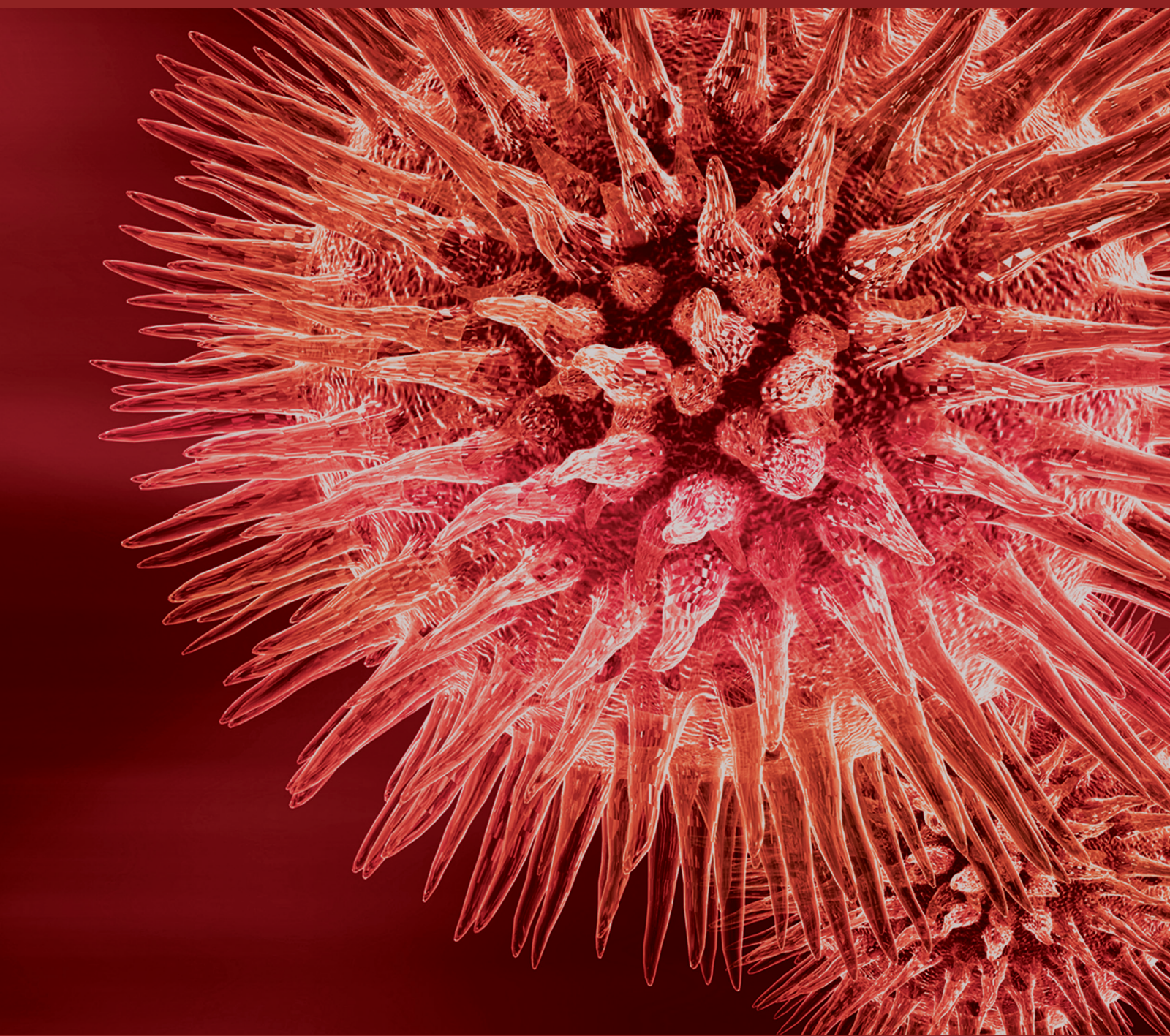
BioMed Research International

Bioinformatic Approaches for Fungal Omics

Guest Editors: Guohua Xiao, Xinyu Zhang, and Qiang Gao



Bioinformatic Approaches for Fungal Omics

BioMed Research International

Bioinformatic Approaches for Fungal Omics

Guest Editors: Guohua Xiao, Xinyu Zhang, and Qiang Gao

Contents

Hindawi
BioMed Research International
Volume 2017, Article ID 7270485, 1 page
http://dx.doi.org/10.1155/2017/7270485

Editorial

Bioinformatic Approaches for Fungal Omics

Guohua Xiao,[1] Xinyu Zhang,[2] and Qiang Gao[3]

[1]*School of Computer Science, Fudan University, Shanghai 200433, China*
[2]*Department of Psychiatry, Yale Medical School, Yale University, New Haven, CT 06511, USA*
[3]*Department of Pharmacology, University of Texas, Southwestern Medical Center, Dallas, TX, USA*

Correspondence should be addressed to Guohua Xiao; guohua.xiao@gmail.com

Received 7 February 2017; Accepted 7 February 2017; Published 15 March 2017

Fungal omics (including genomics, transcriptomics, metabolomics, proteomics, and lipidomics) are now broadly applied to help understand both the basic fungal biology and associated applications. Due to the advantages of the rapid development of Next-Generation Sequencing (NGS) technologies, bioinformatics algorithms, and the relatively smaller sizes of fungal genomes compared with other eukaryotes, sequencing and analysis of fungal genomics become much easier. Along with the acquisition of fungal genomic data, other omic data, such as proteomic data, have been increasingly reported.

Therefore, appropriate data mining of these omic data in depth and the obtained information can benefit our understanding of the complex fungal biological processes from genotype and physiology to phenotype, including cell-cell (microbial) communications and pathogen-host interactions and beyond. The research papers published in this special issue represent recent progress in the aspects, including RNA sequencing analysis, comparative proteomics, metagenomics, developments of new computational workflows, and fungal physiology and biology revealed by bioinformatic approaches. All of these papers provide novel ideas and technologies in the field and stimulate future research for fungal omics.

Xyr1 is one of the main transcription activators of (hemi)cellulases in the well-known cellulase producer *Trichoderma reesei*. L. Ma et al. identified the genes regulated by Xyr1 through RNA sequencing. Their results and analysis might help elucidate the regulation system for synthesis and secretion of (hemi)cellulases in *T. reesei*.

L. H. Tang et al. performed a comparative proteomic analysis of mycelial brown film formation in *Lentinula edodes*. These results provided useful help for future detailed investigations of the proteins linked to brown film formation.

J. Tang et al. analyzed the bacterial communities of two soy sauce aroma liquors. They provided new insights into the bacterial composition of the Chinese liquor Daqu and the fermentation process using high-throughput sequencing technology. At last, 17 phyla species were obtained from the Daqu samples, which were considered to be useful in the manufacture of Chinese liquors.

M. Převorovský et al. developed a computational workflow for the calculation of genome-wide splicing efficiency in *Saccharomyces cerevisiae* using strand-specific RNA-seq data. They demonstrated the functionality of the workflow using RNA-seq datasets from three spliceosome mutants and provided all relevant scripts in a ready-to-use form.

Y. Wang et al. compared the putative pathogenicity-related genes identified by T-DNA insertional mutagenesis (comprising 1024 genes) with the genes by microarray expression profiling (comprising 236 genes) in *Magnaporthe oryzae*. Only 13 genes were overlapped between the two gene lists. Their results of gene knockout mutants were also negative.

D. Lü et al. applied molecular cloning in combination with bioinformatic analysis to explore the function of *Bombyx mori* Lebocin 5 Gene. Their results showed that *Bombyx mori* Lebocin 5 Gene might play an important role in the immune response of silkworm to defend *B. bassiana* infection.

The guest editors are grateful to all authors and reviewers for their contributions to this special issue.

Guohua Xiao
Xinyu Zhang
Qiang Gao

Hindawi
BioMed Research International
Volume 2017, Article ID 6271358, 9 pages
http://dx.doi.org/10.1155/2017/6271358

Research Article

Analysis of the Bacterial Communities in Two Liquors of Soy Sauce Aroma as Revealed by High-Throughput Sequencing of the 16S rRNA V4 Hypervariable Region

Jing Tang,[1,2] Xiaoxin Tang,[1,2] Ming Tang,[1,2] Ximin Zhang,[1,2] Xiaorong Xu,[1,2] and Yin Yi[1,2]

[1] *The Key Laboratory of Plant Physiology and Development in Guizhou Province, Guizhou Normal University, Guiyang, Guizhou Province 550001, China*
[2] *College of Bioscience, Guizhou Normal University, Guiyang, Guizhou Province 550001, China*

Correspondence should be addressed to Yin Yi; gzklppdr@gznu.edu.cn

Received 18 August 2016; Accepted 15 January 2017; Published 28 February 2017

Academic Editor: Xinyu Zhang

Chinese liquor is one of the world's oldest distilled alcoholic beverages and an important commercial fermented product in China. The Chinese liquor fermentation process has three stages: making *Daqu* (the starter), stacking fermentation on the ground, and liquor fermentation in pits. We investigated the bacterial diversity of Maotai and Guotai *Daqu* and liquor fermentation using high-throughput sequencing of the V4 hypervariable region of the 16S rRNA gene. A total of 70,297 sequences were obtained from the *Daqu* samples and clustered into 17 phyla. The composition of the bacterial communities in the *Daqu* from these two soy sauce aroma-style Chinese liquors was the same, although some bacterial species changed in abundance. Between the *Daqu* and liquor fermentation samples, 12 bacterial phyla increased. The abundance of *Lactobacillus* and *Pseudomonas* increased in the liquor fermentation. This study has used high-throughput sequencing to provide new insights into the bacterial composition of the Chinese liquor Daqu and fermentation. Similarities in the distribution of bacteria in the soy sauce aroma-style Chinese liquors *Daqu* suggest that the abundance of bacteria might be generally concerned to other liquor.

1. Introduction

Fermentation is a well-known ancient technique that uses microorganism to process and preserve food. Chinese liquor is one of the six well-known distillates in the world. It has a long history of production and is produced through unique a fermentation process. It is typically produced from cereals, such as sorghum and rice, via the solid-state fermentation of grain. Chinese liquor has five main styles: strong aroma, light aroma, soy sauce aroma, sweet honey, and miscellaneous. *Maotai* and *Guotai* are Chinese liquors famous for their soy sauce aroma. *Maotai-flavor* liquor is as symbolic a drink in China as whisky is in Scotland and brandy in France [1–3].

Microorganisms usually do the main work of degrading biopolymers, producing alcohol, and forming aromatic compounds. The microbial community of Chinese liquor has been analyzed in previous studies using culture-dependent and culture-independent methods. Culture-dependent studies of the microbial community have used methods such as isolation and enumeration on selective media [4–6]. Culture-independent studies have used methods such as polymerase chain reaction-denaturing gradient gel electrophoresis (PCR-DGGE) [7–9], amplified fragment length polymorphism [10], and 16S RNA or 26S RNA clone libraries [11].

Chinese liquors are typically produced via solid-state fermentation using a natural fermentation starter termed *Daqu*. The *Daqu* starter has long been believed to play a key role in the fermentation of Chinese liquor [12]. *Maotai* and *Guotai* liquors share the same unique and complicated spontaneous fermentation process, which includes making *Daqu* (the starter), stacking fermentation, and liquor fermentation (Figure 1) [13]. In the *Daqu*-making stage, the maximum temperature of *Daqu* reaches approximately 65°C,

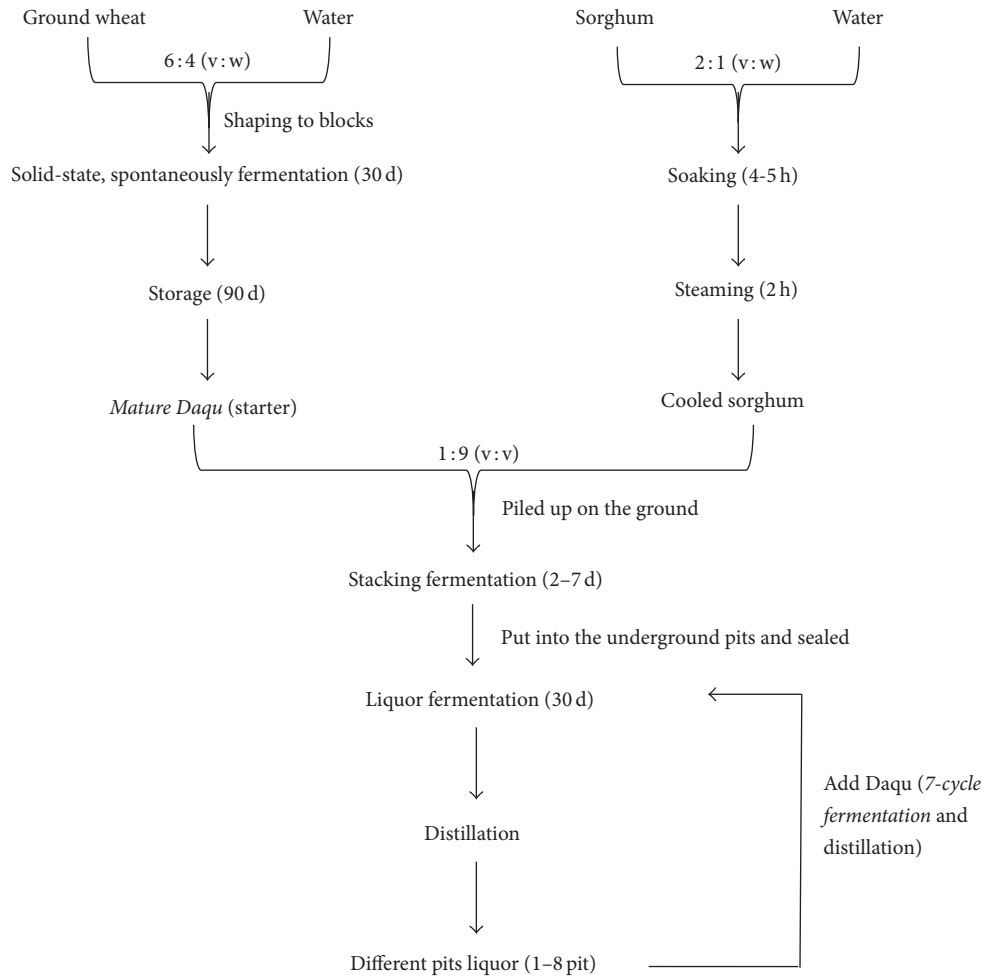


Figure 1: Flow sheet for making soy sauce aroma-style flavor liquor.

which is typical of high-temperature *Daqu*. Several studies have shown the diversity of the microbial community in *Daqu* or fermentation, especially the yeast and fungi [2, 3, 13, 14]. Therefore, it was of interest to analyze the microbial community in *Daqu* and find the bacteria important to *Daqu* properties. Little is known regarding the bacterial community composition in *Maotai Daqu*.

In previous research, culture-independent cloning methods were used for the analysis of microbial communities in *Daqu* or Chinese liquor fermentation. Here, we applied high-throughput sequencing of the V4 hypervariable region of the 16S rRNA gene to examine in-depth microbial communities from soy sauce aroma-style Chinese liquor to gain insight into the specific fermentative microorganisms. The main objective of this study was to (i) analyze the composition of the microbial communities in the *Daqu* of soy sauce aroma-style Chinese liquor; (ii) compare them with two different *Daqu* of soy sauce aroma-style Chinese liquor; and (iii) compare the composition of the microbial communities in the *Daqu* with those in the liquor fermentation.

2. Materials and Methods

2.1. Samples of Daqu and Fermented Grains. Sampling was performed in two different liquor production factories (*Maotai* and *Guotai*) in Guizhou Province, China. Three pits of *Daqu* samples from *Maotai* and *Guotai* were taken in the same years from 2011 to 2013. Samples of fermented grains were obtained from a randomly selected fermentation batch at the same time points (the fourth liquor fermentation) in 2013. The sampling was randomly selected from the mixture of the upper, middle, and bottom stacked layers. All samples were transferred to sterile bags, sealed, and stored at −80°C.

2.2. DNA Extraction and Quantitation. 10 g samples were suspended in 50 mL of sterile PBS buffer (0.1 mol/L, pH 7.2–7.4) and vortexed for 15 min at ambient temperature. The suspension was then centrifuged (500 ×g, 4°C) for 5 min, and the pellet was washed three times in PBS buffer. The supernatants were also collected and centrifuged (10,000 ×g, 4°C) for 10 min; the resulting pellets were washed three times in PBS buffer. The resulting pellets were resuspended in PBS

buffer and stored at −20°C until DNA extraction. DNA was extracted according to the method of Wang et al. [2, 3].

2.3. MiSeq Sequencing of 16S rRNA Gene Amplicons. The communities of bacteria were analyzed using Illumina MiSeq sequencing of the 16S rRNA gene V4 region amplicons, which can yield accurate taxonomic information and shows few biases for various bacterial taxa [15]. The V4 region of the 16S rRNA gene was amplified with the primer set 515f (5′-GTGCCAGCMGCCGCGGTAA-3′)/806r (5′-GGACTACHVGGGTWTCTAAT-3′), and all PCR amplifications were conducted in triplicate for each sample. This short targeted gene region can provide sufficient resolution for the accurate taxonomic classification of microbial sequences [16]. The initial 10 cycles of PCR amplification were performed. The products were then purified with Agencourt® Ampure® XP (Beckman Coulter, Inc., CA, USA) and used as a template for the second PCR amplification of 20 cycles using the same primer set; however, the reverse primer contained the appropriate adapters and different barcodes to distinguish samples. PCR products were visualized using 1% agarose gels stained with ethidium bromide, and negative controls were always performed to confirm the absence of contamination. True positive amplicons were quantified using a PicoGreen dsDNA Assay kit (Invitrogen, CA, USA), combined equally, and then gel purified. The DNA library was sequenced using the Illumina MiSeq platform according to the manufacturer's instructions [17]. Sequences were analyzed with the QIIME [18] software package and UPARSE pipeline [19]. Quality filtering and processing of MiSeq reads were conducted by QIIME. Default settings for Illumina processing in QIIME were used ($r = 3$ $p = 0.75$ total read length; $q = 3$; $n = 0$) (p: minimum number of consecutive high-quality base calls to retain read; r: maximum number of consecutive low-quality base calls allowed before truncating a read; n: maximum number of ambiguous (N) characters allowed in a sequence; q: last quality score considered low quality). We ultimately obtained 10,083–17,973 high-quality sequences from the *Daqu* samples and 12,418–15,302 sequences from the liquor fermentation samples. Then we use UPARSE pipeline to picking operational taxonomic units (OTUs) through making OTU table. Sequences were assigned to OTUs at 97% similarity. We pick representative sequences for each OTU and use the RDP classifier [20] to assign taxonomic data to each representative sequence.

2.4. Statistical Analysis. The datasets generated using 16S rRNA gene sequencing (OTU composition) were further analyzed with the following statistical methods: (i) α-diversity comparison (Chao value and Shannon index) and β-diversity comparison (unweighted UniFrac distances and weighted UniFrac distances); (ii) hierarchical clustering based on the relative abundance of bacteria (the specificity measure (SPM)) and Euclidean distance and complete linkage being used in this clustering analysis; and (iii) significance tests based on unpaired Student's t-tests and Wilcoxon rank-sum test to identify differences between any two compared objects. All statistical analyses described above were performed using the R package vegan/gplot. To quantitatively estimate the relative abundance of a bacterial genus in a sample, the specificity measure (SPM) [21] was introduced as follows and was used in a Heatmap. Each abundance of a bacterial genus was first transformed into vector X:

$$X = (x_1, x_2, \ldots, x_i, \ldots, x_{n-1}, x_n), \quad (1)$$

where n is the number of samples in a profile. At the same time, vector X_i was created to represent the abundance of a bacterial genus in sample i:

$$X_i = (0, 0, \ldots, x_i, \ldots, 0, 0). \quad (2)$$

The SPM of a bacterial genus in a sample was then determined by calculating the cosine value of intersection angle θ between vectors X_i and X in high-dimension feature space:

$$\mathrm{SPM}_i = \cos\theta = \frac{X_i \cdot X}{|X_i| \cdot |X|}, \quad (3)$$

where $|X_i|$ and $|X|$ are the length of vectors X_i and X, respectively. The value of SPM ranges from 0 to 1.0. A SPM value close to 1.0 indicates the major contribution of bacterial abundance in a designated sample (e.g., vector X_i) against that in all samples (vector X).

3. Results

3.1. Composition of Bacterial Communities in Different Daqu Determined Using High-Throughput Sequencing. A total of 70,297 high-quality sequences (approximately 260 bp) were obtained from the 6 *Daqu* samples, with an average of 11,716 sequences per sample. There were no significant differences between the *Maotai Daqu* high-quality sequences with the *Guotai Daqu* (t-test, $p = 0.530$). The OTUs detected in the *Guotai Daqu* largely overlapped with the *Maotai Daqu*; the number of OTUs in each of the 6 samples (average 5,418) was similar (t-test, $p = 0.691$). These OTUs clustered into 17 phyla, 55.5% of which, on average, were classified as Proteobacteria, followed by Firmicutes 39.1% (Figure 2). In terms of relative abundance, Proteobacteria (mainly Gammaproteobacteria) were the most abundant bacterial phylum in the Chinese liquor *Daqu* (accounting for 37.16–64.07% of the different *Daqu* samples), and Firmicutes (mainly Bacilli) were also abundant (accounting for 28.69–61.80%; Figure 3). There were no significant differences (t-test, $p > 0.05$) in the most bacterial relative abundances with two *Daqu* samples. However, Actinobacteria and Bacteroidetes were significantly different (t-test, $p < 0.05$). Of all the bacterial families, Bacillaceae, Enterobacteriaceae, and Pseudomonadaceae were the most abundant.

The Chao value and Shannon index, which reflect the α-diversity of bacterial communities, showed no significant differences between the *Guotai Daqu* and the *Maotai Daqu* (t-test, $p = 0.140$ and $p = 0.117$, and Wilcoxon rank-sum test, $p = 0.210$ and $p = 0.213$, resp.). The unweighted UniFrac distances and weighted UniFrac distances which reflect the β-diversity of bacterial communities showed no significant

Table 1: α-Diversity and β-diversity of bacterial communities. (a) The Chao value and Shannon index; (b) the unweighted UniFrac distances and weighted UniFrac distances. The unweighted UniFrac distances are indicated by $**$, and weighted UniFrac distances are indicated by $*$.

(a)

a-diversity	MJQ1	MJQ2	MJQ3	GJQ1	GJQ2	GJQ3	$p(t-)$	$p(w-)$
Chao value	3655.8	3872.88	3427.75	3038.48	3067.18	3608.63	0.140	0.210
Shannon index	6.782	6.876	6.863	6.199	5.944	6.818	0.117	0.213

(b)

	MJQ1	MJQ2	MJQ3	GJQ1	GJQ2	GJQ3
MJQ1	0	0.119^{*}	0.187^{*}	0.200^{*}	0.190^{*}	0.183^{*}
MJQ2	0.315^{**}	0	0.201^{*}	0.177^{*}	0.178^{*}	0.189^{*}
MJQ3	0.397^{**}	0.418^{**}	0	0.171^{*}	0.141^{*}	0.181^{*}
GJQ1	0.441^{**}	0.434^{**}	0.404^{**}	0	0.081^{*}	0.172^{*}
GJQ2	0.457^{**}	0.465^{**}	0.411^{**}	0.302^{**}	0	0.171^{*}
GJQ3	0.465^{**}	0.469^{**}	0.396^{**}	0.327^{**}	0.308^{**}	0

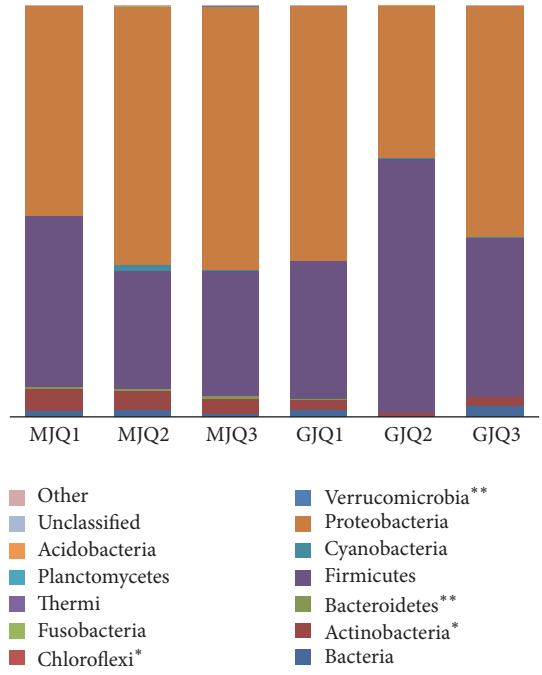


Figure 2: The percentages of OTUs assigned to major bacterial phyla (≥0.01%). Three pits of Maotai-liquor Daqu samples in 2011 to 2013 years are abbreviated as MJQ1, MJQ2, and MJQ3, respectively, and the three Guotai-liquor Daqu samples are GJQ1, GJQ2, and GJQ3 (t-test, $^{*}p < 0.05$, $^{**}p < 0.01$). An asterisk indicates a significant difference with major bacterial phyla between the Maotai-liquor Daqu and Guotai-liquor Daqu.

differences between the *Guotai Daqu* and the *Maotai Daqu* (Wilcoxon rank-sum test, $p = 0.128$ and $p = 0.462$) (Table 1).

However, the 14 genera whose abundances were greater than 0.01% of the total bacteria in the *Guotai* and the *Maotai* were significantly different (t-test, $p < 0.05$; Figure 4). In terms of relative abundance, 13 bacterial genera were significantly higher in the *Maotai Daqu*, and only *Lactobacillus* was significantly higher in the *Guotai Daqu*. Although there were no significant differences in the number of detected bacterial taxa, the samples clustered into two groups based on the relative abundances of the major genera, indicating clear differences between *Maotai Daqu* and *Guotai Daqu* (Figure 4). Including *Thermoactinomyces*, *Saccharopolyspora*, *Acinetobacter,* and *Pseudomonas*, 17 bacterial genera in the *Maotai Daqu* were higher, and 6 genera bacterial (from *Corynebacterium* to *Sebaldella*) in the *Guotai Daqu* were higher.

3.2. Composition of Bacterial Communities in Liquor Fermentation Determined Using High-Throughput Sequencing. A total of 27,720 high-quality sequences were obtained from the 2 liquor fermentation samples, corresponding to 10,200 OTUs. These OTUs clustered into 27 phyla, the most abundant of which was Proteobacteria (67.04% and 70.76%), followed by Firmicutes (20.74% and 24.11%, Figure 5). The OTUs clustered into 17 phyla in the *Daqu* sample and then in the liquor fermentation sample plus an additional 10 phyla. Twelve bacterial phyla were more abundant in the liquor fermentation process, including Acidobacteria, Bacteroidetes, Chlorobi, Chloroflexi, Proteobacteria, and Planctomycetes. However, Actinobacteria and Firmicutes were less abundant in the liquor fermentation process relative to the *Daqu* samples.

These OTUs clustered into 314 genera, and the genera whose abundances were greater than 0.01% of the total bacteria were different between the Daqu and liquor fermentation samples (Figure 6). *Lactobacillus* abundance increased from an average of 1.35% to 19.78% in the liquor fermentation and *Pseudomonas* from 5.07% to 33.52%. In the liquor fermentation process, 18 bacterial genera increased, including *Pseudomonas*, *Lactobacillus*, *Agrobacterium*, *Rhodoplanes*, *Ochrobactrum*, and *Nitrospira*, and 5 bacterial genera decreased, *Lactococcus*, *Enterococcus*, *Saccharopolyspora*, *Bacillus*, and *Pediococcus*.

Taxonomy	MDQ			GDQ		
	MJQ1	MJQ2	MJQ3	GJQ1	GJQ2	GJQ3
Bacteria	1.39	1.69	0.83	1.68	0.09	2.76
*Actinobacteria**	5.51	4.69	3.76	2.56	0.83	2.09
Actinobacteria*	5.5	4.7	3.8	2.6	0.8	2.1
Pseudonocardiaceae*	2.7	2.1	1.6	1	0.4	0.8
*Bacteroidetes***	0.36	0.53	0.49	0.07	0.03	0.04
Sphingobacteriia**	0.3	0.5	0.4	0.1	0	0
Firmicutes	41.63	28.69	30.46	33.51	61.8	38.75
Bacilli	41.4	28.5	30.3	33.4	59.7	38.6
Bacillaceae	13.4	5.3	9.8	18	3.7	22
Staphylococcaceae*	5.8	4.5	3.4	1.5	0.8	2.3
Thermoactinomycetaceae	1.6	1.4	1.3	1.5	0.4	1.5
Enterococcaceae	1.2	0.7	0.7	1	8.7	1.1
Lactobacillaceae	4.3	4.1	3.8	3.4	7.3	3.9
Leuconostocaceae	10	8.7	7.3	3	17.9	2.6
Streptococcaceae	2	1.4	1.4	1.5	18.6	1.7
Clostridia	0.2	0.2	0.2	0.1	2.1	0.1
Cyanobacteria	0.13	1.29	0.24	0.17	0.09	0.05
Chloroplast	0.1	1.3	0.2	0.2	0.1	0
Proteobacteria	50.86	62.95	64.07	61.91	37.16	56.22
Alphaproteobacteria	0.2	1	0.5	0.3	0.1	0.1
Betaproteobacteria	0.10.	0.1	0.1	0.2	0	0.1
Gammaproteobacteria	50.5	61.8	63.8	61.5	37	55.9
Aeromonadaceae	6.4	6.3	6.8	3.2	4.6	7.3
Enterobacteriaceae	12.9	13.3	14.3	29.8	12.1	14.2
Pseudomonadaceae	13.6	24.4	24.3	16.7	7.4	15.4
Unclassified	0.1	0.13	0.11	0.08	0	0.07
Other	0.003	0.001	0.011	0.003	0	0.003

Figure 3: Abundance of the dominant taxonomy (relative abundance of phylum > 0.1% and of family > 1.0%) determined by 16S rRNA gene sequencing. Numbers in cells are percentages of relative abundance, highest abundances are red, middle values are in white, and lowest values are in blue. An asterisk indicates a significant different between the Guotai-liquor Daqu and the Maotai-liquor Daqu. (t-test, $^{*}p < 0.05$;$^{**}p < 0.01$).

4. Discussion

Prior studies on the Chinese liquor fermentation process have focused on a limited number of isolated samples or microbial diversity examined using 16S rRNA gene library analysis and PCR-DGGE. Previously, the higher bacterial diversity as measured by the Shannon index ($H' = 1.19$) was found in a high-temperature Daqu (9-H-S-W) [11]. In the current study, the average Shannon index of the *Daqu* was $H' = 6.58$. Using high-throughput sequencing of the V4 hypervariable region

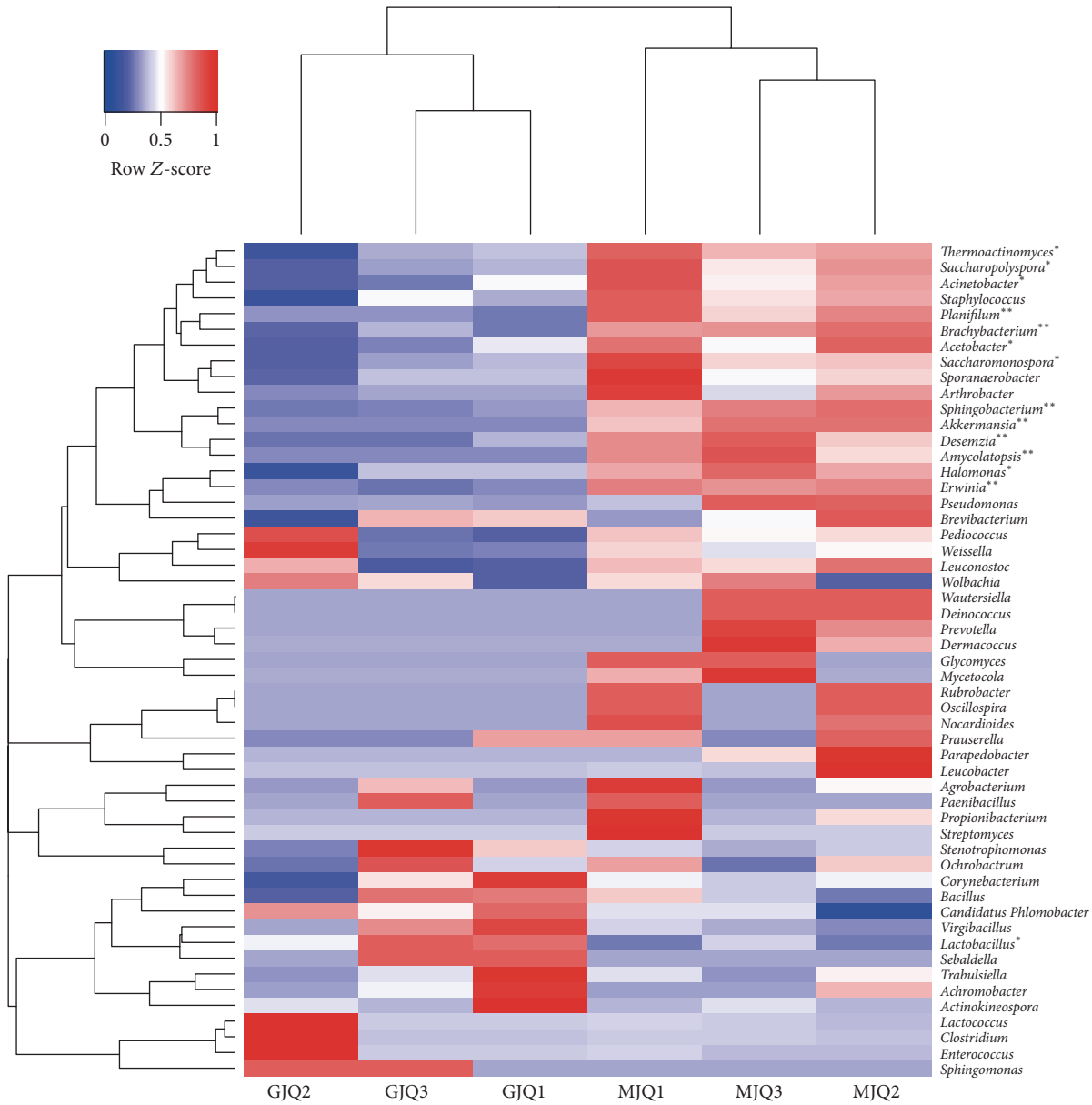


FIGURE 4: Heatmap showing the differences among the six investigated communities based on the abundance of the domain genura. (relative abundance > 0.01%) An asterisk indicates a significant different between Goutai and Maotai Daqu. (t-test, $^{*}P < 0.05$, $^{**}P < 0.01$).

of the 16S rRNA gene to examine microbial communities in-depth, the abundance of the bacterial community in *Daqu* was higher than previously reported [2, 3, 6, 9]. Some of the bacterial genera observed in our study were not previously reported.

Knowledge of the microbiota of Chinese liquor fermentation is still far from complete, especially the microbiota of *Daqu*. Therefore, this study was initiated to understand the composition of the microbial community in two representative soy sauce aroma-style Chinese liquors. In the soy sauce aroma-style Chinese liquor, *Daqu* is made from ground wheat and is produced in Guizhou Province in China, applying high-temperature fermentation conditions for *Daqu* production. It is expected that the relative abundance of several of the microorganisms identified in the *Daqu* correlate with different Chinese liquors. Several studies indicated that the bacterial community in Daqu is affected by certain factors, including raw materials, environmental conditions (e.g., soil and air), moisture content, oxygen condition and "mother *Daqu*" (Daqu that was produced 1 year ago). *Maotai*-flavor and *Guotai*-flavor liquor have the same unique, complicated spontaneous fermentation process and the same environmental conditions. Their *Daqu* may thus have the same microbial communities. The Chao value and Shannon index, which reflect the α-diversity of bacterial communities, of the *Daqu* samples were not significantly

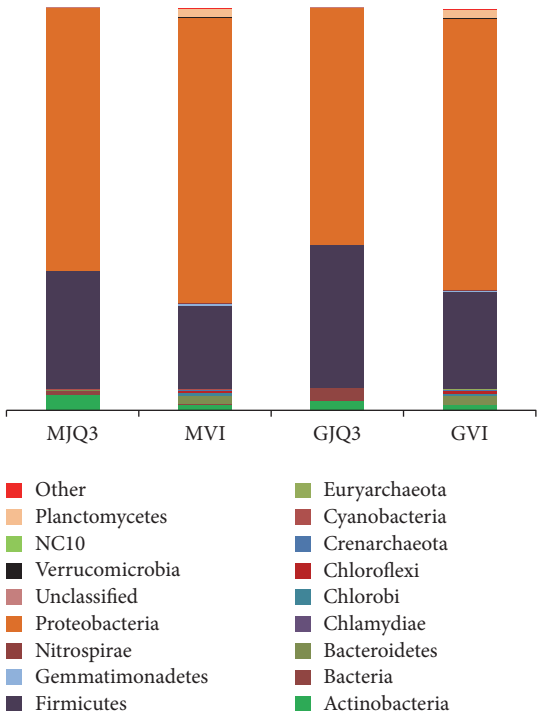


Figure 5: The percentages of OTUs assigned to major bacterial phyla (≥0.01%). The Maotai-liquor fermentation sample are abbreviated as MVI, and the Guotai-liquor fermentation as GVI. The Daqu sample in 2013 are abbreviated as MJQ3 and GJQ3.

different between the *Guotai Daqu* and the *Maotai Daqu*. Furthermore, the same bacterial species were present in the *Daqu* of both liquors. However, the abundance of 13 bacterial genera in the *Maotai Daqu* was significantly higher than in the *Guotai Daqu*: *Thermoactinomyces*, *Erwinia, Saccharopolyspora*, *Acinetobacter*, *Planifilun*, *Brachybacterium*, *Acetobacter*, *Akkermansia*, *Saccharomonospora*, *Sphingobacterium*, *Desemzia*, *Amycolatopsis*, and *Halomonas*. Only the abundance of *Lactobacillus* was significantly higher in the *Guotai Daqu* than in the *Maotai Daqu* (Figure 4). Based on our analysis of different *Daqu* samples from soy sauce aroma-style Chinese liquor, we concluded that the composition of the bacterial communities in the different Chinese liquor were the same. Nevertheless, some bacterial species have significantly different abundances between the different *Daqu* samples. Some types of *Daqu* contained highly similar bacteria species, whereas some bacterial species abundances were significantly different. The relative abundance of bacteria may be important for the Daqu.

