Factors of Fusarium graminearum Including Factors Discovered Using Next Generation Sequencing Technologies and Proteomics. *Microorganisms*, *8*(2), 305. https://doi.org/10.3390/microorganisms8020305

Ren, K. (2021). *rlist: A Toolbox for Non-Tabular Data Manipulation* (Version 0.4.6.2) [R]. https://CRAN.R-project.org/package=rlist

Ristaino, J. B., Anderson, P. K., Bebber, D. P., Brauman, K. A., Cunniffe, N. J., Fedoroff, N. V., Finegold, C., Garrett, K. A., Gilligan, C. A., Jones, C. M., Martin, M. D., MacDonald, G. K., Neenan, P., Records, A., Schmale, D. G., Tateosian, L., & Wei, Q. (2021). The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences*, *118*(23), e2022239118. https://doi.org/10.1073/pnas.2022239118

Sasaki, E., Takahashi, C., Asami, T., & Shimada, Y. (2011). AtCAST, a Tool for Exploring Gene Expression Similarities among DNA Microarray Experiments Using Networks. *Plant and Cell Physiology*, *52*(1), 169–180. https://doi.org/10.1093/pcp/pcq185

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, *3*(3), 430–439. https://doi.org/10.1038/s41559-018-0793-y

Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Connor, R., Feldgarden, M., Fine, A. M., Funk, K., Hoffman, J., Kannan, S., Kelly, C., Klimke, W., Kim, S., Lathrop, S., Marchler-Bauer, A., Murphy, T. D., O'Sullivan, C., Schmieder, E., … Pruitt, K. D. (2025). Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, *53*(D1), D20–D29. https://doi.org/10.1093/nar/gkae979

Shang, Y., Wang, Y., Wang, G., Li, H., Li, P., Guo, J., & Zhang, C. (2024). Transcriptomics reveals gene expression patterns of *Fusarium graminearum* under pH stress. *Acta Microbiologica Sinica*, *64*(11), 4271–4289. https://doi.org/10.13343/j.cnki.wsxb.20240302

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny* [R]. https://plotly-r.com

Sievert, C. (2023). *bsicons: Easily Work with 'Bootstrap' Icons* (Version 0.1.2) [R]. https://CRAN.R-project.org/package=bsicons

Sievert, C., Cheng, J., & Aden-Buie. (2024). *bslib: Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'* (Version 0.8.0) [R]. https://CRAN.R-project.org/package=bslib

Singh, B. K., Delgado-Baquerizo, M., Egidi, E., Guirado, E., Leach, J. E., Liu, H., & Trivedi, P. (2023). Climate change impacts on plant pathogens, food security and paths forward. *Nature Reviews Microbiology*, *21*(10), 640–656. https://doi.org/10.1038/s41579-023-00900-7

Sweeney, M. (1998). Mycotoxin production by Aspergillus, Fusarium and Penicillium species. *International Journal of Food Microbiology*, *43*(3), 141–158. https://doi.org/10.1016/S0168-1605(98)00112-3

Syngenta Biotechnology, Inc. (2011). *Fusarium verticillioides strain 7600 genome assembly (GCF_000149555.1)* [Data set]. NCBI RefSeq. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000149555.1

Tolic, A., Pickering, S., Romanis, M., Daly, N., Blanck, M., & Akkaraju, K. V. K. (2024). *TeaGen—An interactive web application for Tea resources and RNA-Seq analysis* [R].

United Nations, Department of Economic and Social Affairs. (2019). *World Population Prospects 2019: Highlights*. United Nations. https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html

Ushey, K., Allaire, J., & Tang, Y. (2024). *reticulate: Interface to 'Python'* (Version 1.40.0) [R]. https://CRAN.R-project.org/package=reticulate

Váry, Z., Mullins, E., McElwain, J. C., & Doohan, F. M. (2015). The severity of wheat diseases increases when plants and pathogens are acclimatized to elevated carbon dioxide. *Global Change Biology*, *21*(7), 2661–2669. https://doi.org/10.1111/gcb.12899

Verheecke-Vaessen, C., Diez-Gutierrez, L., Renaud, J., Sumarah, M., Medina, A., & Magan, N. (2019). Interacting climate change environmental factors effects on Fusarium langsethiae growth, expression of Tri genes and T-2/HT-2 mycotoxin production on oat-based media and in stored oats. *Fungal Biology*, *123*(8), 618–624. https://doi.org/10.1016/j.funbio.2019.04.008

Verheecke-Vaessen, C., Garcia-Cela, E., Lopez-Prieto, A., Osk Jonsdottir, I., Medina, A., & Magan, N. (2021). Water and temperature relations of Fusarium langsethiae strains and modelling of growth and T-2 and HT-2 mycotoxin production on oat-based matrices. *International Journal of Food Microbiology*, *348*, 109203. https://doi.org/10.1016/j.ijfoodmicro.2021.109203

