# **TITLE:** Fungi.guru: comparative genomic and transcriptomic database for the Fungi kingdom

**AUTHORS:** Jolyn Jia Jia Lim<sup>1</sup>, Jace Koh<sup>1</sup>, Jia Rong Moo<sup>1</sup>, Erielle Marie Fajardo Villanueva<sup>1</sup>, Dhira Anindya Putri<sup>1</sup>, Yuen Shan Lim<sup>1</sup>, Wei Song Seetoh<sup>1</sup>, Sriya Mulupuri<sup>1</sup>, Janice Wan Zhen Ng<sup>1</sup>, Nhi Le Uyen Nguyen<sup>1</sup>, Rinta Reji<sup>1</sup>, Margaret Xuan Zhao<sup>2</sup>, Tong Ling Chan<sup>1</sup>, Edbert Edric Rodrigues<sup>1</sup>, Ryan Kairon<sup>1</sup>, Natasha Cassandra Chee<sup>1</sup>, Ann Don Low<sup>1</sup>, Zoe Chen Hui Xin<sup>1</sup>, Shan Chun Lim<sup>1</sup>, Vanessa Lunardi<sup>1</sup>, Fong Tuck Choy<sup>1</sup>, Cherlyn Xin'Er Chua<sup>1</sup>, Kenny Koh Ting Sween<sup>1</sup>, Jonathan Wei Xiong Ng<sup>1</sup>, Marek Mutwil<sup>1\*</sup>

<sup>1</sup>School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore, 637551, Singapore

<sup>2</sup>College of Medicine and Veterinary Medicine, University of Edinburgh, Old College, South Bridge, Edinburgh EH8 9YL, United Kingdom

\*Corresponding author:

Marek Mutwil

School of Biological Sciences,

Nanyang Technological University, 60 Nanyang Drive,

637551, Singapore,

Singapore

Email: mutwil@ntu.edu.sg

# **ABSTRACT**

The fungi kingdom is composed of eukaryotic heterotrophs, which are responsible for balancing the ecosystem and play a major role as decomposers. They also produce a vast diversity of secondary metabolites, which have antibiotic or pharmacological properties. However, our lack of knowledge of gene function in fungi precludes us from tailoring them to our needs and tapping into their metabolic diversity. To remedy this, we gathered genomic and gene expression data of 19 most widely-researched fungi to build a database, fungi.guru, which contains tools for cross-species identification of conserved pathways, functional gene modules, and gene families. We exemplify how our database can elucidate the molecular function, biological process and cellular component of genes involved in various biological processes, by identifying a secondary metabolite pathway producing gliotoxin in *Aspergillus fumigatus*, the catabolic pathway of cellulose in *Coprinopsis cinerea* and the conserved DNA replication pathway in *Fusarium graminearum* and *Pyricularia oryzae*. The database is available at www.fungi.guru.

## INTRODUCTION

The fungi kingdom comprises an enormous diversity of taxa with varied life cycle strategies, ecologies, and morphologies ranging from unicellular yeasts to large mushrooms. Since ancient times, fungi were used as fermenters and food, but recent mycological research demonstrates yet more novel applications of fungi in food production (1, 2). While fungi can cause diseases and decay of building materials and fibers (3), fungi have immensely contributed to agriculture and medicine. For example, secondary metabolites naturally produced by Penicillium mould are vital antibiotics such as penicillin (4), while in agriculture, fungi could become a sustainable alternative for animal feed and revolutionize food processing (5). The sheer diversity of fungal species, estimated at 2.2 to 3.8 million (6), holds much promise for future research.

As more fungi are being discovered and studied, it is becoming clear how little we know about the gene functions in this kingdom (7). Genome sequence alone often cannot reveal the function of the species-specific genes, or how the genes work together to form a biological pathway (8). Genes with similar expression profiles across environmental perturbations, developmental stages, and organs tend to be functionally related, and the identification of these co-expressed genes is thus a powerful tool to study gene function (9–14). These co-expressed genes can be represented as networks, where nodes (or vertices) correspond to genes and edges (or links) connect and indicate genes that have similar expression profiles (15, 16). The analysis of gene expression data and coexpression networks can reveal groups of functionally related genes (i.e., gene modules), while conservation of these networks over large phylogenetic distances can reveal core components of biological processes (10, 17–20). Thus, coexpression analysis is an invaluable tool to reveal gene function (21), but this resource is missing for fungi.

To facilitate the elucidation of gene functions in fungi, we present fungi.guru. This online database allows analysis of coexpression networks, gene expression profiles, ontology terms, gene families, and conserved coexpression clusters of 19 fungal species. The database provides a plethora of tools that allow functional inferences of uncharacterized fungal genes.

Our database opens up new possibilities to study fungi and to harness their rich metabolomic arsenal to generate novel drugs and other high-value compounds.

### **MATERIALS AND METHODS**

# Used genomic and transcriptomic data

Publicly available RNA sequencing experiments for the 20 unicellular and multicellular fungi were identified through NCBI SRA (22). The SRA runtables obtained included experiments from different organs (for multicellular fungi), gene knockout variants, and samples subjected to various chemical and environmental treatments. Coding sequence (CDS) files for each fungus were downloaded from Ensembl Fungi (Table S1) and used to build Kallisto index files for subsequent gene expression estimation with Kallisto v0.46.1 (23). RNA-sequencing experiments for all the fungal species were streamed as fastq files from the European Nucleotide Archive (ENA) (24) by using the LSTRaP-Cloud pipeline (ref). In total, 40843 samples comprising 40.8 terabytes were downloaded for 20 fungi (Table S2). TPM (transcripts per million) gene expression values were generated from fastq files by running the Kallisto quant command in the LSTRaP-cloud pipeline with default parameters (25). For paired-end reads, only the files containing the first read designated with "\_1" were downloaded. Single-end and paired-end libraries were mapped with an estimated fragment length of 200 bp and an estimated standard deviation of 20.

To annotate the RNA-seq samples, we used information provided within the SRA runtable and annotation of the sequencing experiments. Annotations include the type of medium the samples were cultivated in, organs, location, sexual/asexual types, temperature, and strains. The annotations were used to construct expression profiles on the CoNekT database (26).

### Quality control of RNA-seq experiments and construction of coexpression networks

For each fungal species, RNA-seq samples were quality-controlled by removing samples that showed low absolute number of processed reads (NPR) and low percentage of pseudoaligned reads (PPR) to the CDS (denoted by Kallisto as n\_pseudoaligned and p\_psuedoaligned respectively)(25, 27). The thresholds were set to remove samples with <1 million NPR and a species-specific threshold of PPR. The PPR threshold was estimated by manual inspection of scatter plots, which show NPR and PPR values on the x- and y-axis, respectively (Figure S1). Fungal species *Ramularia collo-cygni* showed low pseudoalignment values and was eliminated from the project. RNA-sequencing samples that passed the thresholds were used to construct expression matrices for each species. In total, 21135 out of 40843 samples passed the quality control (Table 1, Table S2).

# Functional annotation of fungal coding sequences

The protein IDs were obtained from the pep files available on Ensembl Fungi. Interproscan-5.44-79 (28) was used to obtain the Pfam domains and Gene Ontology terms (GO terms) for each protein in the 20 species. The groups of orthologous genes were obtained using Orthofinder v2.3.12 (29) using Diamond (30) with default settings.