Several studies mentioned the importance of lactic acid bacteria (LAB) during the production of *Daqu*, including *Enterococcus*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Streptococcus*, and *Weissella*. In this study, various genera of lactic acid bacteria (LAB) were identified in *Maotai* and *Guotai Daqu*. In the current study, LAB were only found at high abundance during the beginning of the *Fenjiu-Daqu* production process [22], and LAB except *Weissella* were found in low abundance [11]. We observed a high abundance of *Weissella* and *Pediococcus*, and the abundance of LAB except *Lactobacillus* was not significantly different. Some species of *Lactobacillus* inhibit the growth of *Bacillus* [23]. However, in this study, there were no significant differences in *Bacillus* between the two *Daqu* samples. *Bacillus* is a well-known producer of proteases and amylases [24] and produces more than 70 metabolites, most of which are flavor compounds and flavor precursors (Yan et al. 2013). The production of these molecules is important for the aroma of fermented products. This explained that the abundance of *Bacillus* in *Maotai* and *Guotai Daqu* was not significantly different, although *Lactobacillus* was. In the liquor fermentation process, the abundance of LAB excluding *Lactobacillus* decreased, and *Lactobacillus* abundance increased from the average 1.35% to 19.78%. With the dramatic increase of *Lactobacillus*, the abundance of *Bacillus* decreased. At the initial stage of stacking fermentation, *Lactobacillus* quickly propagates and becomes the main bacteria [2, 3].

In this study, the abundance of the phyla Actinobacteria, Bacteria, and Firmicutes decreased between both types of *Daqu* and the fermentation processes, while Proteobacteria and others increased. The genus *Lactobacillus* of the family Lactobacillaceae and the *Pseudomonas* of the family Pseudomonadaceae increased. When the fermentation began, several chemicals were enriched, making the habitat suitable for some bacteria, such as *Lactobacillus* and *Pseudomonas*. During the Fen liquor fermentation process, the bacterial community diversity decreases, and only the family Lactobacillaceae increases [25]. The family Lactobacillaceae is a contributor to the fermentation reaction.

Different microbial communities in Daqu facilitate the selection of starters for creation of unique flavors. Further research is required to gain deeper insight into the microbial communities of the different types of Daqu and the function of unknown microorganisms in these communities. This work may increase liquor producers' understanding of the bacterial community in *Daqu*, and the relative abundance of bacteria in Daqu may be another important topic for liquor production.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Jing Tang and Xiaoxin Tang contributed equally to this work.

Acknowledgments

This work was supported by the Social Development Project of Guizhou province (sz20113068, sy20123035), the International Cooperation Project of Guizhou province (G20137020), Major Projects of Guizhou Education Department (2012005), Program for Changjiang Scholars and Innovative Research Team in University of China (IRT1227), and Science and Technology Platform Construction Project of Guizhou Province (Z20114005).

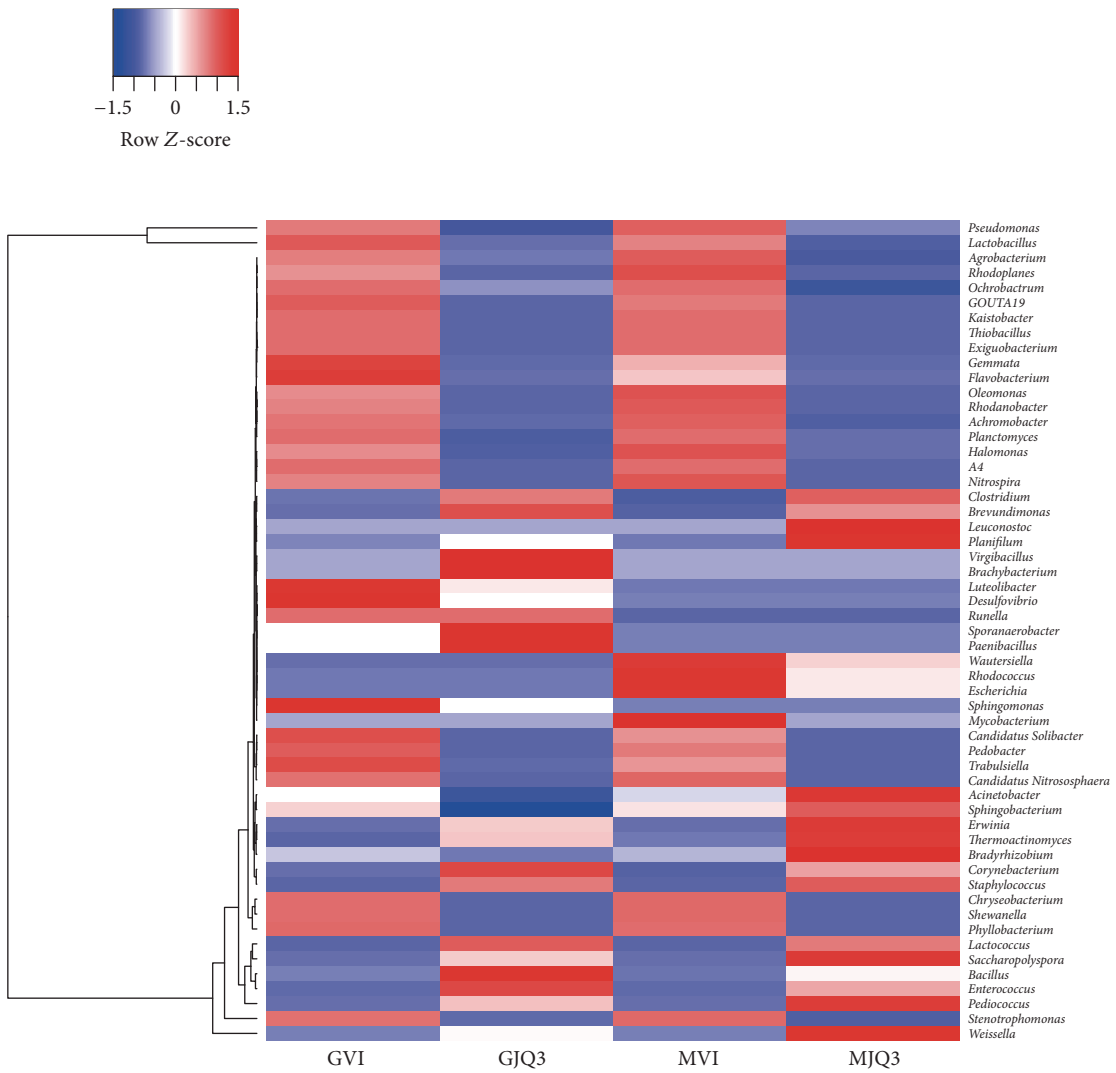


FIGURE 6: Heatmap showing the difference among the liquor fermentation based on the abundance of the domain genera. (relative abundance > 0.01%).

References

[1] W. Fan and M. C. Qian, "Characterization of aroma compounds of Chinese "Wuliangye" and "Jiannanchun" liquors by aroma extract dilution analysis," *Journal of Agricultural and Food Chemistry*, vol. 54, no. 7, pp. 2695–2704, 2006.

[2] C.-L. Wang, D.-J. Shi, and G.-L. Gong, "Microorganisms in Daqu: a starter culture of Chinese Maotai-flavor liquor," *World Journal of Microbiology and Biotechnology*, vol. 24, no. 10, pp. 2183–2190, 2008.

[3] H.-Y. Wang, X.-J. Zhang, L.-P. Zhao, and Y. Xu, "Analysis and comparison of the bacterial community in fermented grains during the fermentation for two different styles of Chinese liquor," *Journal of Industrial Microbiology and Biotechnology*, vol. 35, no. 6, pp. 603–609, 2008.

[4] C. Li, L. Mu, J. Y. Wang, Z. H. Lei, J. Y. Chen, and B. Z. Han, "Physiochemical and microbiological analysis of Fen-type Daqu," *China Brew*, vol. 1, pp. 140–142, 2009 (Chinese).

[5] X.-W. Zheng, Z. Yan, M. J. R. Nout et al., "Microbiota dynamics related to environmental conditions during the fermentative production of Fen-Daqu, a Chinese industrial fermentation starter," *International Journal of Food Microbiology*, vol. 182-183, pp. 57–62, 2014.

[6] X.-W. Zheng, Z. Yan, B.-Z. Han et al., "Complex microbiota of a Chinese 'Fen' liquor fermentation starter (Fen-Daqu), revealed by culture-dependent and culture-independent methods," *Food Microbiology*, vol. 31, no. 2, pp. 293–300, 2012.

[7] H.-Y. Wang, Y.-B. Gao, Q.-W. Fan, and Y. Xu, "Characterization and comparison of microbial community of different typical Chinese liquor Daqus by PCR-DGGE," *Letters in Applied Microbiology*, vol. 53, no. 2, pp. 134–140, 2011.

[8] L. Zhang, C. Wu, X. Ding, J. Zheng, and R. Zhou, "Characterisation of microbial communities in Chinese liquor fermentation starters Daqu using nested PCR-DGGE," *World Journal of Microbiology and Biotechnology*, vol. 30, no. 12, pp. 3055–3063, 2014.

[9] Y. Zheng, X.-W. Zheng, B.-Z. Han, J.-S. Han, M. J. R. Nout, and J.-Y. Chen, "Monitoring the ecology of Bacillus during Daqu incubation, a fermentation starter, using culture-dependent

and culture-independent methods," *Journal of Microbiology and Biotechnology*, vol. 23, no. 5, pp. 614–622, 2013.

[10] L. C. Yan, S. C. Zhang, X. W. Ma, E. J. Tang, Z. X. Huang, and Y. Q. Chen, "Phylogenetic analysis of 16S rDNA sequence and PCR: RFLP of Bacillus from Fumao-flavor Daqu," *Biotechnology*, vol. 22, pp. 54–58, 2012 (Chinese).

[11] X.-W. Zheng, Z. Yan, M. J. R. Nout et al., "Characterization of the microbial community in different types of *Daqu*samples as revealed by 16S rRNA and 26S rRNA gene clone libraries," *World journal of microbiology & biotechnology*, vol. 31, no. 1, pp. 199–208, 2015.

[12] W. X. Zhang, Z. W. Qiao, W. L. Xiang et al., "Micro-ecological research progress of the cellar of Chinese flavour liquor," *Liquor Making*, vol. 31, pp. 31–35, 2004.

[13] Q. Wu, L. Chen, and Y. Xu, "Yeast community associated with the solid state fermentation of traditional Chinese *Maotai*-flavor liquor," *International Journal of Food Microbiology*, vol. 166, no. 2, pp. 323–330, 2013.

[14] Q. Wu, B. Chen, and Y. Xu, "Regulating yeast flavor metabolism by controlling saccharification reaction rate in simultaneous saccharification and fermentation of chinese maotai-flavor liquor," *International Journal of Food Microbiology*, vol. 200, pp. 39–46, 2015.

[15] S. T. Bates, D. Berg-Lyons, J. G. Caporaso, W. A. Walters, R. Knight, and N. Fierer, "Examining the global distribution of dominant archaeal populations in soil," *The ISME Journal*, vol. 5, no. 5, pp. 908–917, 2011.

[16] Z. Liu, C. Lozupone, M. Hamady, F. D. Bushman, and R. Knight, "Short pyrosequencing reads suffice for accurate microbial community analysis," *Nucleic Acids Research*, vol. 35, no. 18, article e120, 2007.

[17] Q. Yan, Y. Bi, Y. Deng et al., "Impacts of the Three Gorges Dam on microbial structure and potential function," *Scientific Reports*, vol. 5, article 8605, 2015.

[18] J. G. Caporaso, J. Kuczynski, J. Stombaugh et al., "QIIME allows analysis of high-throughput community sequencing data," *Nature Methods*, vol. 7, no. 5, pp. 335–336, 2010.

[19] R. C. Edgar, "UPARSE: highly accurate OTU sequences from microbial amplicon reads," *Nature Methods*, vol. 10, no. 10, pp. 996–998, 2013.

[20] J. G. Caporaso, C. L. Lauber, W. A. Walters et al., "Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, S1, pp. 4516–4522, 2011.

[21] J.-B. Pan, S.-C. Hu, H. Wang, Q. Zou, and Z.-L. Ji, "PaGeFinder: quantitative identification of spatiotemporal pattern genes," *Bioinformatics*, vol. 28, no. 11, pp. 1544–1545, 2012.

[22] Z. Lei, "Preliminary analysis of the change of microbes in the fermentation process of Fenjiu Daqu," *Liquor Making Science Technology*, vol. 6, pp. 65–68, 2011 (Chinese).

[23] K. Katina, M. Sauri, H.-L. Alakomi, and T. Mattila-Sandholm, "Potential of lactic acid bacteria to inhibit rope spoilage in wheat sourdough bread," *LWT—Food Science and Technology*, vol. 35, no. 1, pp. 38–45, 2002.

[24] H. Karataş, F. Uyar, V. Tolan, and Z. Baysal, "Optimization and enhanced production of α-amylase and protease by a newly isolated *Bacillus licheniformis* ZB-05 under solid-state fermentation," *Annals of Microbiology*, vol. 63, no. 1, pp. 45–52, 2013.

[25] X.-R. Li, E.-B. Ma, L.-Z. Yan et al., "Bacterial and fungal diversity in the traditional Chinese liquor fermentation process," *International Journal of Food Microbiology*, vol. 146, no. 1, pp. 31–37, 2011.

Hindawi
BioMed Research International
Volume 2017, Article ID 7198614, 7 pages
http://dx.doi.org/10.1155/2017/7198614

Research Article

Comparison and Validation of Putative Pathogenicity-Related Genes Identified by T-DNA Insertional Mutagenesis and Microarray Expression Profiling in *Magnaporthe oryzae*

Ying Wang, Ying Wáng, Qi Tan, Ying Nv Gao, Yan Li, and Da Peng Bao

National Engineering Research Centre of Edible Fungi, Key Laboratory of Edible Fungi Resources and Utilization (South), Ministry of Agriculture, Institute of Edible Fungi, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China

Correspondence should be addressed to Da Peng Bao; baodapeng@saas.sh.cn

Received 26 September 2016; Revised 16 November 2016; Accepted 12 December 2016; Published 14 February 2017

Academic Editor: Xinyu Zhang

High-throughput technologies of functional genomics such as T-DNA insertional mutagenesis and microarray expression profiling have been employed to identify genes related to pathogenicity in *Magnaporthe oryzae*. However, validation of the functions of individual genes identified by these high-throughput approaches is laborious. In this study, we compared two published lists of genes putatively related to pathogenicity in *M. oryzae* identified by T-DNA insertional mutagenesis (comprising 1024 genes) and microarray expression profiling (comprising 236 genes), respectively, and then validated the functions of some overlapped genes between the two lists by knocking them out using the method of target gene replacement. Surprisingly, only 13 genes were overlapped between the two lists, and none of the four genes selected from the overlapped genes exhibited visible phenotypic changes on vegetative growth, asexual reproduction, and infection ability in their knockout mutants. Our results suggest that both of the lists might contain large proportions of unrelated genes to pathogenicity and therefore comparing the two gene lists is hardly helpful for the identification of genes that are more likely to be involved in pathogenicity as we initially expected.

1. Introduction

Rice blast caused by the fungal pathogen *Magnaporthe oryzae* is one of the most destructive diseases of rice threatening the sustainability of global food production. *M. oryzae* attacks various parts of rice including leaves, stems, nodes, and panicles [1, 2]. The infection of *M. oryzae* is a complex process [3, 4]. It begins from the attachment of a three-celled conidium to the rice leaf, followed by the germination of the conidium and the differentiation of the germ tube into a dome-shaped cell called appressorium. The appressorium matures and generates turgor by accumulating high concentrations of compatible solutes, which can directly penetrate the host cuticle, resulting in a disease lesion. After penetration, the bulbous branched infectious hyphae rapidly spread to adjacent cells and form conidiophores to release conidia into the environment so as to initiate new infection. In short, the infection process of *M. oryzae* consists of four stages: attachment, germination, differentiation, and penetration. Obviously, such a complex process must be controlled by a great number of genes.

To prevent and control rice blast efficiently, it is necessary to understand the mechanisms involved in this disease and find out genes related to the pathogenicity of *M. oryzae*. It is known that the interaction between rice and blast fungus complies with the gene-for-gene relationship. Therefore, avirulence genes play important roles in the pathogenicity of *M. oryzae*. However, many other genes involved in the infection process may also contribute to the pathogenicity. In recent years, a few gene manipulation approaches, such as homologous recombination [5–7], T-DNA insertional mutagenesis [8, 9], and RNA interference [10–12], have been employed to study the pathogenicity-related genes in *M. oryzae*. Some key genes for pathogenicity have been identified and cloned [4, 13–15].

As the complete genome sequence of *M. oryzae* has been available [16], large-scale and systematic identification of pathogenicity-related genes in the rice blast fungus

becomes feasible. Jeon et al. [17] obtained a total of 21,070 mutants in *M. oryzae* through large-scale T-DNA insertional mutagenesis, from which over 1,000 genes putatively related to pathogenicity were identified using a high-throughput phenotype screening pipeline. Oh et al. [18] employed a whole genome oligo-DNA microarray of *M. oryzae* to analyze genome-wide gene expression changes during spore germination and appressorium formation. They detected hundreds of differentially expressed genes, which were thought to be related to appressorium formation (and therefore possibly to pathogenicity). These results will be helpful for deep research on the molecular mechanisms of pathogenicity in *M. oryzae*. However, the exact functions of these genes still need to be verified individually. This will be laborious work.

Since the gene list obtained by the T-DNA insertional mutagenesis [17] and that obtained by the microarray analysis [18] are both putatively related to pathogenicity, it would be interesting to see how many genes are overlapped between them and whether the overlapped genes are more likely to be pathogenicity-related. This would allow us to evaluate the reliability and usefulness of the gene lists. For this purpose, in this study, we examined the overlapping between these two gene lists and validated the functions of the overlapped genes by knocking them out using the method of target gene replacement.

2. Materials and Methods

2.1. Sources and Comparison of Gene Lists. Two gene lists of *M. oryzae* were used for the study. One was acquired from the Supplementary Tables 2 and 4 of Jeon et al. [17]. The list contained a total of 1024 putatively mutated genes identified from 547 T-DNA insertion mutant strains. ~90% of the mutants displayed changed phenotypes and/or weakened pathogenicity. There were two types of mutations: Type I is that a gene is disrupted by the T-DNA, which is inserted inside the gene, so that the gene's function is usually lost; Type II is that a gene remains intact in structure, but its function is possibly affected by the T-DNA, which is inserted beside it. Most (~88%) of the genes in this list belonged to Type I, only a small proportion (~12%) belonged to Type II. The other gene list was obtained from the Additional Data File 3 of Oh et al. [18]. The list contained 236 genes that were either upregulated (~2/3) or downregulated (~1/3) in response to the stimulation of hydrophobic surface during appressorium formation according to microarray analysis and therefore were thought to be possibly related to appressorium formation. For convenience, we shall call the two gene lists as MUG (mutated gene) list and DEG (differentially expressed gene) list, respectively. The overlapping between the two gene lists was examined.

2.2. Fungal Strains, DNA Extraction, and Southern Blot Analysis. The wild-type *M. oryzae* strain Guy11, which is pathogenetic to many rice varieties and can infect barley and the model grass species *Brachypodium distachyon* as well [4, 19], was used for the gene-knockout experiment. All *M. oryzae* strains (including Guy11 and its gene-knockout mutants) were cultured on CM medium. Genomic DNA of each strain was extracted using the CTAB method as described [20]. Restriction enzyme digestion and ligation were performed according to Sambrook and Russell [21]. The DIG high prime DNA labeling and Kit I (Roche, Germany) were used for Southern blot analysis.

2.3. Vector Construction and Transformation. Construction of target gene replacement vector and fungal transformation were performed as described [3, 22]. Putative gene deletion mutants were recovered and selected on a complete medium with 200 ug/mL hygromycin and further selected by three rounds of screening using PCR (1st round), Southern blot (2nd round), and RT-PCR (3rd round). Primers used for the gene replacement, PCR, and RT-PCR are listed in Table 1.

2.4. Phenotype Investigation and Pathogenicity Test. The growth characteristic, conidiation, conidial germination, appressorium formation, and conidial morphology of each *M. oryzae* strain were observed following the protocols as described [3, 22]. After a strain was cultured on CM medium for 10 days, conidia were collected and suspended in 0.2% (w/v) gelatin solution. After resuspension to a concentration of 1×10^5 conidia/mL, the conidia were inoculated by spraying on the 14-day-old seedlings of rice (*Oryza sativa* L.) cultivar CO-39. The inoculated rice seedlings were placed in a chamber with a 12–24 h photophase under 25°C and then transferred to a greenhouse with 14 h of light and 10 h of dark for 7 days. Disease severity was rated according to the method of [23].

3. Results

3.1. Overlapping between the Two Gene Lists. Comparison between the MUG list and the DEG list indicated that the overlapping between them was very low (Table 2). Only 13 genes were overlapped between the two lists, accounting for 1.27% (13/1024) in the MUG list and 5.51% (13/236) in the DEG list, respectively. Among the 13 overlapped genes, eleven were upregulated and two were downregulated in response to the stimulation of hydrophobic surface during appressorium formation. Meanwhile, only two genes belonged to Type I, while most belonged to Type II. Noticeably, the proportion of Type I genes in the overlapped genes ($2/13 \approx 15.4\%$) was very close to that in the MUG list (~12%). In addition, eight (61.5%) genes were previously found to be possibly related to pathogenicity according to the phenotypes of their T-DNA insertion mutants. This proportion was smaller than that in the MUG list (~90%).

3.2. Phenotypic Effects of the Overlapped Genes. To evaluate the overlapped genes, we examined five of them, that is, MGG_00623, MGG_00745, MGG_00871, MGG_04068, and MGG_06704 (Table 2), by gene knockout using the method of target gene replacement. Considering that Type II genes were predominant in the MUG list and the function loss of a gene is usually ensured in Type I mutation but not in Type II mutation, the five genes we selected all belonged to Type II. In addition, these five genes were all upregulated in response to the stimulation of hydrophobic surface and were all previously proved to be possibly involved in pathogenicity

Table 1: The primers used for genes replacement and PCR in this study.

Primers name	Sequence of PCR primers	Application of primers
P1UFMGG-00623	ctcgagATTCGGGTCCTTCGTTAT	For genes replacement
P1URMGG-00623	gtcgacCCTCCCTCTGTTGTCTTGT	For genes replacement
P1DFMGG-00623	actagtGACCGTGATCGACCTTCC	For genes replacement
P1DRMGG-00623	gagctcATGCCCTCTTTGACTTGG	For genes replacement
P2UFMGG_00871	ggtacc TCGAGGGTTATCAAGCAA	For genes replacement
P2URMGG_00871	gtcgacAAATAGAAGCCGCCAGAC	For genes replacement
P2DFMGG_00871	gaattcGATGACGAGTTGCGATGT	For genes replacement
P2DRMGG_00871	tctagaGGGACCTGCTCTGTATCA	For genes replacement
P3UFMGG_06704	gggcccCCGTCATCACCTAACCAA	For genes replacement
P3URMGG_06704	ctcgagGAACAGCGTCGTCTCCAT	For genes replacement
P3DFMGG_06704	actagtGACCGTGATCGACCTTCC	For genes replacement
P3DRMGG_06704	gagctcATGCCCTCTTTGACTTGG	For genes replacement
P4UFMGG_00745	ctcgagGCGGGTCAAAGAGTGTATT	For genes replacement
P4URMGG_00745	gagctcGTCGTTGGGTATTGGGTC	For genes replacement
P4DFMGG_00745	gaattcCACTTCTTTCCCTGGTCG	For genes replacement
P4DRMGG_00745	gagctcTCCTCTGGAGCTTTCCTC	For genes replacement
P5UFMGG_04068	gggcccGGGGCAAGGTTCTCAAAG	For genes replacement
P5URMGG_04068	gtcgacAAGCGAGGTGGCAGGTAG	For genes replacement
P5DFMGG_04068	aagcttAGGTCGTAGACATACTGAGGT	For genes replacement
P5DRMGG_04068	gaattcAAGGCTGTAGATGGCTGA	For genes replacement
CP1FMGG-00623	ACTTGATGGCTAACCACTACTT	For PCR screening
CP1RMGG-00623	CCAATATGTCCGAGACGAT	For PCR screening
CP2FMGG_00871	CCCATTGATACTGCGGTTAG	For PCR screening
CP2RMGG_00871	TTGATCGTGCCGTCCTCT	For PCR screening
CP3FMGG_06704	CATCGTGGACATCTTGGAG	For PCR screening
CP3RMGG_06704	CGAAACTTCTGGTGGTGAT	For PCR screening
CP4FMGG_00745	CTCCGTTGCGTCGTCTGT	For PCR screening
CP4RMGG_00745	TCTGGTCCGTCTTGCTGTT	For PCR screening
CP5FMGG_04068	ATCACAACCCTCCGAACCA	For PCR screening
CP5FMGG_04068	GCAAACCTGTCCTCGTAGTCC	For PCR screening
R-P1FMGG-00623	CGCATCCCAAGCCTGAAT	For RT-PCR screening
R-P1RMGG-00623	AGAACGGCGGGTGACAAG	For RT-PCR screening
R-P2FMGG_00871	AAGGGTCCGACGAGCAAA	For RT-PCR screening
R-P2RMGG_00871	CCTCCAACTCCACGGGTAT	For RT-PCR screening
R-P3FMGG_06704	GGAGTGGGAGGACAATGAA	For RT-PCR screening
R-P3RMGG_06704	GTCGCAATGGCAAGAACA	For RT-PCR screening
R-P4FMGG_00745	GGACCCAATACCCAACGAC	For RT-PCR screening
R-P4RMGG_00745	ACGGCTCATACGGCATAAA	For RT-PCR screening

as their T-DNA insertion mutants all exhibited visible phenotypic changes and pathogenicity reduction.

We failed to obtain null mutants of MGG_04068 gene although we performed four independent transformations and screened several hundreds of transformants. A possible reason is that the cell could not survive when MGG_04068 gene is knocked out. But the molecular function of MGG_04068 is not known. According to the annotation, it is a conserved hypothetical protein.

The other four genes were all successfully knocked out, confirmed by Southern blot and RT-PCR (Figure 1). However, the null mutants of the four genes did not show significant phenotypic changes on growth characteristic (GC; Table 3), pigmentation (PG; Figure 2), conidiation (Table 4), conidial germination (Table 4), appressorium formation (Table 4), and conidia and appressorium morphology (Figure 2), suggesting that these genes have little effect on fungal development. Moreover, after inoculation, these mutants all resulted in

Table 2: Information of 13 overlapping genes ("+" indicate change).

Gene name	Regulation	Insert location	Mutant ID	GR[a]	PG[b]	CN[c]	GM[d]	AP[e]	CM[f]	PT[g]
MGG_00450	Up	MGG_00450	0137A2, 0128D5	+	+	+		+		+
MGG_00623	Up	MGG_00623-MGG_00624	0035C2							+
MGG_00745	Up	MGG_00744-MGG_00745	0059A3					+		+
MGG_00871	Up	MGG_00870-MGG_00871	0430D2					+		+
MGG_00994	Up	MGG_00994-MGG_11455	0673D3, 0690A4			+				+
MGG_01778	Up	MGG_01778-MGG_01779	0008C4							
MGG_02763	Up	MGG_02763-MGG_12596	0156D5							
MGG_02817	Down	MGG_02817	0010A5							
MGG_04068	Up	MGG_04068-MGG_04069	0257C4			+			+	
MGG_06704	Up	MGG_06704-MGG_06705	0416A3, 0420C2					+	+	+
MGG_09096	Up	MGG_09096-MGG_09095	0007B2							
MGG_09200	Down	MGG_09200-MGG_11770	0059B3					+		+
MGG_09942	Up	MGG_09942-MGG_09941	0236B3							

[a]GR: growth rate; [b]PG: pigmentation; [c]CN: conidiation; [d]GM: conidial germination; [e]AP: appressorium formation; [f]CM: conidial morphology; [g]PT: pathogenicity.

Table 3: Growth characteristic of *M. oryzae* strains[A].

M. oryzae	Growth days after				
	3	5	7	9	11
Guy11	2.09 ± 0.18[a]	3.43 ± 0.23[a]	5.20 ± 0.00[a]	6.61 ± 0.01[a]	8.02 ± 0.00[a]
ΔMGG_00623	2.08 ± 0.05[a]	3.46 ± 0.05[a]	5.1 ± 0.00[a]	6.55 ± 0.00[a]	7.80 ± 0.20[a]
ΔMGG_00871	2.03 ± 0.17[a]	3.46 ± 0.15[a]	5.00 ± 0.19[a]	6.41 ± 0.05[a]	7.90 ± 0.00[a]
ΔMGG_06704	2.08 ± 0.14[a]	3.45 ± 0.14[a]	5.00 ± 0.17[a]	6.39 ± 0.20[a]	7.87 ± 0.11[a]
ΔMGG_00745	2.18 ± 0.05[a]	3.62 ± 0.07[a]	5.05 ± 0.22[a]	6.53 ± 0.11[a]	7.93 ± 0.15[a]

[A]The diameter of the wild-type Guy11 and mutants was measured after inoculation in CM plates for 3, 5, 7, 9, and 11 days. The targeted genes replacement had no distinguishable effect on growth rate. The letters "a" in each column are not significantly different, as estimated by Duncan's test ($p \leq 0.05$).

typical spindle-like and gray-center lesions in rice seedling leaves with similar disease score to that caused by wild-type Guy11 (Figure 3). The result indicates that these four genes are not necessary for pathogenicity in *M. oryzae*.

4. Discussion

T-DNA insertional mutagenesis and microarray expression profiling are two powerful technologies for functional genomics research. Although the two technologies analyze gene functions from different aspects, it can be expected that the lists of genes identified by them for the same traits would tend to be overlapped or positively correlated. Based on this consideration, in this study, we tried to screen genes that would be more likely to be related to pathogenicity in *M. oryzae* by comparing the MUG and DEG lists. However, we were surprised to find that the overlapping between the two gene lists was terribly low (Table 2).

Based on the genome sequence draft of *M. oryzae*, it was estimated that there are totally 11109 genes in *M. oryzae* [16]. According to this estimate, we can find that the number of overlapped genes between the MUG and DEG lists is expected to be $1024 \times 236/11109 \approx 22$ provided the two gene lists are independent random samples from the whole set of *M. oryzae* genes. Obviously, if the two gene lists are positively correlated, the number of overlapped genes will be much greater than this value. But the actual number we found was 13, even smaller than that. This result suggests that the two gene lists are not positively correlated as expected.

In the subsequent experiment for validating gene function, we successfully knocked out four Type II overlapped genes. The T-DNA insertion mutants of these four genes all displayed phenotypic changes and pathogenicity reduction (Table 1; [17]), but none of the knockout mutants of these four genes we obtained exhibited any phenotypic changes and pathogenicity reduction as expected. Hence, according to our result, these four genes are not related to pathogenicity. Suppose the proportion of pathogenicity-unrelated genes among the Type II overlapped genes is p, then the probability that all the four genes examined are not related to pathogenicity will be p^4. Following the principle that an event of small probability is unlikely to occur in a single experiment, we may require $p^4 > 0.05$, which means $p > 0.473$, or half of the Type II overlapped genes would be pathogenicity-unrelated. As this is a conservative estimate, the real proportion of pathogenicity-unrelated genes might be higher.

Taken together, our study indicates that the MUG list and the DEG list only have a small overlapping, and at least half of the overlapped genes are not related to pathogenicity. These two results are consistent, both of which suggest that

Table 4: Conidiation, conidial germination, and appressorium formation of the wild Guy11 and mutants.

M. oryzae	Conidiation ($\times 10^4$ cm^{-2})[A]	Conidial germination (%)[B] 4 h	Appressorium formation (%)[B] 12 h
Guy11	117.85 ± 27[a]	91.23 ± 2.79[a]	90.62 ± 3.71[a]
ΔMGG_00623	120.71 ± 26[a]	92.93 ± 2.20[a]	90.11 ± 1.62[a]
ΔMGG_00871	105.71 ± 41[a]	90.70 ± 3.84[a]	90.80 ± 4.17[a]
ΔMGG_06704	114.28 ± 48[a]	92.74 ± 2.26[a]	91.08 ± 4.47[a]
ΔMGG_00745	117.85 ± 55[a]	92.10 ± 2.64[a]	92.09 ± 3.08[a]

[A]After incubation for 10 days on CM plates, conidia were collected using three 1 cm diameter discs of mycelium in water and counted with a haemacytometer under a microscope.
[B]Conidial germination and appressorium formation were recorded after different time incubation of the conidial suspension on the hydrophobic surface of film as described previously. The letters "a" in each column are not significantly different, as estimated by Duncan's test ($p \leq 0.05$).

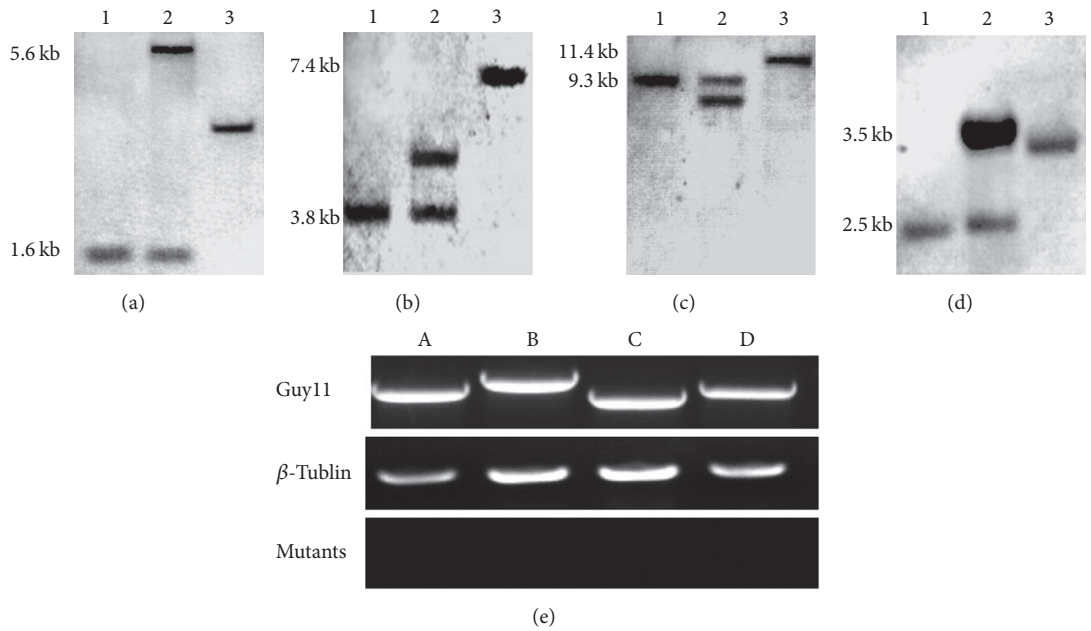


Figure 1: Southern blot and RT-PCR analysis of the transformants. (a) MGG_00623. DNA samples were digested with *Apa*I, probed with a 1.2 kb upstream flanking sequence fragment of the gene replacement vector. The 5.6 kb band was detected in mutant, whereas the 1.6 kb band was detected in wild-type Guy11; (b) MGG_00871. DNA samples were digested with *Kpn*I, probed with a 1.0 kb upstream flanking sequence fragment of the gene replacement vector. The 7.4 kb band was detected in mutant, whereas the 3.8 kb band was detected in wild-type Guy11; (c) MGG_06704. DNA samples were digested with *Sac*I, probed with a 1.3 kb downstream flanking sequence fragment of the gene replacement vector. The 11.4 kb band was detected in mutant, whereas the 9.3 kb band was detected in wild-type Guy11; (d) MGG_00745. DNA samples were digested with *Eco*RV, probed with a 1.0 kb upstream flanking sequence fragment of the gene replacement vector. The 3.5 kb band was detected in mutant, whereas the 2.5 kb band was detected in wild-type Guy11; (a–d) Lane of 1: Guy11; 2: transformant with selection marker but not single copy; 3: mutant with targeted gene had been replaced. (e) Lane of A: ΔMGG_00623; B: ΔMGG_00871; C: ΔMGG_06704; D: ΔMGG_00745.

the two gene lists are basically independent and at least one of the two gene lists is close to a random sample in regard to pathogenicity. We suspect that the DEG list is more likely to be the case. It is possible that in the DEG list most of the genes responding to the stimulation of hydrophobic surface are not involved in pathogenicity. In our experiment, we also knocked out another four genes in the DEG list but not among the overlapped genes, and these four genes also did not show any phenotypic changes and pathogenicity reduction (results not presented). Hence, we totally examined eight genes in the DEG list by gene knockout and these genes all showed negative results. This suggests that there must be a large proportion of genes that are not related to pathogenicity in the DEG list, making the DEG list behave as a random sample approximately.

