FungiExp: A Comprehensive Platform For Exploring Fungal Gene Expression and Alternative Splicing Based On 35,821 RNA-Seq Experiments From 220 Fungi

Jinding Liu
Nanjing Agricultural University
Fei Yin
Nanjing Agricultural University
Kun Lang
Nanjing Agricultural University
Wencai Jie
Nanjing University
Suxu Tan
Michigan State University
Rongjing Duan
Nanjing Agricultural University
Shuiqing Huang (sqhuang@njau.edu.cn)
Nanjing Agricultural University https://orcid.org/0000-0002-1646-9300

Research

Keywords: Fungi, Gene expression, Alternative splicing, Differential expression analysis, Specific expression analysis

Posted Date: June 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-618004/v1

Abstract

Background: RNA-seq has become a standard tool in biology and has produced large and diverse transcriptomic datasets for users to explore fungal expression regulation. Fungal alternative splicing, which is attracting increasing attention because of evolutionary adaptations to changing external conditions has not been thoroughly investigated in previous studies, unlike that of animals and plants. However, the analyses of RNA-seq datasets are made difficult by the heterogeneity of study design and complex bioinformatics approaches. Comprehensive analyses of these published datasets should contribute new insights into fungal expression regulation.

Results: We have developed a web-based platform called FungiExp hosting fungal gene expression levels and alternative splicing profiles in 35,821 curated RNA-seq experiments from 220 species. It allows users to perform retrieval via diverse terms and sequence similarity. Moreover, users can customize experimental groups to perform differential and specific expression analyses. The wide range of data visualization is an additional important feature that should help users intuitively understand retrieval and analysis results.

Conclusions: With its uniform data processing, easy data accessibility, convenient retrieval, and analysis functions, FungiExp is a valuable resource and tool that allows users to (re)use published RNA-seq datasets. It is accessible at http://bioinfo.njau.edu.cn/fungiExp.

Background

Recent studies have revealed that alternative splicing is involved in many fungal biological functions. More than 50% of the multi-exon genes in *Verticillium dahlia* are reported to have alternative isoforms and its development process involves a large number of alternative splicing events comprising multiple types of single or combinatory alternative splicing patterns [1]. In *Schizohpyllum commune*, alternative splicing has been found to influence multiple development stages that affect several important functional gene groups, such as transcription factors, carbohydrate active enzyme, secreted proteins, cytochrome P450s, and metabolic genes [2]. Fungal alternative splicing has also been shown to be regulated by various environmental factors. In *Monascus pilosus*, its nitrogen source affects the alternative splicing patterns of a methyltransferase gene that has multiple transcript isoforms [3]. The extracellular inorganic phosphate concentration and pH are reported to change the alternative splicing of many genes in *Neurospora crassa* and *Aspergillus nidulans* [4-5]. Furthermore, alternative splicing can regulate the pathogenicity and virulence of pathogenic fungi. For example, the average alternative splicing percentage is significantly higher in pathogenic fungi compared with nonpathogenic fungi [6]. Despite its important regulatory roles in fungi, alternative splicing has not been thoroughly investigated in previous studies, unlike in animals and plants, and its regulatory mechanisms remain unknown.

In the last decade, RNA-seq has become a standard tool in biology and has produced large and diverse fungal transcriptomic datasets for users to explore fungal expression regulation. This vast resource is

tremendously useful for researchers who wish to investigate transcriptional regulation, tissue specificity, stress responses, and developmental dynamics of the genes in which they are interested. However, a challenge is faced by most biological researchers in reusing these RNA-seq data because of the lack of bioinformatics skills and computational resources. Some excellent resources have been developed for researchers to explore plant gene expression. For example, ePlant hosts 1,385 sample data of *Arabidopsis thaliana* for visually exploring the spatial and temporal dynamics of gene expression [7]. The ARS hosts more than 20,000 RNA-seq datasets of *A. thaliana* for users to search gene expression with regard to tissue specificity, developmental stage, and stress-related [8]. In animals, VastDB hosts thousands of RNA-seq datasets from human, mouse, and chicken for researchers to view gene expression and alternative splicing profiles in various tissues [9]. ASlive hosts thousands of RNA-seq datasets from several types of major livestock enabling researchers to interrogate gene alternative splicing profiles [10]. In fungi, FungiDB offers a user-friendly web interface with embedded bioinformatics tools for users to explore genomic, transcriptomic, proteomic and other data from nearly 100 fungi, and supports a Galaxy-based workspace enables users to generate custom pipelines for large-scale RNA-seq data analysis [11]. These resources based on RNA-seq datasets are becoming increasingly useful as an exploratory and hypothesis generating tool, but they do not simultaneously combine gene expression and alternative splicing, and not provide analysis functions based on flexibly customized experimental groups.

As one of main branches of biology, studies of fungi need similar gene expression resources for the scientific community. Unfortunately, up to now, no resource is available for collecting RNA-seq datasets on a large scale for users to explore gene expression and alternative splicing in fungi. Here, we present FungiExp, which is committed to the improvement of gene models including more transcripts and alternative splicing events and to the provision of a collection of gene expression levels and alternative splicing profiles in large scale RNA-seq samples. Moreover, it allows users flexibly to customize various experimental groups in order to perform differential and specific expression analyses. To help users intuitively to understand retrieval and analysis results, FungiExp provides various charts that summarize results and is rich in links to external databases.

Material And Methods

Collection of fungi and corresponding RNA-seq datasets

A total of 220 fungi covering 51 orders were collected in FungiExp, whose genome assemblies and annotations are hosted in the Ensembl database (Release 47) [12]. Only RNA-seq data generated from Illumina platforms were considered because of their ubiquity and high base-calling accuracy. The corresponding RNA-seq raw sequence files were retrieved by searching the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) database using terms Platform='Illumina', Strategy='RNA-seq' and Source='Transcriptomic'. The retrieved RNA-seq datasets were gathered with further manual curation of sample information including strain, tissue, development stage, genotype, and treatment conditions. After

being mapped, a total of 35,821 RNA-seq experiments from 3,052 studies containing 67.8 terabytes bases were adopted in FungiExp (Table S1).

Improvement of gene model annotation

The correct structural annotation of genes is critical for downstream analyses, especially for the identification of alternative splicing events based on read alignments on exons. Except for a few model fungi, the gene model annotations of most fungi are largely incomplete. Only one reference transcript is annotated for each gene. We used the following procedure to improve gene models.

1. High-quality RNA-seq data curated manually were mapped to the reference genome with Hisat2 [13] and generated transcript structures based on read alignments with StringTie2 [14].

2. Only novel multi-exonic transcripts of at least 200 bps in length, an average coverage of 2x per transcript, and an average coverage of 1x per exon were retained for further refining.

3. We further removed the novel transcripts not occurring in at least 1/3 of all experiments and at least 3 experiments.

After improvement, the splice junctions and exons were respectively increased by 7.4% and 10.86% on average (Table S2). The multi-exon genes with alternative transcripts were increased from 0.76% to 15.85% on average (Table S2). The average transcript number per multi-exon gene was increased from ~1 to 1.2. Some fungi, such as *Saccharomyces cerevisiae*, did not give additional alternative transcripts, nearly all their genes being single-exonic. *Ustilaginoidea virens* showed the greatest improvement in gene models, with more than 69% of multi-exon genes having alternative transcripts after the process.