# Construction of fungi.guru database

The coexpression networks were constructed by using the Highest Reciprocal Rank metric (31). Conekt database framework with default settings was used to construct the database (26). Coexpression clusters for each species were generated via Heuristic Cluster Chiseling Algorithm (HCCA)(31), where cluster sizes were limited to 100 genes.

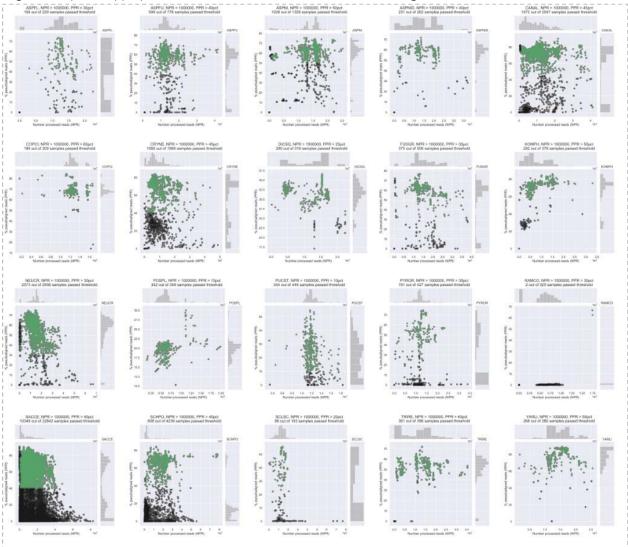
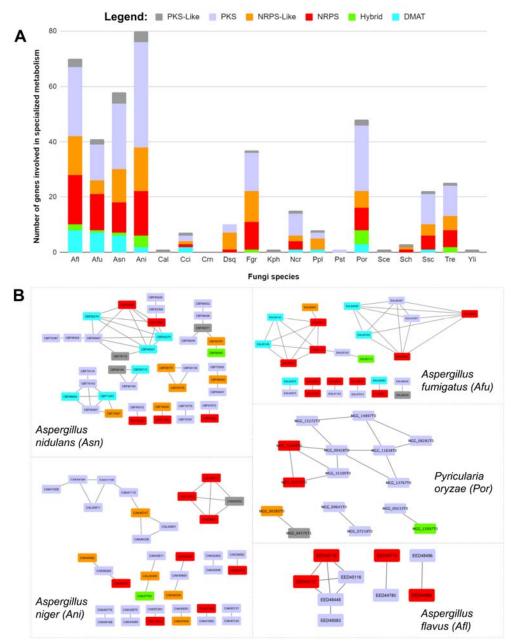


Figure S1. Quality control and data selection for the 20 fungi. The number of processed reads (NPR, x-axis) versus the percentage of pseudoaligned reads (PPR, y-axis) are shown for all 20 species, which are arranged in alphabetical order. Each point represents one sample, where black and green indicate samples that failed and passed the quality control, respectively. The histograms to the right and on top indicate the distributions of the PPR and NPR values, respectively. The sample quality control outcomes are shown in Table S1.

# **RESULTS**

To allow comparison of the coexpression relationships across the fungi kingdom, the genes have been assigned to Pfam domains (via InterproScan)(28) and gene families (via OrthoFinder)(29). Our database offers multiple ways to query it, that can be found in the

'Search' panel on the top of the page. The user can find the gene of interest by using gene IDs (e.g., *CBF69394*), sequence similarity (BLAST), Pfam domains, or Gene Ontology terms.



**Figure S2. Analysis of specialized metabolism in the 19 fungi.** A) The bar chart indicates the fungi (x-axis) and the number of genes (y-axis) predicted to be involved in specialized metabolism by SMURF. The types of metabolites are indicated by color. B) This panel presents five fungal species that have diverse networks of backbone genes as predicted by SMURF.

The database allows viewing the genomic and transcriptomic data on multiple levels. For example, it contains pages for species (<a href="www.fungi.guru/species">www.fungi.guru/species</a>), genes (<a href="www.fungi.guru/species">www.fungi.guru/species</a>), gene families (<a href="www.fungi.guru/family/view/117804">www.fungi.guru/species</a>), gene families (<a href="www.fungi.guru/family/view/117804">www.fungi.guru/family/view/117804</a>), coexpression clusters (<a href="http://www.fungi.guru/cluster/view/741">http://www.fungi.guru/cluster/view/741</a>), neighborhoods

(www.fungi.guru/network/graph/15360), phylogenetic trees of families (www.fungi.guru/tree/view/45802), Pfam domains (www.fungi.guru/interpro/view/5689), Gene Ontology terms (www.fungi.guru/go/view/16530) and others. These pages contain data relevant to the type of the viewed data. For example, gene pages contain the functional annotation, the cDNA and protein sequences, assigned gene family, phylogenetic tree of the family, expression profiles, coexpression neighborhood and cluster, significantly similar neighborhoods, and Gene Ontology information. Conversely, Gene Ontology pages indicate the annotation of the GO term, the number of genes in the 19 fungi that have the GO term, and the co-expressed clusters that are enriched for the genes with this GO. A full description of the features is found at localhost/features.

To exemplify these features, we provide three typical case studies addressing different aspects of fungal biology.

# Example 1: Identification of specialized metabolism hubs by coexpression analysis

Coexpression analysis has been used successfully to study numerous metabolic pathways in plants (32, 33) and fungi (refs). To exemplify how our database can be used to study specialized metabolism in fungi, we first set to identify 'backbone' enzymes responsible for the biosynthesis of the different metabolite classes. To this end, we used Secondary Metabolite Unique Regions Finder (SMURF) database (34), to assign the genes of the 19 fungi to nonribosomal peptide synthases (NRPSs), polyketide synthases (PKSs), NRPS-like enzymes, PKS-like enzymes, Hybrid enzymes and dimethylallyl tryptophan synthase (DMAT). The analysis revealed that the 19 fungi have a diverse set of the backbone enzymes, ranging from 80 enzymes in *Aspergillus niger* (Ani), to 0 in *Cryptococcus neoformans* (Crn), where five (Asn, Ani, Afu, Por, Afl, Table 1) of the 19 fungi were shown to have more than 40 backbone enzymes (Figure S2A).

To identify groups of genes involved in metabolite synthesis and conditions that result in high metabolite production, backbone enzymes were first retrieved from SMURF (34), and entered into 'Tools/Create custom network' tool (<a href="www.fungi.guru/custom\_network/">www.fungi.guru/custom\_network/</a>), which identifies coexpression relationships in the provided list of genes. The custom network analysis of specialized metabolism genes in *Aspergillus fumigatus* produced an interactive network that revealed a high degree of coexpression among the backbone enzymes, indicating that biosynthesis of the various metabolites is transcriptionally coordinated (Figure 1A).

One of the Aspergillus fumigatus backbone enzymes, EAL85149, nonribosomal peptide synthase ftmA, was found to be co-expressed with a large variety of genes involved in secondary metabolite synthesis (Figure 1A), suggesting that the enzyme is part of a transcriptional program that produces numerous metabolites. By entering the EAL85149 into the dedicated search box. arrived gene page EAL85149 а (www.fungi.guru/sequence/view/20959). The expression profile (www.fungi.guru/profile/view/27005) revealed that the gene shows the highest expression in fructose medium (Figure 1B). Since specialized metabolism tends to be transcriptionally regulated (35, 36), this information can be used to identify growth conditions needed to induce high production of specialized metabolites.