The MUG lists may also contain a large proportion of pathogenicity-unrelated genes. We have seen that the MUG list mainly consists of Type II genes (~88%). Generally speaking, Type II genes are resulted from the T-DNA insertion between two adjacent genes. In other words, each outside-gene T-DNA insertion can result in two Type II genes. Hence, in general, in each mutant containing an outside-gene T-DNA insertion, only one of Type II genes would possibly be the cause of the mutant phenotypes, while the other one

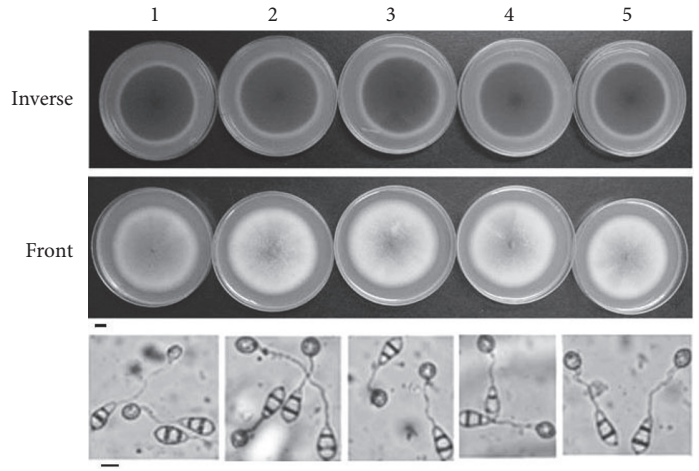


Figure 2: Pigmentation and Morphology of conidia and appressorium of *M. oryzae* strains. The wild-type Guy11 and mutants were grown on CM medium for 11 days, and colonies were photographed; we had not found change on pigmentation. Bar: 1 cm; the pictures of appressorium were photographed after the conidia induced 6 h on film. Bar: 10 μm. Number of 1: Guy11; 2: ΔMGG_00623; 3: ΔMGG_00871; 4: ΔMGG_06704; 5: ΔMGG_00745.

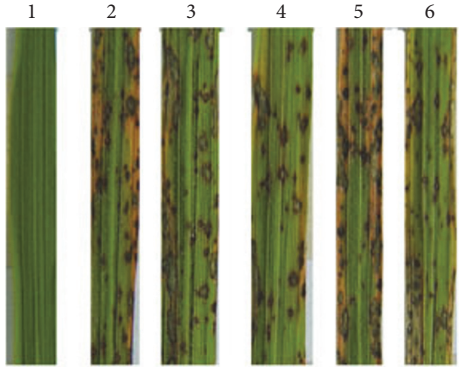


Figure 3: Leaves from CO-39 were spray inoculated individually with conidia. Photographed images 7 days after rice seedlings were inoculated with conidia (1×10^5 conidia/mL) from the wild-type Guy11 and mutants; Number of 1: 0.2% gelatin (control); 2: Guy11; 3: ΔMGG_00623; 4: ΔMGG_00871; 5: ΔMGG_06704; 6: ΔMGG_00745.

would be unrelated. In addition, sometimes there can be multiple copies of T-DNA insertion in a mutant. In this case, the number of unrelated genes among the Type II genes will be increased. Therefore, we can expect that at least half of the Type II genes in the MUG list are not related to pathogenicity.

5. Conclusion

Based on the above discussion, we think that the large proportions of pathogenicity-unrelated genes in the MUG and DEG lists (especially the latter) must be the major reason for the low overlapping or little correlation between the two gene lists. Our results suggest that comparing the two types of gene lists does not facilitate the identification of genes that are more likely to be involved in pathogenicity as we initially expected. Hence, how to efficiently validate the functions of genes identified by T-DNA insertional mutagenesis and microarray expression profiling is still an arduous task. Analysis of double or multiple mutants would be an effective approach for determining the functions of genes discovered by T-DNA random insertion [24].

Competing Interests

There are no competing interests related to this paper.

Authors' Contributions

In this article Ying Wang and Ying Wáng contributed equally to this work.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant no. 31170098) and the Science and Technology Commission of the Shanghai Municipality (Grant no. 13391912202).

References

[1] R. A. Dean, J. A. L. Van Kan, Z. A. Pretorius et al., "The Top 10 fungal pathogens in molecular plant pathology," *Molecular Plant Pathology*, vol. 13, no. 4, pp. 414–430, 2012.

[2] A. Sesma and A. E. Osbourn, "The rice leaf blast pathogen undergoes developmental processes typical of root-infecting fungi," *Nature*, vol. 431, no. 7008, pp. 582–586, 2004.

[3] T.-B. Liu, G.-Q. Chen, H. Min, and F.-C. Lin, "*MoFLP1*, encoding a novel fungal fasciclin-like protein, is involved in conidiation and pathogenicity in *Magnaporthe oryzae*," *Journal of Zhejiang University: Science B*, vol. 10, no. 6, pp. 434–444, 2009.

[4] R. A. Wilson and N. J. Talbot, "Under pressure: investigating the biology of plant infection by Magnaporthe oryzae," *Nature Reviews Microbiology*, vol. 7, no. 3, pp. 185–195, 2009.

[5] J. P. Lu, H. J. Cao, L. L. Zhang, P. Y. Huang, and F. C. Lin, "Systematic analysis of Zn_2Cys_6 transcription factors required

for development and pathogenicity by high-throughput gene knockout in the rice blast fungus," *PLoS Pathogens*, vol. 10, no. 10, Article ID e1004432, 2014.

[6] G. Giaever, A. M. Chu, L. Ni et al., "Functional profiling of the *Saccharomyces cerevisiae* genome," *Nature*, vol. 418, no. 6896, pp. 387–391, 2002.

[7] E. A. Winzeler, D. D. Shoemaker, A. Astromoff et al., "Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis," *Science*, vol. 285, no. 5429, pp. 901–906, 1999.

[8] M. Bevan and S. Walsh, "The Arabidopsis genome: a foundation for plant research," *Genome Research*, vol. 15, no. 12, pp. 1632–1642, 2005.

[9] H. Hirochika, E. Guiderdoni, G. An et al., "Rice mutant resources for gene discovery," *Plant Molecular Biology*, vol. 54, no. 3, pp. 325–334, 2004.

[10] A. J. Andres, "Flying through the genome: a comprehensive study of functional genomics using RNAi in *Drosophila*," *Trends in Endocrinology & Metabolism*, vol. 15, no. 6, pp. 243–247, 2004.

[11] K. Ashrafi, F. Y. Chang, J. L. Watts et al., "Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes," *Nature*, vol. 421, pp. 268–272, 2003.

[12] K. Berns, E. M. Hijmans, J. Mullenders et al., "A large-scale RNAi screen in human cells identifies new components of the p53 pathway," *Nature*, vol. 428, no. 6981, pp. 431–437, 2004.

[13] N. Liu, G.-A. Ning, X.-H. Liu et al., "An autophagy gene, HoATG5, is involved in sporulation, cell wall integrity and infection of wounded barley leaves," *Microbiological Research*, vol. 192, pp. 326–335, 2016.

[14] X.-Q. Zeng, G.-Q. Chen, X.-H. Liu et al., "Crosstalk between SNF1 pathway and the peroxisome-mediated lipid metabolism in *Magnaporthe oryzae*," *PLoS ONE*, vol. 9, no. 8, Article ID e103124, 2014.

[15] G. Mosquera, M. C. Giraldo, C. H. Khang, S. Coughlan, and B. Valent, "Interaction transcriptome analysis identifies magnaporthe oryzae BAS1-4 as biotrophy-associated secreted proteins in rice blast disease," *Plant Cell*, vol. 21, no. 4, pp. 1273–1290, 2009.

[16] R. A. Dean, N. J. Talbot, D. J. Ebbole et al., "The genome sequence of the rice blast fungus Magnaporthe grisea," *Nature*, vol. 434, no. 7036, pp. 980–986, 2005.

[17] J. Jeon, S.-Y. Park, M.-H. Chi et al., "Genome-wide functional analysis of pathogenicity genes in the rice blast fungus," *Nature Genetics*, vol. 39, no. 4, pp. 561–565, 2007.

[18] Y. Oh, N. Donofrio, H. Pan et al., "Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus Magnaporthe oryzae," *Genome Biology*, vol. 9, no. 5, article R85, 2008.

[19] D. Parker, M. Beckmann, D. P. Enot et al., "Rice blast infection of *Brachypodium distachyon* as a model system to study dynamic host/pathogen interactions," *Nature Protocols*, vol. 3, no. 3, pp. 435–445, 2008.

[20] N. J. Talbot, D. J. Ebbole, and J. E. Hamer, "Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus Magnaporthe grisea," *Plant Cell*, vol. 5, no. 11, pp. 1575–1590, 1993.

[21] J. Sambrook and D. W. Russell, *Molecular Cloning: A Laboratory Manual*, CSHL Press, Cold Spring Harbor, NY, USA, 2001.

[22] J.-P. Lu, X.-X. Feng, X.-H. Liu, Q. Lu, H.-K. Wang, and F.-C. Lin, "Mnh6, a nonhistone protein, is required for fungal development and pathogenicity of *Magnaporthe grisea*," *Fungal Genetics and Biology*, vol. 44, no. 9, pp. 819–829, 2007.

[23] J. M. Bonman, T. Vergel de Dios, and M. Khin, "Physiologic specialization of *Pyricularia oryzae* in the Philippines," *Plant Disease*, vol. 70, pp. 767–769, 1986.

[24] F. E. Tax and D. M. Vernon, "T-DNA-associated duplication/translocations in *Arabidopsis*. Implications for mutant analysis and functional genomics," *Plant Physiology*, vol. 126, no. 4, pp. 1527–1538, 2001.

Hindawi
BioMed Research International
Volume 2017, Article ID 9390803, 10 pages
http://dx.doi.org/10.1155/2017/9390803

Research Article

Molecular Cloning, Bioinformatic Analysis, and Expression of *Bombyx mori* Lebocin 5 Gene Related to *Beauveria bassiana* Infection

Dingding Lü,[1,2] Chengxiang Hou,[1] Guangxing Qin,[1] Kun Gao,[1] Tian Chen,[1] and Xijie Guo[1]

[1]*School of Biotechnology and Sericultural Research Institute, Jiangsu University of Science and Technology, Zhenjiang 212018, China*
[2]*Nursing School, Zhenjiang College, Zhenjiang 212003, China*

Correspondence should be addressed to Xijie Guo; guoxijie@126.com

Received 4 October 2016; Revised 21 November 2016; Accepted 19 December 2016; Published 17 January 2017

Academic Editor: Qiang Gao

A full-length cDNA of lebocin 5 (BmLeb5) was first cloned from silkworm, *Bombyx mori*, by rapid amplification of cDNA ends. The BmLeb5 gene is 808 bp in length and the open reading frame encodes a 179-amino acid hydroxyproline-rich peptide. Bioinformatic analysis results showed that BmLeb5 owns an O-glycosylation site and four RXXR motifs as other lebocins. Sequence similarity and phylogenic analysis results indicated that lebocins form a multiple gene family in silkworm as cecropins. Quantitative real-time PCR analysis revealed that BmLeb5 was highest expressed in the fat body. In the silkworm larvae infected by *Beauveria bassiana*, the expression level of BmLeb5 was upregulated in the fat body and hemolymph which are the most important immune tissues in silkworm. The recombinant protein of BmLeb5 was for the first time successfully expressed with prokaryotic expression system and purified. There are no reports so far that the expression of lebocins could be induced by entomopathogenic fungus. Our study suggested that BmLeb5 might play an important role in the immune response of silkworm to defend *B. bassiana* infection. The results also provided helpful information for further studying the lebocin family functioned in antifungal immune response in the silkworm.

1. Introduction

Silkworm (*Bombyx mori*) is one of the important economic insects spinning cocoon. But the silkworm always suffers from diseases in the sericulture production, especially white muscardine caused by infection with entomopathogenic fungi. *Beauveria bassiana* was the first discovered fungus causing the devastating white muscardine disease to the silkworms and is always a troublesome pathogenic fungus in sericulture [1–3]. *B. bassiana* works by direct penetrating into the insect cuticle and colonize *in vivo* [4]. Then the fungi fight with the host's innate immune defense, consisting mainly of cellular and humoral mechanisms [5, 6]. A set of antimicrobial peptides (AMPs) are produced to protect the host against invading microbes in humoral immune system [7]. Insect AMPs are mainly synthesized in the fat body, secreted into the hemolymph, and regulated by Rel family transcription factors via activating the Toll and IMD signaling pathways [8–12].

In the silkworm, the AMPs have been identified into six groups: cecropins [13–15], moricins [16], gloverins [17], attacins [18], enbocins [19], and lebocins [20, 21]. Lebocins are first isolated from silkworm and classified as proline-rich AMPs [20, 22]. Then lebocins also have been identified in *Trichoplusia ni*, *Samia cynthia ricini*, *Pseudoplusia includes*, and *Manduca sexta* [23–26]. Lebocins are generally synthesized in the form of precursors and usually inactive and need to be proteolytically processed at RXXR motifs to generate short active peptides [27]. Unlike the fact that *M. sexta* active lebocin peptides are located at the N-terminal ends, *B. mori* lebocins are close to the C-terminal with 32 residue peptides. These peptides are O-glycosylated at Thr^{15} which is important for the antibacterial activity [22]. But abaecin of the honeybee, whose primary structure and antimicrobial activity resemble lebocins, is not O-glycosylated [28].

Four lebocins 1–4 were isolated from silkworm larvae immunized with *Escherichia coli*. Lebocins 1 and 2 have the same amino acid sequences and consequently named

lebocin 1/2. Lebocin 1/2 occurred tissues-specifically in the fat body of silkworm after bacterial induction [20]. Lebocins 3 and 4 could be induced by lipopolysaccharide (LPS) and represent tissues-specifically in the fat body and hemocytes [21]. Lebocin 1–3 showed lower antimicrobial activity against Gram-negative bacteria than cecropin B (an AMP of silkworm) in physiological condition. But they exhibited a higher bactericidal activity against *E. coli* under low-ionic-strength [22]. The self-defense mechanism of these lebocins was not clear. Lebocin 3 was confirmed that it has a synergistic effect with cecropin D in antibacterial activity [29]. Meanwhile, there are reports that silkworm lebocins could be induced by LPS or *E. coli*, but there are no reports that fungus could induce their expression and they could also not be expressed in vitro so far.

In this study, a new upregulated homologous sequences gene of lebocins was derived from our transcriptome analysis to compare differently expressed genes in the silkworm between *B. bassiana*-infected and normal larvae [30]. In order to elucidate the expression mechanisms of lebocin, the full-length cDNA of BmLeb5 by RACE according to the homologous sequences was cloned. Quantitative real-time PCR was used to analyze the expression characteristics of *BmLeb5* in the whole larvae, in the fat body, and in hemolymph at different periods after *B. bassiana* infection. The recombinant protein of BmLeb5 was successfully expressed with prokaryotic expression system and purified.

2. Materials and Methods

2.1. Silkworm and Fungal Strains. In this study, the experimental silkworm larvae were p50 strain (provided by Sericultural Research Institute of Chinese Academy of Agricultural Sciences) which were fed with fresh mulberry leaves under a photoperiod of 12 h light and 12 h dark at 25°C. The larvae were raised to third or fifth instar for subsequent experiments. Strain BbHN6 of *B. bassiana*, originally isolated from the cadaver of muscardine silkworm, was also provided by Sericultural Research Institute of Chinese Academy of Agricultural Sciences as we reported before [31]. The conidia of *B. bassiana* were gathered by culturing the fungus on potato dextrose agar (PDA) for 14 days at 25 ± 1°C.

2.2. B. bassiana Inoculation. The concentration of conidia of BbHN6 was adjusted to 1×10^8 conidia mL^{-1} in distilled water containing 0.01% (vol vol^{-1}) Tween-20 after getting rid of mycelia. About 100 larvae of third or 300 larvae of fifth instar silkworm larvae were immersed in conidia suspension for 20 s and distributed into groups in a relative humidity of breeding environment for 48 hours (hrs) and afterward at ambient humidity. The larvae of control groups were treated in distilled water containing 0.01% (vol vol^{-1}) Tween-20 and placed in the same breeding environment as the BbHN6-infected groups.

2.3. Tissues Collection. The tissues (cuticle, fat body, hemolymph, Malpighian tubule, midgut, and silk gland) of the fifth instar silkworm were collected from both control and BbHN6-infected groups from 8 hours after inoculation (hpi) to 54 hpi at intervals of four hours. The anatomical larvae tissues were quickly put into distilled water containing 0.15 M NaCl to remove attached mulberry leaf pieces. Then the same tissues of 10 larvae were pooled in a storage tube with 0.5 mL Trizol reagent (TaKaRa) and immediately frozen in liquid nitrogen. But the hemolymph was directly inpoured into the storage tubes from a ruptured proleg. All the samples were stored at −80°C for the following study. There were three repeated samples for every treatment.

2.4. Cloning of BmLeb5 Gene. The extraction of total RNA was performed from third instar silkworm larvae using Trizol reagent (TaKaRa) and detected for the ratio of 28S : 18S by agarose electrophoresis. The open reading frame (ORF) was cloned from the third instar larvae by RT-PCR using the primers Leb5orfF (ATGTACAAGTTTTTAGTATTCAGT) and Leb5orfR (CTATTCTTGGAAAATATCCCTCGG) according to our transcriptome analysis and its homologous gene *Lebocin1/2* (NP_001037468.1). Full-length cDNA sequences of *BmLeb5* were obtained using SMARTer® RACE 5′/3′ Kit (Clontech) according to the instructions of the manufacturer. The specific 5′RACE and 3′RACE primers were designed according to ORF of BmLeb5. GSP1 (GATTACGCCAAGCTTGGTCCCTTGTGTTACGGTGGCTCT) was used for cloning the 5′-untranslated regions (UTR). GSP2 (GATTACGCCAAGCTTAGAGCCACCGTAACACAAGGGACC) was used to amplify the 3′-UTR. The PCR program was conducted using the following parameters: denaturation at 94°C for 5 min, followed by 25 cycles of 94°C for 30 s, 68°C for 30 s, 72°C for 3 min, and then a final extension at 72°C for 7 min. The nested primers NGSP1 (CGGCTGGATGAACCTCTGGCACG) and NGSP2 (ACATGCGTCGGAGGATCAAGAAG) were used for generating good RACE products of 5′- and 3′-UTR using the same PCR programs.

The DNA products were gel-purified according to the instructions of the kit. Then purified DNA was cloned into pMD19-T vector (TaKaRa) and transformed into *E. coli* Top 10. The positive clones were identified by PCR and sequencing. The sequences of BmLeb5 were assembled with the obtained fragments from 5′RACE and 3′RACE using DNASTAR software and analyzed using the National Center for Biotechnology Information (NCBI) BLAST algorithm.

2.5. Bioinformatic Analysis of BmLeb5. BmLeb5 sequence obtained was analyzed using the SeqBuilder program in DNASTAR software package. The signal peptide was predicted with SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/). The potential O-glycosylation sites were predicted with NetOGlyc 4.0 (http://www.cbs.dtu.dk/services/NetOGlyc/). The homologous sequences were found by BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple protein sequences were aligned using Clustalx 2.1 software. The phylogenetic tree was constructed in MEGA 7 [32] using neighbor-joining method [33].

2.6. Quantitative Real-Time PCR (qRT-PCR). The relative expression levels of BmLeb5 were detected between control and BbHN6-infected silkworm using qRT-PCR. Total RNAs

were extracted, respectively, from third instar whole larvae of both the control and the BbHN6-infected groups at 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, and 45 hpi by using a RNApure total RNA rapid extraction kit (Beijing Bolingkewei Biotechnology Co. Ltd.). After being quantified by using a Biophotometer (Eppendorf), the same amounts of total RNAs were prepared for synthesizing the first strand cDNAs using PrimeScript™ RT Reagent Kit (TaKaRa). Then total RNAs of different tissues (cuticle, fat body, hemolymph, Malpighian tubule, midgut, and silk gland) from ten normal larvae of fifth instar were also extracted, respectively, and the cDNA templates were synthesized using the same methods. The relative expression levels of BmLeb5 in hemolymph of the fifth instar larvae at 8, 12, 16, 20, 24, 30, 36, 42, 48, and 54 hpi were also detected using qRT-PCR for the two groups. But the expression was detected for the corresponding fat body from 20 hpi because it was difficult to collect the fat body at the early period.

The qRT-PCR reaction system was 20 μL: 10 μL 2 × SYBR Premix Ex Taq™, 0.4 μL 50 × ROX Reference Dye, 0.5 μL forward primer LebqF (CGTTTAACCCCAAGCCAATA) and reverse primer LebqR (TGCACTCCGAAATCTTTTGT), and 1 μL diluted cDNA template, adding ddH_2O to 20 μL. The qRT-PCR reaction program was denaturation at 95°C for 5 min, followed by 45 cycles of 95°C for 15 s, 50°C for 15 s, and 72°C for 40 s. The *β-actin* (NM_001126254.1) was used as internal control by the primers BmactinF (CCGTATGCGAAAGGAAATCA) and BmactinR (TTGGAAGGTAGAGAGGGAGG). The qRT-PCR assays were run on an ABI Prism 7300 Sequence Detection System (Applied Biosystems) according to the protocol of the SYBR Premix Ex Taq Kit (TaKaRa). Each sample was repeated three times. The results were first analyzed with Applied Biosystems 7300 System SDS software RQ Study Application Version 1.4. The relative expression differences of *BmLeb5* were calculated by $2^{-\Delta\Delta ct}$ method [34]. Then the results were subjected to IBM SPSS Statistics 19 to analyze the statistical significance. All data were given in terms of relative mRNA expression as terms of means ± SE.

2.7. Construction of Expression Vectors and Expression of Recombinant BmLeb5. The ORF fragment of *BmLeb5* without the predicted signal peptide sequence was amplified using the upstream primer (GGATCCCAGGCTTCGTGCCAGAGG) and the downstream primer (AAGCTTCTATTCTTGGAAAATATC) by gradient PCR. The PCR products were gel-purified and cloned into pMD19-T vector (TaKaRa). After being transformed into *E. coli* Top 10, the positive clones containing pMD19-T-BmLeb5 were identified by PCR and sequencing. Then the fragment of *BmLeb5* from pMD19-T-BmLeb5 plasmid was digested with *Bam* HI and *Hind* III enzymes, gel-purified, and ligated into the vector pET-30a(+) (Novagen). The recombinant expression plasmid pET-30a(+)-BmLeb5 was also confirmed as described above. Then the recombinant plasmid was transformed into the cells of *E. coli* BL21.

The *E. coli* BL21 with pET-30a(+)-BmLeb5 was cultured overnight (about 12 h) in Luria-Bertani (LB) broth containing 30 μg/mL kanamycin. The culture was diluted 1 : 100 into fresh LB containing 30 μg/mL kanamycin and further incubated at 37°C on 220 rpm shaker until OD_{600} reached 0.6. The isopropyl β-D-thiogalactoside (IPTG) was added to the culture to a final concentration of 0.2 mM to induce the expression of the recombinant BmLeb5 for 8 h at 11°C or 37°C. Then the bacterial cells were harvested, respectively, by centrifugation at 12000g for 5 min and resuspended using 2 × SDS-PAGE loading buffer to analysis the expression by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.8. Purification of BmLeb5. The induced bacterial cells expressing BmLeb5 from the expanding culture at 11°C were resuspended in 20 mL of Binding Buffer (20 mM Tris-HCl, 0.15 M NaCl, pH 8.0) and subjected to sonication at 400 W for 20 minutes (4 s working, 8 s free) on ice. The ultrasonic mixture was centrifuged at 10,000g for 20 min. Then the centrifugation supernatant was loaded onto a Ni-IDA-Sepharose column (Novagen) which had been preequilibrated with Binding Buffer. The column was washed using Binding Buffer and Washing Buffer (20 mM Tris-HCl, 20 mM imidazole, and 0.15 M NaCl, pH 8.0) until the OD_{600} of eluate achieved baseline, respectively. The recombinant protein of BmLeb5 was detached by Elution Buffer (20 mM Tris-HCl, 250 mM imidazole, and 0.15 M NaCl, pH 8.0). The last eluate was dialysed by putting into Spectra/Por (R) Universal Closure (Millipore) overnight (about 14 h) using phosphate buffer saline (PBS, PH 7.4). The purified protein was detected by 12% SDS-PAGE. The concentration of purified BmLeb5 was determined by a Bicinchoninic Acid Kit (Beijing Bolingkewei Biotechnology Co. Ltd.).

2.9. Western Blot Analysis. For western blot analysis, 0.1 μg of recombinant protein was added in SDS-PAGE gel. Then nonstained SDS-PAGE gel was transferred to polyvinylidene fluoride membrane by electroblotting. After being blocked in PBST (1 × PBS + 0.1% Tween-20) containing 5% (vol vol^{-1}) bovine serum albumin (BSA) for 2 h, the membrane was incubated with mouse anti-His-tag polyclonal antibody in PBST for overnight at 4°C. Then the membrane was washed three times with PBST and incubated with horseradish peroxidase-conjugated anti-mouse IgG for 1 h, followed by immunoreactivity detection with diaminobenzidine (DAB) as a chromogenic substrate.

2.10. Mass Spectrum Identification. As the molecular weight of recombinant BmLeb5 was bigger than the prediction, Learning Content Management System (LC-MS) was performed to test the recombinant protein molecular weight on TripleTOF® 6600 (SCIEX) by Analyst TF 1.7 (AB Sciex) software. After confirmation of the expressed protein containing BmLeb5, Peptide Mass Fingerprinting (PMF) was developed to identify that the peptide sequence was correct using Mascot v2.3 software.

3. Results and Discussion

3.1. Confirmation of Fungal Infection. In this study, all the silkworm larvae were infected with cuticle contact

Figure 1: The disease symptoms of fifth instar silkworm infected by *B. bassiana*. (a) The oily spots in the body of diseased silkworm larvae. The control silkworms were kept healthy. (b) The large oily spots on the cuticle and the mycelia growing in cuticle junctions. (c) The mycelia and conidia of *B. bassiana* on dead silkworm body.

inoculation. The successful infection was first confirmed by the appearance of oily spots on the cuticle as typical symptoms and further observation of the blastospores in hemolymph under microscope (Figures 1(a) and 1(b)). For the third instar silkworm, the typical symptoms appeared about 48 hpi. The same phenomenon also happened to the fifth instar larvae. Finally, all the inoculated larvae died, and 2 days later the cadavers were all rigid and mycosed (Figure 1(c)), while the larvae of the control groups were kept uninfected and cocooned normally.

3.2. cDNA Cloning and Bioinformatic Analysis. The full-length cDNA of *BmLeb5* was obtained by RACE method according to the manufacturer. After sequencing and assembling, the cDNA of *BmLeb5* was deposited in GenBank (Accession number KX100575). It contained 808 bp consisting of a 39 bp 5′-untranslated region (UTR), a 540 bp ORF, and a 229 bp 3′-UTR (Figure 2). The ORF encodes a proline-rich polypeptide of 179 amino acids with a calculated molecular weight of about 21.1 kDa. The isoelectric point was 7.57 at PH 7.0. SignalP 4.1 server predicted that *BmLeb5* had a putative N-terminal signal peptide of 20 amino acids, suggesting that it is a secreted protein. Following the signal peptide, there were a putative prosegment of 100 amino acids, a mature peptide of 32 amino acids, and an additional segment of 27 amino acids at the carboxyl terminus (Figure 2). A putative polyadenylation signal of AATAAA was located at 162 nucleotides after translation stop codon and a poly (A) tail was detected in 3′-UTR (Figure 2).

One of the characteristics of *B. mori* lebocins is that they own a unique O-glycosylated threonine residue and the modification seemed to be important for the antimicrobial activity [21, 22]. Synthetic lebocins did not show obvious inhibition zones, probably due to their lack of necessary glycosylations [22, 27]. By NetOGlyc 4.0 Server, an O-glycosylation site was also predicted at Thr^{15} of BmLeb5 mature peptide as in other peptides of lebocins, suggesting that BmLeb5 is also an O-glycosylated antimicrobial peptide and could also play an important role to defend the microbial invasion (Figure 2) [35].

BmLeb5 is also one of the proline-rich AMPs and owns multiple RXXR motifs which could be recognized by intracellular processing enzymes (Figure 2) [36, 37]. There are four RXXR motifs in the prosegment of BmLeb5: $R^{39}TVR^{42}$, $R^{83}YVR^{86}$, $R^{117}NTR^{120}$, and $R^{151}YRR^{154}$, following a hydrophilic residue (Figure 2). Most AMPs are synthesized as precursor proteins and then cleaved into some small active peptides to achieve functions. The structure of BmLeb5 implies that the final active peptides of lebocins may be generated from preproteins in the secretory pathway.

Through the BLASTp search, the results of sequence comparison showed that BmLeb5 was 98% identical to *B. mori* BmLeb3 (NP_001119732.1), 91% identical to *B. mori* BmLeb4 (NP_001119731.1), 96% identical to *B. mori* BmLeb1/2 (NP_001037468.1), 56% identical to *Antheraea mylitta* AmLeb (ABG72704.1) which was the first prorich apidaecin isolated [28], 47% identical to *Pieris rapae* PrLeb (AEO21919.1), and 46% identical to *Antheraea pernyi* ApLeb (ACB45567.1). So far, 5 lebocin genes were identified in the silkworm. Sequence comparison results implied that these lebocin genes had a very close structural relationship and may be evolved from an ancestral gene. Furthermore, the phylogenetic tree was constructed using the amino acid sequences of BmLeb5 homolog proteins. All the lebocin proteins were derived from lepidoptera insects, implying that lebocins were important AMPs in insects (Figure 3). The phylogenetic relationship of silkworm lebocin family showed that an original ancestral gene first divided into two groups, lebocin 4 gene and a common gene for other 3 genes, and then the latter group was duplicated into lebocin 1/2 and lebocin 3 and 5. Lebocin 5 and 3 owned the highest homology and formed a branch (Figure 3). The discovery of BmLeb5 suggested that the lebocin underwent a further gene evolution in silkworm along with environmental threats. The results also suggested

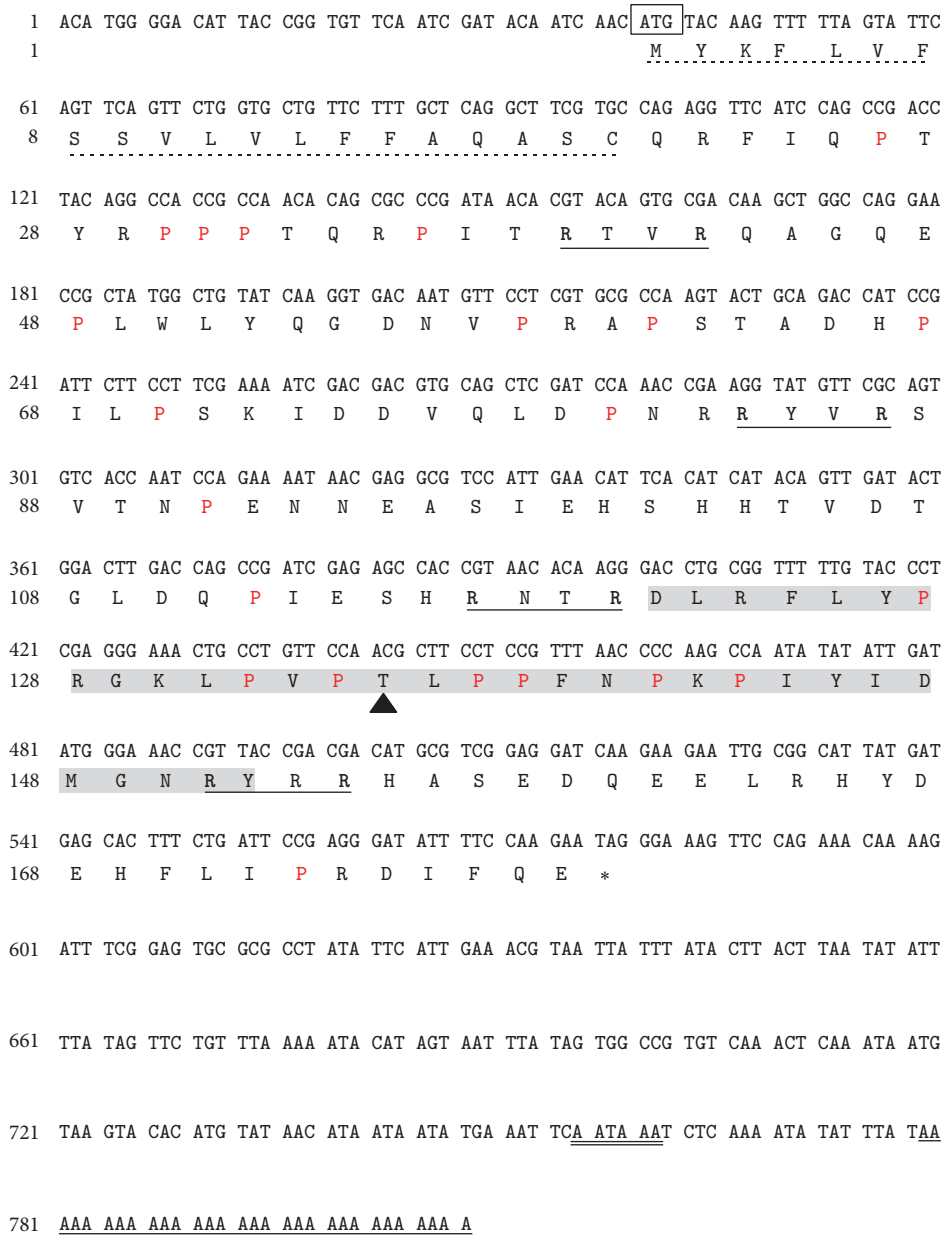


Figure 2: The nucleotide sequence of *BmLeb5* full-length cDNA and deduced amino acid sequence. The predicted signal sequence at the N-terminus is marked with a dotted underline. The start codon is boxed. The asterisk (∗) represents the stop codon. The polyadenylation signal (AATAAA) is marked with double underlines and polyadenylation tail is underlined. The mat peptide of BmLeb5 is marked in gray bars. The O-glycosylation site is indicated by triangle (▲). The proline-rich polypeptides are marked in red. RXXR motifs are marked in bold arginine and underlined. The numeric positions of the nucleotide and amino acid sequences are shown on the left.

that lebocin owned a multiple gene family as cecropins in the silkworm [13, 15, 38].