In order to explore cross-species conservation conveniently, we identified orthologous gene groups based on the longest protein sequences of genes by using orthofinder [15].

Identification of alternative splicing events

Fungal alternative splicing events are evolutionary adaptations to changing external conditions and widely interest the fungus research community [16]. In our experiments, rMATs was used to call five types of alternative splicing events according to RNA-seq reads alignments on exons (Figure 1A) [17]. As in plants, RI (retained intron) was the most popular type of alternative splicing event detected in all collected RNA-seq samples, accounting for 58% on average (Figure 1B). The intron chains of 74.35% alternative splicing events conformed to known transcripts, and the remaining events were deduced by rMATS according to read mapping on exons (Figure 1C).

Estimate of gene expression and alternative splicing profiles

StringTie2 was used to detected gene and transcript expression levels based on RNA-seq read alignments. Two expression level metrics such as TPM (transcripts per million) and FPKM (fragments per kilo base per million mapped reads) were adopted in FungiExp.

PSI (percentage spliced in) is used to estimate alternative splicing profiles [18]. It refers to the ratio between reads including or excluding exons and indicates how efficiently sequences of interest are spliced into transcripts. Two metrics are used to estimate alternative splicing including JCEC and JC. For JCEC, PSI values are calculated based on both reads that span junctions defined by rMATS (Junction Counts) and reads that do not cross an exon boundary (Exon Counts). For JC, PSI values are calculated based only on reads that span junctions (Junction Counts) [17].

Differential and specific expression analyses

Two popular methods, DESeq2 [19] and edgeR [20], are integrated into FungiExp for users to perform differential expression analysis. The former supports comparison with sample replicates, whereas the latter supports comparison both with and without sample replicates.

Three statistical models, including rMATS_unpaired, rMATS_paired, and MATS_LRT are used to detect differential alternative splicing by means of read counts on events. In the MATS_LRT model, read counts of all replicates in a group are added together. The rMATS_unpaired model is used for replicates without pairs between groups. The rMATS_paired is employed when each replicate is paired with another between the two sample groups.

For specific expression analysis, users need to customize at least 4 experimental groups to detect specifically expressed genes. If a gene/transcript's expression level is significantly higher/lower in a group than all other groups, the gene/transcript is supposed to be specifically expressed in this group. A similar definition is used for a specific alternative splicing event.

Enrichment analysis

The two enrichment analysis methods, namely the Hypergeometry test and Gene Set Enrichment Analysis, were integrated into FungiExp for users to summarize the function distribution of differentially/specifically expressed genes [21-22]. The former will find enrichment among genes where the difference is large, but it will not detect a situation in which the difference is small but evidenced in a coordinated way in a set of related genes. The latter method addresses this limitation.

Utility

The FungiExp platform was developed based on improved gene models, alternative splicing events, gene/transcript expression levels, and alternative splicing profiles in all collected RNA-seq samples

(Figure 2). The 220 fungi are listed in an interactive table in the home page for users to access. For each fungus, users can download all of its data in the summary page, perform retrieval operations in the search and blast pages, and perform expression analyses in the comparison and specificity pages.

Retrieval

In the search page, users can query genes to display gene/transcript expression levels and profiles of associated alternative splicing events by gene identifier, symbol, and functional annotation terms (Figure 3A). Users can also search genes by a pathway through which relationships among genes can be displayed visually (Figure 3B). In addition, users can specify metrics (TPM or FPKM) to show expression levels and specify a metric (JCEC or JC) to display alternative splicing profiles. Alternatively, genes can be queried by sequence similarity in the blast page (Figure 3C). This is useful when looking for homologous or orthologous genes if they are not easily identified by an identifier or symbols. The retrieved genes are displayed in a concise table with basic information and links to a gene page showing expression levels (Figure 3D).

Display of gene expression and alternative splicing

By means of the query and analysis modules, the genes are listed in an interactive table that includes links for users to open pages showing gene/transcript expression levels and alternative splicing profiles (Figure 4A). The gene, transcript, and alternative splicing pages consist of similar structures.

The basic information in them will be listed at the top of page, including genome position, symbol, and orthologous and various functional annotation terms. Moreover, by clicking the orthologous group ID, users can open a pop-up page to explore orthologous gene expression levels in other species (Figure 4B).

Subsequently, a genome browser allows users to explore gene, transcript, and alternative splicing in the genome context (Figure 4C). Notably, the sequence of interest can be directly obtained from the genome browser.

As a transcript page, the next section is a functional structure graph that illustrates the positions of Pfam domains (Figure 4D).

This is followed by a drop-down list box containing all collected studies and a hierarchical column chart displaying expression levels or alternative splicing profiles in a chosen study. Study information, such as study title, RNA-seq information, and experimental group number, is listed in the drop-down list box. The average expression levels or alternative splicing profiles in the experimental groups are shown in the 1st level column chart. The sample information of the experimental group, such as strain, tissue, genotype, developmental stage and treatment, are displayed in the horizontal ordinate. A click on an experimental

group opens the 2nd column chart, which displays the expression level or alternative splicing profile in each experiment (Figure 4E).

Finally, two tables respectively list the associated transcripts and alternative splicing events with links to pages showing corresponding transcript expression levels or alternative splicing profiles. In transcript and alternative splicing pages, users can open the impact page to show the way that alternative splicing events act on the transcript if available (Figure 4D).

Differential expression analysis

In the comparison page, users can flexibly customize two experimental groups to perform differential expression analyses. In addition, users can select different differential expression/splicing analysis methods and downstream enrichment analysis methods and corresponding parameters. FungiExp supports offline analysis and email notification functions. Users can open the result page to view or download results of analysis by clicking on the link in the notification email.

On the result page, three sections display the differentially expressed genes. The first section contains the principal components analysis (PCA) result and heat map diagrams illustrating experiment clustering based on gene expression levels (Figure 5A). Then, two diagrams present the results for GO and pathway enrichment analysis in differentially expressed genes (Figure 5B). Finally, the differentially expressed genes are listed in an interactive table (Figure 5C). In the table, users can directly obtain important information such as mean expression levels in the control and treatment groups, fold-changed, Q-value, and links for users to open new pages to explore the details of gene expression levels.

Similarly, three sections display the differentially alternative splicing events. The PCA and heat map diagrams are based on alternative splicing PSI values. The GO and pathway enrichment analysis diagrams are based on differentially spliced genes. The differentially alternative splicing events are also listed in an interactive table with links for users to open new pages to explore the details of gene alternative splicing.

Specific expression analysis

In addition to the analysis methods and parameters that are similar in the comparison page, users should customize at least 4 experimental groups to discover differentially expressed or alternative splicing genes. For example, users may want to find the specifically expressed genes in different developmental stages. Indeed, specific expression analysis consists of multiple differential expression analyses. Therefore, compared with differential expression analysis, it tends to consume more time.

On the result page, four sections display the specifically expressed genes. The first section contains the PCA result and heat map diagrams illustrating experiment clustering based on gene expression levels (Figure 6A). Then, a column chart summarizes the specifically expressed gene numbers in each group

(Figure 6B). Next, two tables respectively show enriched GO and pathways for specifically expressed genes (Figure 6C). Finally, the specifically expressed genes are listed in an interactive table (Figure 6D). In the table, users can click links to open new pages to explore the details of gene expression. Similarly, four sections display specifically alternative splicing events.