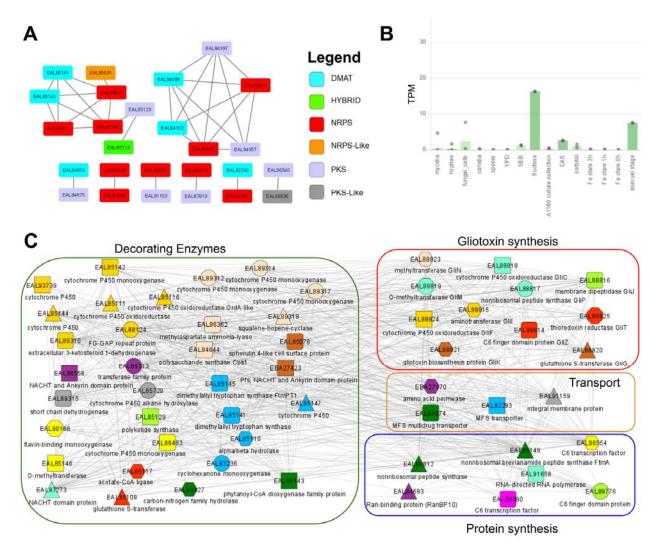


Figure 1. Coexpression analysis of specialized metabolic hubs in *Aspergillus fumigatus*.

A) Coexpression network of the backbone enzymes involved in specialized metabolism. Nodes represent genes, while edges connect co-expressed genes. For brevity, enzymes that are not co-expressed are not shown. The color of the node indicates the type of backbone enzyme. The plot was generated by exporting the network file and opening the file in standalone Cytoscape. B) Expression profile of *EAL85149*, a nonribosomal peptide synthase ftmA that shows the highest connection to other backbone enzymes. The different samples are represented on the x-axis, while the gene expression values in the form of Transcripts Per Million (TPM) are indicated on the y-axis. The bars and the dots indicate the average and the minimum/maximum values of the RNA-seq experiments in the sample, respectively. C) Coexpression network of *EAL85149*. Nodes represent genes, while edges connect co-expressed genes. Colored shapes indicate which genes contain Pfam domains and orthogroups in common. For brevity, only part of the network is shown. The network contains four types of genes: gliotoxin family genes (indicated in the red rectangle), genes involved in protein synthesis (blue rectangle), metabolite transporters (orange rectangle), and decorating enzymes (green rectangle). The function of the genes can be inferred from the legend and by clicking on the nodes.

EAL85149 is involved in the biosynthesis of toxins such as gliotoxins and brevianamide F, a precursor bioactive prenylated alkaloid (37). Brevianamide F acts as a precursor for fungal toxins fumitremorgin C and trypacidin that can prevent phagocytosis of fungal conidia, and serves an essential protective role for the pathogenesis of fungi (38). To identify genes that are functionally related to EAL85149, we viewed its coexpression network available at the gene page (www.fungi.guru/network/graph/15360)(Figure 1C). 11 out of 13 gliotoxin biosynthesis genes are found in the gliotoxin gli coexpression cluster, with the exception of GliA, responsible for the exportation of extracellular gliotoxin essential for gliotoxin tolerance (39) and GliH whose function is still unknown (40). Interestingly, the gliotoxin biosynthesis genes found in the coexpression cluster are also part of the genomic gli cluster, where the gli biosynthetic genes are found on the same chromosomal locus (41). The coexpression cluster also contains a variety of transporters such as Major Facilitator Superfamily (MFS) transporter that could be used to secrete gliotoxin and other metabolites potentially synthesized by the coexpression cluster. Additionally, the coexpression cluster contains large numbers of cytochrome P450s, which are decorative enzymes essential for secondary metabolite synthesis (42).

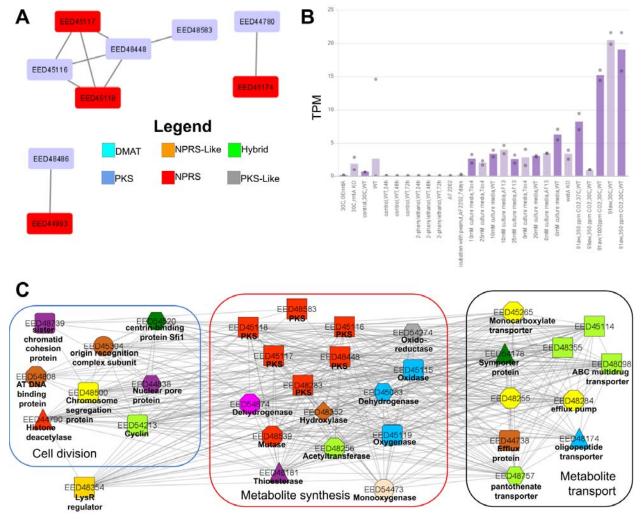


Figure S3. Analysis of specialized metabolic hubs in *Aspergillus flavus*. A) Coexpression network of the backbone enzymes involved in specialized metabolism. Nodes represent genes,

while co-expressed genes are connected by edges. For brevity, enzymes that are not co-expressed are not shown. The color of the node indicates the type of backbone enzyme. B) Expression profile of *EED48448*, a putative polyketide synthase that shows the highest connection to other backbone enzymes. The different samples are represented on the x-axis, while the gene expression values in the form of Transcripts Per Million (TPM) are indicated on the y-axis. The bars and the dots indicate the average and the minimum/maximum values of the RNA-seq experiments in the sample, respectively. C) Coexpression network of *EED48448*. Nodes represent genes, while edges connect co-expressed genes. Colored shapes indicate which genes contain Pfam domains and orthogroups in common. For brevity, only part of the network is shown. The network contains three types of genes: metabolic enzymes (indicated in the red rectangle), cell division genes (blue rectangle), and metabolite transporters (black rectangle). The function of the genes can be inferred from the legend and by clicking on the nodes.

In addition to the discussed genes, we observed multiple genes coding for hypothetical proteins in the cluster. Since these genes are co-expressed with the gli pathway, they could also play a crucial role in secondary metabolite synthesis and merit further study. Thus, this example shows how fungi.guru can be easily used to reveal conditions that are likely to control the production of metabolites and reveal novel genes involved in the metabolite production (Figure S3 shows a similar analysis performed in *Aspergillus flavus*).

# Example 2: Gene Ontology search reveals clusters important for cellobiose degradation

Many fungi, especially saprobic fungi, obtain energy from decaying organic matter. Their main source of energy comes from digesting decomposing plant material such as cellulose and lignin (43). Basidiomycetous fungi, for example, *Coprinopsis cinerea (Cci)*(44), are one of the most potent degraders of cellulose as they often grow in environments that are cellulose-abundant, such as dead wood and plants (45, 46). Plant cell walls are multi-layered and made up of different materials which include cellulose, hemicelluloses, and lignins (47), leading to the need for fungi to produce various types of hydrolytic enzymes (endoglucanase, cellobiohydrolase, and beta-glucosidase) to break down these biomolecules into substances that can be used as sources of energy for cellular activity (48). Thus, depending on the composition of the organic matter decomposed, different fungal species utilize their own unique set of enzymes and metabolic pathways for cellobiose degradation (49).