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., Cumming, M., Kelley, L. A., Sternberg, M. J., Krishnakumar, V., Ferlanti, E., Miller, J., Town, C., Stuerzlinger, W., & Provart, N. J. (2017). ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *The Plant Cell*, *29*(8), 1806–1821. https://doi.org/10.1105/tpc.17.00073

Wang, J., Vanga, S., Saxena, R., Orsat, V., & Raghavan, V. (2018). Effect of Climate Change on the Yield of Cereal Crops: A Review. *Climate*, *6*(2), 41. https://doi.org/10.3390/cli6020041

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* [R]. https://ggplot2.tidyverse.org

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations* (Version 1.5.1) [R]. https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* (Version 1.1.4) [R]. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2023). *purrr: Functional Programming Tools* (Version 1.0.2) [R]. https://CRAN.R-project.org/package=purrr

Wickham, H., Hester, J., & Ooms, J. (2023). *xml2: Parse XML* (Version 1.3.6) [R]. https://CRAN.R-project.org/package=xml2

Wickham, H., & Müller, K. (2024). *DBI : R Database Interface* (Version 1.2.3) [R]. https://CRAN.R-project.org/package=DBI

Wu, Q., Pan, Y.-B., Su, Y., Zou, W., Xu, F., Sun, T., Grisham, M. P., Yang, S., Xu, L., & Que, Y. (2022). WGCNA Identifies a Comprehensive and Dynamic Gene Co-Expression Network That Associates with Smut Resistance in Sugarcane. *International Journal of Molecular Sciences*, *23*(18), 10770. https://doi.org/10.3390/ijms231810770

Xie, Y., Cheng, J., & Tan, X. (2024). *DT: A Wrapper of the JavaScript Library 'DataTables'* (Version 0.33) [R]. https://CRAN.R-project.org/package=DT

Zhang, H., Zhang, F., Yu, Y., Feng, L., Jia, J., Liu, B., Li, B., Guo, H., & Zhai, J. (2020). A Comprehensive Online Database for Exploring ~20,000 Public Arabidopsis RNA-Seq Libraries. *Molecular Plant*, *13*(9), 1231–1233. https://doi.org/10.1016/j.molp.2020.08.001

Zouine, M., Maza, E., Djari, A., Lauvernier, M., Frasse, P., Smouni, A., Pirrello, J., & Bouzayen, M. (2017). TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. *The Plant Journal*, *92*(4), 727–735. https://doi.org/10.1111/tpj.13711

Appendices

Appendix A Ethical approval letter

4 June 2025

Dear Miss Stead ,

Reference: CURES/25518/2025
Project ID: 28885
Title: Development of a central data repository and visualisation framework for Fusarium transcriptomic data

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/25518/2025 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Appendix B Supplementary Tables

B.1 List of supplementary tables

Table B-1: List of *F.langsethiae* TRI genes with corresponding gene IDs, identified in Goodisman (2023). 46

Table B-2: List of trichothecene biosynthesis and regulatory genes, identified by searching for the keyword 'trichothecene' in the dataset. 47

Table B-1: List of *F.langsethiae* TRI genes with corresponding gene IDs, identified in Goodisman (2023).

Gene ID	Gene Name
TRINITY_DN4286_c1_g1_i1	TRI3
TRINITY_DN4286_c0_g1_i1	TRI4
TRINITY_DN17003_c0_g1_i2	TRI5
TRINITY_DN43694_c0_g1_i1	TRI6
TRINITY_DN34130_c1_g1_i1	TRI7
TRINITY_DN34130_c0_g1_i1	TRI8
TRINITY_DN35002_c0_g1_i1	TRI9
TRINITY_DN17003_c3_g1_i1	TRI10
TRINITY_DN19291_c0_g1_i1	TRI12
TRINITY_DN11167_c1_g1_i1	TRI13
TRINITY_DN11167_c0_g2_i1	TRI14
TRINITY_DN25107_c0_g1_i1	TRI15
TRINITY_DN29911_c0_g1_i1	TRI16
TRINITY_DN29267_c4_g1_i1	TRI101

Table B-2: List of trichothecene biosynthesis and regulatory genes, identified by searching for the keyword 'trichothecene' in the dataset. Genes originally identified and functionally annotated by Shang et al (2024).

Gene ID	Functional Annotation
FG3G00030	related to Tri201-trichothecene 3-O-acetyltransferase
FG2G26150	3-acetyltrichothecene 4-O-acetyltransferase
FG2G26240	trichothecene efflux pump
FG4G22330	trichothecene 3-O-acetyltransferase
FG2G26140	trichothecene C-15 esterase
FG1G32130	tri7-trichothecene biosynthesis
FG4G20450	trichothecene c-15 hydroxylase