A case study to explore fungal gene expression changes induced by heat stress

Heat stress affects a broad range of cellular processes that result in cell cycle arrest, damage to membranes and cytoskeletal structures, and accumulation of misfolded proteins [23-24]. We applied FungiExp to a set of RNA-seq datasets (SRP071140) that were obtained from *Fusarium graminearum* under normal and high temperature and released by Duc-Cuong et. al.. FungiExp reported a total of 3,268 differentially expressed genes after heat stress. Among these genes, the change in gene expression level of FGRAMPH1_01G04861 (*FgHSP90*) was validated by quantitative real-time (qRT)-PCR [25]. FungiExp showed that these differentially expressed genes were enriched with respect to unfolded protein binding (GO:0051082), transmembrane transporter activity (GO:0022857), transmembrane transport (GO: 0055085), mismatch repair (GO:0006298), and protein folding (GO:0006457) (Figure 7A-1).

We further explored gene splicing changes after heat stress. From the PCA result and heat maps, the RNA-seq samples were clustered into control and treatment groups based on alternative splicing profiles (Figure 7B). This implied that gene splicing was also affected by heat stress in *F. graminearum*. A total of 235 differential alternative splicing events in 189 genes were identified in which RI events accounted for more than 87% (Figure 7C). In the 189 differentially spliced genes, only 59 genes were significantly changed in their expression levels (Figure 7D). FungiExp showed that the differentially spliced genes were enriched with regard to protein kinase activity (GO:0004672) and protein phosphorylation (GO:0006468) (Figure 7A-2) and to MAPK signaling pathway – yeast (ko04011). Some studies have shown that, in eukaryotic species from yeast to human, stress-activated protein kinases, i.e., members of the MAPK subfamily, regulate the transcriptional response to various environmental stresses [26].

In the differentially spliced genes, the protein product of FGRAMPH1_01G26803 contained two RNA binding domains (PF00076) that are highly abundant in eukaryotes and are found in proteins involved in post-transcriptional gene expression processes including mRNA and rRNA processing, RNA export, and RNA stability [27]. The expression level of this gene was not changed significantly, but the PSI value of an associated alternative splicing event FGRAPH1RI0000000595 was significantly increased by 0.516 after heat stress (Figure 7E). The alternative splicing removed 54 nucleotides from the associated transcript FGRAMPH1_01T26803 and finally caused a shorter distance between the two domains in the corresponding protein product (Figure 7F). The findings of alternative splicing in the regulation gene may provide some cues for further investigation of fungal response to external stress. We believe that the discovery of these events demonstrates the value of FungiExp as an important and necessary complement to the existing RNA-seq studies.

Discussion

In this study, we created the FungiExp platform by integrating 35,821 RNA-seq datasets from 220 fungi. Although the frequency of alternative splicing is not as high in fungi as it is in animals and plants, alternative splicing is increasingly being shown to be important for the development and pathogenicity of fungi or for their response to environmental stress [28]. The ratio of alternative splicing events has been reported to range from 0.2% in the non-pathogenic yeast *Saccharomyces cerevisiae* to 9.5% in the filamentous fungus *Aspergillus niger* [29] and from 2.3% to 18.2% in human pathogenic fungi such as *Rhizopus oryzae* and *Cryptococcus neoformans* [30]. Nevertheless, only one reference transcript is annotated for multi-exon genes in most reference gene models. We have improved gene models to provide users with additional alternative transcripts that should facilitate research on fungal alternative splicing (Table S2). Identification of alternative splicing events tends to be easier than the assembly of full-length transcripts. Hence, we gathered a large number of alternative splicing events to complement alternative transcript annotation (Table S2). Furthermore, with FungiExp, users can intuitively view the impacts of an alternative splicing event on transcripts or the occurrence of associated transcripts (Figure 4D).

Because RNA-seq has been applied for over one decade, the obtained data are large and extremely rich. To enable the sharing or reuse of the data, public repositories have been set up in NCBI, EMBL, and DDBJ to store the Sequence Read Archive (SRA). The rapid growth of RNA-seq data is providing opportunities for their comprehensive use and is leading to the increasing generation of novel hypotheses. To gain better leverage of the data, SRAdb has been built to facilitate access to sequencing data in the SRA [31]. Here, we mainly have curated the RNA-seq samples to provide information with regard to strain, tissue, genotype, development stage, and stress treatment for users to perform comparison between different sample groups. Moreover, in addition to the title given by the releaser, we have also assigned some sample keywords relevant to each study for users to learn the purposes of that study. With these sample information, users can explore gene expression changes under certain contexts in order to deduce their associations with some biological processes.

The rapid processing and analysis of RNA-seq raw data are difficult tasks without bioinformatic skills and abundant computer resources. Despite differences among research aims, many common features exist within data processing and analysis. Hence, FungiExp provides users with common analysis functions, such as differential/specific expression analysis functions and enrichment analysis functions. In addition, multiple tools with similar function are integrated into FungiExp to meet demands of different users and to deal with different data states. For example, in addition to DESeq2, edgeR is integrated into FungiExp to process RNA-seq datasets from early studies without replicates. Not only the Hypergeometry test, but also Gene Set Enrichment Analysis has been adopted to perform enrichment analysis. This means that genes with small differences can be analyzed, and evidence can be presented in a coordinated way for a set of related genes.

RNA-seq data is expected to be increasingly produced from various fungal studies. We will continue to collect new data that will be used regularly to update FungiExp. With the increasing availability of third-generation sequencing technology, we will use long read sequences to improve transcript and alternative splicing annotation further. In addition, we will expand FungiExp to cover more fungi. Finally, another direction that we foresee is the integration of WGS data into our platform to facilitate the interpretation of the association between gene expression and genetic variations.

Conclusions

In summary, FungiExp is a most comprehensive data and analysis platform for fungal gene expression analysis up to now. It allows users to not only search gene expression and alternative splicing profiles, but also perform comparison by flexibly customizing experimental groups to find genes or alternative splicing events with biological significances. Furthermore, it offers various visual charts to help users understand their retrieval and analysis results.

Declarations

Acknowledgements

We acknowledge the works of all the fungal genome and RNA-seq data producers.

Availability of data and materials

All data sets are available at: https://bioinfo.njau.edu.cn/fungiExp.

Authors' contributions

J.L. and S.H. conceived and designed the study; J.L. developed tools, performed analyses, and developed the platform; F.Y., K.L., W.J., S.T., and R.D. curated sample information; J.L. and S.H. wrote the manuscript with input from all authors. All authors read and approved the manuscript.