To gain insight into cellobiose degradation in *Cci*, we entered 'cellulose catabolism' into 'Tools/Find enriched clusters' (<a href="www.fungi.guru/search/enriched/clusters">www.fungi.guru/search/enriched/clusters</a>), which revealed Gene Ontology term GO:0030245 (GO term for Cellulose Catabolic Process). The database identified six clusters significantly enriched (p-value for enrichment < 0.05) for this GO process in six species (<a href="mailto:Aspergillus nidulans">Aspergillus niger</a>, *Coprinopsis cinerea*, *Neurospora crassa*, *Pyricularia oryzae*, and *Sclerotinia sclerotiorum*). For *Cci*, we focused on cluster 21, which was enriched for GO:0030245 (p-value for enrichment < 0.001). Using the database, we arrived at a page that details the cluster analysis of genes involved in cellobiose degradation in *Cci* (<a href="www.fungi.guru/cluster/view/741">www.fungi.guru/cluster/view/741</a>)(Figure 2A).

The coexpression cluster contained 150 genes, which were involved in different processes (alcohol metabolism, cellulose catabolism, cell division, fungal growth, lipolysis, proteolysis, and transporters)(Figure 2A). We observed that cellobiohydrolases (50), cellulose-

binding domains (51), glucanases (52), glycoside hydrolases (53), xylanases (54), and sugar transporters are the main genes present in the cluster, supporting the function of this cluster in cellobiose degradation and uptake of sugars. The cluster page also contains information about the enriched GO terms, whether similar clusters can be found in other fungi, presence of InterPro domains, enriched GO terms, and presence of gene families. Additionally, we also investigated the expression profile for this cluster (Figure 2B) and observed that the gene expression is the highest in the mycelium, suggesting that these structures are most active in degrading cellulose in *Cci*.

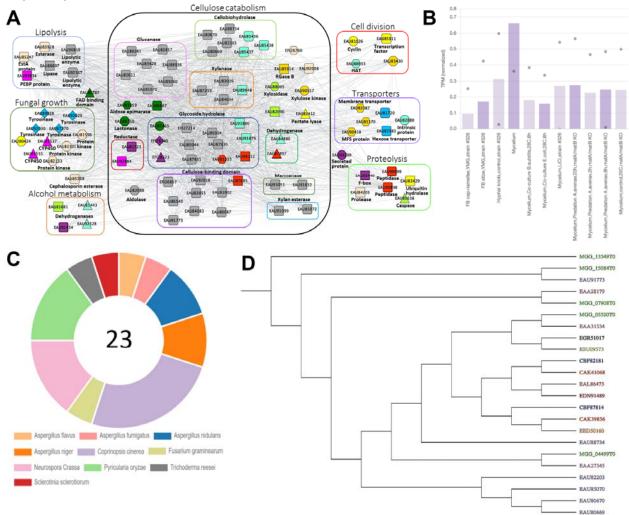


Figure 2. Cluster analysis involved in cellulose catabolic process in *Coprinopsis cinerea*. A) Coexpression network of genes identified as enriched for GO:0030245 (GO term for Cellulose Catabolic Process). Nodes represent genes, while co-expressed genes are connected by edges. For brevity, novel genes (hypothetical proteins) are not shown in the network. Colored

by edges. For brevity, novel genes (hypothetical proteins) are not shown in the network. Colored shapes indicate which genes contain Pfam domains and orthogroups in common. B) Average expression profile of cluster 21 for *Cci*. Only part of the expression profile containing the discussed sample is shown. The different samples are represented on the x-axis, while the gene expression values in the form of Transcripts Per Million (TPM) are indicated on the y-axis. The bars and the dots indicate the average and minimum/maximum values of the RNA-seq

experiments in the sample. C) The proportion of sequences in the cellobiohydrolases Gene Family: OG\_03\_0001111. There are a total of 23 sequences with this label, from 10 different species. D) Phylogenetic tree of cellobiohydrolases gene family OG\_03\_0001111.

The most abundant gene family in cluster 21 was OG 03 0001111, which contains genes annotated as cellobiohydrolases. We clicked on OG\_03\_0001111 to arrive at a page dedicated to the family (www.fungi.guru/family/view/117804). The gene family page revealed that Cci contains the highest proportion of cellobiohydrolases (Figure 2C), suggesting that it is one of the more potent degraders of cellobiose and its cellobiohydrolase gene family may have undergone a higher degree of expansion compared to the other species. The page also contains link the phylogenetic tree generated to by (www.fungi.guru/tree/view/45802). The tree (Figure 2D) shows us that the cellobiohydrolase family has expanded in Cci and this could hypothetically be attributed to the further enhancement of cellobiose degradation ability in Cci.

Since we identified multiple such clusters in different fungi (please see a similar analysis done in *Postia placenta*, Figure S4), we propose that our database is a good starting point to study cellulose degradation in fungi. Thus, the database allows easy identification of coexpression clusters, genes, and gene families relevant to the biological process of interest.

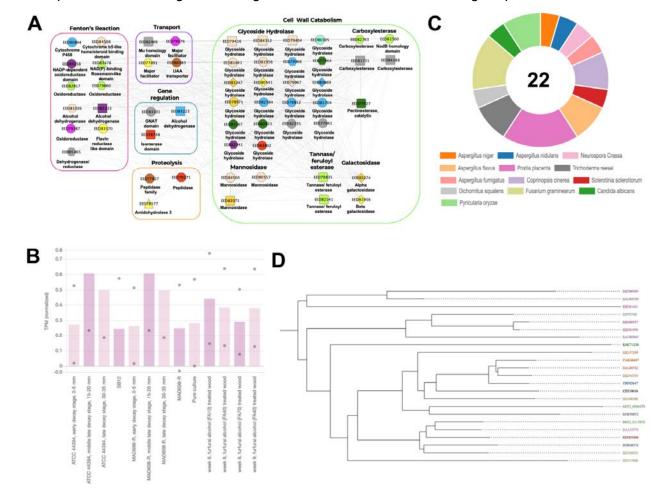


Figure S4. Cluster analysis involved in hydrolase activity in *Postia placenta*. A) Coexpression network of cluster 59 (<a href="www.fungi.guru/go/view/10627">www.fungi.guru/go/view/10627</a>). Nodes represent genes, while edges connect coexpressed genes. Colored shapes indicate which genes contain Pfam domains and orthogroups in common. The network contains five types of genes: cell wall catabolism genes (indicated in a green rectangle), genes encoding for Fenton's Reaction (pink rectangle), genes involved in transport (purple rectangle), gene-regulating genes (blue rectangle) and proteolysis associated proteins (orange rectangle). For brevity, only discussed nodes are shown. B) Expression profile of the cluster 59 from POSPL shows that the cluster has the highest expression during the middle-late decay stage in ATCC 44394 and MAD698-R strains C) Proportion of genes in the gene family OG\_03\_0001073, which is highly abundant in the cluster and which contains genes annotated as glycoside hydrolases and mannosidases.D) Phylogenetic tree of hydrolase gene family OG\_03\_0001073.