3.3. Expression Profiling of BmLeb5. The relative expression of *BmLeb5* was detected by qRT-PCR between the control and infected silkworm larvae. In the third instar whole larvae inoculated with BbHN6, the relative expression of *BmLeb5* was upregulated as compared to the normal ones. *B. bassiana* always could penetrate into silkworm cuticle in 8 hpi. So at the early stage of 9 hpi, the expression of *BmLeb5* had been upregulated about 5-fold than that in the normal groups. Afterwards, it was upregulated about 12-, 13-, 33-, 59-, 95-, 62-, 36-, and 32-fold than that in control ones at 24, 27, 30, 33, 36, 39, 42, and 45 hpi, respectively (Figure 4(a)). The results showed that the expression of *BmLeb5* was rapidly induced by *B. bassiana* challenge and reached a maximum difference level of 95-fold at 36 hpi. This was consistent with the results of our previous transcriptome analysis [30]. Then, the tissue distribution of *BmLeb5* was further investigated in the fifth instar silkworm larvae with qRT-PCR. The tested tissues included cuticle, fat body, hemolymph, Malpighian tubule, midgut, and silk gland. The *BmLeb5* mRNA could be detected

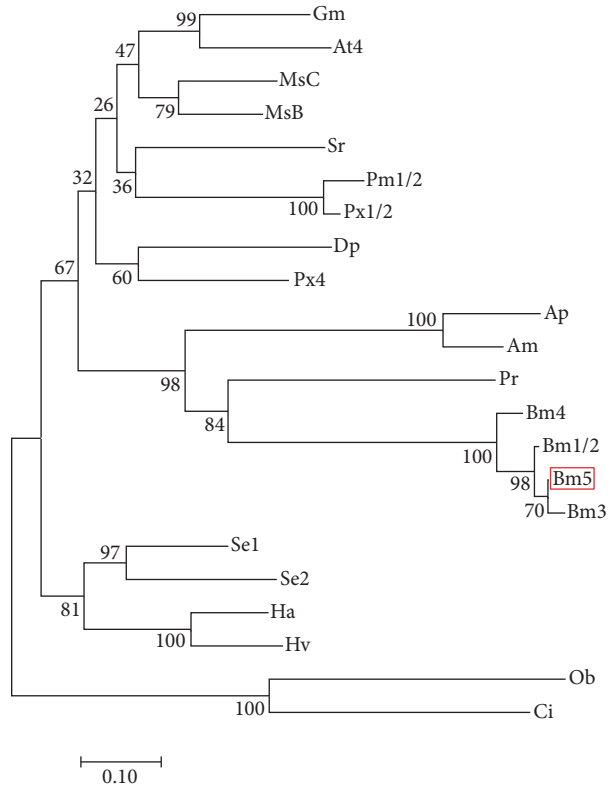


Figure 3: Unrooted phylogenetic tree of BmLeb5 and other homologous proteins. The tree was constructed by the neighbor-joining method within the package MEGA 7. Bootstrap majority consensus values on 1,000 replicates are indicated at each branch point (%). The scale bar represents branch length (number of amino acids substitutions/100 residues). Accession numbers for sequences used in this alignment are as follows (GenBank ID): Bm5 (*Bombyx mori*, KX100575); Bm3 (*Bombyx mori*, NP_001119732.1); Bm4 (*Bombyx mori*, NP_001119731.1); Bm1/2 (*Bombyx mori*, NP_001037468.1); Se1 (*Spodoptera exigua*, AKJ54499.1); Gm (*Galleria mellonella*, ACQ99193.1); Ha (*Helicoverpa armigera*, ALT16900.1); At4 (*Amyelois transitella*, XP_013188814.1); Ap (*Antheraea pernyi*, ACB45567.1); MsB (*Manduca sexta*, ADE20197.1); MsC (*Manduca sexta*, ADE20198.1); Hv (*Heliothis virescens*, ACR78447.1); Pm1/2 (*Papilio machaon*, KPJ06168.1); Px1/2 (*Papilio xuthus*, KPI91575.1); Se2 (*Spodoptera exigua,* AKJ54500.1); Sr (*Samia ricini*, BAD84189.1); Dp (*Danaus plexippus*, EHJ64534.1); Px4 (*Papilio xuthus*, KPI91576.1); Pr (*Pieris rapae*, AEO21919.1); Am (*Antheraea mylitta*, ABG72704.1); Ob (*Operophtera brumata*, KOB78261.1); Ci (*Chrysodeixis includens*, AAS48093.1).

in all the tested tissues of normal larvae but the highest in the fat body (Figure 4(b)). While the relative expressions of *BmLeb5* were only upregulated in fat body and hemolymph in the inoculated larvae, no obvious changes were detected in the other four tissues. The maximum upregulation of expression was approximately 20-fold in fat body as compared to the normal ones (Figure 4(c)). Significant difference in hemolymph achieved 22-fold at 20 hpi (Figure 4(d)). When the internal control of β-actin was changed to GAPDH, the general trend of expression changes was the same.

Previous studies had demonstrated that *B. mori* lebocin 1/2 could be induced and strongly expressed in fat body by bacterial injection [20]. *B. mori* lebocin 3 and lebocin 4 were showed to be induced in the fat body and hemocytes by lipopolysaccharide (LPS) [21]. Our qRT-PCR results for lebocin 5 were consistent with these reports. Without exception, the lebocins are expressed the highest in fat body of silkworm. *B. mori* lebocin 1–3 could be induced by *E. coli* and *B. mori* lebocin 4 could be induced by LPS [20, 22, 29]. Furthermore, the purified peptides of lebocin 1–3 showed antibacterial activity against some Gram-negative bacteria [22]. So far, there are no reports indicating that fungi could induce *B. mori* lebocin expression. Our results suggested that BmLeb5 not only could be induced to the upregulated expression by fungi, but also maybe play an important role in the silkworm to defend *B. bassiana* infection.

3.4. Expression and Purification of Recombinant BmLeb5. The recombinant expression vector pET-30a(+)-BmLeb5 was constructed by inserting the BmLeb5 ORF fragment (without signal peptide sequence) and transformed into *E. coli* BL21 for expression of BmLeb5. After IPTG induction at 11°C and 37°C, respectively, the expression quantity of fusion protein was detected by SDS-PAGE. Though the expression was commonly induced at the temperature of 37°C, the induced expression efficiency for BmLeb5 at 11°C was significantly higher than at 37°C (Figure 5(a)). As we know, there are no reports about the successful in vitro expression of any *B. mori* lebocin genes. In *Manduca sexta*, only fusion protein of lebocin B was expressed at 16°C after IPTG induction [27].

After being purified using a Ni-IDA-Sepharose column, the recombinant BmLeb5 exhibited a significant

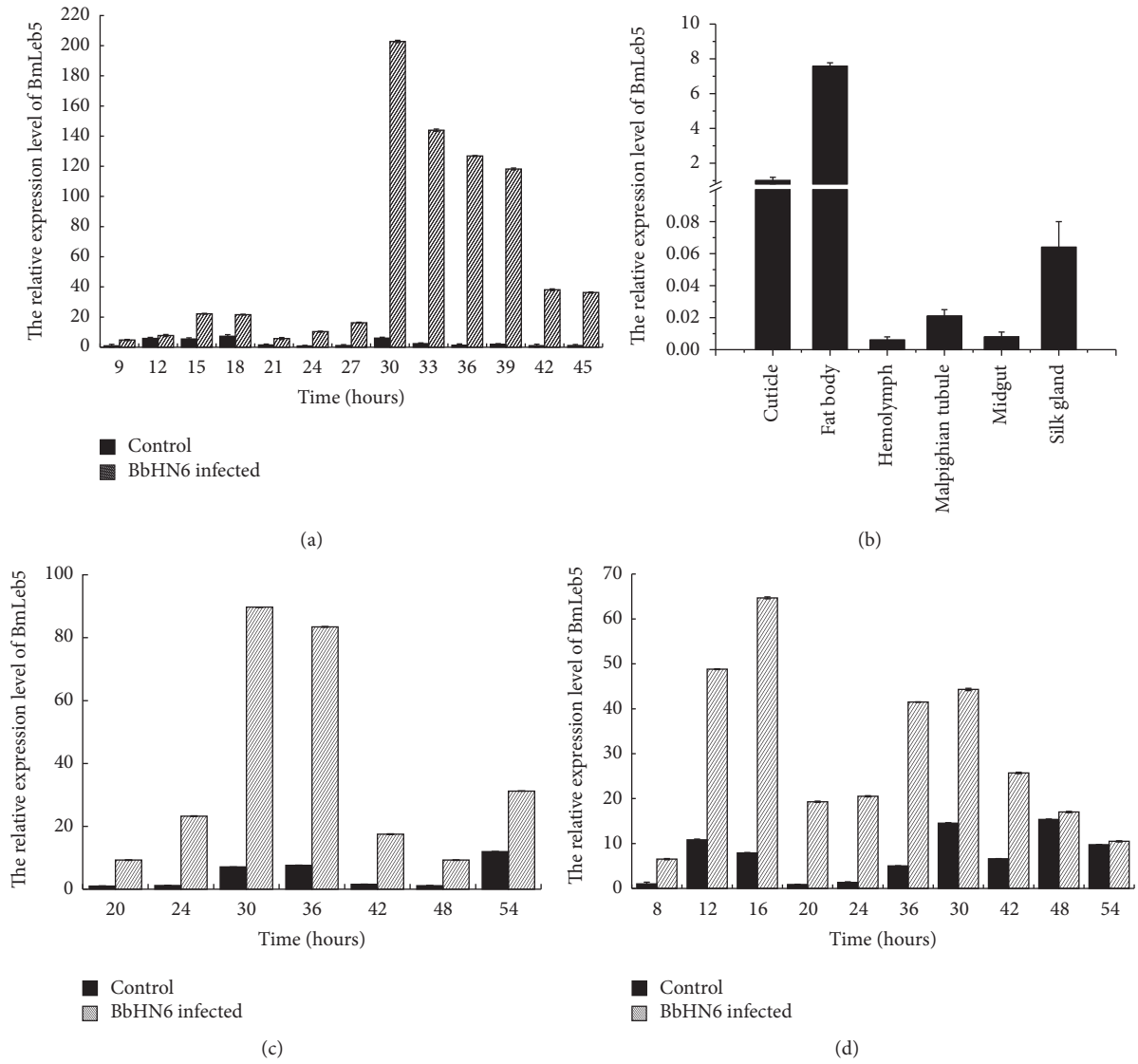


Figure 4: The relative mRNA expression levels of BmLeb5. (a) In the third instar whole larvae of BbHN6-infected and control groups at different times. (b) In different tissues of normal fifth instar larvae. (c) In the fat body of BbHN6-infected and control fifth instar larvae at different times. (d) In the hemolymph of BbHN6-infected and control fifth instar larvae at different times. The y-axis indicates the relative expression level of BmLeb5 mRNA transcripts. Vertical bars represent the mean ± SE ($n = 3$).

fusion protein band of approximately 32 kDa on 12% SDS-PAGE followed by staining with Coomassie Brilliant Blue R-250. Western blot assay was used to further confirm the purified recombinant protein BmLeb5 using anti-histidine antibody. As shown in Figure 5(b), the single protein band revealed that the prokaryotic expression system achieved an excellent inducible expression of BmLeb5. The expressed recombinant protein would be used for the detection of antifungal activity of BmLeb5 and its antifungal mechanism.

3.5. Confirmation of the Recombinant Protein of BmLeb5. By SDS-PAGE and western blot analysis, the molecular mass of recombinant BmLeb5 was approximately 32 kDa, higher than the predicted 24.6 kDa. In order to confirm the target protein, LC-MS was used to detect the size of molecular mass of the purified recombinant protein. Two peaks of molecular mass, corresponding to 24.6 kDa and 31.8 kDa (Figure 6), were presented. The result could judge that the two substances of 24.6 kDa and 31.8 kDa were main ingredients in the solution of recombinant protein. The 24.6 kDa peak was consistent with the predicted results. After trypsin enzymolysis and secondary mass spectrum identification, the sequences of peptide fragments were blasted with database of UniProt by Mascot v2.3 software. 12 peptides were consistent with the sequence of BmLeb5. After assembling, the vast majority of fragments were completely in accordance with the recombinant protein of BmLeb5. The other peptide of 31.8 kDa was blasted as a protein kinase which probably was reassembled from serine residues of the added tryptase in the solution. Therefore, we established successfully for the first time the expression system for BmLeb5 and the expressed protein

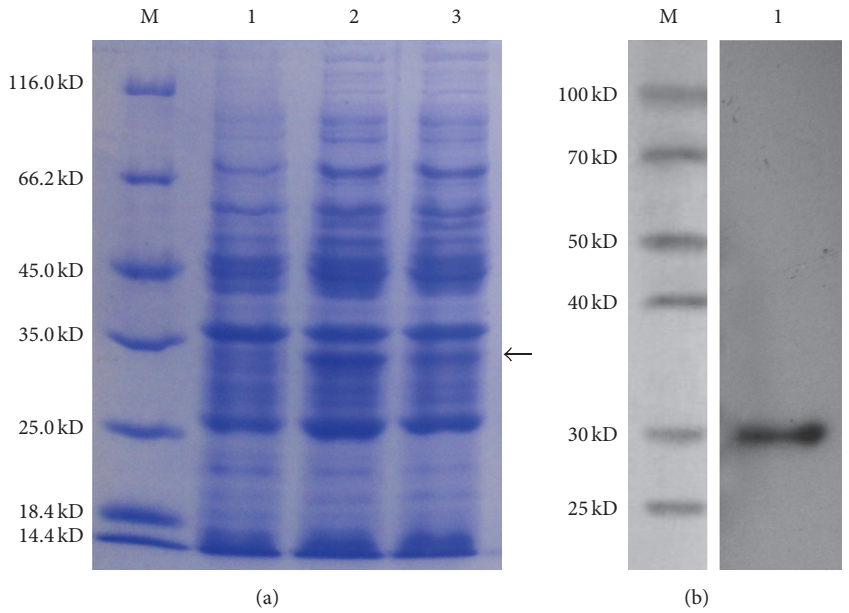


Figure 5: SDS-PAGE and western blot analysis of recombinant BmLeb5. (a) SDS-PAGE analysis of recombinant BmLeb5 expressed at different induction temperatures. Lane M: protein marker (14.4–116 kDa); lane 1: uninduced; lane 2: induced at 11°C; lane 3: induced at 37°C. (b) Western blot analysis of purified recombinant BmLeb5. Lane M: protein marker (25–100 kDa); lane 1: 0.1 μg recombinant protein of BmLeb5.

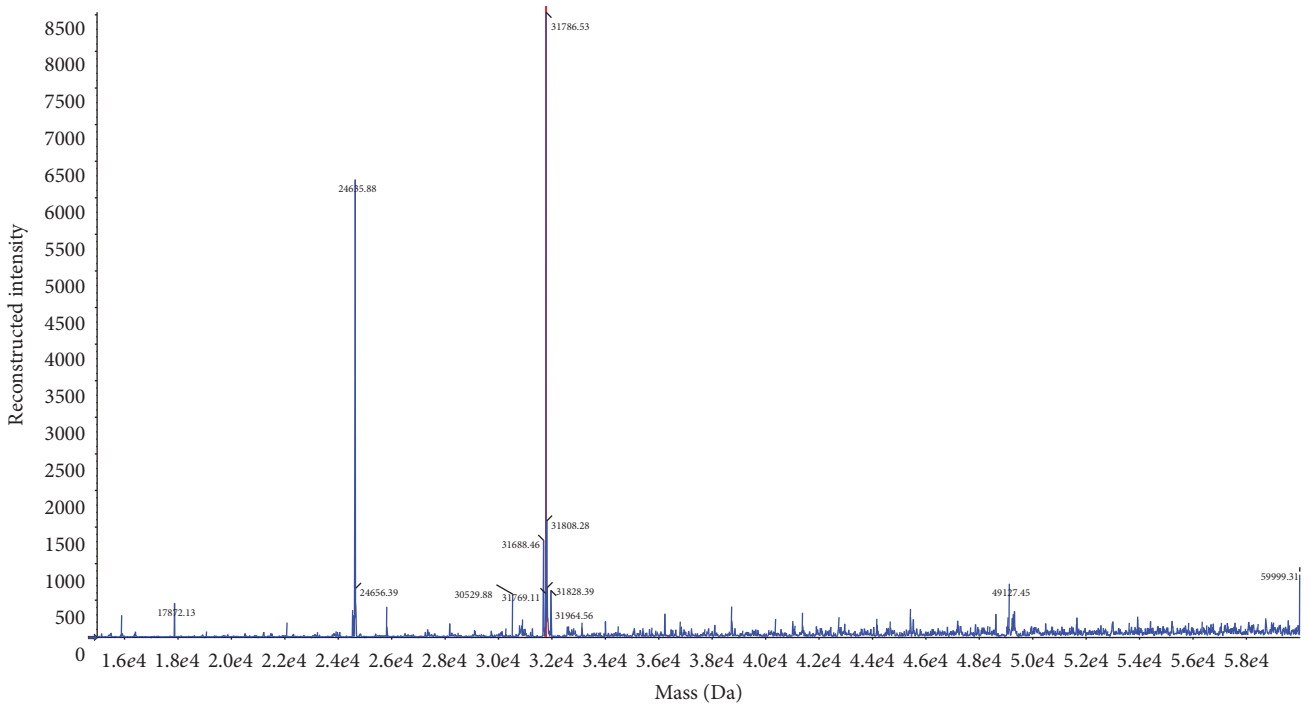


Figure 6: Quality analysis of purified recombinant BmLeb5 with mass spectrometry. There were two substances of 24.6 kDa and 31.8 kDa in the solution of recombinant protein. The 24.6 kDa peak was consistent with the predicted result.

could be used for further study of the roles of lebocins both in vitro and in vivo in the silkworm.

In conclusion, we successfully cloned the full-length cDNA of BmLeb5 gene from silkworm using RACE and the sequence was submitted to GenBank. The deduced amino acid sequence contains an O-glycosylation site and four RXXR motifs. The phylogenetic tree suggests that lebocins form a multiple gene family in silkworm. The upregulated expression of BmLeb5 in the whole silkworm larvae, in the fat body, and in hemolymph indicates that it might play an important role in the immune response of silkworm to defend the *B. bassiana* infection. The recombinant protein

of BmLeb5 was for the first time successfully expressed with prokaryotic expression system and then purified with Ni-IDA-Sepharose column. We also confirmed the molecular mass and sequences of the expressed protein by LC-MS and PMF. Details of antifungal mechanism of the BmLeb5 *in vivo* remain to be explored.

Competing Interests

The authors declare there is no conflict of interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 31272498), the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant no. 20123220110003), the Graduate Science and Technology Innovation Plan of Jiangsu Province (Grant no. KYZZ16_0205), and the Natural Science Foundation of Jiangsu Province (Grant no. BK20161364).

References

[1] C.-X. Hou, G.-X. Qin, T. Liu et al., "Differentially expressed genes in the cuticle and hemolymph of the silkworm, *Bombyx mori*, injected with the fungus *Beauveria bassiana*," *Journal of Insect Science*, vol. 13, article 138, 14 pages, 2013.

[2] X. Chen, C. Huang, L. He, S. Zhang, and Z. Li, "Molecular tracing of white muscardine in the silkworm, *Bombyx mori* (Linn.) II. Silkworm white muscardine is not caused by artificial release or natural epizootic of *Beauveria bassiana* in China," *Journal of Invertebrate Pathology*, vol. 125, pp. 16–22, 2015.

[3] E. A. Steinhaus, "Microbial diseases of insects," *Annual review of microbiology*, vol. 11, pp. 165–182, 1957.

[4] D. Lu, M. Pava-Ripoll, Z. Li, and C. Wang, "Insecticidal evaluation of Beauveria bassiana engineered to express a scorpion neurotoxin and a cuticle degrading protease," *Applied Microbiology and Biotechnology*, vol. 81, no. 3, pp. 515–522, 2008.

[5] M. D. Lavine and M. R. Strand, "Insect hemocytes and their role in immunity," *Insect Biochemistry and Molecular Biology*, vol. 32, no. 10, pp. 1295–1309, 2002.

[6] N. A. Ratcliffe, "Invertebrate immunity—a primer for the non-specialist," *Immunology Letters*, vol. 10, no. 5, pp. 253–270, 1985.

[7] P. Bulet, R. Stöcklin, and L. Menin, "Anti-microbial peptides: from invertebrates to vertebrates," *Immunological Reviews*, vol. 198, no. 1, pp. 169–184, 2004.

[8] C. Luna, X. Wang, Y. Huang, J. Zhang, and L. Zheng, "Characterization of four Toll related genes during development and immune responses in *Anopheles gambiae*," *Insect Biochemistry and Molecular Biology*, vol. 32, no. 9, pp. 1171–1179, 2002.

[9] P. Bulet, C. Hetru, J.-L. Dimarcq, and D. Hoffmann, "Antimicrobial peptides in insects; structure and function," *Developmental and Comparative Immunology*, vol. 23, no. 4-5, pp. 329–344, 1999.

[10] B. Lemaitre and J. Hoffmann, "The host defense of *Drosophila melanogaster*," *Annual Review of Immunology*, vol. 25, no. 1, pp. 697–743, 2007.

[11] C. Hetru and J. A. Hoffmann, "NF-kappaB in the immune response of *Drosophila*," *Cold Spring Harbor perspectives in biology*, vol. 1, no. 6, p. a000232, 2009.

[12] T. Cheng, P. Zhao, C. Liu et al., "Structures, regulatory regions, and inductive expression patterns of antimicrobial peptide genes in the silkworm Bombyx mori," *Genomics*, vol. 87, no. 3, pp. 356–365, 2006.

[13] Y. Yamano, M. Matsumoto, K. Sasahara, E. Sakamoto, and I. Morishima, "Structure of genes for cecropin A and an inducible nuclear protein that binds to the promoter region of the genes from the silkworm, *Bombyx mori*," *Bioscience, Biotechnology and Biochemistry*, vol. 62, no. 2, pp. 237–241, 1998.

[14] K. Taniai, K. Kadono-Okuda, Y. Kato et al., "Structure of two cecropin B-encoding genes and bacteria-inducible DNA-binding proteins which bind to the 5′-upstream regulatory region in the silkworm, *Bombyx mori*," *Gene*, vol. 163, no. 2, pp. 215–219, 1995.

[15] J. Yang, S. Furukawa, A. Sagisaka et al., "cDNA cloning and gene expression of cecropin D, an antibacterial protein in the silkworm, Bombyx mori," *Comparative Biochemistry and Physiology—B Biochemistry and Molecular Biology*, vol. 122, no. 4, pp. 409–414, 1999.

[16] S. Hara and M. Yamakawa, "Moricin, a novel type of antibacterial peptide isolated from the silkworm, *Bombyx mori*," *Journal of Biological Chemistry*, vol. 270, no. 50, pp. 29923–29927, 1995.

[17] S. Kawaoka, S. Katsuma, T. Daimon et al., "Functional analysis of four gloverin-like genes in the silkworm, *Bombyx mori*," *Archives of Insect Biochemistry and Physiology*, vol. 67, no. 2, pp. 87–96, 2008.

[18] M. Sugiyama, H. Kuniyoshi, E. Kotani et al., "Characterization of a *Bombyx mori* cDNA encoding a novel member of the attacin family of insect antibacterial proteins," *Insect Biochemistry and Molecular Biology*, vol. 25, no. 3, pp. 385–392, 1995.

[19] S. H. Kim, B. S. Park, E. Y. Yun et al., "Cloning and expression of a novel gene encoding a new antibacterial peptide from silkworm, *Bombyx mori*," *Biochemical and Biophysical Research Communications*, vol. 246, no. 2, pp. 388–392, 1998.

[20] S. Chowdhury, K. Taniai, S. Hara et al., "cDNA cloning and gene expression of lebocin, a novel member of antibacterial peptides from the silkworm, *Bombyx mori*," *Biochemical and Biophysical Research Communications*, vol. 214, no. 1, pp. 271–278, 1995.

[21] S. Furukawa, K. Taniai, J. Ishibashi, S. Hara, T. Shono, and M. Yamakawa, "A novel member of lebocin gene family from the silkworm, *Bombyx mori*," *Biochemical and Biophysical Research Communications*, vol. 238, no. 3, pp. 769–774, 1997.

[22] S. Hara and M. Yamakawa, "A novel antibacterial peptide family isolated from the silkworm, *Bombyx mori*," *Biochemical Journal*, vol. 310, no. 2, pp. 651–656, 1995.

[23] G. Liu, D. Kang, and H. Steiner, "Trichoplusia ni lebocin, an inducible immune gene with a downstream insertion element," *Biochemical and Biophysical Research Communications*, vol. 269, no. 3, pp. 803–807, 2000.

[24] Y. Bao, Y. Yamano, and I. Morishima, "A novel lebocin-like gene from eri-silkworm, *Samia cynthia ricini*, that does not encode the antibacterial peptide lebocin," *Comparative Biochemistry and Physiology—B Biochemistry and Molecular Biology*, vol. 140, no. 1, pp. 127–131, 2005.

[25] M. D. Lavine, G. Chen, and M. R. Strand, "Immune challenge differentially affects transcript abundance of three antimicrobial peptides in hemocytes from the moth *Pseudoplusia includens*," *Insect Biochemistry and Molecular Biology*, vol. 35, no. 12, pp. 1335–1346, 2005.

[26] S. Rayaprolu, Y. Wang, M. R. Kanost, S. Hartson, and H. Jiang, "Functional analysis of four processing products from multiple

precursors encoded by a lebocin-related gene from *Manduca sexta*," *Developmental and Comparative Immunology*, vol. 34, no. 6, pp. 638–647, 2010.

[27] X.-J. Rao, X.-X. Xu, and X.-Q. Yu, "Functional analysis of two lebocin-related proteins from *Manduca sexta*," *Insect Biochemistry and Molecular Biology*, vol. 42, no. 4, pp. 231–239, 2012.

[28] K. Casteels-Josson, W. Zhang, T. Capaci, P. Casteels, and P. Tempst, "Acute transcriptional response of the honeybee peptide-antibiotics gene repertoire and required post-translational conversion of the precursor structures," *Journal of Biological Chemistry*, vol. 269, no. 46, pp. 28569–28575, 1994.

[29] S. Y. Hara and M. Yamakawa, "Cooperative antibacterial relationship between lebocin and cecropin D, antibacterial peptides isolated from the silkworm, *Bombyx mori* (Lepidoptera: Bombycidae)," *Applied Entomology and Zoology*, vol. 30, no. 4, pp. 606–608, 1995.

[30] C. Hou, G. Qin, T. Liu et al., "Transcriptome analysis of silkworm, Bombyx mori, during early response to Beauveria bassiana challenges," *PLoS ONE*, vol. 9, no. 3, Article ID e91189, 2014.

[31] D. Lu, T. Geng, C. Hou, Y. Huang, G. Qin, and X. Guo, "*Bombyx mori* cecropin A has a high antifungal activity to entomopathogenic fungus *Beauveria bassiana*," *Gene*, vol. 583, no. 1, pp. 29–35, 2016.

[32] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.

[33] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406–425, 1987.

[34] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method," *Methods*, vol. 25, no. 4, pp. 402–408, 2001.

[35] I. B. H. Wilson, Y. Gavel, and G. Von Heijne, "Amino acid distributions around O-linked glycosylation sites," *Biochemical Journal*, vol. 275, no. 2, pp. 529–534, 1991.

[36] L. Devi, "Consensus sequence for processing of peptide precursors at monobasic sites," *FEBS Letters*, vol. 280, no. 2, pp. 189–194, 1991.

[37] J. A. Veenstra, "Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors," *Archives of Insect Biochemistry and Physiology*, vol. 43, no. 2, pp. 49–63, 2000.

[38] Y. Yamano, M. Matsumoto, K. Inoue, T. Kawabata, and I. Morishima, "Cloning of cdnas for cecropins a and b, and expression of the genes in the silkworm, *bombyx mori*," *Bioscience, Biotechnology and Biochemistry*, vol. 58, no. 8, pp. 1476–1478, 1994.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 4841756, 20 pages
http://dx.doi.org/10.1155/2016/4841756

Research Article

RNA Sequencing Reveals Xyr1 as a Transcription Factor Regulating Gene Expression beyond Carbohydrate Metabolism

Liang Ma, Ling Chen, Lei Zhang, Gen Zou, Rui Liu, Yanping Jiang, and Zhihua Zhou

Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

Correspondence should be addressed to Zhihua Zhou; zhouzhihua@sippe.ac.cn

Received 31 August 2016; Accepted 6 November 2016

Academic Editor: Muhammad I. Rajoka

Xyr1 has been demonstrated to be the main transcription activator of (hemi)cellulases in the well-known cellulase producer *Trichoderma reesei*. This study comprehensively investigates the genes regulated by Xyr1 through RNA sequencing to produce the transcription profiles of *T. reesei* Rut-C30 and its *xyr1* deletion mutant (Δ*xyr1*), cultured on lignocellulose or glucose. *xyr1* deletion resulted in 467 differentially expressed genes on inducing medium. Almost all functional genes involved in (hemi)cellulose degradation and many transporters belonging to the sugar porter family in the major facilitator superfamily (MFS) were downregulated in Δ*xyr1*. By contrast, all differentially expressed protease, lipase, chitinase, some ATP-binding cassette transporters, and heat shock protein-encoding genes were upregulated in Δ*xyr1*. When cultured on glucose, a total of 281 genes were expressed differentially in Δ*xyr1*, most of which were involved in energy, solute transport, lipid, amino acid, and monosaccharide as well as secondary metabolism. Electrophoretic mobility shift assays confirmed that the intracellular *β*-glucosidase *bgl2*, the putative nonenzymatic cellulose-attacking gene *cip1*, the MFS lactose transporter *lp*, the *nmrA-like* gene, *endo T*, the acid protease *pepA*, and the small heat shock protein *hsp23* were probable Xyr1-targets. These results might help elucidate the regulation system for synthesis and secretion of (hemi)cellulases in *T. reesei* Rut-C30.

1. Introduction

Lignocelluloses have long been recognized as the most abundant sustainable resources for the production of biofuels and other biomaterials [1, 2]. Cellulose must be hydrolyzed by cellulases to soluble carbohydrates to facilitate fermentation. The typical cellulase system consists of endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91), and *β*-glucosidase (BGL; EC 3.2.1.21), which act synergistically to hydrolyze cellulose to glucose [3]. The ascomycete *Trichoderma reesei* (anamorph of *Hypocrea jecorina*) has been used widely as a cellulose source since its discovery during World War II [4]. However, the cost of lignocellulolytic enzyme preparation is still among the major limitations in the development of an acceptable technology to convert lignocellulose to biofuels and other chemicals [5, 6]. Therefore, *T. reesei* should be further genetically engineered to acquire an improved strain for cellulase production.

Several transcription factors (TFs) involved in the regulation of cellulase gene expression have been identified in *T. reesei*, including the activators Ace2, Xyr1, and Hap2/3/5, as well as the repressor Ace1 and the carbon catabolite repressor Cre1 [7]. Xyr1, a homolog to XlnR in *Aspergillus niger*, is a zinc binuclear cluster protein binding to a 5′-GGCTAA-3′ motif arranged as an inverted repeat [8]. It was demonstrated to play an essential role in transcriptional regulation of cellulolytic and xylanolytic genes, such as *xyn1*, *xyn2*, *bxl1*, *abf2*, *cbh1*, *cbh2*, *egl1*, and *bgl1* [9, 10]. In addition, Xyr1 was reported to receive the lactose induction signal and regulate lactose metabolism by directly activating xylose reductase 1 transcription and indirectly influencing transcription of *β*-galactosidase 1 *(bga1)* [11]. Unlike *xyr1* in *T. reesei*, the deletion of xlnR (ortholog to xlnR in *Aspergillus*) in *Fusarium oxysporum* affects only xylanase activity [12].

Recently, the TF xylan degradation regulator 1 (XLR-1), an ortholog to XlnR/Xyr1 in *A. niger* and *T. reesei*, was identified in *Neurospora crassa*. Deletion of *xlr-1* in *N. crassa* prevented growth on xylan and xylose, but its cellulolytic activity was only slightly affected, indicating a different role from *xyr1* in *T. reesei* [13]. Besides, secretome analyses of wild type and

the *xlnR*/*xlr1*/*xyr1* deletion mutants of five fungi showed that *T. reesei* Xyr1 has a different regulatory pattern compared to its orthologs in other fungi [14]. The above findings, combined with the demonstration that Xyr1 in *T. reesei* could bind not only to the 5′-GGCTAA-3′ motif, but also to the 5′-GGC(A/T)$_3$-3′ motif [15], suggest that Xyr1 behaves as a pleiotropic regulator in *T. reesei*.

Recently, transcription profiling of the *T. reesei* Qm 9414 and its Δ*xyr1* mutant grown on cellulose, sophorose, and glucose were performed and defined the role of the transcriptional factor Xyr1 during cellulose degradation [16]. *T. reesei* mutant Rut-C30 is a hyperproducer of cellulolytic enzymes with its genome has been released [3, 17, 18]. Rut-C30 was obtained through several rounds of random mutagenesis from wt Qm6a. The rearrangement of chromosomes carrying genes encoding cellulolytic enzymes [19] and the missing >100 kb of genomic DNA [20] including the truncation of carbon catabolite repressor *cre1* [17] may contribute to its high protein secretory ability and cellulase production. Portnoy et al. reported challenging results indicating that the full transcription of *xyr1* required Cre1 in *T. reesei* Qm9414 under induction conditions [21]. Due to the special genetic background of Rut-C30, we assumed that its Xyr1 harbored rather special regulatory mechanisms compared to *T. reesei* Qm9414.

Wheat (*Triticum aestivum* L.) bran, which contains lignocelluloses as a major component, is rich in hemicelluloses, cellulose, and lignin [22, 23]. Therefore, wheat bran behaves as an inducer for lignocellulolytic enzymes. In this study, RNA sequencing (RNA-seq) was performed to investigate the functions of Xyr1 through comparison between a wild-type strain Rut-C30 and an *xyr1* disruptant, under lignocellulose and glucose conditions. Our results shed new light on the mechanism by which Xyr1 controls cellulose and hemicellulose utilization and determines the pleiotropic functions of Xyr1. These new findings could offer strategies for strain improvement of *T. reesei* Rut-C30.

2. Materials and Methods

2.1. Fungal Strains and Cultivation Conditions. *T. reesei* Rut-C30 (ATCC 56765) was purchased from ATCC and its *xyr1* deletion mutant strain Δ*xyr1* was constructed as following. Plasmids were propagated in *Escherichia coli* DH5*α*. The vector backbone used in constructing the plasmids was binary vector pCambia1300 (CAMBIA, Canberra, Australia). The *E. coli* cultivations were performed overnight at 37°C in Luria-Bertani (LB) medium plus kanamycin (100 μg ml^{-1}) as selective agent.

The *xyr1* deletion vector was constructed using the binary vector pCambia1300 as a recipient. Primers used are given in Table S1 in Supplementary Material available online at http://dx.doi.org/10.1155/2016/4841756. First, the plasmid pSilent-1 [24] was used as template to amplify the ORF and terminator of hygromycin B phosphotransferase gene *hph* (conferring hygromycin B resistance) with primers HygxhoI and Orfhyg. Second, the pyruvate kinase (*pki*, GenBank accession number L07060) promoter was amplified by primers Ppki and Bampki with *T. reesei* Rut-C30 genome as template. Overlap PCR was conducted to fuse the *pki* promoter to *hph*ORF, and the PCR product was digested with *Bam*HI and *Xho*I and ligated into the *Bam*HI and *Xho*I digested pCambia1300, resulting in pPH. Then the 1.3 kb fragment for 5′ region of *xyr1* was amplified with primers xyr1xhf and xyr1xhr. After digestion by *Xho*I, it was inserted to the *Xho*I digested pPH to generate pPH5X. At last, the 3′ region of *xyr1* was amplified with primers xyr1bam and xyr1sal and digested by *Bam*HI and *Sal*I before inserting into the corresponding restriction enzymes digested product of pPH5X to yield the *xyr1* knockout vector pPHX. The resultant pPHX was transformed into *T. reesei* by *Agrobacterium*-mediated transformation as described previously [25]. Transformants were then subjected to verification of homologous recombination event. Primers X5 and TtrpC were used to amplify a fragment of 2.2 kb in the *xyr1* knockout strain, and similarly primers X3 and Ppki were used to amplify a fragment of 2.3 kb in the *xyr1* knockout strain and the results were further verified by sequencing the PCR products, while random insertion could not yield any specific PCR products.

The *xyr1* recomplementation strain was constructed as: the phleomycin resistance gene *(ble)* [26] was used as a selection marker. The binary vector pPB was constructed in a similar way except replacing the *hph* ORF with *ble*. The *xyr1* gene was amplified with primers X3BamH and X5BamH, and the terminator of *xyr1* was amplified with primers Txyr5Xho and Txyr3Xho (Table S1). The PCR products were digested by *BamH*I and *Xho*I, respectively, and then inserted into corresponding site of the pPB to yield pPBReX. The transformation into Δ*xyr1* strain was performed as described above and transformants were selected on PDA containing phleomycin (4 μg ml^{-1}) as selection agent. The homologous integration of pPBReX at the *xyr1* locus of the Δ*xyr1* strain was verified by PCR. Primers B3homo and pki-phe were used to amplify a fragment of 2.6 kb in the *xyr1* homologous integration retransformant, and similarly, primers X5 and Ppki (Table S1) were used to amplify a fragment of 5.7 kb and the results were further verified by sequencing the PCR products, while random insertion could not yield any specific PCR products.