Funding

This work has been supported by the Fundamental Research Funds for the Central Universities (KYXK2021006) and the National Nature Science Foundation of China (31301691).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Jin L, Li G, Yu D, Huang W, Cheng C, Liao S, Wu Q, Zhang Y. Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus Verticillium dahliae. BMC Genom. 2017;18(1):130.
2. Gehrmann T, Pelkmans J, Lugones L, Wosten H, Abeel T, Reinders M. Schizophyllum commune has an extensive and functional alternative splicing repertoire. SciRep. 2016;6:33640.
3. Zhang M, Miyake T. Development and media regulate alternative splicing of a methyltransferase pre-mRNA in Monascus pilosus. J Agric Food Chem. 2009;57(10):4162–7.
4. Mendes N, Silva P, Silva-Rocha R, Martinez-Rossi N, Rossi A. Pre-mRNA splicing is modulated by antifungal drugs in the filamentous fungus Neurospora crassa. FEBS Open Bio. 2016;6(4):358–68.
5. Rossi A, Cruz A, Santos R, Silva P, Silva E, Mendes N, Martinez-Rossi N. Ambient pH sensing in filamentous fungi: pitfalls in elucidating regulatory hierarchical signaling networks. IUBMB Life. 2013;65(11):930–5.
6. Gr€utzmann K, Szafranski K, Pohl M, Voigt K, Petzold A, Schuster S. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. DNA Res. 2014;21(1):27–39.
7. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, et al. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell. 2017;29(8):1806–21.
8. Zhang H, Zhang F, Yu Y, Feng L, Jia J, Liu B, Li B, Guo H, Zhai J. A Comprehensive Online Database for Exploring approximately 20,000 Public Arabidopsis RNA-Seq Libraries. Mol Plant. 2020;13(9):1231–3.
9. Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, Quesnel-Vallieres M, Permanyer J, Sodaei R, Marquez Y, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res. 2017;27(10):1759–68.
10. Liu J, Tan S, Huang S, Huang W. ASlive: a database for alternative splicing atlas in livestock animals. BMC Genom. 2020;21(1):97.

11. Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert CJJ, Kissinger JC, Roos DS, Hertz-Fowler C. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. JFungi(Basel). 2018;4(1):39.
12. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. 2016;44(D1):D574–80.
13. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
14. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20(1):278.
15. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
16. Grutzmann K, Szafranski K, Pohl M, Voigt K, Petzold A, Schuster S. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. DNA Res. 2014;21(1):27–39.
17. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A. 2014;111(51):E5593–601.
18. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). Curr Protoc Hum Genet. 2015; 87:11 16 11–11 16 14.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
21. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015;31(4):608–9.
22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50.
23. Vidair CA, Huang RN, Doxsey SJ. Heat shock causes protein aggregation and reduced protein solubility at the centrosome and other cytoplasmic locations. Int J Hyperthermia. 1996;12(5):681–95.
24. Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death. Mol Cell. 2010;40(2):253–66.
25. Bui DC, Lee Y, Lim JY, Fu M, Kim JC, Choi GJ, Son H, Lee YW. Heat shock protein 90 is required for sexual and asexual development, virulence, and heat shock response in Fusarium graminearum. Sci Rep. 2016;6:28154.

26. Nguyen AN, Shiozaki K. Heat-shock-induced activation of stress MAP kinase is regulated by threonine- and tyrosine-specific phosphatases. Genes Dev. 1999;13(13):1653–63.
27. Birney E, Kumar S, Kariner AR. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res. 1993;21(25):5803–16.
28. Kempken F. Alternative splicing in ascomycetes. Appl Microbiol Biotechnol. 2013;97(10):4235–41.
29. Sibthorp C, Wu H, Cowley G, Wong PW, Palaima P, Morozov IY, Weedall GD, Caddick MX. Transcriptome analysis of the filamentous fungus Aspergillus nidulans directed to the global identification of promoters. BMC Genom. 2013;14:847.
30. Sieber P, Voigt K, Kammer P, Brunke S, Schuster S, Linde J. Comparative Study on Alternative Splicing in Human Fungal Pathogens Suggests Its Involvement During Host Invasion. Front Microbiol. 2018;9:2313.
31. Zhu Y, Stephens RM, Meltzer PS, Davis SR. SRAdb: query and use public next-generation sequencing data from within R. BMC Bioinformatics. 2013;14:19.

Figures

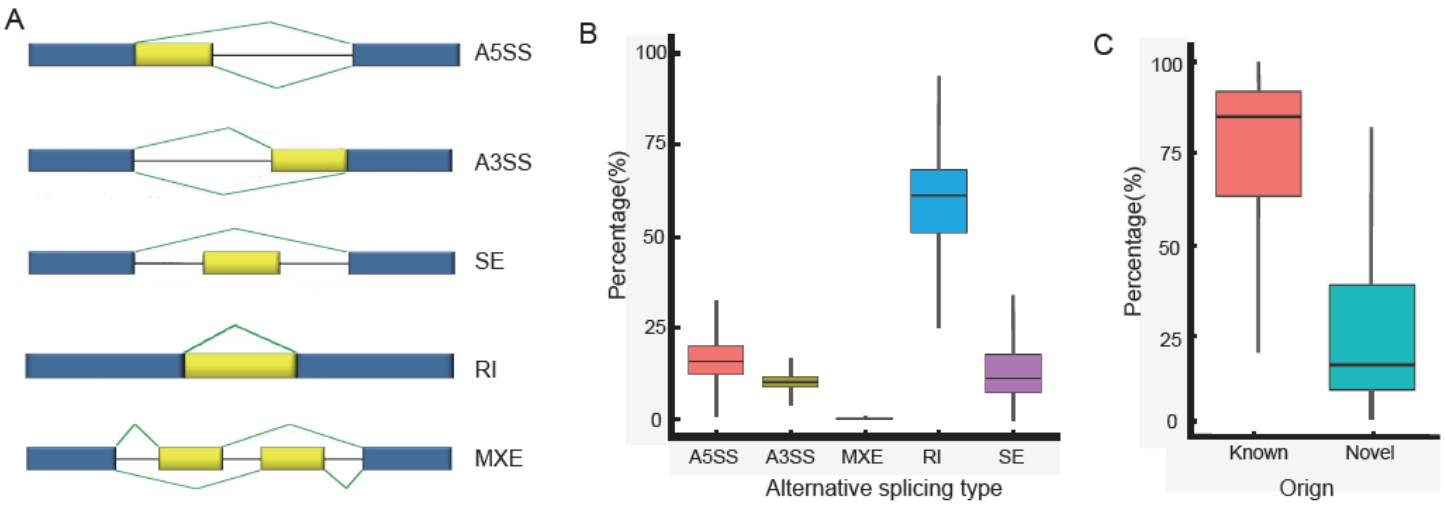


Figure 1

Summary of alternative splicing events. (A) Structure of five alternative splicing types. A5SS: Alternative 5' splice sites (A5SS), A3SS: Alternative 3' splice sites, RI: Retained intron, SE: Skipped exon, MXE: Mutually exclusive exon. (B) Percentage of five alternative splicing events in collected RNA-seq experiments. (C) Percentage of known alternative splicing events from GTF annotation (Known) and novel events derived from read mapping (Novel).