# **Example 3. Comparative analyses of DNA replication in fungi**

Coexpression relationships can be conserved across species, where coexpression clusters containing the same gene families and Pfam domains are found in multiple species (10, 11, 55). Identifying these conserved coexpression clusters thus enables the identification of conserved biological pathways and the core genetic components of these pathways (ref). Fungi.guru contains multiple tools that allow the identification of the conserved pathways.

To exemplify these comparative features, we examined the process of DNA replication between 2 fungi known for their infectious capabilities - *Fusarium graminearum* and *Pyricularia oryzae*. Under 'Tools/Find enriched clusters', we searched for enriched clusters using the GO term "DNA replication" with *F. graminearum* selected as target (<a href="www.fungi.guru/search/enriched/clusters">www.fungi.guru/search/enriched/clusters</a>). By clicking "Show cluster", the database revealed that Cluster\_10 in *F. graminearum* is enriched for this term (<a href="www.fungi.guru/cluster/view/1203">www.fungi.guru/cluster/view/1203</a>). Under 'Similar Clusters' table, we identified *P. oryzae* cluster 10 as being similar to *F. graminearum* cluster 10 with Jaccard index value of 0.202 (26). Clicking on the 'Compare' button shows the coexpression networks of these two conserved modules, where only genes that have a homolog or Pfam domain in both clusters are shown (Figure 3A).

As expected for clusters significantly enriched for GO term DNA replication, the genes in the conserved clusters are involved in chromatin regulation and DNA replication. The genes belong to multiple conserved gene families, including histone acetyltransferase, minichromosome maintenance protein complex (MCM), DNA polymerase, and others (Figure 3A). Histone acetyltransferases are involved in the loosening of the condensed chromatin structure to allow DNA replication machinery access to the template (56). Cyclin-B activates downstream cyclin-dependent kinases involved in the M phase of the cell cycle, while other cyclin types control kinases for other stages in the cell cycle (57). MCM is known to be involved in the assembly of the replication initiation complex at the origin of replication (58). Cdc45 form the catalytic core of DNA helicase that serves to unwind the double-stranded DNA (59). Type 2 DNA topoisomerase introduce negative supercoils in the DNA to release the stress on the DNA set by the unwinding action of DNA Helicase (60).

In addition to identifying conserved clusters, our database can also reveal which gene coexpression neighborhoods are similar. For example, *ESU05693* 

(www.fungi.guru/sequence/view/89556) is a DNA replication licensing factor mcm4 that belongs to OG 03 0001507 gene family (www.fungi.guru/family/view/118200). The gene page for ESU05693 contains a table 'Expression context conservation' which lists other genes from OG 03 0001507 gene family that have a coexpression neighborhood that contains similar content of gene families to ESU05693. By clicking on 'View ECC as graph', the database indicated that 17 genes from the OG\_03\_0001507 gene family belonging to 17 different species have similar coexpression neighborhoods (Figure 3B). This is not surprising as DNA replication is a well-conserved process. The contents of the co-expressed neighborhoods can be viewed by clicking on the 'View ECC pair as graph' in the 'Expression context conservation' table. Similarly to the comparative analysis of the clusters (Figure 3A), the database displays the contents of the two neighborhoods and highlights the conserved gene families and Pfam domains (Figure 3C). The contents of the two coexpression neighborhoods are also related to DNA replication. Thus, fungi.guru allows identifying conserved coexpression clusters and neighborhoods across species, which can be invaluable to identify novel genes and pathways relevant biological to pathway of interest.

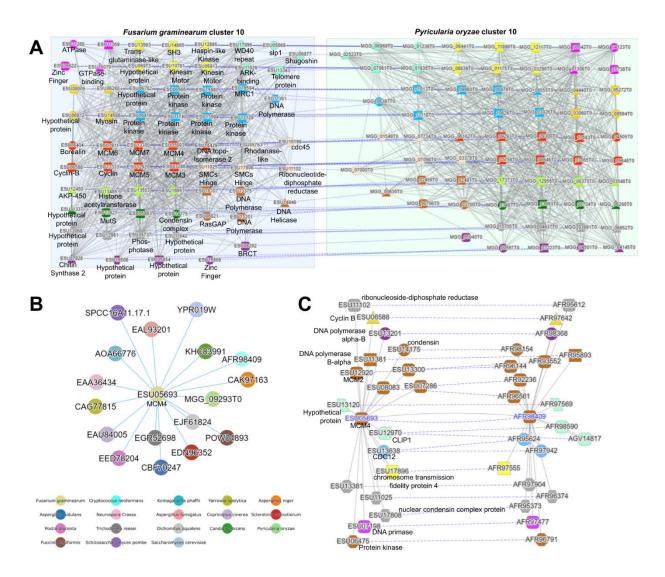


Figure 3. Comparative analysis of gene coexpression networks of DNA replication in *F. graminearum* and *P. oryzae*. A) Comparative analysis of conserved coexpression clusters. Nodes represent genes, while co-expressed genes are connected by edges. The color and shape of the nodes indicates the gene family and pfam domains of the genes. The dashed edges between the conserved clusters connect genes that belong to the same gene families, while gray edges connect co-expressed genes. B) Expression context conservation network of *ESU0569*. Each node indicates a gene co-expression neighborhood, while blue edges connect conserved neighborhoods. The color of the nodes indicates the fungal species. C) Conserved co-expression neighborhoods. Please see text in panel 3A for legend.

### CONCLUSION

To remedy the paucity of functional information for fungi, we constructed www.fungi.guru, a user-friendly database that facilitates the analysis of genomic and transcriptomic data. The tools available in the database allow for the identification and analysis of co-expressed gene neighborhoods, clusters, and gene families. While our examples addressed specialized metabolism, cellulose degradation, and cell division, the database can be used to study molecular function, biological process, and cellular component of all genes expressed in the 19 fungi present in the database. We envision that the gene function prediction tools present in fungi.guru will enhance the selection of relevant genes to further functional analyses.

### **DATA AVAILABILITY**

The expression matrices, RNA-seq sample annotation and gene families are available from the Supplementary Data. The co-expression networks, coding and protein sequences can be downloaded from <a href="https://www.fungi.guru.">www.fungi.guru.</a>

# **ACKNOWLEDGEMENTS**

We would like to acknowledge Dr. Irene Julca nad Dr. Riccardo Delli-Ponti for their help in setting up OrthoFinder and Interproscan analyses. William Goh is acknowledged for helping in the initial steps of the project. Finally, we would like to thank Ryan Chee Khiang Ng for techical assistance.

## **FUNDING**

Nanyang Technological University Start-Up Grant. Funding for open access charge: Start-Up Grant.

Conflict of interest statement.
None declared.

#### **REFERENCES**

- 1. Pensupa, N., Jin, M., Kokolski, M., Archer, D.B. and Du, C. (2013) A solid state fungal fermentation-based strategy for the hydrolysis of wheat straw. *Bioresour. Technol.*, 10.1016/j.biortech.2013.09.061.
- 2. Valverde, M.E., Hernández-Pérez, T. and Paredes-López, O. (2015) Edible mushrooms: Improving human health and promoting quality life. *Int. J. Microbiol.*, 10.1155/2015/376387.