All of the fungal strains were maintained on potato-dextrose agar (PDA). Then, 10^7 conidia of both strains were inoculated in 50-ml shake flasks containing 10 ml of Sabouraud's dextrose broth (SDB) at 28°C for two days, and then the pregrown mycelia were collected by filtration through miracloth and washed with 0.9% NaCl, thoroughly. Then, the mycelia of both strains were transferred into the cellulase-inducing medium described by Ma et al. [25] which contains 2% wheat bran, 3% microcrystalline cellulose (Avicel), and the cellulase-repressing medium containing 2% glucose in place of wheat bran and Avicel, respectively. The flasks were incubated on an orbital shaker at 200 rpm, 28°C. For RNA isolation, the samples of mycelium were collected after 15 hours of cultivation and then subjected to RNA isolation. For fungal growth and protein secretion analysis, 500 μl samples were collected after being induced for 1 day, 2 days, 3 days, and 7 days. After centrifugation, the culture supernatant was subjected to electrophoresis,

extracellular protein concentration assay, enzymatic assays, and the mycelia which were used to quantify the biomass.

2.2. RNA Isolation, Sequencing, and Data Analysis. Fungal mycelia of four samples were harvested by filtration and centrifugation, frozen, and ground under liquid nitrogen. Total RNA were isolated by TRIzol (Invitrogen) according to the instruction, and RNA extracts were monitored by electrophoresis and quantified using a spectrophotometer.

Total RNA were provided to Chinese National Human Genome Center at Shanghai for sequencing. mRNA purified from total RNA using the MicroPoly(A)Purist kit (Ambion) was used for library preparation and latter sequencing using the Illumina Hiseq2000 platform. The quality control was performed with FASTX-Toolkit (Version 0.0.13), and the reads were filtrated when more than 20% of bases with PHRED score were lower than 30. The cleaned RNA-seq reads were mapped to the *T. reesei* Rut-C30 genome using Tophat V1.3.2 [27] with Bowtie V0.12.7.0 [28] with a minimum intron size of 20 bp and a maximum intron size of 100,000 bp; the mean inner distance between mate pairs and the standard deviation for the distribution on inner distances between mate pairs were set to 150 and 75, respectively. Transcript abundance indicated as the number of fragments per kilobase per million fragments mapped (FPKM) was estimated with the program Cuffdiff from the package Cufflinks v0.83 [29]. Differential expression values were determined using DEseq [30, 31]. FDR (False Discovery Rate) control (the cutoff was 0.01) was used to identify the significance of differences for multiple tests [32]. The genome sequence and gene structure information for *T. reesei* QM6a (http://genome.jgi.doe.gov/Trire2/Trire2.home.html) and Rut-C30 (http://genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html) were obtained from the JGI genome Portal site [18, 33]. The FunCat [34] annotation for QM6a was downloaded from the PENDANT genome database [35]. The FunCat annotation for Rut-C30 was performed using a blastn [36] search between QM6a and Rut-C30. The differentially expressed genes were further fine-sorted after being classified according to their FunCat annotations.

2.3. Fungal Growth and Protein Secretion Analysis. The culture filtrate was collected by centrifugation at 4°C, 8,000 *g* for 10 min. For SDS-PAGE analysis, 10 μl of culture supernatants of *T. reesei* Rut-C30, its *xyr1* deletion strain and *xyr1* recomplementation transformant were subjected to SDS-PAGE. Protein concentration of each culture supernatant was determined using Modified Bradford Protein Assay Kit (Sangon, Shanghai, China) according to the manufacturer's guidelines. The absorbance at 595 nm was measured with a Varioskan Flash microplate reader (Thermo electron, Finland).

For enzymatic activity measurement, the culture supernatants from the parent strain and different transformants were prepared for filter paper assay (FPA) which has been widely used to determine the total cellulase activity secreted by fungi [37]. Filter paper activity was measured according to the absorbance at 540 nm [38] in a Varioskan Flash microplate reader. Xylanase activities of different samples were determined according to Turunen et al. [39] with modifications. The diluted supernatants (50 μl) were incubated with 50 μl of 1% beechwood xylan (Sigma) dissolved in acetate buffer (100 mM, pH 4.8) at 50°C for 10 min. Then 100 μl of DNS was added and incubated at 95°C for 10 min, and the absorbance at 540 nm was measured. One unit of activity was defined as the amount of enzyme required to release 1 μmol of reducing sugars per minute.

For fungal growth determination, the mycelia of different samples were used. And the biomass assays were performed as Zhao et al. reported [40] with some modifications. Briefly, the mycelia contained in 500 μl samples were collected and washed by distilled water twice and transferred to a new 1.5 ml EP tube, suspended by 100 μl distilled water. Then 1 ml of diphenylamine reagent was added into the tube, votex thoroughly, and reaction in 60°C for 1 hour. After centrifugation for 10 min at 10000 *g*, 200 μl supernatant was transferred to the ELISA plate and the absorbance at 595 nm was measured with a Varioskan Flash microplate reader. The biomass was indicated by the amount of DNA (mg) per ml sample. Three independent measurements were performed for all quantification assays and the data are averages of the three independent determinations. Student's two-tailed *t*-test was performed using Excel 2007 (Microsoft, WA).

2.4. Quantitative Real-Time PCR (qRT-PCR). The RNA samples obtained from *T. reesei* Rut-C30 and Δ*xyr1* strain on either cellulose or glucose were reverse transcribed into cDNA using PrimeScript® RT reagent Kit with gDNA Eraser (TaKaRa, Dalian, China) according to the manufacturer's protocol. All real-time PCR were carried out on a Mastercycler (Eppendorf, Hamburg, Germany) with twin.tech real-time PCR plate 96 (Eppendorf) and Masterclear real-time PCR Film (Eppendorf). For the reaction the SYBR® Premix Ex Taq™ (TaKaRa, Dalian, China) was used for 25 μl assays. Primers used were given in Additional File, Table S1. Three replicates were performed per experiment. The amplification protocol consisted of an initial denature step (30 sec at 95°C) followed by 40 cycles of denaturation (5 sec at 95°C), annealing (30 sec at 58°C), and elongation (15 sec at 72°C). The data analysis was done using Realplex software (Eppendorf, Germany). The *gpda* gene measurement was performed for reference calculation. Two independent experiments were performed and the data are averages of the duplicate determinations. For every experiment, two biological replicates were carried out with three technical replicates each. Student's two-tailed *t*-test was performed using Excel 2007 (Microsoft, WA).

2.5. In Silico Analysis of Xyr1 Binding Sites of the 5′-Upstream Region of Genes. Gene models in the *T. reesei* Rut-C30 genome were downloaded from the *T. reesei* genome database at the Joint Gemome Institute website (http://genome.jgi-psf.org/TrireRUTC30_1/TrireRUTC30_1.home.html). 5′ upstream regions (1 kb) for each gene were extracted from the scaffolds. The occurrence of 5′-GGC(A/T)$_3$-3′ motifs and 5′-GGC(A/T)$_4$-3′ motifs in the 5′-upstream regions was determined for both strands.

2.6. Overexpression and Purification of the DNA Binding Domain of Xyr1. The DNA binding domain (residues 55–195) of Xyr1 was expressed by the pGEX system according to the manufacturer's guidelines. The first-strand cDNA was used as a template to amplify the fragment encompassing the ORF of DNA binding domain of Xyr1 using the primers indicated in Table S1. The fragment was then ligated into plasmid pGEX-4T-1 *via BamH*I and *Xho*I double digestion to produce pGEX-4T-Xyr1-Binding and subsequently introduced into *E. coli* BL21 (DE3) for protein production. Purification and verification of the GST-fused proteins were performed according to the methods described previously [41].

2.7. Electrophoretic Mobility Shift Assays (EMSAs). A universal primer (5′-ACTAACTCGCGTACTG-3′) was labeled at 5′-terminal with Cyanine 5 (Cy5) (Sangon, Shanghai, China). Cy5-labeled DNA probes were generated by two steps PCR amplification using the primers as shown in Table S1: first, double-stranded DNA fragments were amplified from the genomic DNA of *T. reesei* Rut-C30 using specific primer pairs with universal primer sequence in their 5′-terminals; second, Cy5-tag was added to the above DNA fragments by PCR reaction using the universal primer labeled with Cy5. The resulting Cy5-labeled probes were recovered by agarose gel electrophoresis. EMSA was performed using a constant amount (10 ng) of labeled DNA probe. The purified protein of Xyr1 binding domain was preincubated with Cy5-labeled probe and then subjected to electrophoresis according to Ren et al. [42]. The gel was visualized using Starion FLA-9000 Scanner (FujiFilm, Japan).

3. Results and Discussion

3.1. Deletion of xyr1 in T. reesei Abolished Lignocellulolytic Enzyme Production. We constructed an *xyr1* deletion strain *(Δxyr1)* by replacing the *xyr1* open reading frame (ORF) with the hygromycin B resistance gene in *T. reesei* Rut-C30. We also constructed an *xyr1* recomplementation strain *(xyr1-rec)* by replacing the hygromycin B resistance gene with a phleomycin resistance gene and the *xyr1* gene in Δ*xyr1* (Table S1; Fig. S1). Rut-C30 and *xyr1-rec* displayed normal profiles of secreted proteins when cultured on cellulose-inducing media, whereas Δ*xyr1* produced significantly less detectable secreted protein under the same conditions, as shown in the polyacrylamide gel electrophoresis profile (Figure 1(a)) and the assay of extracellular protein concentrations (Figure 1(b)).

We measured the cellulase and xylanase activities of the secreted proteins from Rut-C30 and Δ*xyr1* under inducing conditions. The cellulase activity of Δ*xyr1* was almost completely abolished, whereas the secreted proteins of the parent strain Rut-C30 had an FPA (filter paper activity) of 14.4 IU ml^{-1} (Figure 1(c)). Although the culture supernatants of Rut-C30 had a xylanase activity of 529.2 IU ml^{-1}, Δ*xyr1* had nearly no detectable xylanase activity (Figure 1(d)). A similar phenomenon was reported previously in *T. reesei* Qm9414 and its *xyr1* deletion mutant [9]. The role of Xyr1 in regulating cellulase and xylanase gene expression is strain-independent.

During 7 d of cultivation on lignocelluloses, the biomass of Rut-C30 increased from 54 $\mu g\,ml^{-1}$ on the first day to 87 $\mu g\,ml^{-1}$ on the seventh day, whereas the biomass of Δ*xyr1* did not increase detectably (Figure 1(e)). Δ*xyr1* appeared to be unable to produce lignocellulolytic enzymes that hydrolyzed the lignocellulolytic substrates into monosaccharaides, for further utilization as a carbon source for mycelia growth. *T. reesei* Qm9414 and its Δ*xyr1* strain showed similar growth rates on plates containing xylan or cellulose [9]. This discrepancy might be due to the different methods used to measure growth rate. Similar to our findings, XLR-1 was previously recognized as a homolog of Xyr1/xlnR and regulated some hemicellulase gene expression in *N. crassa*. Deletion of *xlr1* resulted in minimal growth on xylan, whereas the parental strain grew well [13].

To exclude the growth difference in measuring protein secretion, we normalized protein secretion by the biomass of Rut-C30 and Δ*xyr1*. Extracellular protein concentration of the parent strain Rut-C30 was also significantly higher than that of Δ*xyr1* (Figure 1(f)). By contrast, no significant difference in growth, secretive protein concentrations, or activities of FPase and xylanase was observed between Rut-C30 and Δ*xyr1* cultured in medium with glucose as the sole carbon source. All values were similar to those of Δ*xyr1* cultured on cellulose-inducing media (Figures 1(b)–1(f)).

3.2. RNA-Seq Data Processing and FunCat Analysis. In preparation for RNA-seq, the *T. reesei* Rut-C30 and Δ*xyr1* strains were precultured in SDB (Sabouraud's dextrose broth) for 48 h and then the mycelium was collected, washed, and transferred into medium containing lignocellulose or glucose for another 15 h, and then samples were prepared for RNA isolation and further sequencing. The 26.8–49.8 M reads generated corresponded to different samples. After sequence quality control and mapping, the number of properly paired reads per sample ranged from 13.96 to 26.56 M (Table S2). The transcription levels of 22 genes, which were expressed differently in the parent strain Rut-C30 and its *xyr1* deletion strain, according to RNA-seq data, were further analyzed by quantitative reverse-transcriptase polymerase chain reaction (Table S3). The results were consistent with the results from transcription profiling.

When cultured on lignocellulosic medium, 467 genes were expressed differentially in the Δ*xyr1* strain compared with the parental strain Rut-C30 (Table S4). Among these genes, 177 were found to be downregulated and the other 290 were upregulated in Δ*xyr1*. Polysaccharide metabolism, transport, cell rescue/defense and virulence, lipid metabolism, interaction with the environment, protein fate, energy, secondary metabolism, biogenesis of cellular components, transcription, amino acid metabolism, the signal transduction mechanism, cell fate, protein synthesis, nucleotide metabolism, and aromate metabolism were the main functional categories (fine-sorted after being classified by FunCats) to which the 467 differentially expressed genes were allocated (Figure 2).

We also examined the expression of genes affected by *xyr1* deletion when strains were cultured on glucose. A total of 281 genes were found to be significantly affected; 186 genes were downregulated in the Δ*xyr1* strain, and 95 genes displayed significantly higher expression levels in

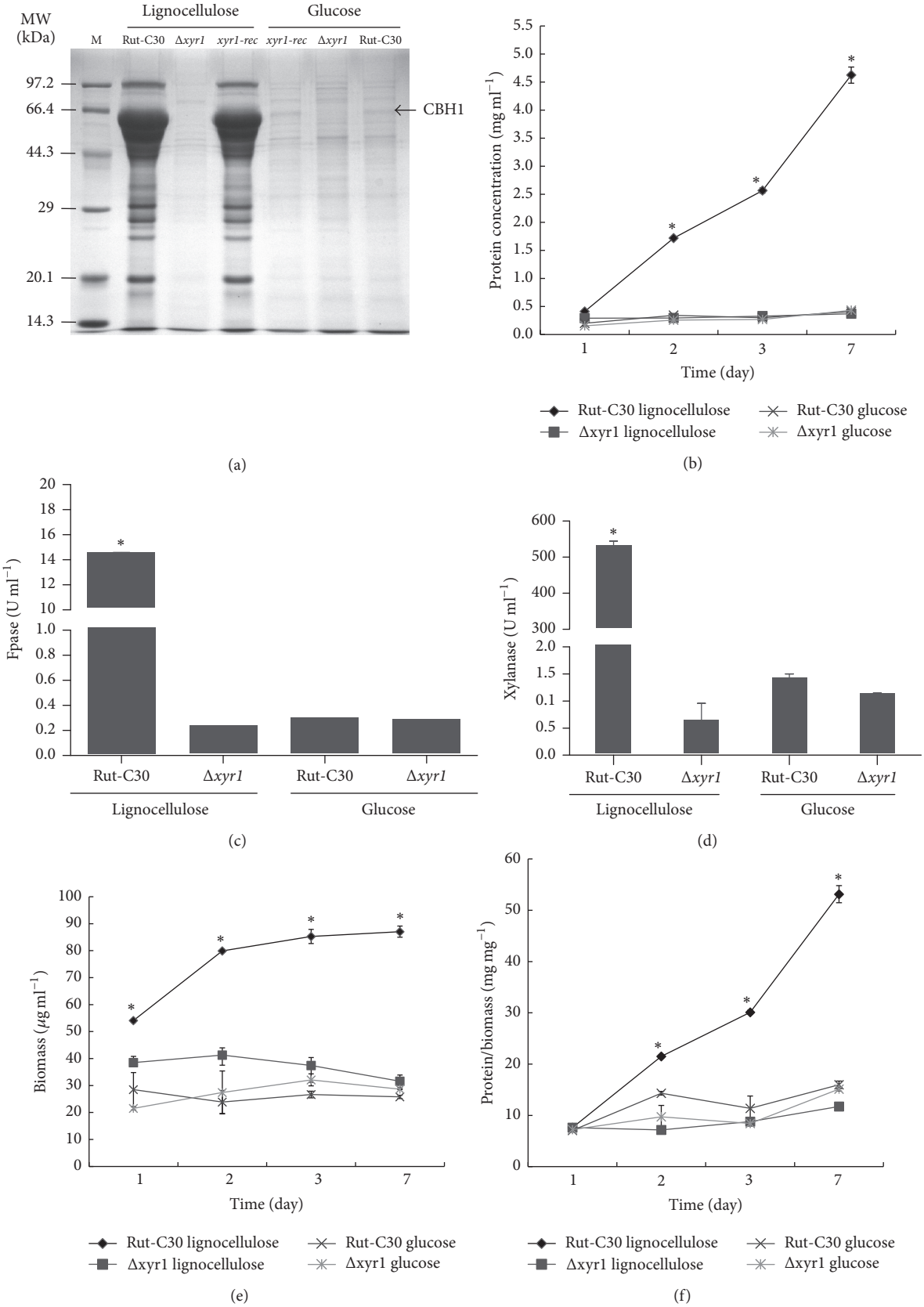


Figure 1: Analysis of fermentation liquor of strains Rut-C30, Δ*xyr1*, and *xyr1-rec* in lignocellulosic and glucose medium. (a) SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of culture supernatants of Rut-C30, Δ*xyr1*, and *xyr1-rec*. For each sample, 10 μl supernatant was loaded. (b) Protein concentration, (c) FPase activity, (d) xylanase activity, (e) biomass, and (f) protein concentration (normalized by biomass of samples) of Rut-C30 and Δ*xyr1* in lignocellulosic and glucose medium. For (b), (e), and (f), samples were collected after cultivation for 1, 2, 3, and 7 d. For SDS-PAGE (a) and characterization of FPase activity (c) and xylanase activity (d), samples collected 7 d after fermentation were used. All values presented in (b)–(f) are means of three independent measurements; error bars indicate standard deviations. $^{*}p < 0.01$.

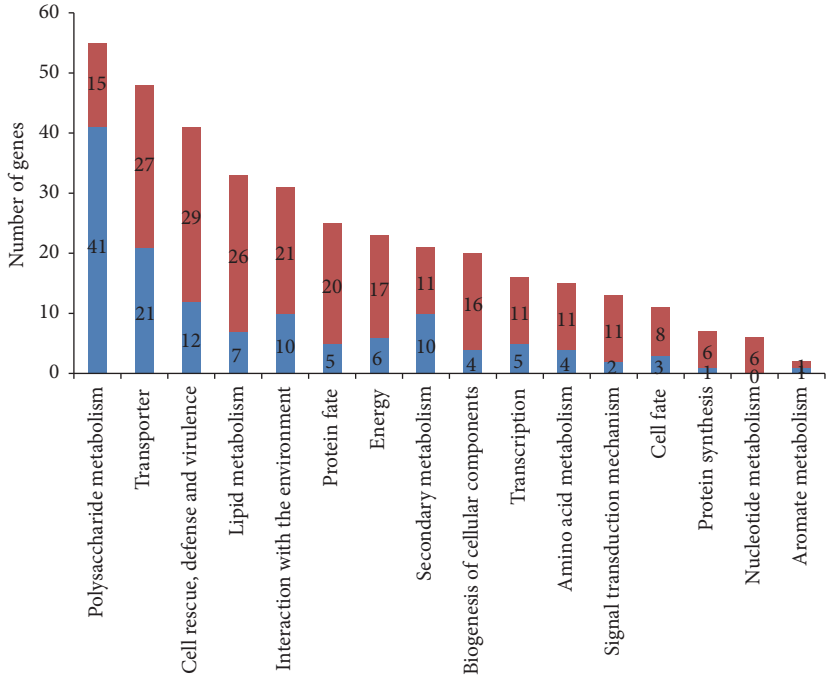


Figure 2: Significant examples of functional categories of genes differentially expressed under the lignocellulose condition. The blue bars indicate genes downregulated, and the red bars indicate genes upregulated, in the *xyr1* deletion strain Δ*xyr1* compared with Rut-C30.

Δ*xyr1* (Table S5). Among the 281 genes, 47 also appeared with the genes transcriptionally affected by *xyr1* deletion under induced conditions (Table S6). The main categories represented among the 186 downregulated genes were those involved in cell rescue, defense, virulence, energy, transport, lipid metabolism, secondary metabolism, energy, amino acid metabolism, interaction with the environment, and polysaccharide metabolism (Figure 3). On the other hand, the main categories represented among the 95 upregulated genes were involved in energy, transport, amino acid metabolism, and protein fate and synthesis (Figure 3; Table S5).

In the parental strain Rut-C30, 338 genes were upregulated and 741 genes were downregulated under induced (lignocellulose) compared with repressed (glucose) conditions (Table S7). Notably, there were many more genes with changed expression levels than the Δ*xyr1* strain under the induced and repressed conditions, which may be because (1) full activation of the function of Xyr1 required special culture conditions and (2) changing the culture conditions affected not only Xyr1, but also other regulators. Analysis of genes existing in the intersection of Rut-C30/Δ*xyr1* under induced conditions and the Rut-C30 (lignocellulose)/Rut-C30 (glucose) would facilitate understanding of the regulation mechanisms of these genes (Table S8).

3.3. Functional Xyr1 Stimulated the Expression of Lignocellulose Degradation-Related Genes. When cultured on lignocellulosic medium, most of the 177 downregulated genes in the Δ*xyr1* strain were involved in carbohydrate metabolism, compared with the parent strain Rut-C30 (Table S4). In addition to previously reported *cbh1*, *cbh2*, *egl1*, and *bgl1* [9], expression of the other functional cellulase genes *egl2*, *egl3*, *egl4* (*cel61a*), *egl5*, *cel61b*, and *bgl2* was significantly impaired in the Δ*xyr1* strain (Table 1).

Additional nonenzymatic cellulose-attacking proteins were also regulated in a coordinated fashion with other cellulose-degrading enzymes (Table 1). These proteins include swollenin SWO1, which is a protein carrying a cellulose-binding domain and an expansin-like domain that disrupts the crystalline cellulose structure [43], and *cip1* and *cip2*, which encode a CE15 glucuronyl esterase [44]. Both *cip1* and *cip2* contain cellulose-binding domains and signal sequences [45].

A previous report indicated that Xyr1 is not involved in the activation of *bgl2* (TrireRUTC30: 127115) expression [9]. Although the results from our study showed that it was significantly downregulated in the *xyr1* deletion mutant, the function of *cip1* remains unknown. To further determine whether Xyr1 could directly regulate *bgl2* and *cip1*, the DNA fragments from promoter regions of *bgl2* and *cip1* were chosen as the candidates in electrophoretic mobility shift assays (EMSAs). Substantial gel shifts were observed for the labeled probes corresponding to the promoter regions of *bgl2* and *cip1* and typical protein concentration-dependent binding trends appeared (Figures 4(a) and 4(b)). The results suggest that Xyr1 activated the expression of *bgl2* and *cip1* directly, by binding to their promoter regions, just as it was previously reported to bind to the promoter regions of cellulase genes [15]. As each of the two promoters has more than one $5'$-GGC(A/T)$_3$-$3'$ motif, various Xyr1-DNA complexes could be formed, which was reflected in different low mobility bands (Figures 4(a) and 4(b)).

Similarly, deletion of *xyr1* affected expression of hemicellulase genes on the inducing medium (Table 1). In *T. reesei*,

Table 1: Changes in CAZome gene expression between *T. reesei* Rut-C30 and Δ*xyr1* under the lignocellulose condition.

	Protein ID	$\log_2$ ratio Rut-C30/Δ*xyr1*	*Q*-value
Cellulases			
Endoglucanase 1 cel7b	5304	9.64332	0
Endoglucanase 2 cel5a	72489	10.4187	0
Endoglucanase 3 cel12a	124438	8.46247	$1.07E-07$
Endoglucanase 5 cel45a	25940	9.22161	$2.18E-08$
Exoglucanase 1 cel7a	125125	10.8384	0
Exoglucanase 2 cel6a	122470	10.6842	0
Cellulase enhancing protein cel61a	139633	10.1565	0
Cellulase enhancing protein cel61b	122518	10.063	0
β-Glucosidase 1 cel3a	136547	5.89937	0
β-Glucosidase 2 cel1a	127115	1.83786	$3.43E-05$
β-Glucosidase cel1b	77989	2.52453	$5.64E-11$
β-Glucosidase cel3d	122639	4.9912	0
β-Glucosidase cel3e	74305	1.24505	0.002047
β-Glucosidase	109567	−1.18554	0.004262
Hemicellulases			
Acetylxylan esterase axe1	139631	10.4907	0
Acetylxylan esterase	88887	8.921338	$3.99E-08$
α-Galactosidase	71638	4.96894	$1.37E-08$
α-N-Arabinofuranosidase abf1	102517	3.19423	$1.50E-07$
α-Xylosidase	134448	6.166894	$2.22E-08$
β-Xylosidase	77521	4.332579	$4.09E-07$
Endo-1,4-β-xylanase 1 xyn1	38418	3.793969	0.000173
Endo-1,4-β-xylanase 2 xyn2	124931	10.536	0
Endo-1,4-β-xylanase 3 xyn3	23616	10.5316	0
Mannan endo-1,6-α-mannosidase	126869	1.63426	0.001508
α-L-Arabinofuranosidase abf2	118070	6.129569	$6.81E-07$
β-Galactosidase	101346	2.59714	$5.30E-11$
β-Mannosidase	67432	5.82358	0
Exo-1,4-β-xylosidase	140746	11.148	0
Mannan endo-1,4-β-mannosidase	122377	7.256397	$4.35E-05$
Xyloglucanase	111943	6.34823	0
Glucuronoxylanase xynC	93498	10.5903	$5.56E-11$
α-Glucuronidase	90302	8.110081	$5.93E-6$
Nonenzymatic cellulose attacking enzymes			
Cip1	121449	10.417	0
Cip2	125575	11.0829	0
Swollenin	104220	8.73037	0
Chitinases			
Chitinase	94061	−1.583	0.009666
Chitinase	142298	−1.60663	$3.00E-05$
Chitinase	7503	−1.74855	0.00568
Endochitinase	33168	−1.71966	0.002243
Endochitinase	33886	−2.0161	$4.09E-07$
Endochitinase	124526	−1.52294	$5.15E-05$
Endochitinase	142123	2.15448	$5.61E-08$
Exochitinase	104242	−2.13026	$4.56E-09$
Glucan 1,3-β-glucosidase			
Glucan endo-1,3-β-glucosidase	125426	−1.71506	0.00101449
Glucan endo-1,3-β-glucosidase	125427	−2.13884	0.00104273
Glucan 1,3-β-glucosidase	25104	−2.27709	$3.96E-08$
Glucan endo-1,3-β-glucosidase	103726	−2.34806	$2.49E-05$

Table 1: Continued.

	Protein ID	$\log_2$ ratio Rut-C30/Δ*xyr1*	*Q*-value
Other glycoside hydrolases			
Probable exopolygalacturonase X	91667	1.79769*e* + 308	9.37*E* − 09
Glucuronoxylanase xynC	90847	10.4524	0
Endo-β-1,6-galactanase	11580	4.68839	4.37*E* − 09
1,3-β-Glucanosyltransferase gel4	103899	3.38036	0
Probable endopolygalacturonase	133383	3.07815	6.93*E* − 06
1,4-α-Glucan-branching enzyme	114146	1.22399	0.00350323
Putative endo-1,3(4)-β-glucanase 2	100837	−1.52054	0.000245344
α-1,2-Mannosidase	122299	−1.59948	3.01*E* − 05
Glucan endo-1,3-α-glucosidase agn1	94877	−3.05265	1.76*E* − 13

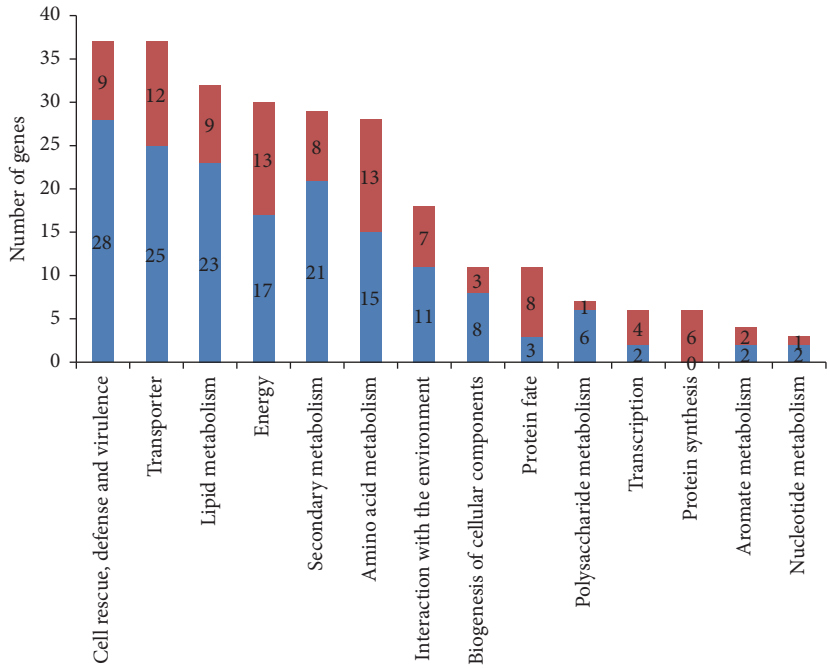


Figure 3: Significant examples of functional categories of genes differentially expressed under the glucose condition. The blue bars indicate genes downregulated, and the red bars indicated genes upregulated, in Δ*xyr1* compared with Rut-C30.

hemicellulase genes *xyn1*, *xyn2*, and *bxl1* were previously reported to be transcriptionally activated by Xyr1 when the inducer was supplied [9]. According to the RNA-seq data in this study, the β-1,4-xylan main chain degrading enzyme gene *xyn3* and the exo-1,4-β-xylosidase gene, which hydrolyzes D-xylose from the nonreducing end of xylan, were found to be downregulated in a coordinated fashion with *xyn1*, *xyn2*, and *bxl1*. Other genes encoding hemicellulases, which hydrolyze the side groups linked to the β-1,4-xylan main chain, were also reported to be activated by Xyr1. These genes included α-arabinofuranosidases (*abf1* and *abf2*), which remove the arabinose side chain, β-galactosidases, which catalyze the hydrolysis of β-galactosidase from the side groups, β-mannosidases, which hydrolyze the β-1,4-manno-oligomers, acetylxylan esterases, which remove an acetyl group from the xylan backbone, α-glucuronidases, which cleave the α-1,2-glycosidic bond of the 4-*O*-methy-D-glucuronic acid side chain of xylans, and xyloglucanase, which releases glucose from xyloglucan [46–48]. Thus, almost all functional cellulase and hemicellulase genes in *T. reesei* were coregulated by the TF Xyr1, and the disruption of *xyr1* severely blocked their expression under inducing conditions.

In addition, we analyzed the transcription of genes potentially involved in the xylose metabolism pathway, which may be crucial for the expression of xylanases [49]. Two genes encoding D-xylose and L-xylulose reductase were also found to be downregulated in Δ*xyr1* on lignocellulose-inducing media. Both D-xylose and L-xylulose reductase may participate in xylose metabolism. This result was consistent with results from previous studies [11, 26]. In *T. reesei*, D-xylose reductase was required to metabolize D-xylose to achieve full induction of xylanase expression [49]. The L-xylulose reductase transcript was reported to be absent in a *T. reesei* Δ*xyr1* strain, which severely affects its growth on D-xylose as the sole carbon source [9]. The expression

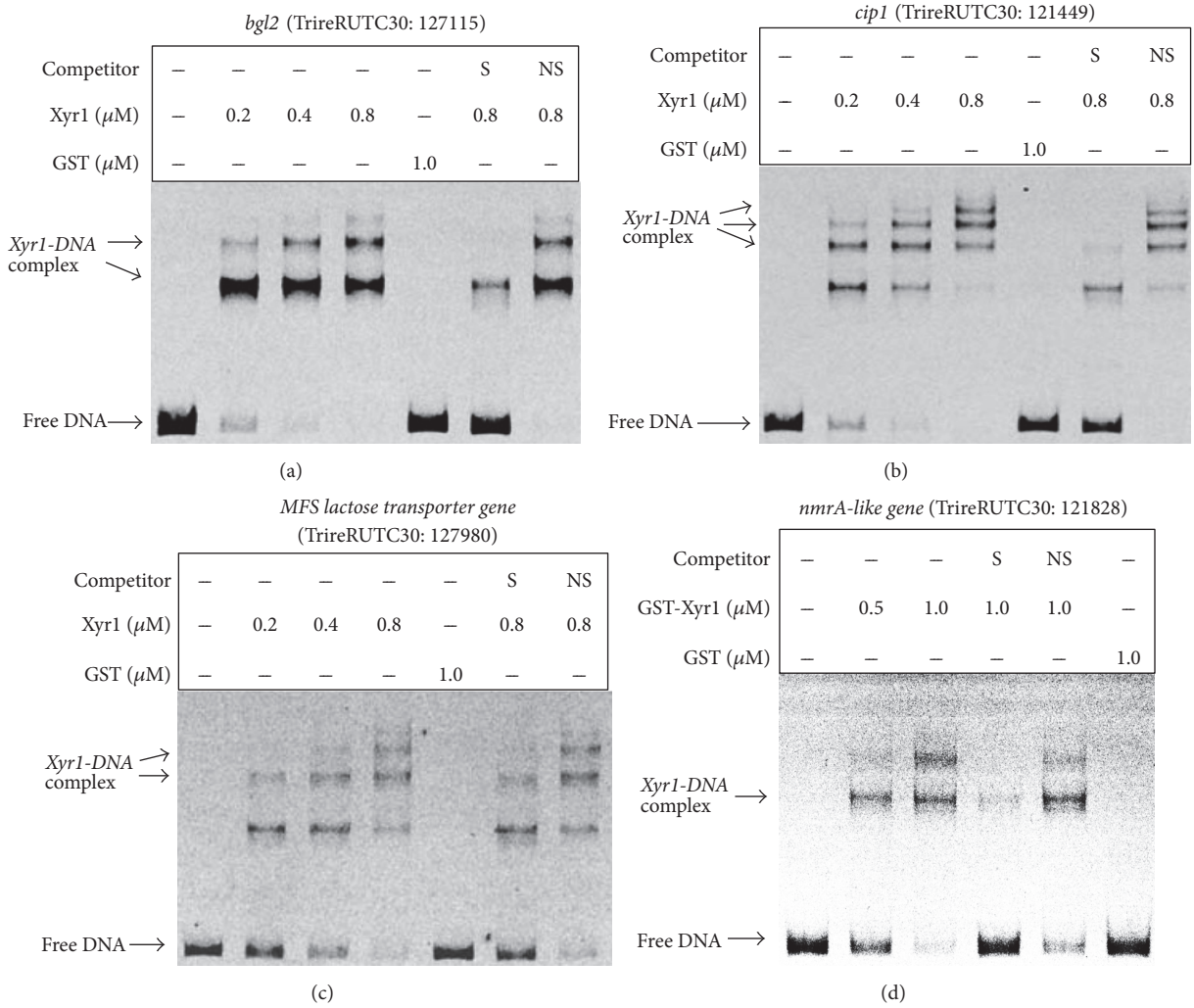


Figure 4: Xyr1 functionally binds to the upstream regions of four genes downregulated in Δ*xyr1* on inducing medium. Binding of Xyr1 to the promoter regions of the intracellular *β*-glucosidase-encoding gene *bgl2* (a), the nonenzymatic cellulose-attacking enzyme-encoding gene *cip1* (b), the putative MFS lactose transporter-encoding gene *lp* (c), and the *nmr*-like gene (d). The amounts of purified Xyr1 (*μ*M) used were as indicated and about 10 ng of Cy5-labeled probes was added to each reaction. The specificity of shifts was verified by adding 100-fold excess unlabeled specific (S) and nonspecific (NS) competitor DNA. Purified GST (1 *μ*M) was used as the negative control to exclude nonspecific binding by GST.

of xylose reductase genes in *A. niger* is also dependent on XlnR, a homolog to Xyr1 [50]. These results indicate that Xyr1 regulates xylan-decomposing and xylose-metabolizing genes, suggesting that these genes are subjected to concerted evolution (Fig. S2).

Although lignin is another main component of lignocellulose, *T. reesei* is not considered to be a potential lignin-degrading fungus [51, 52], and no ligninolytic activity has been reported. However, a recent study demonstrated that some lignin depolymerizing oxidoreductases, including laccase, glyoxal oxidase, peroxidase/catalase, L-ascorbate peroxidase, copper/zinc superoxide dismutase, and several other oxidoreductases, are expressed when *T. reesei* Rut-C30 is grown on natural lignocellulosic biomass [53].