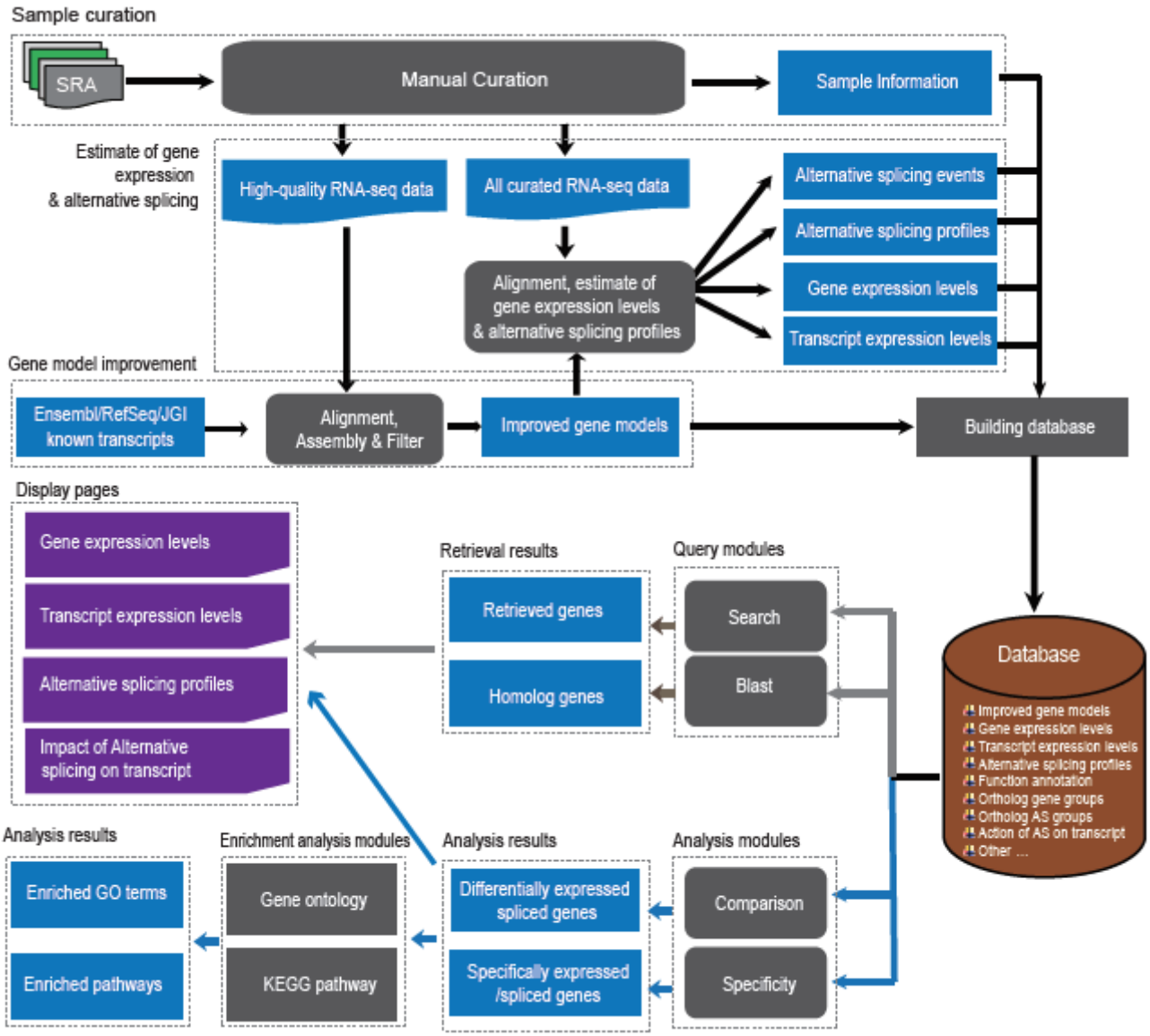


Figure 2

Flowchart of construction of platform. (1) RNA-seq samples are curated manually. (2) High-quality RNA-seq data are used for gene model improvement. (3) Estimate of gene/transcript expression levels and gene alternative splicing. (4) Database includes gene model, gene/transcript expression levels, alternative splicing events, alternative splicing profiles and other information. (5) Two query modules are developed for gene retrieval. (6) Two analysis modules are developed to perform differential and specific gene expression analyses. (7) Four pages are designed to display gene/transcript expression levels, alternative splicing profiles, and action of alternative splicing on transcript. (8) Enrichment analysis models are developed to summarize function distribution of differentially and specifically expressed genes.

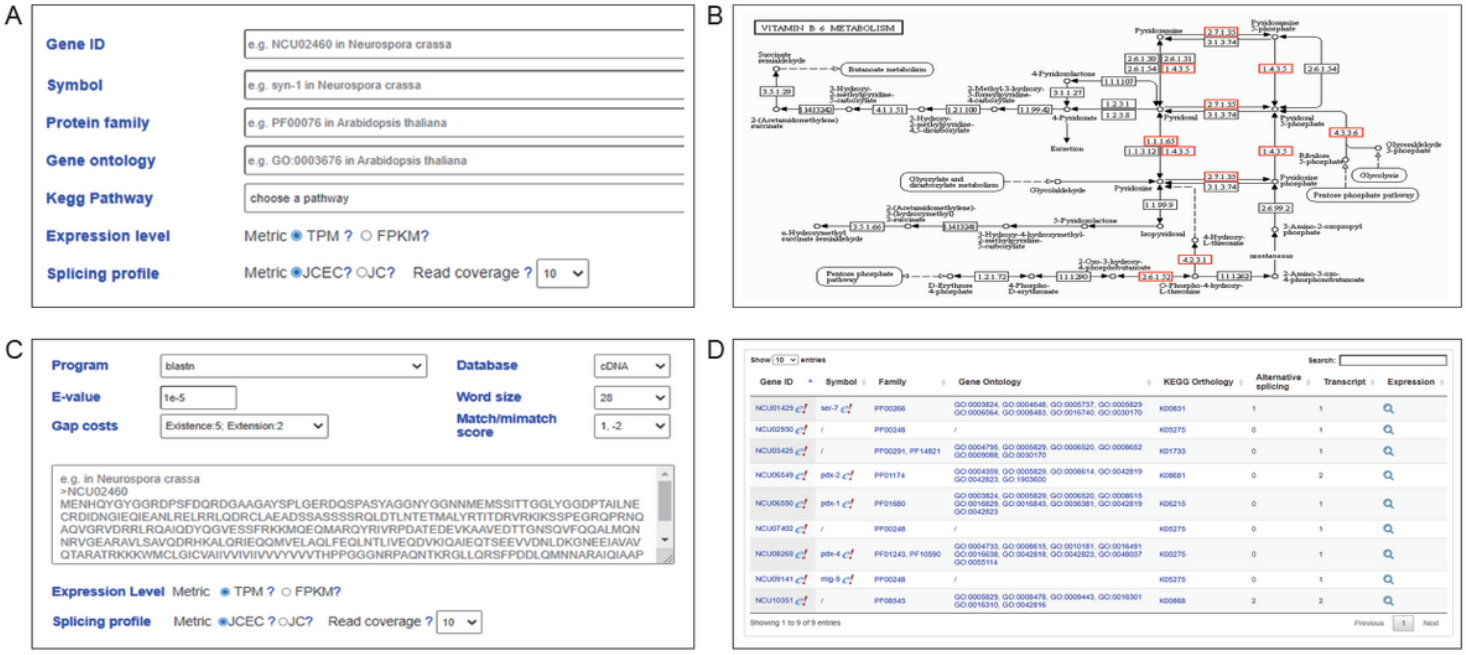


Figure 3

Search entrances and result. (A) Search genes by identifier, symbol, and functional annotation terms. (B) Diagram of a retrieved pathway. Red nodes indicate mapped genes. (C) Search genes by Blast program. (D) Retrieval genes listed in a table with links to gene pages showing gene expression details.