- 3. Jones, D. and Brischke, C. (2017) Performance of Bio-based Building Materials.
- Laich,F., Fierro,F. and Martín,J.F. (2002) Production of penicillin by fungi growing on food products: Identification of a complete penicillin gene cluster in Penicillium griseofulvum and a truncated cluster in Penicillium verrucosum. *Appl. Environ. Microbiol.*, 10.1128/AEM.68.3.1211-1219.2002.
- Ahmed,S., Mustafa,G., Arshad,M. and Rajoka,M.I. (2017) Fungal Biomass Protein Production from Trichoderma harzianum Using Rice Polishing. *Biomed Res. Int.*, 10.1155/2017/6232793.
- 6. Fungal Diversity Revisited: 2.2 to 3.8 Million Species (2017) In The Fungal Kingdom.
- 7. Meyer, V., Andersen, M.R., Brakhage, A.A., Braus, G.H., Caddick, M.X., Cairns, T.C., de Vries, R.P., Haarmann, T., Hansen, K., Hertz-Fowler, C., *et al.* (2016) Current challenges of research on filamentous fungi in relation to human welfare and a sustainable bio-economy: a white paper. *Fungal Biol. Biotechnol.*, 10.1186/s40694-016-0024-8.
- 8. Rhee,S.Y. and Mutwil,M. (2014) Towards revealing the functions of all genes in plants. *Trends Plant Sci.*, **19**, 212–221.
- 9. Stuart, J.M., Segal, E., Koller, D. and Kim, S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* (80-. )., **302**, 249–255.
- 10. Mutwil, M., Klie, S., Tohge, T., Giorgi, F.M., Wilkins, O., Campbell, M.M., Fernie, A.R., Usadel, B., Nikoloski, Z. and Persson, S. (2011) PlaNet: Combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell*, **23**, 895–910.
- 11. Ruprecht, C., Vaid, N., Proost, S., Persson, S. and Mutwil, M. (2017) Beyond Genomics: Studying Evolution with Gene Coexpression Networks. *Trends Plant Sci.*, **22**, 298–307.
- 12. Ferrari, C., Proost, S., Ruprecht, C. and Mutwil, M. (2018) PhytoNet: Comparative co-expression network analyses across phytoplankton and land plants. *Nucleic Acids Res.*, **46**, W76–W83.
- 13. Ruprecht, C., Mutwil, M., Saxe, F., Eder, M., Nikoloski, Z. and Persson, S. (2011) Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Front. Plant Sci.*, **2**.
- 14. Ruprecht, C., Proost, S., Hernandez-Coronado, M., Ortiz-Ramirez, C., Lang, D., Rensing, S.A., Becker, J.D., Vandepoele, K. and Mutwil, M. (2017) Phylogenomic analysis of gene coexpression networks reveals the evolution of functional modules. *Plant J.*, **90**, 447–465.
- 15. Lee,T., Yang,S., Kim,E., Ko,Y., Hwang,S., Shin,J., Shim,J.E., Shim,H., Kim,H., Kim,C., *et al.* (2015) AraNet v2: An improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res.*, **43**, D996–D1002.
- 16. Usadel,B., Obayashi,T., Mutwil,M., Giorgi,F.M., Bassel,G.W., Tanimoto,M., Chow,A., Steinhauser,D., Persson,S. and Provart,N.J. (2009) Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant, Cell Environ.*, **32**, 1633–1651.
- 17. Mutwil, M., Obro, J., Willats, W.G.T. and Persson, S. (2008) GeneCAT--novel webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res.*, **36**.
- 18. Hansen,B.O., Vaid,N., Musialak-Lange,M., Janowski,M. and Mutwil,M. (2014) Elucidating gene function and function evolution through comparison of co-expression networks of plants. *Front. Plant Sci.*, **5**, 1–9.
- 19. Ng,J.W.X., Tan,Q.W., Ferrari,C. and Mutwil,M. (2019) Diurnal.plant.tools: Comparative Transcriptomic and Co-expression Analyses of Diurnal Gene Expression of the Archaeplastida Kingdom. *Plant Cell Physiol.*, 10.1093/pcp/pcz176.
- 20. Ferrari, C., Shivhare, D., Hansen, B.O., Pasha, A., Esteban, E., Provart, N.J., Kragler, F., Fernie, A.R., Tohge, T. and Mutwil, M. (2020) Expression Atlas of Selaginella moellendorffii Provides Insights into the Evolution of Vasculature, Secondary Metabolism, and Roots. *Plant Cell*, 10.1105/tpc.19.00780.
- 21. Cairns, T.C., Feurstein, C., Zheng, X., Zhang, L.H., Zheng, P., Sun, J. and Meyer, V. (2019)

- Functional exploration of co-expression networks identifies a nexus for modulating protein and citric acid titres in Aspergillus niger submerged culture. *Fungal Biol. Biotechnol.*, 10.1186/s40694-019-0081-x.
- 22. Leinonen,R., Sugawara,H. and Shumway,M. (2011) The sequence read archive. *Nucleic Acids Res.*, **39**.
- 23. Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, **34**, 525–527.
- 24. Leinonen,R., Akhtar,R., Birney,E., Bower,L., Cerdeno-Tárraga,A., Cheng,Y., Cleland,I., Faruque,N., Goodgame,N., Gibson,R., *et al.* (2011) The European nucleotide archive. *Nucleic Acids Res.*, **39**, D28–D31.
- 25. Tan,Q.W., Goh,W. and Mutwil,M. (2020) Lstrap-cloud: A user-friendly cloud computing pipeline to infer coexpression networks. *Genes (Basel).*, 10.3390/genes11040428.
- 26. Proost,S. and Mutwil,M. (2018) CoNekT: An open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic Acids Res.*, **46**, W133–W140.
- 27. Tan,Q.W. and Mutwil,M. (2019) Inferring biosynthetic and gene regulatory networks from Artemisia annua RNA sequencing data on a credit card-sized ARM computer. *Biochim. Biophys. acta. Gene Regul. Mech.*, 10.1016/j.bbagrm.2019.194429.
- 28. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., *et al.* (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.
- 29. Emms, D.M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, **16**.
- 30. Buchfink,B., Xie,C. and Huson,D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, **12**, 59–60.
- 31. Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O. and Persson, S. (2010) Assembly of an interactive correlation network for the Arabidopsis genome using a novel Heuristic Clustering Algorithm. *Plant Physiol.*, **152**, 29–43.
- 32. Sibout,R., Proost,S., Hansen,B.O., Vaid,N., Giorgi,F.M., Ho-Yue-Kuang,S., Legée,F., Cézart,L., Bouchabké-Coussa,O., Soulhat,C., *et al.* (2017) Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon. *New Phytol.*, **215**, 1009–1025.
- 33. Ruprecht, C., Mutwil, M., Saxe, F., Eder, M., Nikoloski, Z. and Persson, S. (2011) Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Front. Plant Sci.*, **2**, 1–13.
- 34. Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H. and Fedorova, N.D. (2010) SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.*, 10.1016/j.fgb.2010.06.003.
- 35. Mutwil, M. (2020) Computational approaches to unravel the pathways and evolution of specialized metabolism. *Curr. Opin. Plant Biol.*, 10.1016/j.pbi.2020.01.007.
- 36. Wisecaver, J.H., Borowsky, A.T., Tzin, V., Jander, G., Kliebenstein, D.J. and Rokas, A. (2017) A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell*, **29**, 944–959.
- 37. Stack, D., Neville, C. and Doyle, S. (2007) Nonribosomal peptide synthesis in Aspergillus fumigatus and other fungi. *Microbiology*, 10.1099/mic.0.2006/006908-0.
- 38. Mattern, D.J., Schoeler, H., Weber, J., Novohradská, S., Kraibooj, K., Dahse, H.M., Hillmann, F., Valiante, V., Figge, M.T. and Brakhage, A.A. (2015) Identification of the antiphagocytic trypacidin gene cluster in the human-pathogenic fungus Aspergillus fumigatus. *Appl. Microbiol. Biotechnol.*, 10.1007/s00253-015-6898-1.
- 39. Wang, D.N., Toyotome, T., Muraosa, Y., Watanabe, A., Wuren, T., Bunsupa, S., Aoyagi, K., Yamazaki, M., Takino, M. and Kamei, K. (2014) GliA in Aspergillus fumigatus is required for its tolerance to gliotoxin and affects the amount of extracellular and intracellular gliotoxin.