In this study, expression levels of two laccase genes (TrireRUTC30: 104519; TrireRUTC30: 36885) were higher on lignocellulose medium than on glucose-repressing medium in Rut-C30 (Table S7). In the Δ*xyr1* strain, one of the laccase genes (TrireRUTC30: 104519) could also be induced on lignocellulosic medium (Table S9). However, other lignin-degrading genes, such as glyoxal oxidase and Cu/Zn superoxide dismutase (TrireRUTC30: 26844, TrireRUTC30: 112797), were not induced on lignocellulose in either strain (Tables S7 and S9). No gene encoding typical lignin-degrading enzymes, including the six laccase-like multicopper oxidase-encoding genes [51], was downregulated in Δ*xyr1*. Although several lignin degradation-related enzymes could be induced by lignocelluloses, their expression was possibly Xyr1-independent, which inferred a specific regulatory pattern in controlling the expression of the lignin-degrading enzyme genes.

When cultured on glucose, *xyr1* deletion in Rut-C30 resulted in the downregulated expression of six genes

encoding lignocellulolytic enzymes (Table S5). For example, transcriptional levels of *cbh1* and *cbh2* were reduced by 5.4- and 8-fold, respectively, in Δ*xyr1* compared with the transcription levels in Rut-C30 (Table S5). At the protein level, the band of CBH1 could be detected in Rut-C30 and the *xyr1* recomplementation strain *xyr1-rec* cultured on glucose, but not obvious in the *xyr1* deletion strain Δ*xyr1* (Figure 1(a)). Although other lignocellulolytic enzymes did not show statistically significant differences, their expression levels decreased similarly (data not shown). These results suggest that, for the carbon catabolite derepression strain Rut-C30, Xyr1 was partially functional in controlling the expression of lignocellulolytic enzymes, even under repressed conditions.

Among the 1079 differently expressed genes (338 upregulated and 741 downregulated genes) of the parent strain Rut-C30 cultured under the induced (lignocellulose) and repressed (glucose; Table S7) conditions, a total of 69 genes (53 upregulated and 16 downregulated) were categorized as involved in carbohydrate metabolism, according to the FunCat classification (Table S10). Among the 53 upregulated genes, 35 were detected among the downregulated genes in Δ*xyr1* under lignocellulose conditions compared with those of the parental strain Rut-C30 (Tables 1, S8, and S10), which included almost all functional cellulase and hemicellulase genes. These 35 putative Xyr1-regulated target genes may have been activated under the lignocellulose condition to participate in carbohydrate metabolism, while the remaining 18 genes were potentially coregulated by other unidentified regulators. By contrast, only one cellulase gene, TrireRUTC30: 109567 (BGL, GH3 family; FPKM [the number of fragments per kilobase per million fragments mapped]_Δ*xyr1*_lignocellulose: 107.295, FPKM_Δ*xyr1*_glucose: 5.3187), was detected to be expressed differently in Δ*xyr1* under the induced and repressed conditions.

Further in vivo or in vitro assays would be needed to detect whether any of the 53 upregulated genes were directly regulated by Xyr1. For example, *bgl2* (TrireRUTC30: 127115) was not detected as upregulated in Rut-C30 under lignocellulose compared with glucose conditions. Because of these results, *bgl2* was previously regarded as not regulated by Xyr1 [9]. *bgl2* could also bind with Xyr1 in the gel-retardation assays (Figure 4(a)). In this case, *bgl2* might be subjected to the combined reactions of Xyr1 and several other coregulators.

Overall, the above results supported the global role of Xyr1 as an essential regulator in activating lignocellulose degradation-related genes, including cellulose- and hemicellulose-encoding genes and genes participating in xylose metabolism, but not lignin degradation-related genes. Nonenzymatic cellulose-attacking enzymes and two glycoside hydrolase family AA9 protein (Cel61a and Cel61b)-encoding genes harbored coordinated transcription changes with (hemi)cellulase genes, which suggested that they were under the same regulation of Xyr1 and played key roles in lignocellulose degradation. Even under glucose conditions, Xyr1 plays a partial role in regulation of the expression of cellulase genes in *T. reesei* Rut-C30, which implies its direct or indirect interactions with Cre1 in *T. reesei* wild-type strains.

3.4. Transcription Levels of Transporter Genes Affected by xyr1 Deletion in T. reesei Rut-C30. Transporter was the second largest category of genes downregulated in the Δ*xyr1* strain when cultured on lignocellulose, which is comprised of 21 transporter-encoding genes (Table S11; Figure 2), among which 18 genes belonged to the major facilitator superfamily (MFS). The MFS transporters are single-polypeptide secondary carriers capable of transporting only small solutes [54], and they are distributed ubiquitously throughout virtually all currently recognized organismal phyla [55]. The 18 MFS genes were classified into seven families, according to the Transporter Classification Database (http://www.tcdb.org/), which included the sugar porter (SP), fructose H^+ symporter, nitrate/nitrite porter families, the phosphate H^+ symporter, monocarboxylate porter, anion:cation symporter, and vacuolar basic amino acid transporter families (Figure 5(a)).

Among the 18 MFS genes, 11 genes belong to the SP family (Table S11), which is the largest MFS family transporting sugars such as glucose, fructose, mannose, galactose, xylose, maltose, lactose, α-glucoside, and quinate [56]. In addition, the expression levels of 9 of the 11 SP family members were significantly reduced in the parental strain Rut-C30, which was cultured from inducing (lignocellulose) to repressing (glucose) conditions (Tables S8 and S11; Figure 5). Our EMSA experiments, by testing the purified Xyr1 and a putative lactose transporter gene TrireRUTC30: 127980 belonging to these SP family members (Figure 4(c)), confirmed that the transcription of SP transporters might be directly regulated by Xyr1 in *T. reesei* or its homolog in other organisms.

A sophorose-inducible β-diglucoside permease was previously reported to be involved in the induction of the cellulases in *T. reesei* [57], indicating that some sugar transporters have an important role in induction of cellulase expression. The induced expression of sugar transporters might be the previous and necessary step in the induction of cellulose-encoding genes, by promoting the transportation of inducible glucose into the cell. Galazka et al. [58] introduced the cellodextrin transporter from *N. crassa* into *Saccharomyces cerevisiae,* which has led to efficient growth of this yeast on cellodextrins. However, the functions of the Xyr1-dependent transporters have not yet been characterized. Recently, one of these Xyr1-dependent transporters (TrireRUTC30: 109243, Trire2: 3405) was identified as essential for lactose uptake and cellulase induction by lactose [59]. In another study, the deletion strain of Trire2: 3405, *crt1*, showed severe growth defects on Avicel [60]. These results indicated that Xyr1's role, as a major activator of cellulases, was achieved by regulating the transcription of some transporters. Likewise, 11 carbohydrate transporter genes were identified to depend on a functional XLR-1 for increased expression levels when exposed to xylan in *N. crassa* [13], indicating that the impact of Xyr1 or its homolog on the expression of transporter genes is not a phenomenon exclusive to *T. reesei*. As the uptake of cellulose oligosaccharides could play an important role in cellulase

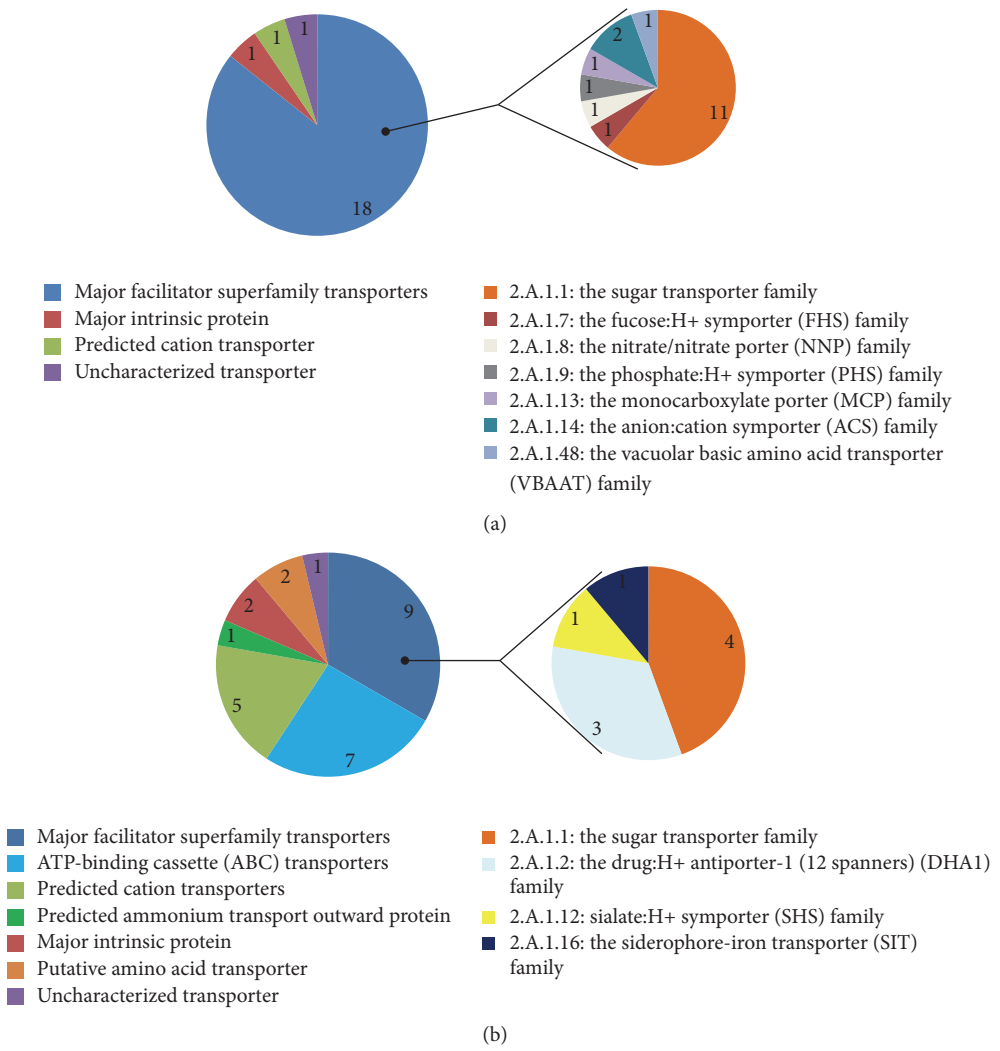


FIGURE 5: Distribution of putative transporters differentially expressed in Δ*xyr1* compared with Rut-C30 when cultured on lignocellulose. Distribution of putative transporters downregulated (a) and upregulated (b) in Δ*xyr1* compared with Rut-C30 when cultured under the lignocellulose culture condition. Details of major facilitator superfamily (MFS) transporter analysis are provided in the right panels of (a) and (b).

formation [57], identification of the characteristics of various sugar transporters is important and would further assist in the identification of the cellulase induction mechanism.

ATP-binding cassettes (ABCs) containing ABC transporters are also reported to be important sugar transporters. Our RNA-seq data demonstrated that no ABC transporter gene was downregulated by deletion of *xyr1* or by culture under the repressing condition (Tables S11; S12). However, seven genes encoding ABC transporters were upregulated after *xyr1* deletion under lignocellulose-inducing conditions (Table S11; Figure 5(a)). These results implied that the ABC transporters are inconsistent for the expression trends of (hemi)cellulase genes that might be coregulated by Xyr1. By contrast, two ABC transporters (Tr 2687 and Tr 58366) were reported downregulated under the cellulose or glucose condition in the Δ*xyr1* mutant strain compared to its parental strain Qm9414 [16]. Another nine MFS transporter genes also exhibited increased expression levels after deletion of *xyr1* under lignocellulose-inducing conditions (Table S11).

Whether the increased expression of these transporters contributed to nutrient uptake under starvation, due to an inability to utilize lignocelluloses in Δ*xyr1*, is not known. Thus, two MFS transporter-encoding genes—the putative MFS glucose transporter *rco3* (TrireRUTC30: 136988) and the predictedMFS siderophore iron transporter *sit* (TrireRUTC30: 115870)—were selected for EMSAs. However, no gel retardation was observed for the probes corresponding to the promoters of these two genes (Fig. S3A and B). These results suggest that the upregulated transporter genes were not regulated directly by Xyr1.

3.5. Several Genes Relevant to Basal Metabolism Were Inclined to Be Repressed by Xyr1. More genes encoding enzymes related to basal metabolism, such as lipid metabolism, protein fate, amino acid metabolism, and nucleotide metabolism,

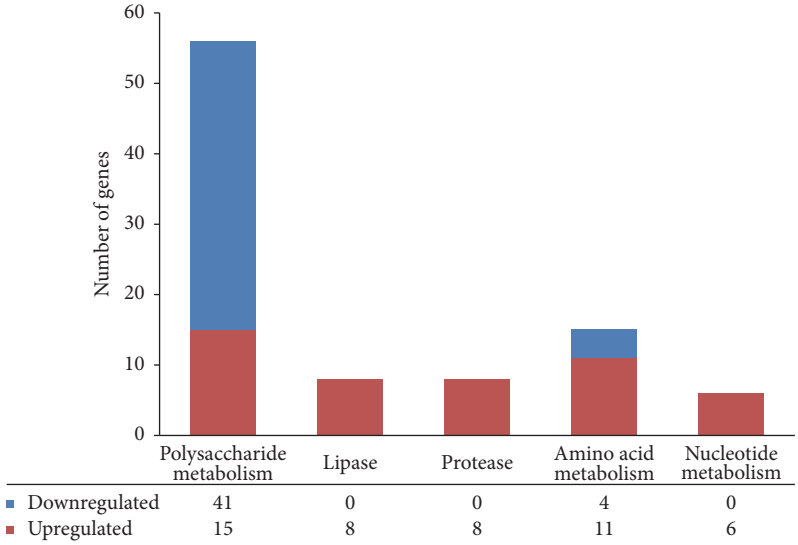


Figure 6: Distribution of differentially expressed genes involved in basal metabolism after *xyr1* deletion under lignocellulosic medium.

were detected to be upregulated in Δ*xyr1* than in Rut-C30 under inducing conditions (Figure 2). For example, *xyr1* deletion resulted in increased expression of 11 genes encoding enzymes participating in amino acid metabolism and 6 genes participating in nucleotide metabolism when cultured on lignocelluloses (Figure 6). In addition, eight lipase genes and eight protease genes were upregulated (five of them were predicted to be secreted using SignalP V4.0 program [http://www.cbs.dtu.dk/services/SignalP/]; Figure 6).

Chitinases participate in fungal cell-wall morphogenesis, including spore germination, hyphal elongation, hyphal branching, and autolysis of mycelium [61–63]. Members of the fungal genus *Trichoderma* are known to produce chitinase, but very little is known about the regulation of expression of these chitinase genes. According to our RNA-seq data, six genes (TrireRUTC30: 94061 (chi18-12), 142298, 7503 (chi18-15), 33168 (chi18-6), 124526 (chi18-5), and 104242) encoding secreted chitinases [64] were found to be upregulated on lignocellulose after *xyr1* deletion (Table 1). By contrast, the transcription levels of these genes did not differ between strains in the medium with glucose as a carbon source. In accordance with this, as Bischof et al. reported, one of the major differences in the *T. reesei* transcriptome between wheat straw and lactose is the wheat straw specific chitinases and mannosidase, which were significantly higher expressed in an *xyr1*-deletion strain [65]. Seidl et al. found that chi18-12 belonged to Group B, chi18-6, and chi18-5 belonged to Group A, while chi18-15 did not belong to any group. In this case, we could predicted that TrireRUTC30: 94061, 33168, and 124526 were Chitinases possibly involved in mycoparasitism [64]. Induction of chitinase genes could be influenced not only by the presence of chitin, but also by carbon catabolite repression, the N source, and starvation [66]. As the strain Δ*xyr1* might be subjected to a shortage of nutrients, as reflected in retarded mycelia growth (Figure 1(e)), these chitinase genes seemed to be induced by carbon starvation.

One of the upregulated putative chitinase genes, *endo T* (TrireRUTC30: 142298), was shown to be not involved in chitin degradation, but it has mannosyl-glycoprotein endo-*N*-acetyl-*β*-D-glucosaminidase activity [67]. It was reported to be responsible for *N*-deglycosylation of proteins expressed and secreted by *T. reesei* [68]. According to our RNA-seq data, the *endo T* in Rut-C30 was significantly upregulated under lignocellulose conditions (FPKM = 42.87) compared with glucose conditions (FPKM = 3.74). However, *endo T* further displayed significantly increased expression levels in the *xyr1* deletion mutant Δ*xyr1* compared with Rut-C30 when cultured on lignocelluloses, with FPKM values of 130.57 and 42.87, respectively. These controversial results suggest that the regulatory mechanism of *endo T* is very complex.

Foreman et al. [45] reported that *endo T* is not coregulated with the expression of cellulase genes in *T. reesei* Qm6a or RL-P37. Whether the increased expression of *endo T* was in response to carbon starvation remains unknown, as does the mechanism of regulation by Xyr1. As a result, EMSAs were performed and strong gel retardation was observed for the probe corresponding to the upstream region of the putative *endo T* with Xyr1 (Figure 7(a)). In *T. reesei* RutC-30, the expression of *endo T* was potentially activated by Xyr1, as well as lignocellulose degradation-related genes, when cultured on lignocellulose. Thus, the resulted single GlcNAc on the heterologous expressed TrBglS benefitted from its enzymatic activity and thermostability, compared with PpBglS [69]. After deletion of *xyr1*, the upregulation of *endo T* might be due to starvation of carbon sources, which might be an effort to deglycosylate the glycan coat of the glycoprotein composed of the cell wall, and contribute to further protease degradation [67].

In filamentous fungi, extracellular proteases are usually produced in response to carbon or N derepression [70]. These phenomena mean that genes related to nonpreferred N source utilization were stimulated in the Δ*xyr1* strain,

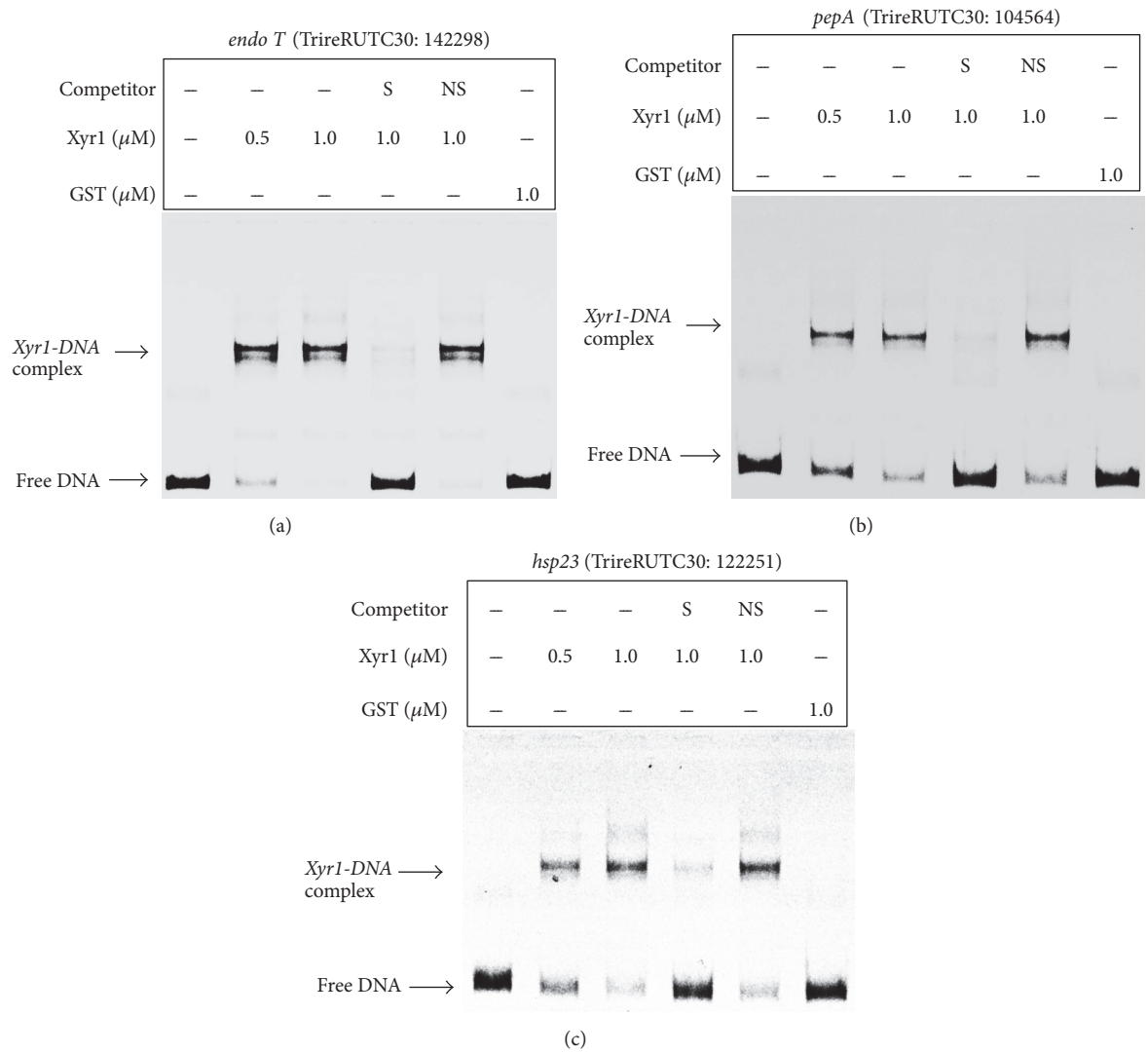


Figure 7: DNA binding of Xyr1 to the upstream regions of three genes upregulated after *xyr1* deletion. Xyr1 functionally binds to the promoter regions from the mannosyl-glycoprotein endo-*N*-acetyl-*β*-D-glucosaminidase, the putative acid aspergillopepsin I, and the possible heat shock protein Hsp23-encoding genes *endo T* (a), *pepA* (b), and *hsp23* (c), which were upregulated in Δ*xyr1* when cultured on the inducing medium. Strong gel shifts were observed after Xyr1 was added to the reactions. The protein-DNA complex increased with protein concentration. The amounts of purified Xyr1 (*μ*M) used were as indicated and about 10 ng Cy5-labeled probe was added to each reaction. The specificity of shifts was verified by adding 100-fold excess unlabeled specific (S) and nonspecific (NS) competitor DNA. Purified GST (1 *μ*M) was used as the negative control to exclude nonspecific binding by GST.

possibly due to the inability to utilize lignocelluloses. Among the five upregulated secreted proteases, the *pepA* (TrireRUTC30: 104564)-encoding putative acid aspergillopepsin I, but not the *tryp* (TrireRUTC30: 94189)-encoding putative trypsin with an optimal operating pH of approximately 7.5–8.5, was a putative Xyr1 target in Rut-C30 (Figures 7(b) and S3C). These results imply that Xyr1 also tends to repress the acid peptidases directly to reduce the degradation of cellulases in medium with a pH value of about 5. In addition, an *α*-1,2-mannosidase gene *mds1* (TrireRUTC30: 122299), which participated in *N*-glycosylation modification of glycoprotein, showed increased expression levels in Δ*xyr1* (FPKM values ranged from 31.51 in Rut-C30 to 95.51 in Δ*xyr1*). The *α*-1,2-mannosidase was reported to readily convert $Man_8GlcNAc_2$, or a mixture of $Man_{(6-9)}GlcNAc_2$ oligosaccharides, to the respective Man_5 structures [71], and was suggested to be localized in the Golgi apparatus [72]. No interaction between Xyr1 and the promoter region of this *α*-1,2-mannosidase gene *mds1* was detected in EMSAs (Fig. S3D). In this case, the upregulated *α*-1,2-mannosidase expression level in Δ*xyr1* might be attributed largely to its carbon starvation status.

3.6. Transcription of Genes Related to Energy Metabolism. In our study, three genes predicted to be alcohol dehydrogenase genes (TrireRUTC30: 128036, TrireRUTC30: 26479, and TrireRUTC30: 133809) were also upregulated when *xyr1* was deleted in the lignocellulose condition (Table S4). In

the glucose condition, however, their expression exhibited no significant difference between Rut-C30 and Δ*xyr1* (Table S5). In *Aspergillus nidulans*, an alcohol dehydrogenase (ADHII) was previously shown to be induced by carbon starvation stress [73]. We speculated that the transcription of these alcohol dehydrogenase genes could be sensitive to carbon starvation, but with no relationship to the existence of Xyr1. In agreement with this speculation, the expression of TrireRUTC30: 133809 was upregulated significantly in Rut-C30, as well as the *xyr1* deletion strain Δ*xyr1*, under induced conditions compared with repressed conditions (Tables S7, S9).

The expression of TrireRUTC30: 26479 was also stimulated in Δ*xyr1* when transferred from the glucose to the lignocellulose condition (Table S9). In addition, TrireRUTC30: 87029, the *Podospora anserine* Pa*ATG1* ortholog, was upregulated in Δ*xyr1* compared with Rut-C30 under a lignocellulose condition (Table S4). TrireRUTC30: 87029 is a serine-threonine kinase composed of an N-terminal kinase domain and a C-terminal domain with unknown function. The Pa*ATG1* mutant displayed developmental defects characteristic of abrogated autophagy in *P. anserine* [74]. Autophagy is a process of protein and organelle degradation by the vacuole (lysosome). This process is conserved in organisms, and functions as a cell survival mechanism during nutrient starvation. Therefore, elevated *ATG1* gene expression probably contributed to accommodating the Δ*xyr1* strain during carbon starvation, reflecting an *xyr1* deletion that totally eliminated the capability to utilize lignocelluloses as a carbon source.

The glycolysis and citric acid cycles are among the most important energy-releasing pathways in glucose metabolism. Comparing the gene expression levels between Rut-C30 and Δ*xyr1* under a glucose condition, deletion of *xyr1* led to significant variation in a succinyl-CoA ligase gene (TrireRUTC30: 135123) and a glycerol-3-phosphate dehydrogenase gene (TrireRUTC30: 116453; Table S5). The AD-lactate dehydrogenase (TrireRUTC30: 73249) and NAD-dependent malic enzyme (TrireRUTC30: 69465) genes (Table S5) were also influenced by *xyr1* deletion, and these two enzymes were involved in yielding the important carbohydrate metabolic intermediate pyruvate. The binding sites of Xyr1 were detected in the promoter regions of the genes mentioned above in *xyr1* deletion mutants. However, more investigation is required to confirm whether Xyr1 also participates in the glucose metabolism and energy-releasing pathways.

3.7. Transcription Changes of Putative Heat Shock Proteins. Heat shock proteins are often induced when organisms respond to extreme temperature and other stresses, such as starvation [75, 76]. Based on transcription profiling of Rut-C30 and Δ*xyr1* under induced and repressed conditions, we attempted to explain the roles that heat shock proteins play in lignocellulose degradation, as well as their transcriptional regulation mechanism.

The transformation of *hsp23* (small heat shock protein-encoding gene of *Trichoderma virens*) into *Trichoderma harzianum* was reported to confer thermotolerance and result in higher biomass accumulation under thermal stress [77]. Under lignocellulose-inducing conditions, a gene homologous to *hsp23*, TrireRUTC30: 122251, was upregulated in Δ*xyr1* relative to its parent strain (Table S4). These results illustrate the process of lignocellulose degradation by *T. reesei* Rut-C30, in which the heat shock protein-encoding genes harbored expression profiles similar to profiles of the lignocellulose degradation-related genes.

Due to the putative Xyr1-binding consensuses in the upstream region of this *hsp23* homolog, EMSAs were performed. The results indicated that Xyr1 could bind to the probe corresponding to the upstream region of this *hsp23* homolog (Figure 7(c)). In this case, we speculated that, under induced conditions, Xyr1 was inclined to repress the expression of heat shock proteins to maintain the balance of the parent strain Rut-C30 and to ensure maximum lignocellulose degradation. In agreement with these results, even under the glucose condition, deletion of *xyr1* caused upregulation of a heat shock protein Hsp78-encoding gene (TrireRUTC30: 73724; Table S5). However, another heat shock protein Hsp70-coding gene, TrireRUTC30: 25176, was downregulated in Δ*xyr1*, suggesting that the regulation of heat shock proteins is much more complex in Rut-C30 under the glucose condition. Two putative heat shock protein-encoding genes, Hsp70 (TrireRUTC30: 97499) and DnaJ (Hsp40, TrireRUTC30: 137482), were downregulated in Rut-C30 under the lignocellulose condition compared with the glucose condition (Table S7).

3.8. Expression Profiles of Transcription Factors. With the exception of genes participating in lignocellulose degradation and transporters, the expression profiles of characterized TF genes were analyzed. We found that the Ace3 (TrireRUTC30: 98455)-encoding gene (FPKM = 137.3–68.3) and the gene encoding the NmrA-like family domain-containing protein (TrireRUTC30: 121828; FPKM = 27.84–5.88) were also downregulated in Δ*xyr1* compared with the parent strain under a lignocellulose-inducing condition (Table S4).

Ace3 is a recently identified transcription activator of cellulase genes [78]. In Δ*xyr1*, its transcription level decreased by about 50%. Häkkinen et al. [78] reported that the expression level of *xyr1* was lower in the *ace3* deletion strain than in the parent strain Qm9414. In this case, Xyr1 and Ace3 might be cross-regulated, which requires further study by protein-DNA and protein-protein interaction assays. In *T. reesei*, Kap8 was identified to be essential for the nuclear transport of the Xyr1, and deletion of *kap8* completely abolished the transcription of 42 CAzymes (including all the cellulases and hemicellulases), which resemble the phenomenon of a *xyr1* loss-of-function mutant [79]. However, the transcription of *xyr1* itself is not affected by *kap8* deletion, implying that its transcription is hardly dependent on Xyr1-autoregulation. On the contrary, Ace3 exhibited significant reduced expression level after *kap8* deletion under the induced condition, indicated that its expression is partially Xyr1-dependent [79].

NmrA is a negative transcriptional regulator involved in N metabolite repression. In *A. nidulans*, deletion of *nmrA* resulted in partial derepression of activities, with utilization of nonpreferred N sources in the case of N repression [80]. This phenomenon was also observed in other filamentous

Table 2: The transcription of several identified TFs in Rut-C30 under the induced compared with repressed condition.

Encoding genes of TFs	Protein ID TrireRUTC30	FPKM (glucose)	FPKM (lignocellulose)	Significant
xyr1	98788	43.73	42.93	No
cre1-1	23706	498.07	802.49	No
ace1	122363	337.92	200.20	No
ace2	32395	23.7	21.07	No
ace3	98455	81.56	137.33	No
bglR	91236	225.27	128.07	No
hap2	93466	39.57	21.40	No
hap3	24298	71.16	75.36	No
clr-1	68701	11.14	9.83	No
clr-2	76250	3.21	2.80	No
nmrA	121828	21.5	27.85	No
areA	140814	52.95	36.58	No

fungi, such as *N. crassa* [81]. We speculated that the decreased expression of this *nmrA*-like gene might facilitate the utilization of an N source in response to carbon starvation in the *xyr1*-deleted strain or be directly regulated by Xyr1. EMSA was employed to detect the potential interaction between Xyr1 and this *nmrA*-like gene, and a strong gel shift of the probe was observed when Xyr1 was added to the reaction (Figure 4(d)). The results suggested that, under a lignocellulose-inducing condition, Xyr1 activates the expression of *nmrA* to repress utilization of nonpreferred N sources, possibly through repression of the activity of the *A. nidulans* AreA [82] homolog in *T. reesei* Rut-C30, and ensure degradation of lignocellulose. However, the precise characteristic of this gene product remains to be determined.

Similar to the effect of *xyr1* deletion on the expression of the *nmrA*-like gene when cultured on lignocellulose, Δ*xyr1* showed a significantly lower level of the *nmrA*-like gene when both strains were cultured on glucose. Obviously, decreased expression of this *nmrA*-like gene was not a response to carbon starvation. As described in the gel-retardation assay, the *nmr*-like gene was a putative target of Xyr1 (Figure 4(d)). The results further imply that the *nmrA*-like gene was a putative target downstream of Xyr1 and repressed by Xyr1, regardless of the culture conditions.

The transcription levels of genes encoding transcription regulators for cellulase or hemicellulase genes were also compared in Rut-C30, under induced and repressed conditions. These genes included *cre1-1*, *ace1*, *ace2*, *ace3*, *bglR*, *hap2*, *hap3*, *clr-1*, and *clr-2*. However, none of these genes showed a significant response (Table 2). These results imply that all of these genes had constitutive expression and were not induced when transferred to lignocellulose-induced conditions. The two transcription regulators involved in N metabolism, *nmrA*-like and *areA*, were also analyzed and the same expression mode was detected (Table 2), suggesting that complex posttranslational regulation, such as phosphorylation on Cre1 [83], acted on these transcription regulators when *T. reesei* Rut-C30 was transferred from repressed to induced conditions. Considering the lack of a significant difference in *xyr1* transcription levels of Rut-C30 under induced (FPKM = 42.93) and repressed (FPKM = 43.73) conditions, these results imply that the function of Xyr1 was severely repressed under repressive culture conditions.

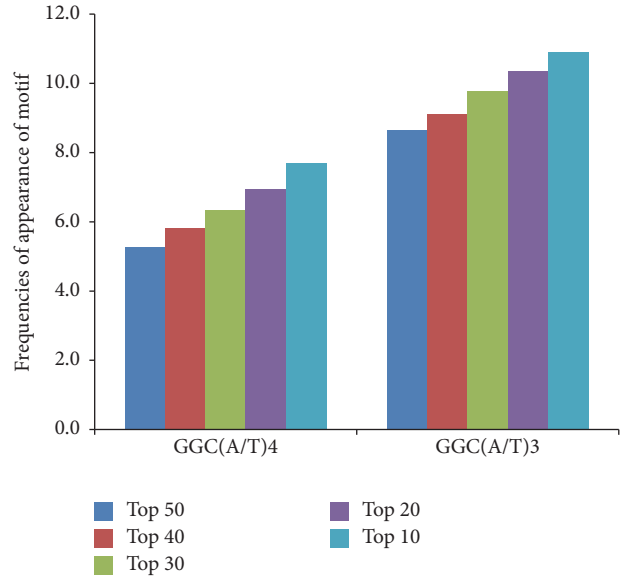


Figure 8: Frequencies of appearance of 5′-GGC(A/T)$_3$-3′ and 5′-GGC(A/T)$_4$-3′ motifs in the upstream regions of Xyr1-upregulated genes.

3.9. Putative Regulation Mechanism of Xyr1 in T. reesei. As the 5′-GGC(A/T)$_3$-3′ and 5′-GGC(A/T)$_4$-3′ motifs play important roles as functional Xyr1-binding sites in *T. reesei* [15], we searched for these motifs in the 1-kb 5′-upstream regions of all genes. We identified an obvious tendency of increasing occurrence of Xyr1 binding sites in the upstream regions of Xyr1-upregulated genes under the lignocellulolytic condition (Table S7). An average of 2.3 5′-GGC(A/T)$_4$-3′ or 5.4 5′-GGC(A/T)$_3$-3′ motifs was detected in the 1-kb 5′-upstream regions of 177 genes. An average of 5.3 5′-GGC(A/T)$_4$-3′ motifs was detected in the top 50 Xyr1-upregulated genes, and the occurrence increased to 7.7 in the top 10 Xyr1-upregulated genes (Figure 8). Similarly, the

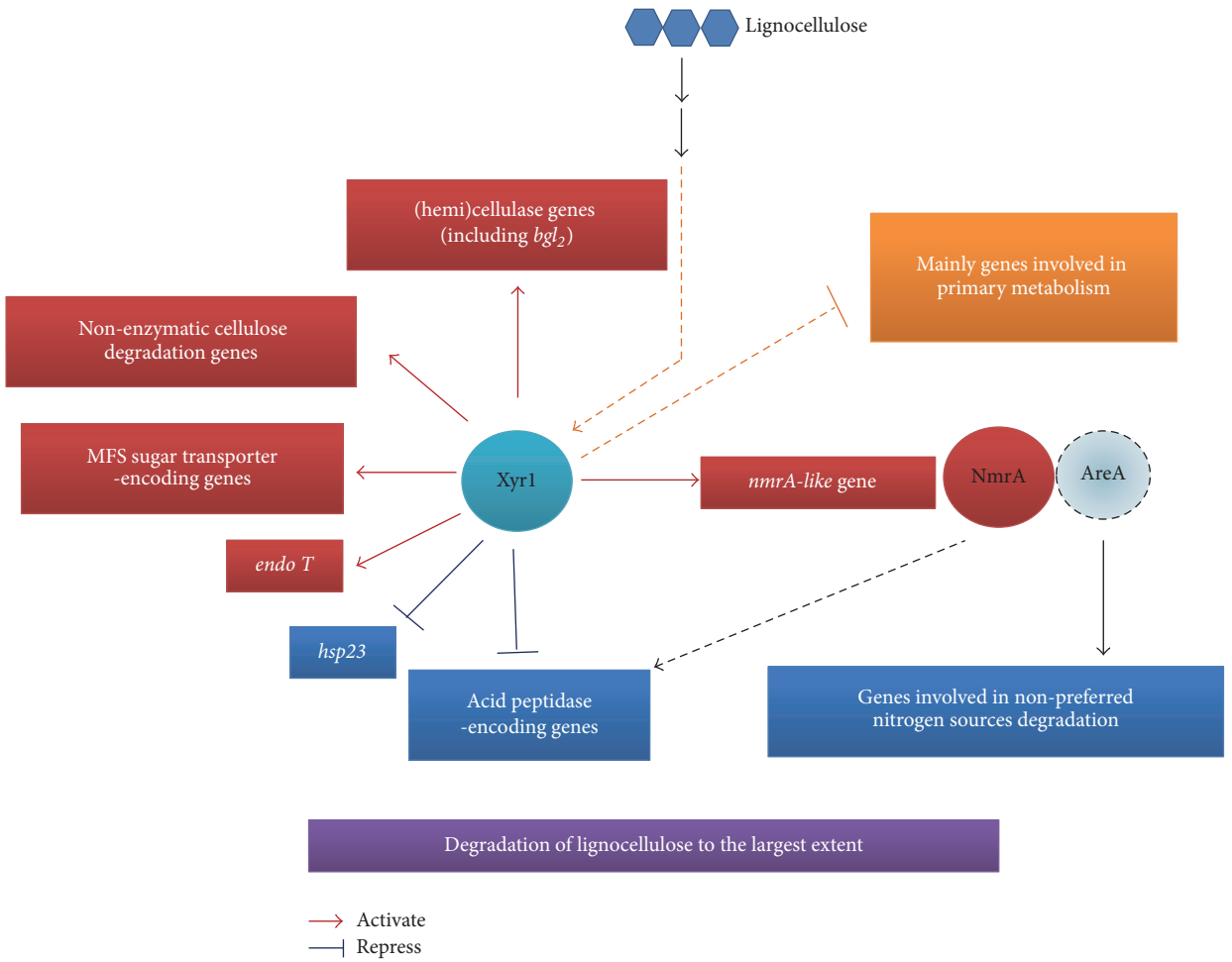


FIGURE 9: The predicted model of Xyr1-mediated gene regulation in *T. reesei* Rut-C30 under lignocellulose-inducing conditions. Xyr1 plays pleiotropic regulatory roles in the process of lignocellulose degradation and may be involved in primary metabolism. It is also implicated in interaction with the nonpreferred N source degradation repressor NmrA. → positive transcriptional control; ⊣ negative transcriptional control; ⇢ indirect regulation.