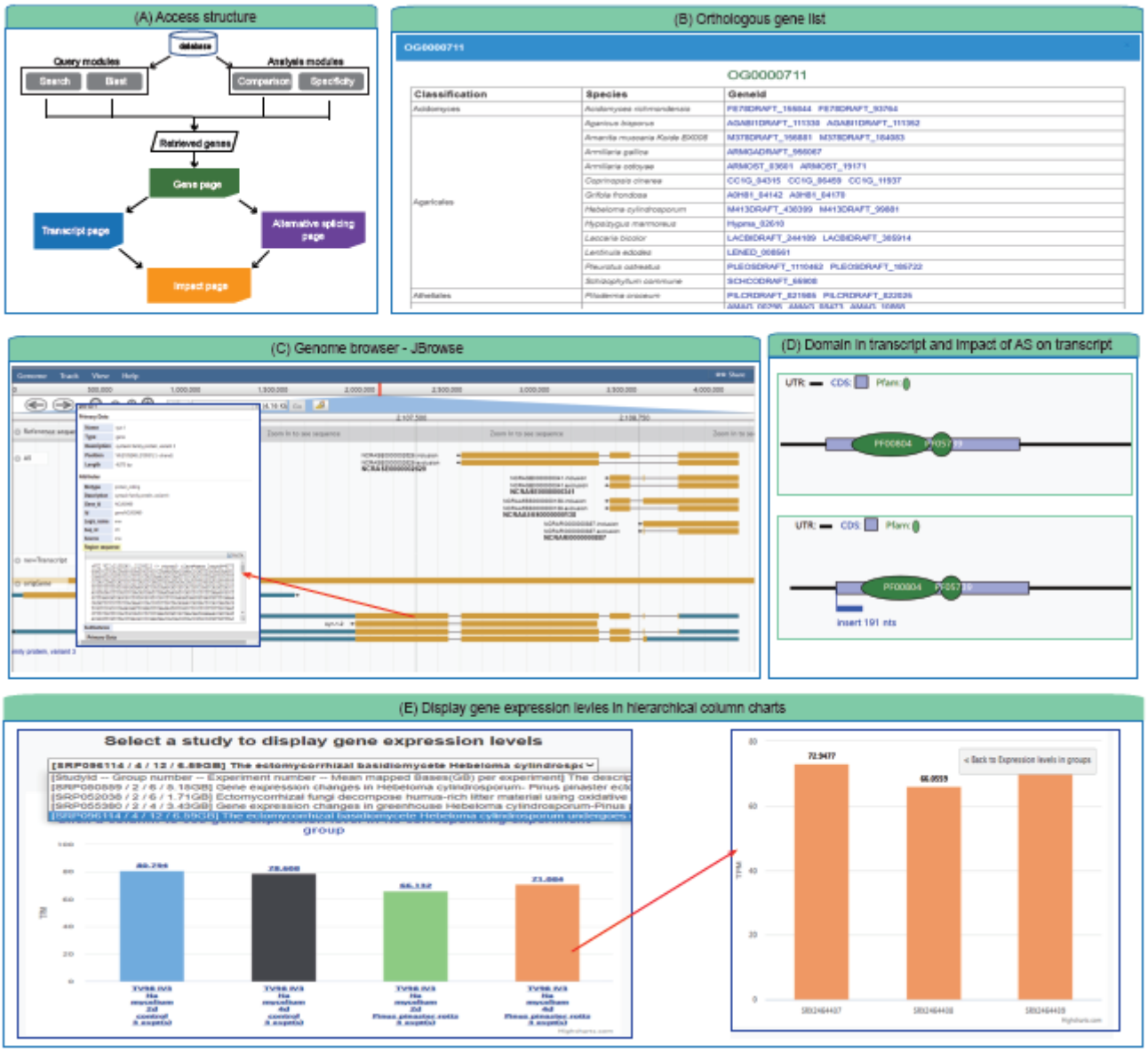
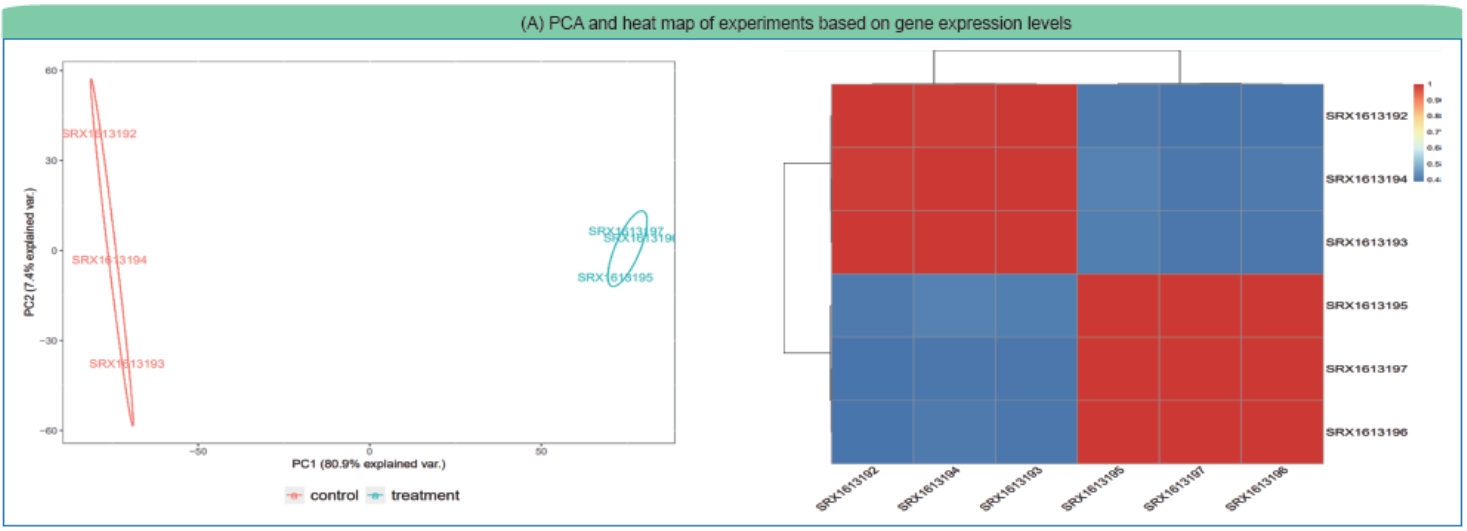
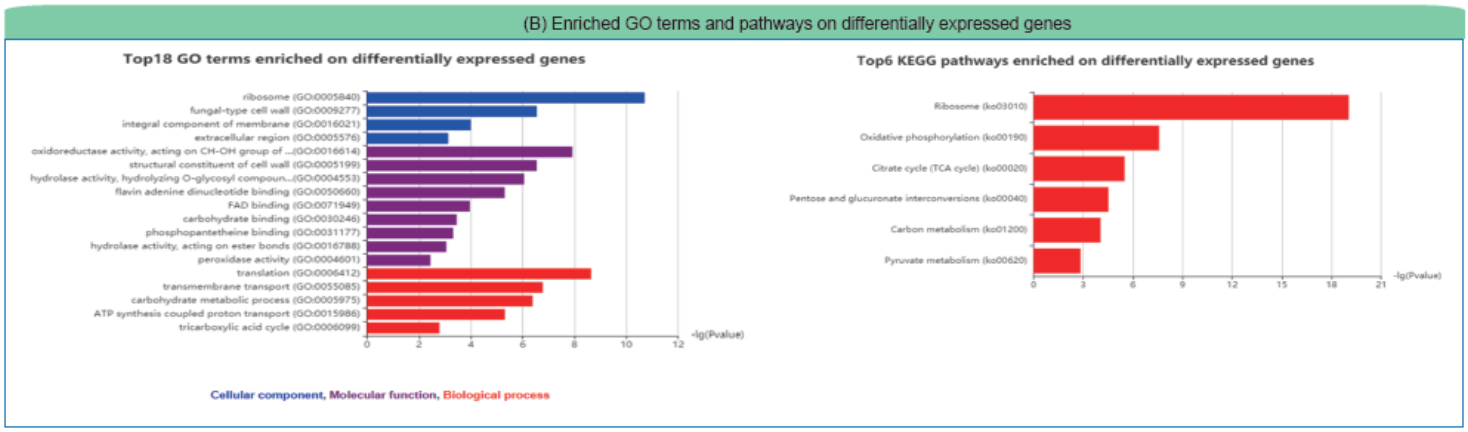