- Med. Mycol., 10.1093/mmy/myu007.
- 40. Huang, Z.L., Ye, W., Zhu, M.Z., Kong, Y.L., Li, S.N., Liu, S. and Zhang, W.M. (2020) Interaction of a novel zn2cys6 transcription factor dcgliz with promoters in the gliotoxin biosynthetic gene cluster of the deep-sea-derived fungus dichotomomyces cejpii. *Biomolecules*, 10.3390/biom10010056.
- 41. Gardiner, D.M. and Howlett, B.J. (2005) Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of Aspergillus fumigatus. *FEMS Microbiol. Lett.*, 10.1016/j.femsle.2005.05.046.
- 42. Shin,J., Kim,J.E., Lee,Y.W. and Son,H. (2018) Fungal cytochrome p450s and the p450 complement (Cypome) of fusarium graminearum. *Toxins (Basel)*., 10.3390/toxins10030112.
- 43. Andlar, M., Rezić, T., Marđetko, N., Kracher, D., Ludwig, R. and Šantek, B. (2018) Lignocellulose degradation: An overview of fungi and fungal enzymes involved in lignocellulose degradation. *Eng. Life Sci.*, 10.1002/elsc.201800039.
- 44. Hashimoto,K., Yoshida,M. and Hasumi,K. (2011) Isolation and characterization of CcAbf62A, a GH62 α-L- arabinofuranosidase, from the basidiomycete Coprinopsis cinerea. *Biosci. Biotechnol. Biochem.*, 10.1271/bbb.100434.
- 45. Baldrian, P. and Valášková, V. (2008) Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiol. Rev.*, 10.1111/j.1574-6976.2008.00106.x.
- 46. Sudarson, J., Ramalingam, S., Kishorekumar, P. and Venkatesan, K. (2014) Expeditious Quantification of Lignocellulolytic Enzymes from Indigenous Wood Rot and Litter Degrading Fungi from Tropical Dry Evergreen Forests of Tamil Nadu. *Biotechnol. Res. Int.*, 10.1155/2014/127848.
- 47. Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., et al. (2004) Toward a systems approach to understanding plant cell walls. *Science (80-.).*, 10.1126/science.1102765.
- 48. Dashtban, M., Schraft, H. and Qin, W. (2009) Fungal bioconversion of lignocellulosic residues; Opportunities & perspectives. *Int. J. Biol. Sci.*, 10.7150/ijbs.5.578.
- 49. Rytioja, J., Hildén, K., Yuzon, J., Hatakka, A., de Vries, R.P. and Mäkelä, M.R. (2014) Plant-Polysaccharide-Degrading Enzymes from Basidiomycetes. *Microbiol. Mol. Biol. Rev.*, 10.1128/mmbr.00035-14.
- 50. Teeri, T.T. (1997) Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends Biotechnol.*, 10.1016/S0167-7799(97)01032-9.
- 51. Carrard, G., Koivula, A., Söderlund, H. and Béguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. U. S. A.*, 10.1073/pnas.160216697.
- 52. Yang,Z., Peng,Z., Wei,S. and Yu,Y. (2013) Cloning and characterization of endo-β-1,4-glucanase genes in the common wheat line three pistils. *Genet. Mol. Biol.*, 10.1590/S1415-47572013000300015.
- 53. Méndez-Líter, J.A., Gil-Muñoz, J., Nieto-Domínguez, M., Barriuso, J., De Eugenio, L.I. and Martínez, M.J. (2017) A novel, highly efficient β-glucosidase with a cellulose-binding domain: Characterization and properties of native and recombinant proteins. *Biotechnol. Biofuels*, 10.1186/s13068-017-0946-2.
- 54. Jampala,P., Tadikamalla,S., Preethi,M., Ramanujam,S. and Uppuluri,K.B. (2017) Concurrent production of cellulase and xylanase from Trichoderma reesei NCIM 1186: enhancement of production by desirability-based multi-objective method. *3 Biotech*, 10.1007/s13205-017-0607-y.
- 55. Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A.R., Nikoloski, Z., Persson, S. and Mutwil, M. (2016) Famnet: A framework to identify multiplied modules driving pathway expansion in plants. *Plant Physiol.*, **170**, 1878–1894.
- 56. Li,Q., Zhou,H., Wurtele,H., Davies,B., Horazdovsky,B., Verreault,A. and Zhang,Z. (2008)

- Acetylation of Histone H3 Lysine 56 Regulates Replication-Coupled Nucleosome Assembly. *Cell*, 10.1016/j.cell.2008.06.018.
- 57. Dahmann, C., Diffley, J.F.X. and Nasmyth, K.A. (1995) S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. *Curr. Biol.*, 10.1016/S0960-9822(95)00252-1.
- 58. Leatherwood, J. (1998) Emerging mechanisms of eukaryotic DNA replication initiation. *Curr. Opin. Cell Biol.*, 10.1016/S0955-0674(98)80117-8.
- 59. Mimura,S. and Takisawa,H. (1998) Xenopus Cdc45-dependent loading of DNA polymerase α onto chromatin under the control of S-phase cdk. *EMBO J.*, 10.1093/emboj/17.19.5699.
- 60. Champoux, J.J. (2001) DNA Topoisomerases: Structure, Function, and Mechanism. *Annu. Rev. Biochem.*, 10.1146/annurev.biochem.70.1.369.

# TABLE AND FIGURES LEGENDS TABLES

Table 1. Fungal species included in the database.