$5'$-GGC(A/T)$_3$-$3'$ motifs displayed an increasing frequency of appearance, from 8.6 to 10.9, corresponding to the top 50 and top 10 Xyr1-upregulated genes, respectively.

In a previous study [15], the frequency of appearance of the $5'$-GGC(A/T)$_3$-$3'$ motifs in the 16 Xyr1-regulated genes was compared with those of all other annotated and predicted ORFs in the *T. reesei* genome database. The $5'$-GGC(A/T)$_3$-$3'$ motifs were present at a higher frequency in the Xyr1-upregulated genes. This finding, combined with ours, suggests the involvement of motifs in Xyr1-mediated gene expression. However, Xyr1-downregulated genes, such as *pepA* and *hsp23*, cultured on lignocellulose showed no such frequency of motif appearance in their promoter regions (data not shown). In this case, the mechanism of Xyr1 in fine-tuning the expression of these genes requires further study.

4. Conclusions

In *T. reesei*, all ever-identified (hemi)cellulase genes, the intracellular *bgl2*, genes encoding nonenzymatic cellulose-attacking enzymes, and some MFS transporter genes (most belonging to the SP family) responsible for transporting glucose, fructose, mannose, galactose, xylose, maltose, lactose, α-glucosides, and quinate, were activated by Xyr1 under the lignocellulose condition. Transcription levels of most of these Xyr1 targets were significantly decreased in the parent strain Rut-C30 cultured under the inducing (lignocellulose) condition compared with the repressing (glucose) condition.

The genes encoding TFs NmrA and Ace3, Endo T (TrireRUTC30: 142298) with mannosyl-glycoprotein endo-N-acetyl-β-D-glucosaminidase activity and the acid peptidase *pepA*, as well as a small heat shock protein-encoding gene *hsp23*, were putative Xyr1 targets. All of these transcriptional regulations might contribute to the production of (hemi)cellulases and ensure the digestion of lignocellulose to the largest extent (Figure 9). On the other hand, the transcription levels of most genes relevant to basal and energy metabolism were potentially affected by the starvation of carbon sources.

Furthermore, deletion of *xyr1* also affected various genes, even when cultured on glucose, indicating that Xyr1 has pleiotropic functions in biological processes. In this study,

lignocellulose-inducing conditions do not induce expression level of *xyr1* but activate its function, possibly through posttranslational modifications. These findings not only improve our understanding of the regulatory roles of Xyr1, including lignocellulose degradation, transport, and interaction with N metabolism, but also provide useful information for further exploration of genes involved in lignocellulose degradation in *T. reesei*.

Abbreviations

SP:	Sugar porter
MFS:	Major facilitator superfamily
EMSA:	Electrophoretic mobility shift assay
TF:	Transcription factor
RNA-seq:	RNA sequencing
FPKM:	The number of fragments per kilobase per million fragments mapped
FPA:	Filter paper activity
ORF:	Open reading frame
ABC:	ATP-binding cassette
qRT-PCR:	Quantitative real-time PCR
SDB:	Sabouraud's dextrose broth
PDA:	Potato-dextrose agar
N:	Nitrogen.

Disclosure

The RNA-seq data discussed in this study have been deposited in the Sequence Read Archive of NCBI (http://www.ncbi.nlm.nih.gov/Traces/sra) with the accession number SRP028306. All the datasets supporting the results and conclusions of this article are included within the article and the additional files.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Liang Ma constructed the Δ*xyr1* and *xyr1-rec* mutant strains, prepared the RNA samples, and drafted the manuscript. Ling Chen performed electrophoretic mobility shift assays and prepared the figures. Lei Zhang analyzed the transcription profiling data and performed the annotations and statistics. Gen Zou carried out real-time PCR analysis. Rui Liu and Yanping Jiang performed the FunCat analysis. Zhihua Zhou designed the study and revised the manuscript. All authors read and approved the final manuscript. Liang Ma and Ling Chen contributed equally to this work.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (31300073 and 31470201) and High-Tech Research and Development Program of China (863:2013AA102806).

References

[1] L. R. Lynd, C. E. Wyman, and T. U. Gerngross, "Biocommodity engineering," *Biotechnology Progress*, vol. 15, no. 5, pp. 777–793, 1999.

[2] M. E. Himmel, S.-Y. Ding, D. K. Johnson et al., "Biomass recalcitrance: engineering plants and enzymes for biofuels production," *Science*, vol. 315, no. 5813, pp. 804–807, 2007.

[3] B. S. Montenecourt and D. E. Eveleigh, "Selective screening methods for the isolation of high yielding cellulase mutants of *Trichoderma reesei*," in *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis*, vol. 181 of *Advances in Chemistry*, chapter 14, pp. 289–301, 1979.

[4] E. T. Reese, "History of the cellulase program at the U.S. army Natick Development Center," *Biotechnology and Bioengineering Symposium*, vol. 6, no. 6, pp. 9–20, 1976.

[5] C. E. Wyman, "What is (and is not) vital to advancing cellulosic ethanol," *Trends in Biotechnology*, vol. 25, no. 4, pp. 153–157, 2007.

[6] L. R. Lynd, M. S. Laser, D. Bransby et al., "How biotech can transform biofuels," *Nature Biotechnology*, vol. 26, no. 2, pp. 169–172, 2008.

[7] C. P. Kubicek, M. Mikus, A. Schuster, M. Schmoll, and B. Seiboth, "Metabolic engineering strategies for the improvement of cellulase production by Hypocrea jecorina," *Biotechnology for Biofuels*, vol. 2, article 19, 2009.

[8] R. Rauscher, E. Würleitner, C. Wacenovsky et al., "Transcriptional regulation of xyn1, encoding xylanase I, in Hypocrea jecorina," *Eukaryotic Cell*, vol. 5, no. 3, pp. 447–456, 2006.

[9] A. R. Stricker, K. Grosstessner-Hain, E. Würleitner, and R. L. Mach, "Xyr1 (Xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*," *Eukaryotic Cell*, vol. 5, no. 12, pp. 2128–2137, 2006.

[10] E. Akel, B. Metz, B. Seiboth, and C. P. Kubicek, "Molecular regulation of arabinan and L-Arabinose metabolism in hypocrea jecorina (Trichoderma reesei)," *Eukaryotic Cell*, vol. 8, no. 12, pp. 1837–1844, 2009.

[11] A. R. Stricker, M. G. Steiger, and R. L. Mach, "Xyr1 receives the lactose induction signal and regulates lactose metabolism in *Hypocrea jecorina*," *FEBS Letters*, vol. 581, no. 21, pp. 3915–3920, 2007.

[12] F. Calero-Nieto, A. Di Pietro, M. I. G. Roncero, and C. Hera, "Role of the transcriptional activator XlnR of Fusarium oxysporum in regulation of xylanase genes and virulence," *Molecular Plant-Microbe Interactions*, vol. 20, no. 8, pp. 977–985, 2007.

[13] J. Sun, C. Tian, S. Diamond, and N. L. Glassa, "Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa," *Eukaryotic Cell*, vol. 11, no. 4, pp. 482–493, 2012.

[14] S. Klaubauf, H. M. Narang, H. Post et al., "Similar is not the same: differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi," *Fungal Genetics and Biology*, vol. 72, pp. 73–81, 2014.

[15] T. Furukawa, Y. Shida, N. Kitagami et al., "Identification of specific binding sites for XYR1, a transcriptional activator of cellulolytic and xylanolytic genes in Trichoderma reesei," *Fungal Genetics and Biology*, vol. 46, no. 8, pp. 564–574, 2009.

[16] L. dos Santos Castro, R. G. de Paula, A. C. Antoniêto, G. F. Persinoti, R. Silva-Rocha, and R. N. Silva, "Understanding the role of the master regulator XYR1 in *Trichoderma reesei* by global transcriptional analysis," *Frontiers in Microbiology*, vol. 7, article 175, 2016.

[17] M. Ilmén, C. Thrane, and M. Penttilä, "The glucose repressor gene cre1 of Trichoderma: isolation and expression of a full-length and a truncated mutant form," *Molecular and General Genetics*, vol. 251, no. 4, pp. 451–460, 1996.

[18] J. A. Serrato, L. A. Palomares, A. Meneses-Acosta, and O. T. Ramírez, "Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures," *Biotechnology and Bioengineering*, vol. 88, no. 2, pp. 176–188, 2004.

[19] A. L. Mäntylä, K. H. Rossi, S. A. Vanhanen, M. E. Penttilä, P. L. Suominen, and K. M. H. Nevalainen, "Electrophoretic karyotyping of wild-type and mutant Trichoderma longibrachiatum (reesei) strains," *Current Genetics*, vol. 21, no. 6, pp. 471–477, 1992.

[20] S. Le Crom, W. Schackwitz, L. Pennacchio et al., "Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei* using massively parallel DNA sequencing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16151–16156, 2009.

[21] T. Portnoy, A. Margeot, V. Seidl-Seiboth et al., "Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in trichoderma reesei strains producing high and low levels of cellulase," *Eukaryotic Cell*, vol. 10, no. 2, pp. 262–271, 2011.

[22] A. S. Sandberg, H. Andersson, B. Hallgren, K. Hasselblad, B. Isaksson, and L. Hultén, "Experimental model for in vivo determination of dietary fibre and its effect on the absorption of nutrients in the small intestine," *British Journal of Nutrition*, vol. 45, no. 2, pp. 283–294, 1981.

[23] F. Woldesenbet, A. P. Virk, N. Gupta, and P. Sharma, "Effect of microwave irradiation on xylanase production from wheat bran and biobleaching of eucalyptus kraft pulp," *Applied Biochemistry and Biotechnology*, vol. 167, no. 1, pp. 100–108, 2012.

[24] H. Nakayashiki, S. Hanada, N. B. Quoc, N. Kadotani, Y. Tosa, and S. Mayama, "RNA silencing as a tool for exploring gene function in ascomycete fungi," *Fungal Genetics and Biology*, vol. 42, no. 4, pp. 275–283, 2005.

[25] L. Ma, J. Zhang, G. Zou, C. Wang, and Z. Zhou, "Improvement of cellulase activity in Trichoderma reesei by heterologous expression of a β-glucosidase gene from Penicillium decumbens," *Enzyme and Microbial Technology*, vol. 49, no. 4, pp. 366–371, 2011.

[26] A. Miettinen-Oinonen and P. Suominen, "Enhanced production of Trichoderma reesei endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric," *Applied and Environmental Microbiology*, vol. 68, no. 8, pp. 3956–3964, 2002.

[27] C. Trapnell, L. Pachter, and S. L. Salzberg, "TopHat: discovering splice junctions with RNA-Seq," *Bioinformatics*, vol. 25, no. 9, pp. 1105–1111, 2009.

[28] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biology*, vol. 10, no. 3, article no. R25, 2009.

[29] C. Trapnell, B. A. Williams, G. Pertea et al., "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nature Biotechnology*, vol. 28, no. 5, pp. 511–515, 2010.

[30] S. Anders and W. Huber, "Differential expression of RNA-Seq data at the gene level—the DESeq package," 2012.

[31] L. Wang, Z. Feng, X. Wang, X. Wang, and X. Zhang, "DEGseq: an R package for identifying differentially expressed genes from RNA-seq data," *Bioinformatics*, vol. 26, no. 1, pp. 136–138, 2010.

[32] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 57, no. 1, pp. 289–300, 1995.

[33] R. A. Ogert, M. K. Lee, W. Ross, A. Buckler-White, M. A. Martin, and M. W. Cho, "N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism," *Journal of Virology*, vol. 75, no. 13, pp. 5998–6006, 2001.

[34] D. Surmeli, O. Ratmann, H.-W. Mewes, and I. V. Tetko, "FunCat functional inference with belief propagation and feature integration," *Computational Biology and Chemistry*, vol. 32, no. 5, pp. 375–377, 2008.

[35] M. C. Walter, T. Rattei, R. Arnold et al., "PEDANT covers all complete RefSeq genomes," *Nucleic Acids Research*, vol. 37, no. 1, pp. D408–D411, 2009.

[36] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[37] T. M. Wood and K. M. Bhat, "Methods for measuring cellulase activities," *Methods in Enzymology*, vol. 160, pp. 87–112, 1988.

[38] Z. Xiao, R. Storms, and A. Tsang, "Microplate-based filter paper assay to measure total cellulase activity," *Biotechnology and Bioengineering*, vol. 88, no. 7, pp. 832–837, 2004.

[39] O. Turunen, M. Vuorio, F. Fenel, and M. Leisola, "Engineering of multiple arginines into the Ser/Thr surface of Trichoderma reesei endo-1,4-β-xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH," *Protein Engineering*, vol. 15, no. 2, pp. 141–145, 2002.

[40] Y. Zhao, S. Xiang, X. Dai, and K. Yang, "A simplified diphenylamine colorimetric method for growth quantification," *Applied Microbiology and Biotechnology*, vol. 97, no. 11, pp. 5069–5077, 2013.

[41] L. Chen, G. Zou, L. Zhang et al., "The distinctive regulatory roles of PrtT in the cell metabolism of *Penicillium oxalicum*," *Fungal Genetics and Biology*, vol. 63, pp. 42–54, 2014.

[42] C. Ren, Y. Gu, Y. Wu et al., "Pleiotropic functions of catabolite control protein CcpA in Butanol-producing Clostridium acetobutylicum," *BMC Genomics*, vol. 13, no. 1, article no. 349, 2012.

[43] M. Saloheimo, M. Paloheimo, S. Hakola et al., "Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials," *European Journal of Biochemistry*, vol. 269, no. 17, pp. 4202–4211, 2002.

[44] X.-L. Li, S. Špániková, R. P. de Vries, and P. Biely, "Identification of genes encoding microbial glucuronoyl esterases," *FEBS Letters*, vol. 581, no. 21, pp. 4029–4035, 2007.

[45] P. K. Foreman, D. Brown, L. Dankmeyer et al., "Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei," *Journal of Biological Chemistry*, vol. 278, no. 34, pp. 31988–31997, 2003.

[46] Y. Noguchi, M. Sano, K. Kanamaru et al., "Genes regulated by AoXlnR, the xylanolytic and cellulolytic transcriptional regulator, in *Aspergillus oryzae*," *Applied Microbiology and Biotechnology*, vol. 85, no. 1, pp. 141–154, 2009.

[47] D. Shallom and Y. Shoham, "Microbial hemicellulases," *Current Opinion in Microbiology*, vol. 6, no. 3, pp. 219–228, 2003.

[48] S. G. Grishutin, A. V. Gusakov, A. V. Markov, B. B. Ustinov, M. V. Semenova, and A. P. Sinitsyn, "Specific xyloglucanases as a new class of polysaccharide-degrading enzymes," *Biochimica et Biophysica Acta*, vol. 1674, no. 3, pp. 268–281, 2004.

[49] A. R. Mach-Aigner, M. E. Pucher, and R. L. Mach, "D-xylose as a repressor or inducer of xylanase expression in Hypocrea jecorina (*Trichoderma reesei*)," *Applied and Environmental Microbiology*, vol. 76, no. 6, pp. 1770–1776, 2010.

[50] A. A. Hasper, J. Visser, and L. H. de Graaff, "The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression," *Molecular Microbiology*, vol. 36, no. 1, pp. 193–200, 2000.

[51] A. Levasseur, M. Saloheimo, D. Navarro et al., "Exploring laccase-like multicopper oxidase genes from the ascomycete Trichoderma reesei: a functional, phylogenetic and evolutionary study," *BMC Biochemistry*, vol. 11, , article no. 32, 2010.

[52] T. Kirk, "Degradation of lignin," in *Microbial Degradation of Organic Compounds*, Marcel Dekker, New York, NY, USA, 1984.

[53] S. S. Adav, L. T. Chao, and S. K. Sze, "Quantitative secretomic analysis of *Trichoderma reesei* strains reveals enzymatic composition for lignocellulosic biomass degradation," *Molecular and Cellular Proteomics*, vol. 11, no. 7, 2012.

[54] S. S. Pao, I. T. Paulsen, and M. H. Saier Jr., "Major facilitator superfamily," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 1, pp. 1–34, 1998.

[55] V. S. Reddy, M. A. Shlykov, R. Castillo, E. I. Sun, and M. H. Saier Jr., "The major facilitator superfamily (MFS) revisited," *FEBS Journal*, vol. 279, no. 11, pp. 2022–2035, 2012.

[56] M. H. Saier Jr., "Families of transmembrane sugar transport proteins," *Molecular Microbiology*, vol. 35, no. 4, pp. 699–710, 2000.

[57] C. P. Kubicek, R. Messner, F. Gruber, M. Mandels, and E. M. Kubicek-Pranz, "Triggering of cellulase biosynthesis by cellulose in Trichoderma reesei. Involvement of a constitutive, sophorose-inducible, glucose-inhibited β- diglucoside permease," *Journal of Biological Chemistry*, vol. 268, no. 26, pp. 19364–19368, 1993.

[58] J. M. Galazka, C. Tian, W. T. Beeson, B. Martinez, N. L. Glass, and J. H. D. Cate, "Cellodextrin transport in yeast for improved biofuel production," *Science*, vol. 330, no. 6000, pp. 84–86, 2010.

[59] C. Ivanova, J. A. Bååth, B. Seiboth, and C. P. Kubicek, "Systems analysis of lactose metabolism in trichoderma reesei identifies a lactose permease that is essential for cellulase induction," *PLOS ONE*, vol. 8, no. 5, Article ID e62631, 2013.

[60] W. Zhang, Y. Kou, J. Xu et al., "Two major facilitator superfamily sugar transporters from *Trichoderma reesei* and their roles in induction of cellulase biosynthesis," *The Journal of Biological Chemistry*, vol. 288, no. 46, pp. 32861–32872, 2013.

[61] G. W. Gooday, W.-Y. Zhu, and R. W. O'Donnell, "What are the roles of chitinases in the growing fungus?" *FEMS Microbiology Letters*, vol. 100, no. 1–3, pp. 387–391, 1992.

[62] N. Takaya, D. Yamazaki, H. Horiuchi, A. Ohta, and M. Takagi, "Cloning and Characterization of a Chitinase-encoding Gene (chiA) from Aspergillus nidulans, Disruption of Which Decreases Germination Frequency and Hyphal Growth," *Bioscience, Biotechnology and Biochemistry*, vol. 62, no. 1, pp. 60–65, 1998.

[63] H. Yamazaki, D. Yamazaki, N. Takaya, M. Takagi, A. Ohta, and H. Horiuchi, "A chitinase gene, chiB, involved in the autolytic process of Aspergillus nidulans," *Current Genetics*, vol. 51, no. 2, pp. 89–98, 2007.

[64] V. Seidl, B. Huemer, B. Seiboth, and C. P. Kubicek, "A complete survey of Trichoderma chitinases reveals three distinct subgroups of family 18 chitinases," *FEBS Journal*, vol. 272, no. 22, pp. 5923–5939, 2005.

[65] R. Bischof, L. Fourtis, A. Limbeck, C. Gamauf, B. Seiboth, and C. P. Kubicek, "Comparative analysis of the *Trichoderma reesei* transcriptome during growth on the cellulase inducing substrates wheat straw and lactose," *Biotechnology for Biofuels*, vol. 6, no. 1, article 127, 2013.

[66] V. Seidl, "Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions," *Fungal Biology Reviews*, vol. 22, no. 1, pp. 36–42, 2008.

[67] I. Stals, B. Samyn, K. Sergeant et al., "Identification of a gene coding for a deglycosylating enzyme in Hypocrea jecorina," *FEMS Microbiology Letters*, vol. 303, no. 1, pp. 9–17, 2010.

[68] I. Stals, S. Karkehabadi, S. Kim et al., "High resolution crystal structure of the Endo-N-Acetyl-β-D-glucosaminidase responsible for the Deglycosylation of Hypocrea jecorina cellulases," *PLoS ONE*, vol. 7, no. 7, Article ID e40854, 2012.

[69] W. Wei, L. Chen, G. Zou et al., "N-glycosylation affects the proper folding, enzymatic characteristics and production of a fungal β-glucosidase," *Biotechnology and Bioengineering*, vol. 110, no. 12, pp. 3075–3084, 2013.

[70] M. E. Katz, S. M. Bernardo, and B. F. Cheetham, "The interaction of induction, repression and starvation in the regulation of extracellular proteases in *Aspergillus nidulans*: Evidence for a role for CreA in the response to carbon starvation," *Current Genetics*, vol. 54, no. 1, pp. 47–55, 2008.

[71] M. Maras, N. Callewaert, K. Piens et al., "Molecular cloning and enzymatic characterization of a Trichoderma reesei 1,2-α-D-mannosidase," *Journal of Biotechnology*, vol. 77, no. 2-3, pp. 255–263, 2000.

[72] F. Van Petegem, H. Contreras, R. Contreras, and J. Van Beeumen, "Trichoderma reesei α-1,2-mannosidase: structural basis for the cleavage of four consecutive mannose residues," *Journal of Molecular Biology*, vol. 312, no. 1, pp. 157–165, 2001.

[73] I. G. Jones, V. Fairhurst, and H. M. Sealy-Lewis, "ADHII in Aspergillus nidulans is induced by carbon starvation stress," *Fungal Genetics and Biology*, vol. 32, no. 1, pp. 33–43, 2001.

[74] B. Pinan-Lucarré, A. Balguerie, and C. Clavé, "Accelerated cell death in Podospora autophagy mutants," *Eukaryotic Cell*, vol. 4, no. 11, pp. 1765–1774, 2005.

[75] M. G. Santoro, "Heat shock factors and the control of the stress response," *Biochemical Pharmacology*, vol. 59, no. 1, pp. 55–63, 2000.

[76] J. J. Cotto and R. I. Morimoto, "Stress-induced activation of the heat-shock response: cell and molecular biology of heat-shock factors," *Biochemical Society Symposium*, vol. 64, pp. 105–118, 1999.

[77] M. Montero-Barrientos, R. E. Cardoza, S. Gutiérrez, E. Monte, and R. Hermosa, "The heterologous overexpression of hsp23, a small heat-shock protein gene from *Trichoderma virens*, confers thermotolerance to *T. harzianum*," *Current Genetics*, vol. 52, no. 1, pp. 45–53, 2007.

[78] M. Häkkinen, M. J. Valkonen, A. Westerholm-Parvinen et al., "Screening of candidate regulators for cellulase and hemicellulase production in Trichoderma reesei and identification of a factor essential for cellulase production," *Biotechnology for Biofuels*, vol. 7, no. 1, article 14, 2014.

[79] S. Ghassemi, A. Lichius, F. Bidard et al., "The β-importin KAP8 (Pse1/Kap121) is required for nuclear import of the cellulase

transcriptional regulator XYR1, asexual sporulation and stress resistance in *Trichoderma reesei*," *Molecular Microbiology*, vol. 96, no. 2, pp. 405–418, 2015.

[80] A. Andrianopoulos, S. Kourambas, J. A. Sharp, M. A. Davis, and M. J. Hynes, "Characterization of the Aspergillus nidulans nmrA gene involved in nitrogen metabolite repression," *Journal of Bacteriology*, vol. 180, no. 7, pp. 1973–1977, 1998.

[81] N. S. Dunn-Coleman, A. B. Tomsett, and R. H. Garrett, "The regulation of nitrate assimilation in *Neurospora crassa*: biochemical analysis of the *nmr*-1 mutants," *MGG Molecular & General Genetics*, vol. 182, no. 2, pp. 234–239, 1981.

[82] M. Macios, M. X. Caddick, P. Weglenski, C. Scazzocchio, and A. Dzikowska, "The GATA factors AREA and AREB together with the co-repressor NMRA, negatively regulate arginine catabolism in Aspergillus nidulans in response to nitrogen and carbon source," *Fungal Genetics and Biology*, vol. 49, no. 3, pp. 189–198, 2012.

[83] A. Cziferszky, R. L. Mach, and C. P. Kubicek, "Phosphorylation positively regulates DNA binding of the carbon catabolite repressor Cre1 of Hypocrea jecorina (Trichoderma reesei)," *Journal of Biological Chemistry*, vol. 277, no. 17, pp. 14688–14694, 2002.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 4783841, 9 pages
http://dx.doi.org/10.1155/2016/4783841

Research Article

Workflow for Genome-Wide Determination of Pre-mRNA Splicing Efficiency from Yeast RNA-seq Data

Martin Převorovský, Martina Hálová, Kateřina Abrhámová, Jiří Libus, and Petr Folk

Department of Cell Biology, Faculty of Science, Charles University, Viničná 7, 128 43 Prague 2, Czech Republic

Correspondence should be addressed to Martin Převorovský; prevorov@natur.cuni.cz

Received 8 September 2016; Accepted 2 November 2016

Academic Editor: Guohua Xiao

Pre-mRNA splicing represents an important regulatory layer of eukaryotic gene expression. In the simple budding yeast *Saccharomyces cerevisiae*, about one-third of all mRNA molecules undergo splicing, and splicing efficiency is tightly regulated, for example, during meiotic differentiation. *S. cerevisiae* features a streamlined, evolutionarily highly conserved splicing machinery and serves as a favourite model for studies of various aspects of splicing. RNA-seq represents a robust, versatile, and affordable technique for transcriptome interrogation, which can also be used to study splicing efficiency. However, convenient bioinformatics tools for the analysis of splicing efficiency from yeast RNA-seq data are lacking. We present a complete workflow for the calculation of genome-wide splicing efficiency in *S. cerevisiae* using strand-specific RNA-seq data. Our pipeline takes sequencing reads in the FASTQ format and provides splicing efficiency values for the 5′ and 3′ splice junctions of each intron. The pipeline is based on up-to-date open-source software tools and requires very limited input from the user. We provide all relevant scripts in a ready-to-use form. We demonstrate the functionality of the workflow using RNA-seq datasets from three spliceosome mutants. The workflow should prove useful for studies of yeast splicing mutants or of regulated splicing, for example, under specific growth conditions.

1. Introduction

In eukaryotes, coding parts of genes, the exons, are interrupted by noncoding parts, the introns. The process through which introns are removed and exons are joined together is called splicing. It occurs via two consecutive transesterification reactions which are catalysed by the spliceosome, a large dynamic ribonucleoprotein complex composed of five snRNP particles (U1, U2, U4/U6, and U5) and other associated protein complexes, like the Nineteen Complex (NTC in yeast; CDC5L in mammals) (reviewed in [1]). Splicing must occur very precisely as even a single nucleotide shift may lead to a frameshift, which could cause many disorders, including cancer [2, 3]. Therefore, regulation of splicing has an important role in gene expression.

Introns are defined by core sequences comprising the 5′ splice site, branch site, and the 3′ splice site. In metazoans, additional sequences are needed for recruiting various *trans*-acting regulatory factors, which modulate the binding of spliceosome subunits and splice site choice and efficiency, deciding on the splicing outcome. This is important especially for alternative splicing (reviewed in [4]).

In contrast to higher eukaryotes, whose genes typically contain multiple short exons alternating with introns up to several kilobases long [5], gene structure in the budding yeast *Saccharomyces cerevisiae* is much simpler. Only five percent of the almost 6000 yeast genes contain introns, usually just one [6]. However, the intron containing genes are very highly expressed. As a result, about one-third of all transcripts are spliced [7]. The consensus sequences of the yeast core spliceosome signals are well defined (GUAUGU for the 5′ splice site, UACUA**A**C for the branch site with the branching A in bold, and AG for the 3′ splice site) [6]. Also, there are only few cases of alternative splicing in budding yeast (reviewed in [8]), and regulation of splicing efficiency plays a more prominent role, for example, during meiosis [9] or under environmental stress [10, 11]. For example, the constitutively transcribed intron containing genes *REC107*, *AMA1*, *SPO22*, and *MER3* are efficiently spliced and processed to form functional mRNAs only during meiosis, when

Table 1: RNA-seq datasets used in this study.

Genotype	ArrayExpress acc. number[a]	ENA acc. number[b]	Read length (nt)	Total reads	Reads with MAPQ ≥ 10	% reads with MAPQ ≥ 10
WT	E-MTAB-5149	ERR1709739	100	27 789 829	25 329 092	91,2%
		ERR1709740	100	22 000 062	20 402 556	92,7%
prp45(1-169)	E-MTAB-5149	ERR1709737	100	27 842 215	25 491 566	91,6%
		ERR1709738	100	25 156 639	23 359 541	92,9%
WT	E-GEOD-44219	SRX233529	100	21 012 048	17 127 536	81,5%
prp4-1	E-GEOD-44219	SRX233535	100	17 142 559	14 457 015	84,3%
WT	E-GEOD-49966	SRR953535	101	35 203 753	7 655 225	21,8%
prp40-1	E-GEOD-49966	SRR953537	101	17 326 529	3 596 304	20,8%

[a]Accession number for the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/).
[b]Accession number for the European Nucleotide Archive (http://www.ebi.ac.uk/ena).

the Mer1 splicing factor is expressed [9, 12–14]. Also, various environmental stresses can lead to differential changes in splicing efficiency in specific groups of genes: amino acid starvation inhibits splicing of the ribosomal protein genes, while ethanol stress has no effect on this group of genes but alters splicing in another group [10, 11].

The yeast spliceosome consists of ~90 proteins, that is, around half of the proteins identified in the human spliceosome. However, nearly all of the yeast spliceosome components have counterparts in human. Therefore, it was suggested that the *S. cerevisiae* spliceosome represents an evolutionarily conserved core of the splicing machinery. Accordingly, many of the human-specific spliceosomal proteins are needed for the regulation of alternative splicing, a feature almost missing in the budding yeast [15]. This, together with the ease of cultivation and genetic manipulation, led to the establishment of the budding yeast as a favourite model for studying the basic mechanisms of pre-mRNA splicing.

To study the influence of genetic perturbations or environmental conditions on splicing, it is important to quantify the splicing efficiency. Splicing efficiency is traditionally calculated as the amount of mRNA divided by the amount of pre-mRNA. The gold standard for mRNA and pre-mRNA quantification is the use of quantitative PCR (RT-qPCR) with primers spanning exon-intron and exon-exon junctions (e.g., [16]). However, this approach is feasible for measuring mRNA and pre-mRNA levels for only a limited number of genes. By contrast, ultrahigh-throughput sequencing of RNA (RNA-seq) allows comprehensive splicing analysis at the genome-wide scale [17–19]. There are multiple paradigms for calculating splicing efficiency from RNA-seq data, which are based on comparing sequencing read counts from intronic and exonic regions or also take into account exon-exon junction reads (transreads). The methods also vary in the length of the window considered (e.g., 25 bp around a splice site versus a whole exon) [17, 18, 20–24]. RNA-seq is a simple, robust, and affordable technique and there is now a wealth of publicly available RNA-seq datasets [25, 26]. Together, this makes RNA-seq a convenient method for genome-wide determination of splicing efficiency, although the bioinformatics analyses involved might be challenging for nonspecialists. Here, we present a complete and documented up-to-date workflow for semiautomatic calculation of genome-wide splicing efficiencies from strand-specific RNA-seq data in *S. cerevisiae*.

2. Materials and Methods

2.1. RNA-seq Datasets. Sequencing reads from strand-specific transcriptome profiling of splicing mutants (*prp45*(1-169), *prp4-1*, and *prp40-1*) and their corresponding wild-type *S. cerevisiae* strains [17, 24] were downloaded from the European Nucleotide Archive (http://www.ebi.ac.uk/ena) in FASTQ format (more information about the various file formats used in this study can be found at https://genome.ucsc.edu/FAQ/FAQformat.html) and experiment metadata were obtained from ArrayExpress (https://www.eabi.ac.uk/arrayexpress/). The relevant database accession numbers are given in Table 1. Importantly, to simplify downstream analyses, only "read 1" FASTQ files were used from paired-end sequencing datasets.

The pipeline requires that strand-specific sequencing read data are used (true for most currently generated RNA-seq datasets). As pervasive antisense transcription has been reported in many eukaryotic organisms, including yeasts [27], strand specificity of sequencing reads helps to separate the contributions of potential overlapping sense/antisense transcripts.

Sequencing read quality and potential contamination by adapters and/or PCR primers was checked with FastQC 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All datasets were found suitable for further analyses. However, if needed, contaminating and low-quality sequences can be filtered and/or trimmed using tools such as Trimmomatic [28].

2.2. Read Mapping. Reads were aligned to *S. cerevisiae* genome (Ensembl R64-1-1) with the fast, splice-aware HISAT2 aligner (version 2.0.4) using *S. cerevisiae* genome index containing transcript structures [29]. Minimum and maximum intron length parameters were set to 20 and 10000 nt, respectively. More details on HISAT2 parameter settings can be found in the "workflow.sh" shell script in the Supplementary Material available online at http://dx.doi.org/10.1155/2016/4783841.

Subsequently, samtools 1.3.1 were used to filter reads by their mapping quality score (MAPQ ≥ 10) to keep only reads that aligned unambiguously to a single locus and to sort and index the resulting BAM files [30]. Read mapping was also assessed visually in the IGV browser 2.3.69 [31].

We also tried aligning reads from the *PRP45*-related datasets with TopHat2 [32], the widely used predecessor of HISAT2. Compared to TopHat2, the HISAT2 alignment step was ~2-fold faster, and the mapping supported identification of ~11% more transreads and calculation of splicing efficiencies for ~4% more splice sites.

2.3. Splice Site Identification and Counting of Transreads. Putative splicing events were detected by regtools 0.2.0 (https://regtools.readthedocs.io/en/latest/). These tools look for presumed transreads (reads spanning exon-exon junction) in BAM files, compile a table of all identified putative splice sites and their characteristics, and provide the counts of transreads spanning these splice sites. Minimum and maximum intron length parameters were set to 20 and 10000 nt, respectively. Detected splice sites were also annotated and classified as known or novel using regtools and *S. cerevisiae* genome annotation (Ensembl R64-1-1) in GTF format. The output from regtools was further processed and coordinates of bases at the very 5′ and 3′ ends of each known intron were extracted (into BED format) by a custom R script (R version 3.3.1, https://www.r-project.org/). See the "workflow.sh" and "junctions.R" scripts in the Supplementary Material for more details.

2.4. Determination of 5′ and 3′ Intron End Coverage. Read coverage of the very first (5′ end) and the very last (3′ end) base of each known intron was determined from all BAM files in parallel using bedtools 2.25.0 (bedtools multicov) [33]. It is critical to set the *-split* parameter to avoid including transreads in the intronic read counts. It is also important to correctly set the *-s*/*-S* parameters according to the sequencing library preparation protocol employed to ensure that only reads mapped to the "sense" strand will be counted (see the "workflow.sh" script in the Supplementary Material).