Figure 4

Access to gene/transcript expression or alternative splicing page. (A) Access to each other from gene, transcript, alternative splicing and impact pages. (B) Orthologous genes listed by orders. (C) Genome browsing for gene, transcript, and AS in JBrowse. (D) Functional domain in transcript. (E) Study list and display of gene/transcript expression or alternative splicing profiles in hierarchical column charts.





(C) List of differentially expressed genes

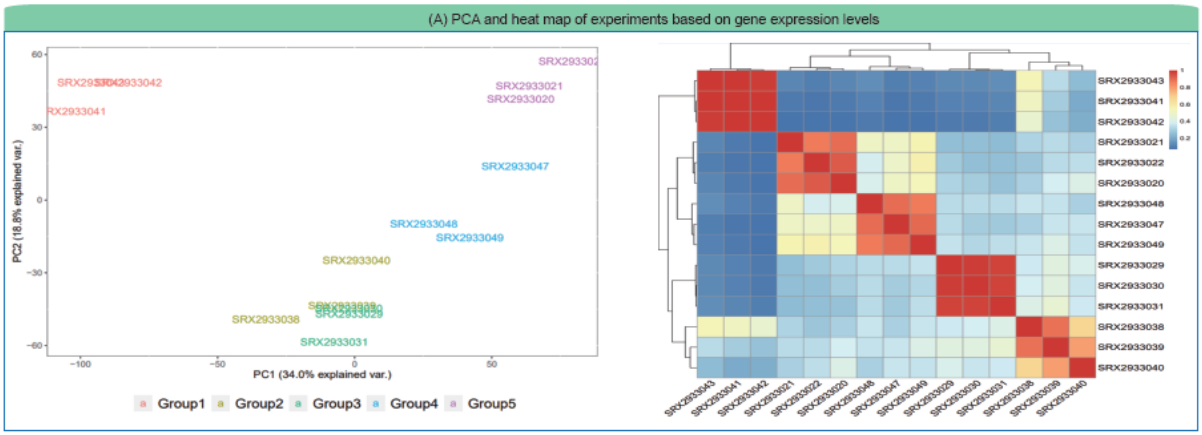
Show 10 entries Search:

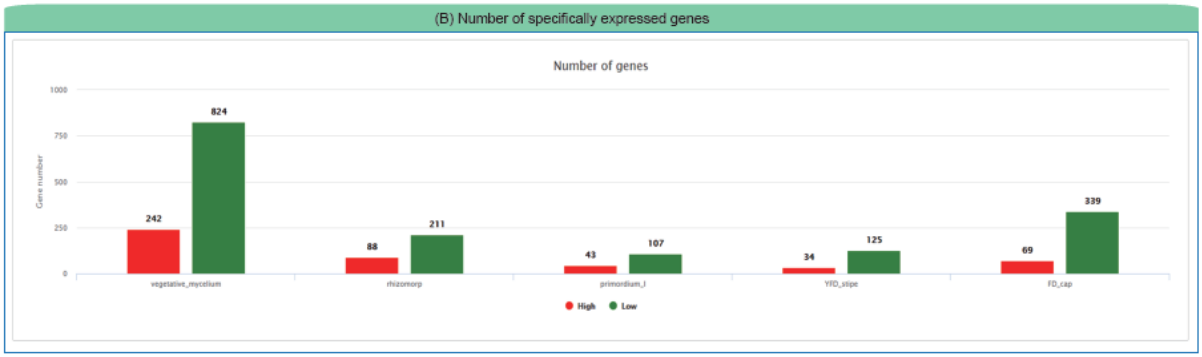
Gene	Symbol	Family	Gene ontology	KEGG	Control	Treatment	Log2FC	Q-value	Comparison
ARMOST_09560	/	PF07690	GO:0016021, GO:0055085	/	29.537	0.000	-12.881	0.0000	
ARMOST_08540	/	PF00199	GO:0004096, GO:0020037, GO:0055114	/	21.268	0.000	-12.617	0.0000	
ARMOST_11428	/	/	/	/	10.858	0.000	-11.952	0.0000	
ARMOST_01275	/	/	/	/	124.407	0.104	-10.253	0.0000	
ARMOST_01386	/	/	/	/	1379.577	1.379	-10.220	0.0000	
ARMOST_18883	/	/	/	/	132.971	0.121	-10.111	0.0000	
ARMOST_22359	/	PF12937	GO:0005515	/	19.343	0.019	-9.886	0.0000	
ARMOST_02377	/	PF05699	GO:0046983	/	60.110	0.072	-9.827	0.0000	
ARMOST_20028	/	/	/	/	32.883	0.087	-8.716	0.0000	
ARMOST_17402	/	/	/	/	18.839	0.043	-8.588	0.0000	

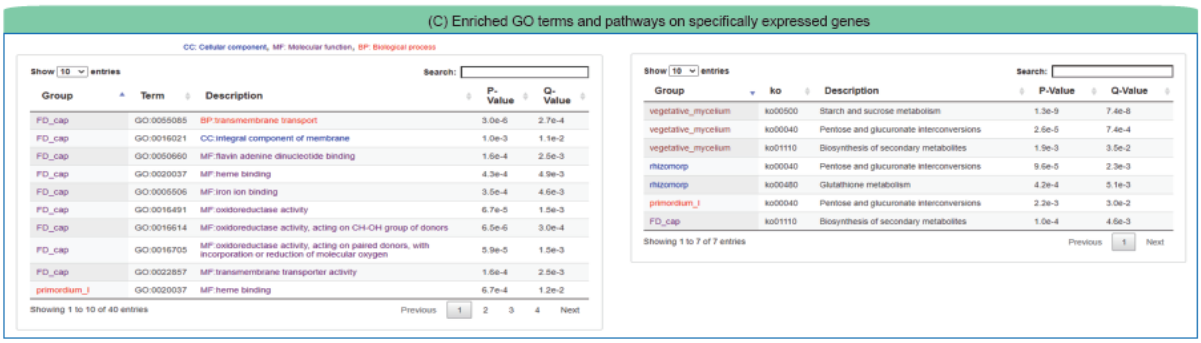
Showing 1 to 10 of 3,460 entries Previous 1 2 3 4 5 ... 346 Next

Figure 5

Comparison between control and treatment experiment groups. (A) PCA and heat map of experiments grouped by gene expression levels. (B) Enriched GO terms and pathway for differentially expressed genes. (C) List of differentially expressed genes.