Species	3-letter code	Division	Use	# samples passed QC / all samples
Aspergillus flavus	Afl	Ascomycot a	Study of preharvest and postharvest infections, as well as the production of compounds that are toxic to mammals	158 / 220
Aspergillus fumigatus	Afu	Ascomycot a	Study of invasive aspergillosis (IA), which results in high mortality in AIDS and organ transplant patients	599 / 778
Aspergillus nidulans	Asn	Ascomycot a	Model organism for studying eukaryotic cell processes such as DNA repair and cell cycle control	1028 / 1328
Aspergillus niger	Ani	Ascomycot a	Study of black mold, which is a common contaminant of certain fruits and vegetables	1028 / 1328
Candida albicans	Cal	Ascomycot a	Model organism for fungal pathogens and the study of HIV	1572 / 2547
Coprinopsis cinerea	Cci	Basidiomyc ota	Model organism for studying fungal sex and mating types	199 / 209
Cryptococcus neoformans	Crn	Basidiomyc ota	Model organism for encapsulated fungal organisms and the study of diseases in	1068 / 1866

			immunocompromised hosts	
Dichomitus squalens	Dsq	Basidiomyc ota	Model organism for wood- degrading white-rot fungi that infect both softwood and hardwood plant species	280 / 316
Fusarium graminearum	Fgr	Ascomycot a	Study of fusarium head blight, a devastating disease affecting wheat and barley	375 / 506
Komagataella phaffii	Kph	Ascomycot a	Model organism for genetic study and the study of protein production	292 / 379
Neurospora crassa	Ncr	Ascomycot a	Elucidation of molecular events involved in circadian rhythms, epigenetics and gene silencing, cell polarity, cell fusion and development	2573 / 2936
Postia placenta	Ppl	Basidiomyc ota	Model organism for studying various patterns of wood decay caused by fungi	342 / 348
Puccinia striiformis	Pst	Basidiomyc ota	Study of the effect of stripe rust on wheat and grasses, as well as their potential resistance to stripe rust	344 / 449
Pyricularia oryzae	Por	Ascomycot a	Study of rice blast, the most important disease concerning the rice crop in the world	191 / 427
Ramularia collo- cygni	Rco	Ascomycot a	Study of Ramularia leaf spot, a disease affecting barley fields throughout temperate regions worldwide	6 / 330
Saccharomyces cerevisiae	Sce	Ascomycot a	Model organism for unicellular eukaryotes	10348 / 22842
Schizosaccharo myces pombe	Sch	Ascomycot a	Study of cellular responses to DNA damage and the process of DNA replication	808 / 4236
Sclerotinia sclerotiorum	Ssc	Ascomycot a	Study of white mold, which is known to infect 408 different plant species	96 / 193
Trichoderma	Tre	Ascomycot	Study of cellulase production	361 / 396

reesei		а	and cellulose hydrolysis	
Yarrowia lipolytica	Yli	Ascomycot a	Study of the synthesis of valuable metabolites and enzymes, as well as the production of specialty lipids	268 / 280

### SUPPLEMENTAL TABLES

- Table S1. Genome versions used to build the database
- Table S2. Pseudoaligning statistics of the 20 fungal species.
- **Table S3. Expression profile of ASPFL genes across RNA-Seq experiments** Rows contain transcripts per million (TPM) expression values for a particular gene, while columns contain all expression values captured by an RNA-Seq experiment. The first row contains all the RNA-Seq experiment IDs while the first column contains all the gene IDs of the fungi.
- Table S4. Expression profile of ASPFU genes across RNA-Seq experiments
- Table S5. Expression profile of ASPNI genes across RNA-Seq experiments
- Table S6. Expression profile of ASPNID genes across RNA-Seq experiments
- Table S7. Expression profile of CANAL genes across RNA-Seq experiments
- Table S8. Expression profile of COPCI genes across RNA-Seg experiments
- Table S9. Expression profile of CRYNE genes across RNA-Seq experiments
- Table S10. Expression profile of DICSQ genes across RNA-Seq experiments
- Table S11. Expression profile of FUSGR genes across RNA-Seg experiments
- Table S12. Expression profile of KOMPH genes across RNA-Seq experiments
- Table S13. Expression profile of NEUCR genes across RNA-Seq experiments
- Table S14. Expression profile of POSPL genes across RNA-Seq experiments
- Table S15. Expression profile of PUCST genes across RNA-Seq experiments
- Table S16. Expression profile of PYROR genes across RNA-Seq experiments
- Table S17. Expression profile of SACCE genes across RNA-Seq experiments
- Table S18. Expression profile of SCHPO genes across RNA-Seq experiments
- Table S19. Expression profile of SCLSC genes across RNA-Seq experiments
- Table S20. Expression profile of TRIRE genes across RNA-Seq experiments
- Table S21. Expression profile of YARLI genes across RNA-Seq experiments

## SUPPLEMENTAL FIGURES

Figure S1. Quality control and data selection for the 20 fungi. The number of processed reads (NPR, x-axis) versus the percentage of pseudoaligned reads (PPR, y-axis) are shown for all 20 species, which are arranged in alphabetical order. Each point represents one sample, where black and green indicate samples that failed and passed the quality control, respectively. The histograms to the right and on top indicate the distributions of the PPR and NPR values, respectively. The sample quality control outcomes are shown in Table S1.

**Figure S2. Analysis of specialized metabolism in the 19 fungi.** A) The bar chart indicates the fungi (x-axis) and the number of genes (y-axis) predicted to be involved in specialized metabolism by SMURF. The types of metabolites are indicated by color. B) This panel presents five fungal species that have diverse networks of backbone genes as predicted by SMURF.

Figure S3. Analysis of specialized metabolic hubs in *Aspergillus flavus*. A) Coexpression network of the backbone enzymes involved in specialized metabolism. Nodes represent genes, while co-expressed genes are connected by edges. For brevity, enzymes that are not co-expressed are not shown. The color of the node indicates the type of backbone enzyme. B) Expression profile of *EED48448*, a putative polyketide synthase that shows the highest connection to other backbone enzymes. The different samples are represented on the x-axis, while the gene expression values in the form of Transcripts Per Million (TPM) are indicated on the y-axis. The bars and the dots indicate the average and the minimum/maximum values of the RNA-seq experiments in the sample, respectively. C) Coexpression network of *EED48448*. Nodes represent genes, while edges connect co-expressed genes. Colored shapes indicate which genes contain Pfam domains and orthogroups in common. For brevity, only part of the network is shown. The network contains three types of genes: metabolic enzymes (indicated in the red rectangle), cell division genes (blue rectangle), and metabolite transporters (black rectangle). The function of the genes can be inferred from the legend and by clicking on the nodes.

Figure S4. Cluster analysis involved in hydrolase activity in *Postia placenta*. A) Coexpression network of cluster 59 (<a href="www.fungi.guru/cluster/view/1745">www.fungi.guru/go/view/10627</a>). Nodes represent genes, while edges connect coexpressed genes. Colored shapes indicate which genes contain Pfam domains and orthogroups in common. The network contains five types of genes: cell wall catabolism genes (indicated in a green rectangle), genes encoding for Fenton's Reaction (pink rectangle), genes involved in transport (purple rectangle), gene-regulating genes (blue rectangle) and proteolysis associated proteins (orange rectangle). For brevity, only discussed nodes are shown. B) Expression profile of the cluster 59 from POSPL shows that the cluster has the highest expression during the middle-late decay stage in ATCC 44394 and MAD698-R strains C) Proportion of genes in the gene family OG\_03\_0001073, which is highly abundant in the cluster and which contains genes annotated as glycoside hydrolases and mannosidases.D) Phylogenetic tree of hydrolase gene family OG\_03\_0001073.