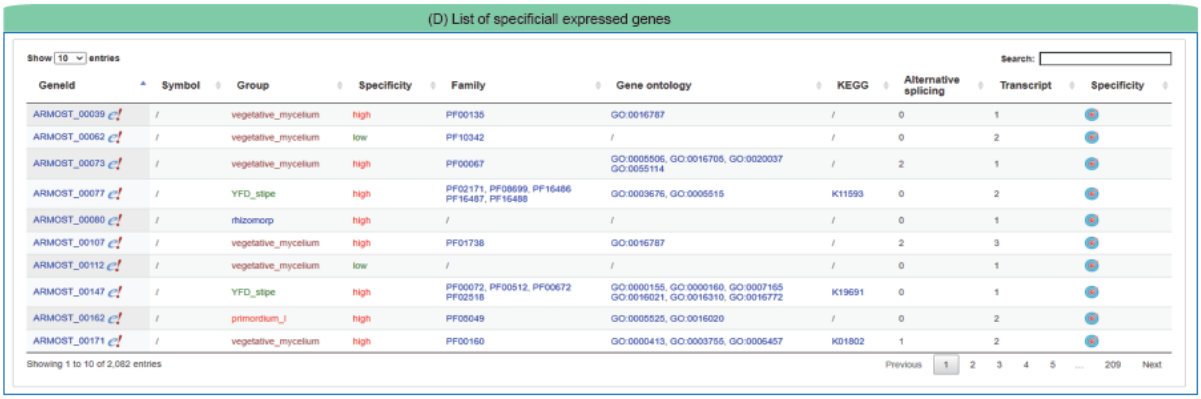


Figure 6

Specific expression analysis. (A) PCA and heat map of experiments grouped by gene expression levels. (B) Column chart summarizing specifically expressed genes in each experimental group. (C) List of enriched GO terms and pathways in specifically expressed genes. (D) List of specifically expressed genes.

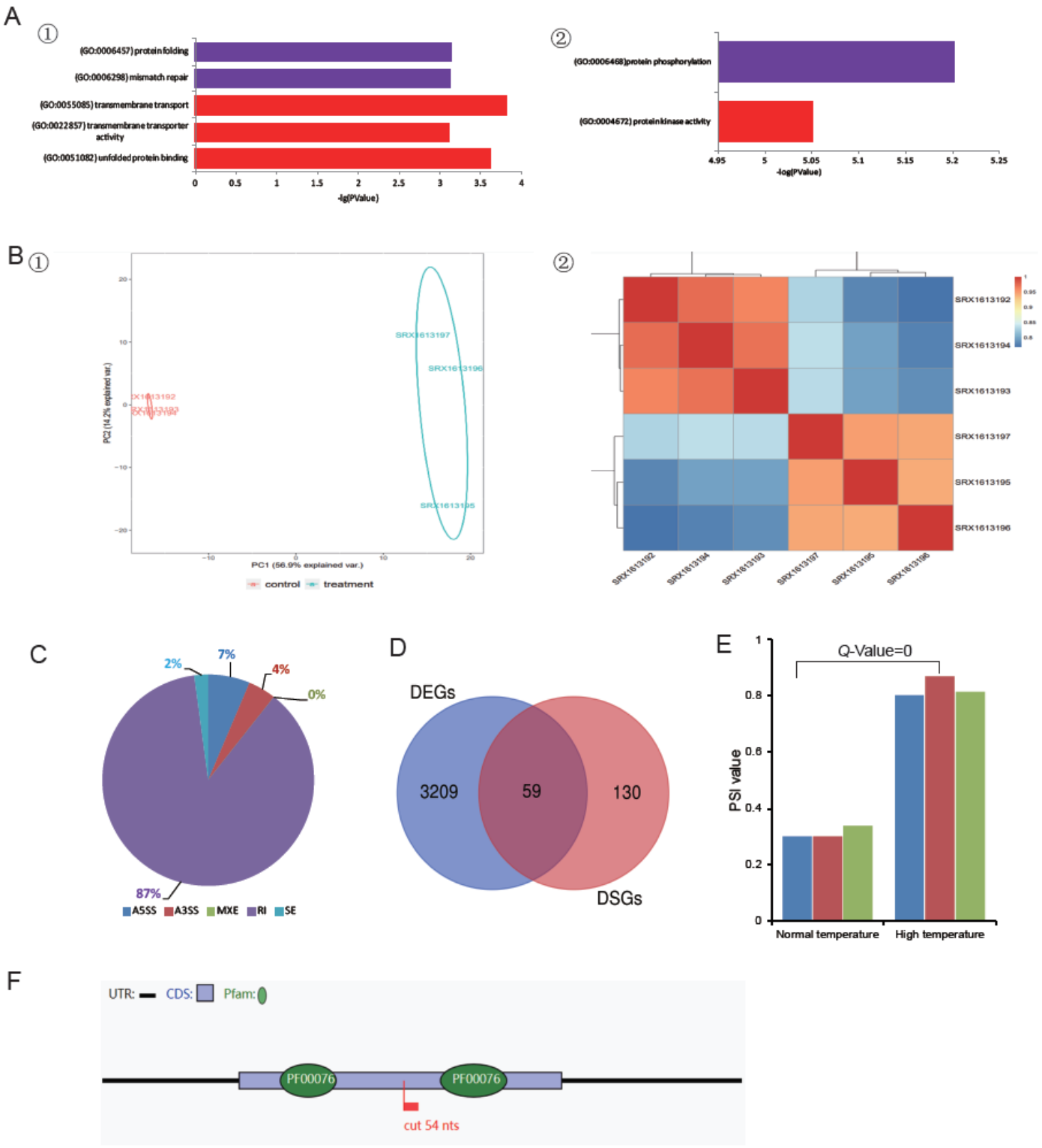


Figure 7

Differential expression analyses of RNA-seq datasets from F. graminearum with heat stress. (A) Enriched GO terms for differentially expressed genes (DEGs) (●1) and differentially spliced genes (DSGs) (●2). Purple indicates molecular function. Red indicates biological process. (B) Experimental clusters based on alternative splicing profiles by using PCA (●1) and heat maps (●2). (C) Percentage of event types of differential alternative splicing events after heat stress. (D) The Venn diagram shows the number of

genes shared between differentially expressed and spliced genes. (E)The PSI values of the alternative splicing event FGRAPH1RI0000000595 in normal and high temperature samples. (F) Impact of the alternative splicing event FGRAPH1RI0000000595 on its associated transcript.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tables12.xlsx