2.5. Splicing Efficiency Calculation. The method for splicing efficiency calculation was derived from the "3′ splice site ratio" method described in [20]. For each intron, splicing efficiency was determined separately for the 5′ splice site and 3′ splice site using the following formulas:

$$\text{Efficiency } 5' = \frac{\text{transread count}}{5' \text{ intron end first base coverage}},$$
$$\text{Efficiency } 3' = \frac{\text{transread count}}{3' \text{ intron end last base coverage}}. \quad (1)$$

Transreads only arise from spliced transcripts and thus reflect directly the abundance of mRNA molecules in which the particular intron has been spliced out. By definition, transreads must cover at least the very last base of exon X and the very first base of exon X + 1. To match this single-base resolution while counting intronic reads, only those reads covering the very first (for 5′ splice site) and the very last (for 3′ splice site) base of the corresponding intron are counted. This allows direct comparison of pre-mRNA and spliced mRNA molecule levels (an approach most similar to the RT-qPCR gold standard; see below) and separate calculation of splicing efficiencies at the 5′ splice site and 3′ splice site of each intron, reflecting the efficiencies of the two splicing steps.

We note that, for some genes, our method may use only a fraction of all available intronic reads (e.g., for mitochondrial genes with long introns). Importantly, the introns of some genes harbour nested snoRNAs or are predicted to contain potential stable structures, suggesting the presence of so far uncharacterised nested noncoding RNAs. It was shown that such introns can be maintained in the cell after splicing [34], potentially affecting the determination of splicing efficiency when sequencing reads along the whole intron are considered.

Depending on the sequencing library size, strain genotype, or cultivation conditions, many introns typically have very low read coverage, which might produce unreliable splicing efficiency data. Therefore, in the final spreadsheet, we do not report splicing efficiency values for junctions with read counts below an arbitrary threshold of 5 reads for both transread count and intron end read count. Users may change this threshold if required; see the "efficiency.R" script in the Supplementary Material for more details.

2.6. Yeast Strains and Cell Cultivation. Wild-type (BY4741-MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and the *prp45*(1-169) mutant (AVY17-MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 prp45*(1-169)-3-HA::NatMX6) strains were grown in the YPAD medium (2% peptone, 1% yeast extract, 0.01% adenine, and 2% glucose) at 30°C to 1.5×10^7 cells/mL. Two millilitres of each culture was harvested by centrifugation and cell pellets were stored at −80°C.

2.7. RNA Isolation, Reverse Transcription, and RT-qPCR Analysis. Total RNA was isolated with the MasterPure Yeast RNA Isolation Kit (Epicentre) according to the manufacturer's protocol. cDNA was prepared from 2 μg of the total RNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) with random hexamer primers. RT-qPCR was performed using LightCycler 480 II (Roche). Each reaction (total volume of 10 μL) was performed in triplicate and consisted of 5 μL of the MESA GREEN qPCR MasterMix Plus for SYBR Assay, No-ROX (Eurogentec), 4 μL of 100-fold diluted cDNA, and a pair of primers 0.3 mM each (for sequences, see the Supplementary Material). Primer pairs were designed to specifically amplify either the spliced or the unspliced transcripts of the *ECM33*, *ACT1*, *COF1*, *RPL22A*, and *RPL22B* genes. Amplicons from unspliced transcripts covered the 3′ splice junction in *ECM33*, while the 5′ splice junction was covered in the other four genes. Four to six biological replicates were analysed for each gene, and the *TOM22* and *SPT15* genes were used as reference controls. Relative pre-mRNA and mRNA quantities were calculated by the ΔΔCt method [35].

3. Results and Discussion

3.1. Pipeline for Determination of Splicing Efficiency. We have put together a workflow for calculating splicing efficiency of each intron from standard RNA-seq data in *S. cerevisiae* (Figure 1). The pipeline consists mostly of established open-source tools for manipulation and analysis of next-generation sequencing data (FastQC, HISAT2, samtools, regtools, and bedtools) and simple custom scripts (Linux shell and R). All scripts including parameter settings for all tools are available in the Supplementary Material.

Briefly, after input quality control (FastQC), reads are mapped into *S. cerevisiae* reference genome (HISAT2, [29]) and filtered, keeping only uniquely mapped reads (samtools, [30]; MAPQ ≥ 10). Positions of splice junctions (both known and novel) and transread counts for these junctions are extracted (regtools). Read coverage of the 5′ and 3′ intron end base is determined (bedtools, [33]) and splicing efficiency calculated for each intron at both 5′ and 3′ splice junction, corresponding to the efficiency of the first and the second step of splicing, respectively.

The description of folder structure and content, both required before starting the analysis and produced during the analysis, is given in the file "folders_readme" in the Supplementary Material. Users need to supply genome sequence (in FASTA format), annotation (in GTF format), and HISAT2 genome index containing transcript structures (in the "genome" folder). A ready-made *S. cerevisiae* genome index can be downloaded from the HISAT2 website ("genome_tran"; https://ccb.jhu.edu/software/hisat2/). Users further need to supply RNA-seq reads in FASTQ format (in the "FASTQ" folder); the pipeline is designed for single-end, strand-specific sequencing data. Depending on the protocol for sequencing library preparation that was used, users might need to adjust strandness-related parameters for HISAT2 and bedtools (see the Materials and Methods and the "workflow.sh" file in the Supplementary Material) to obtain correct read counts and splicing efficiency values.

The main outputs of the pipeline are

(1) a table (CSV format) of transread counts from all samples for known splice junctions (folder "transreads"; file "splice_junctions_coverage.known.csv"; note that only junctions of expressed genes for which transreads were detected are reported in all analyses);

(2) tables of read coverage from all samples for 5′ and 3′ terminal bases of known introns (folder "introns"; files "introns_known_5ss.bed.counts.csv" and "introns_known_3ss.bed.counts.csv");

(3) tables of read threshold-filtered splicing efficiencies in all samples calculated separately for 5′ and 3′ splice sites of known introns (folder "efficiency"; files "splicing_efficiency_5ss_conf.csv" and "splicing_efficiency_3ss_conf.csv").

Users can also define a set of sample (FASTQ filename) pairs to be compared (e.g., mutant versus wild type; see the "efficiency.R" script in the Supplementary Material). The pipeline then produces a table of "relative" splicing efficiencies (folder "efficiency"; files "relative_splicing_efficiency_5ss_conf.csv" and "relative_splicing_efficiency_3ss_conf.csv") and scatterplot images (folder "images"; PDF format; see Figure 2 for an example).

Multiple approaches have been used in the past to determine splicing efficiency from yeast RNA-seq data, for example, [17, 18, 21, 24]. However, the bioinformatics tools used are often outdated now in terms of speed of processing and their abilities to work with split reads (transreads). Also, the complete workflow, including all relevant scripts, is usually not provided. By contrast, the pipeline presented in this study is based upon the latest tools with advanced capabilities for fast and accurate split read processing, suitable for studies of splicing efficiency [29, 30, 33]. The pipeline also allows convenient processing of multiple FASTQ files (samples) with very limited input from the user required. In case of the *PRP45*-related datasets mentioned below, ~55 million 100 nt reads were processed by the pipeline in 38 minutes on a standard desktop PC (quad-core AMD A8-3870 APU CPU with 8 GB RAM) running 64-bit Ubuntu 16.04.

While the scripts provided in the Supplementary Material are customized for analysis of *S. cerevisiae* RNA-seq data, the pipeline can be easily adapted for other yeast species by providing the corresponding genome sequence, genome annotation, and genome index (built using hisat2-build [29]) files and altering the relevant filename variables in the "workflow.sh" script accordingly. The Supplementary Material contains one such example adaptation of the pipeline for the fission yeast *Schizosaccharomyces pombe*.

3.2. Analysis of Splicing Efficiency in Spliceosome Mutants. As we are primarily interested in studying altered splicing patterns in various spliceosome mutants, we selected three publicly available RNA-seq datasets for *S. cerevisiae* spliceosome mutants, which show global reductions in splicing efficiency, to demonstrate the function of our workflow. All datasets contain samples with roughly similar amounts of sequencing reads (~17–35 million) of very similar length (~100 nt) but differ markedly in data quality in terms of the percentage of uniquely mappable reads (Table 1).

The first example dataset is focused on Prp45, an essential splicing factor and a component of the so-called NTC-related complex [36]. Based on cryo-EM structural information, Prp45 stabilizes the catalytic centre of the spliceosome through interactions with many proteins and with all three spliceosomal snRNAs [37]. Cells bearing the C-terminally truncated *prp45*(1-169) allele are temperature-sensitive and have deformed shapes [38]. We analysed two biological replicates of RNA-seq data for wild-type and *prp45*(1-169) cells grown at permissive temperature (30°C). First, we used the data pooled by genotype and found a global decrease of splicing efficiency in the mutant (Figures 2(a) and 2(b)). Next, we calculated relative splicing efficiencies (i.e., mutant normalized to wild type) in each replicate separately to assess reproducibility. Results for introns with sufficient splice junctions coverage (≥5 transreads and ≥5 reads covering intron end base) showed good agreement between the two independent replicates (Figures 2(c) and 2(d)). Finally, to validate the results of our pipeline, we used

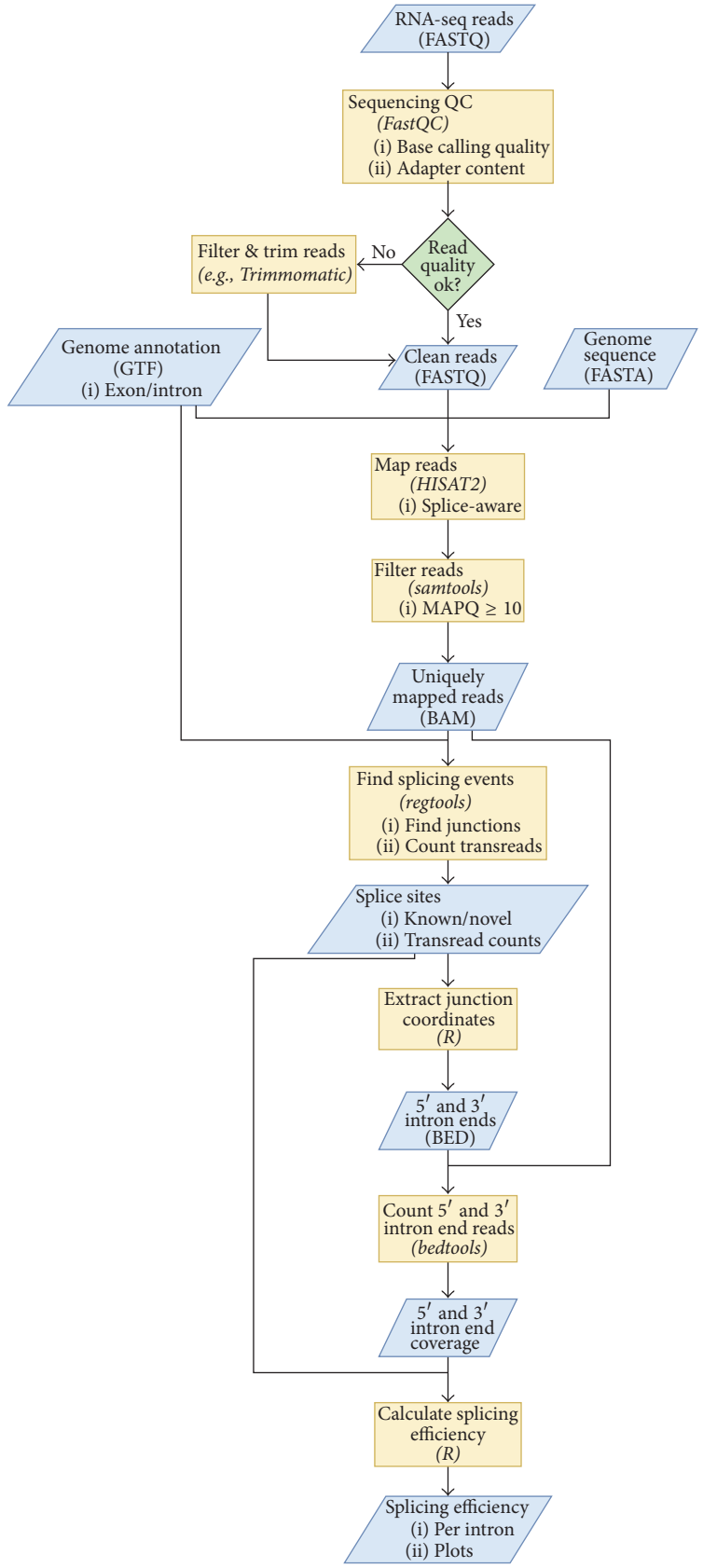


Figure 1: Workflow for calculating splicing efficiency from RNA-seq data. Files and datasets are represented by blue parallelograms (file formats given in parentheses), and processing steps are represented by orange rectangles (tool names given in parentheses). Some files/datasets are used repeatedly in several steps of the workflow as signified by multiple flow lines going from these files/datasets. The diagram was created using draw.io (https://www.draw.io/).

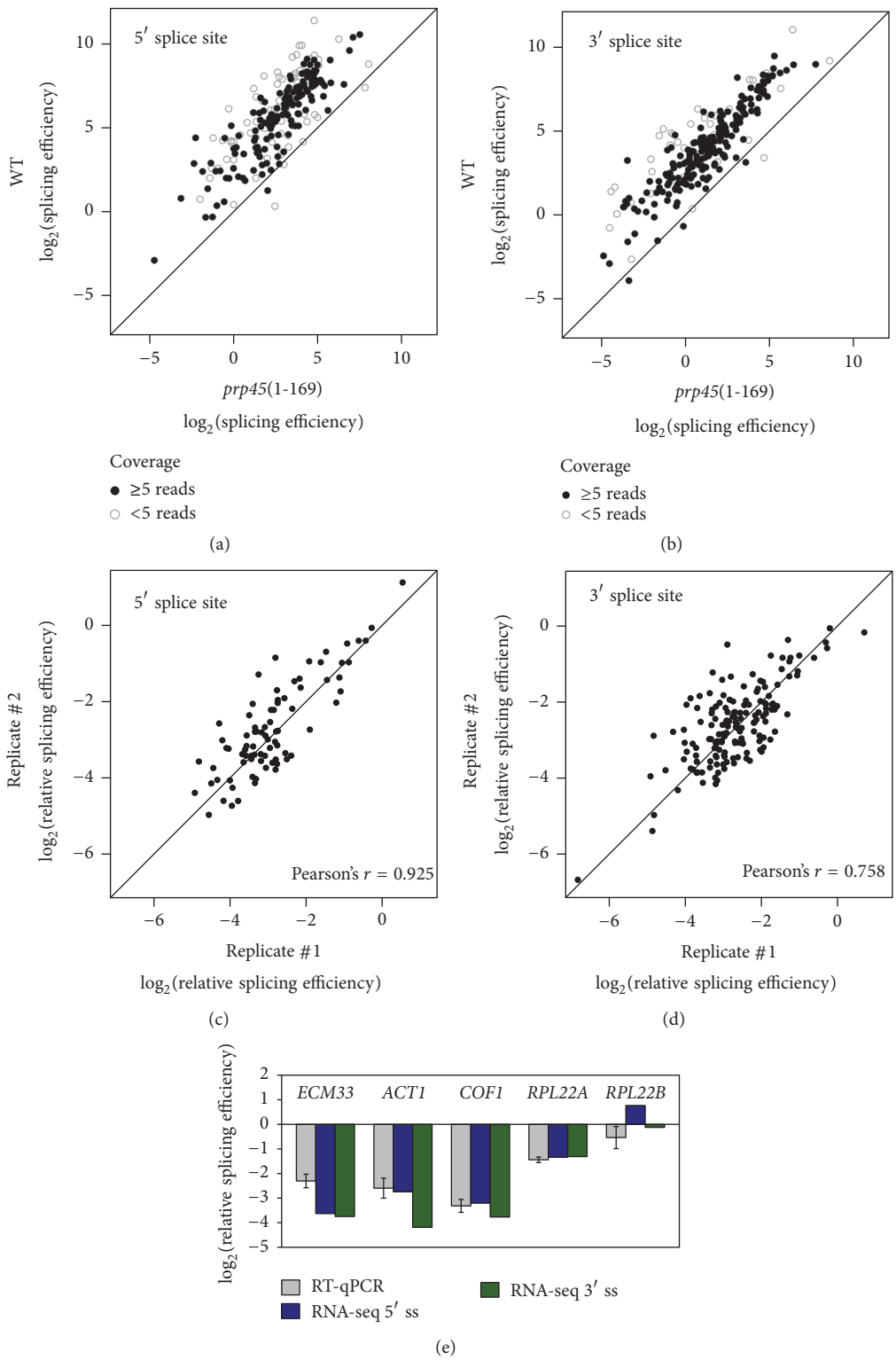


Figure 2: Splicing efficiency in the *prp45*(1-169) mutant. Splicing efficiencies for all known introns (for 5′ and 3′ splice sites separately) were calculated using the pipeline described in Figure 1. (a, b) Results for two pooled biological replicates of the *prp45*(1-169) mutant and its corresponding wild-type strain. Higher values correspond to more efficient splicing. Full circles represent values for introns with sufficient coverage (≥5 transreads and ≥5 reads covering intron end base); open circles represent low-confidence values for introns with low sequencing read coverage. (c, d) Relative splicing efficiencies (*prp45*(1-169) normalized to wild type) at the 5′ and 3′ splice sites were calculated for each biological replicate separately. Only introns with sufficient read coverage were considered. Pearson's *r* values for the two replicates are indicated. (e) Comparison of relative splicing efficiencies at the 5′ and 3′ splice sites of selected genes calculated from the pooled RNA-seq data with relative splicing efficiencies determined by RT-qPCR (means of 4–6 independent RT-qPCR experiments ± SD).

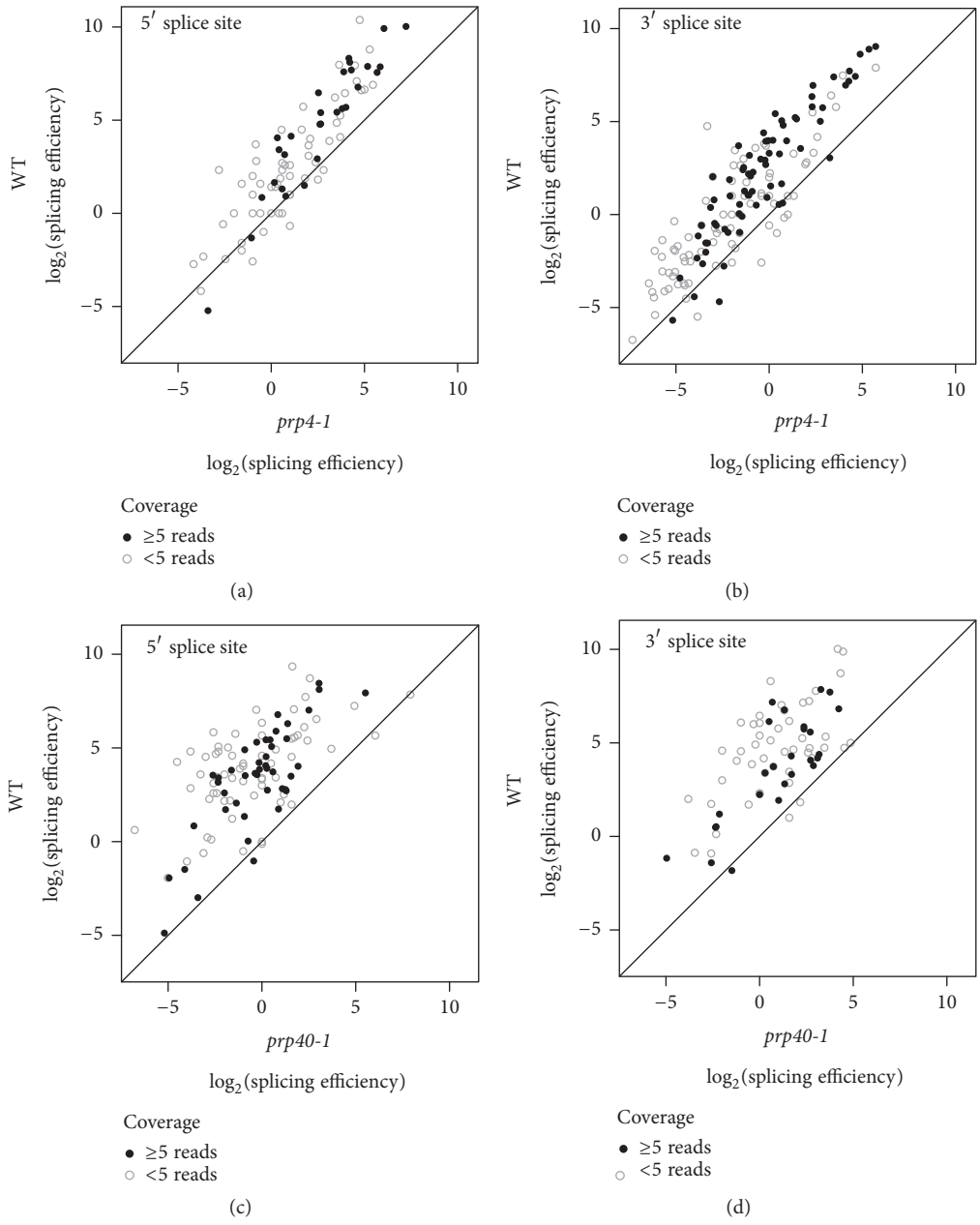


Figure 3: Splicing efficiency in the *prp4-1* and *prp40-1* mutants. Splicing efficiencies for all known introns (for 5′ and 3′ splice sites separately) were calculated using the pipeline described in Figure 1. (a, b) Results for the *prp4-1* mutant and its corresponding wild-type strain [17]. (c, d) Results for the *prp40-1* mutant and its corresponding wild-type strain [24]. Higher values correspond to more efficient splicing. Full circles represent values for introns with sufficient coverage (≥5 transreads and ≥5 reads covering intron end base); open circles represent low-confidence values for introns with low sequencing read coverage.

RT-qPCR to measure the levels of spliced and unspliced transcripts for five selected genes showing various degrees of splicing impairment in the *prp45*(1-169) mutant RNA-seq dataset. Reassuringly, the splicing efficiencies calculated by our pipeline were concordant with those determined by RT-qPCR, pointing to possible differential requirements for Prp45 in pre-mRNA splicing of specific genes (Figure 2(e)).

Next, we analysed data for Prp4, a structural component of the U4/U6 di- and U4/U6-U5 tri-snRNP particles [39, 40]. The *prp4-1* temperature-sensitive allele blocks U4 snRNP dissociation during the catalytic activation of the spliceosome [41, 42]. Using splicing-sensitive microarrays, it was shown that this mutation causes a genome-wide splicing defect even at the permissive temperature of 26°C [43]. We used the *prp4-1* RNA-seq dataset from [17] and we confirmed the global impairment of splicing efficiency in this mutant at permissive temperature (Figures 3(a) and 3(b)). However, for a number of introns, splicing efficiency could not be

convincingly determined (especially at the 5′ splice site) due to low read coverage at the splice junctions. This was unexpected since the *prp4-1* sequencing library sizes and mappability were comparable to the *prp45*(1-169) dataset (Table 1). Furthermore, the whole distribution of splicing efficiencies at the 3′ splice site (for both *prp4-1* and its wild-type control) was shifted compared to the other two splicing mutants we analysed. Visual inspection of mapped *prp4-1* reads revealed decreasing coverage towards 5′ ends of genes, where introns are typically located in *S. cerevisiae*, suggesting possible problems with RNA sample preparation and/or processing.

The third example relates to Prp40, a component of the U1 snRNP particle [44]. Through interactions with many spliceosome subunits and with the phosphorylated C-terminal domain of the RNA polymerase II, the Prp40 protein is important for cotranscriptional spliceosome assembly (reviewed in [45]). The *prp40-1* temperature-sensitive mutant [44] exerts a global splicing defect when shifted to nonpermissive temperature [24], which we confirmed by running the published RNA-seq dataset through our pipeline (Figures 3(c) and 3(d)). It should be noted that this dataset had poor mappability (Table 1) and yielded relatively low read coverage, decreasing the number of junctions for which splicing efficiency could be calculated convincingly.

Thus, these three proof-of-principle scenarios demonstrated that our workflow is able to recapitulate previously identified genome-wide decreases in splicing efficiency using RNA-seq data. The results also highlight a critical requirement for sufficient sequencing library quality and size for successful analysis.

4. Conclusions

We present a complete bioinformatics workflow for determining splicing efficiency in the budding yeast *S. cerevisiae* using data produced by the simple and affordable RNA-seq technique. Starting with strand-specific sequencing reads in the FASTQ format, our pipeline is able to calculate splicing efficiency at the 5′ and 3′ splice junctions of each intron, with very limited input required from the user. All relevant scripts are provided in a documented and ready-to-use form. The workflow should prove useful for studies of yeast splicing mutants or of regulated splicing, for example, under various growth conditions.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Czech Science Foundation Grant 14-19002S and Charles University Grants SVV 260310, UNCE 204013, and GAUK 8214.

References

[1] C. L. Will and R. Lührmann, "Spliceosome structure and function," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 7, pp. 1–23, 2011.

[2] A. J. Ward and T. A. Cooper, "The pathobiology of splicing," *Journal of Pathology*, vol. 220, no. 2, pp. 152–163, 2010.

[3] R. K. Singh and T. A. Cooper, "Pre-mRNA splicing in disease and therapeutics," *Trends in Molecular Medicine*, vol. 18, no. 8, pp. 472–482, 2012.

[4] Z. Wang and C. B. Burge, "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code," *RNA*, vol. 14, no. 5, pp. 802–813, 2008.

[5] M. Deutsch and M. Long, "Intron-exon structures of eukaryotic model organisms," *Nucleic Acids Research*, vol. 27, no. 15, pp. 3219–3228, 1999.

[6] M. Spingola, L. Grate, D. Haussler, and A. Manuel Jr., "Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae," *RNA*, vol. 5, no. 2, pp. 221–234, 1999.

[7] M. Ares, L. Grate, and M. H. Pauling, "A handful of intron-containing genes produces the lion's share of yeast mRNA," *RNA*, vol. 5, no. 9, pp. 1138–1139, 1999.

[8] F. Kempken, "Alternative splicing in ascomycetes," *Applied Microbiology and Biotechnology*, vol. 97, no. 10, pp. 4235–4241, 2013.

[9] J. Engebrecht, K. Voelkel-Meiman, and G. S. Roeder, "Meiosis-specific RNA splicing in yeast," *Cell*, vol. 66, no. 6, pp. 1257–1268, 1991.

[10] M. Bergkessel, G. B. Whitworth, and C. Guthrie, "Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast," *RNA*, vol. 17, no. 8, pp. 1461–1478, 2011.

[11] J. A. Pleiss, G. B. Whitworth, M. Bergkessel, and C. Guthrie, "Rapid, transcript-specific changes in splicing in response to environmental stress," *Molecular Cell*, vol. 27, no. 6, pp. 928–937, 2007.

[12] T. Nakagawa and H. Ogawa, "The Saccharomyces cerevisiae MER3 gene, encoding a novel helicase-like protein, is required for crossover control in meiosis," *EMBO Journal*, vol. 18, no. 20, pp. 5714–5723, 1999.

[13] E. M. Munding, A. H. Igel, L. Shiue, K. M. Dorighi, L. R. Treviño, and M. Ares Jr., "Integration of a splicing regulatory network within the meiotic gene expression program of *Saccharomyces cerevisiae*," *Genes & Development*, vol. 24, no. 23, pp. 2693–2704, 2010.

[14] C. A. Davis, L. Grate, M. Spingola, and M. Ares Jr., "Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast," *Nucleic Acids Research*, vol. 28, no. 8, pp. 1700–1706, 2000.

[15] P. Fabrizio, J. Dannenberg, P. Dube et al., "The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome," *Molecular Cell*, vol. 36, no. 4, pp. 593–608, 2009.

[16] S. Hao and D. Baltimore, "RNA splicing regulates the temporal order of TNF-induced gene expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 29, pp. 11934–11939, 2013.

[17] E. M. Munding, L. Shiue, S. Katzman, J. P. Donohue, and M. Ares, "Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing," *Molecular Cell*, vol. 51, no. 3, pp. 338–348, 2013.

[18] T. Kawashima, S. Douglass, J. Gabunilas, M. Pellegrini, and G. F. Chanfreau, "Widespread use of non-productive alternative splice sites in saccharomyces cerevisiae," *PLoS Genetics*, vol. 10, no. 4, Article ID e1004249, 2014.

[19] D. A. Bitton, C. Rallis, D. C. Jeffares et al., "LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq," *Genome Research*, vol. 24, no. 7, pp. 1169–1179, 2014.

[20] L. Herzel and K. M. Neugebauer, "Quantification of co-transcriptional splicing from RNA-Seq data," *Methods*, vol. 85, pp. 36–43, 2015.

[21] S. B. Livesay, S. E. Collier, D. A. Bitton, J. Bähler, and M. D. Ohi, "Structural and functional characterization of the N terminus of *Schizosaccharomyces pombe* Cwf10," *Eukaryotic Cell*, vol. 12, no. 11, pp. 1472–1489, 2013.

[22] C. J. Grisdale, L. C. Bowers, E. S. Didier, and N. M. Fast, "Transcriptome analysis of the parasite Encephalitozoon cuniculi: an in-depth examination of pre-mRNA splicing in a reduced eukaryote," *BMC Genomics*, vol. 14, no. 1, article 207, 2013.

[23] G. M. Gould, J. M. Paggi, Y. Guo et al., "Identification of new branch points and unconventional introns in Saccharomyces cerevisiae," *RNA*, vol. 22, no. 10, pp. 1522–1534, 2016.

[24] A. Volanakis, M. Passoni, R. D. Hector et al., "Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast," *Genes and Development*, vol. 27, no. 18, pp. 2025–2038, 2013.

[25] F. Ozsolak and P. M. Milos, "RNA sequencing: advances, challenges and opportunities," *Nature Reviews Genetics*, vol. 12, no. 2, pp. 87–98, 2011.

[26] J. A. Reuter, D. V. Spacek, and M. P. Snyder, "High-throughput sequencing technologies," *Molecular Cell*, vol. 58, no. 4, pp. 586–597, 2015.

[27] T. H. Jensen, A. Jacquier, and D. Libri, "Dealing with pervasive transcription," *Molecular Cell*, vol. 52, no. 4, pp. 473–484, 2013.

[28] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: a flexible trimmer for Illumina sequence data," *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.

[29] D. Kim, B. Langmead, and S. L. Salzberg, "HISAT: a fast spliced aligner with low memory requirements," *Nature Methods*, vol. 12, no. 4, pp. 357–360, 2015.

[30] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.

[31] H. Thorvaldsdóttir, J. T. Robinson, and J. P. Mesirov, "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration," *Briefings in Bioinformatics*, vol. 14, no. 2, pp. 178–192, 2013.

[32] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg, "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions," *Genome Biology*, vol. 14, no. 4, article R36, 2013.

[33] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, Article ID btq033, pp. 841–842, 2010.

[34] K. B. Hooks, S. Naseeb, S. Parker, S. Griffiths-Jones, and D. Delneri, "Novel intronic RNA structures contribute to maintenance of phenotype in Saccharomyces cerevisiae," *Genetics*, vol. 203, no. 3, pp. 1469–1481, 2016.

[35] T. D. Schmittgen and K. J. Livak, "Analyzing real-time PCR data by the comparative C_T method," *Nature Protocols*, vol. 3, no. 6, pp. 1101–1108, 2008.

[36] M. D. Ohi, A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald, and K. L. Gould, "Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs," *Molecular and Cellular Biology*, vol. 22, no. 7, pp. 2011–2024, 2002.

[37] R. Wan, C. Yan, R. Bai, G. Huang, and Y. Shi, "Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution," *Science*, vol. 353, no. 6302, pp. 895–904, 2016.

[38] O. Gahura, K. Abrhámová, M. Skružný et al., "Prp45 affects Prp22 partition in spliceosomal complexes and splicing efficiency of non-consensus substrates," *Journal of Cellular Biochemistry*, vol. 106, no. 1, pp. 139–151, 2009.

[39] J. Banroques and J. N. Abelson, "PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle," *Molecular and Cellular Biology*, vol. 9, no. 9, pp. 3710–3719, 1989.

[40] S. P. Bjørn, A. Soltyk, J. D. Beggs, and J. D. Friesen, "PRP4 (RNA4) from *Saccharomyces cerevisiae*: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle," *Molecular and Cellular Biology*, vol. 9, no. 9, pp. 3698–3709, 1989.

[41] L. H. Hartwell, C. S. McLaughlin, and J. R. Warner, "Identification of ten genes that control ribosome formation in yeast," *MGG Molecular & General Genetics*, vol. 109, no. 1, pp. 42–56, 1970.

[42] L. Ayadi, M. Miller, and J. Banroques, "Mutations within the yeast U4/U6 snRNP protein Prp4 affect a late stage of spliceosome assembly," *RNA*, vol. 3, no. 2, pp. 197–209, 1997.

[43] T. A. Clark, C. W. Sugnet, and M. Ares Jr., "Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays," *Science*, vol. 296, no. 5569, pp. 907–910, 2002.

[44] H.-Y. Kao and P. G. Siliciano, "Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle," *Molecular and Cellular Biology*, vol. 16, no. 3, pp. 960–967, 1996.

[45] S. Becerra, E. Andrés-León, S. Prieto-Sánchez, C. Hernández-Munain, and C. Suñé, "Prp40 and early events in splice site definition," *Wiley Interdisciplinary Reviews: RNA*, vol. 7, no. 1, pp. 17–32, 2016.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 5837293, 8 pages
http://dx.doi.org/10.1155/2016/5837293

Research Article

Comparative Proteomic Analysis of Light-Induced Mycelial Brown Film Formation in *Lentinula edodes*

Li Hua Tang,[1,2] Qi Tan,[1] Da Peng Bao,[1] Xue Hong Zhang,[2] Hua Hua Jian,[2] Yan Li,[1] Rui heng Yang,[1] and Ying Wang[1]

[1]*National Engineering Research Centre of Edible Fungi, Key Laboratory of Edible Fungi Resources and Utilization (South), Ministry of Agriculture, Institute of Edible Fungi, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China*
[2]*State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China*

Correspondence should be addressed to Ying Wang; wyhrx@126.com

Received 11 July 2016; Accepted 21 September 2016

Academic Editor: Qiang Gao

Light-induced brown film (BF) formation by the vegetative mycelium of *Lentinula edodes* is important for ensuring the quantity and quality of this edible mushroom. Nevertheless, the molecular mechanism underlying this phenotype is still unclear. In this study, a comparative proteomic analysis of mycelial BF formation in *L. edodes* was performed. Seventy-three protein spots with at least a twofold difference in abundance on two-dimensional electrophoresis (2DE) maps were observed, and 52 of them were successfully identified by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry (MALDI-TOF/TOF/MS). These proteins were classified into the following functional categories: small molecule metabolic processes (39%), response to oxidative stress (5%), and organic substance catabolic processes (5%), followed by oxidation-reduction processes (3%), single-organism catabolic processes (3%), positive regulation of protein complex assembly (3%), and protein metabolic processes (3%). Interestingly, four of the proteins that were upregulated in response to light exposure were nucleoside diphosphate kinases. To our knowledge, this is the first proteomic analysis of the mechanism of BF formation in *L. edodes*. Our data will provide a foundation for future detailed investigations of the proteins linked to BF formation.

1. Introduction

The edible medicinal mushroom species *Lentinula edodes* ranks second in production among all mushrooms in the world and is an important source of not only food but also lentinan that has medicinal value. During cultivation of this mushroom, the formation of brown film (BF) by the mycelia is an important stage before primordia and fruiting formation, and it is closely related to mushroom quality. BF formation is affected by many environmental factors, including light. However, the mechanism of light-induced brown film formation is poorly understood, and the associated changes in protein expression have not been explored.

The rapidly developing field of proteomics has become increasingly relevant to the study of fungal biology. Proteomics is a powerful tool used for the sensitive detection and rapid identification of changes in protein expression in response to various biotic and abiotic stresses [1–3] and has been utilized in many studies on the physiology and development of filamentous fungi [4]. In addition, proteomics has been applied to study the growth and development of various types of mushrooms such as *L. edodes* [5]. Other proteomic studies have evaluated the effects of low-temperature stress on protein expression in the mycelium and fruiting body of *Flammulina velutipes* [6], the effects of light on the growth and development of the cap and stipe of *F. velutipes* [7], differential protein expression in dual-core and single-core mycelial hyphae and in the fruit body of *Flammulina velutipes* [8], and protein expression during the fruiting of *Hericium erinaceus* and *Sparassis crispa* [9]. However, proteomics has not been applied to investigate light-induced BF formation in the *L. edodes* mycelium.

Although the biology and physiology of light-induced BF formation have been studied previously [10], little is known about the changes in protein expression associated with this process. In this study, we used two-dimensional electrophoresis (2-DE) and matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry (MALDI-TOF/TOF/MS) to identify proteins differentially expressed in mycelia with and without BF formation. The results of this study provide insights into the molecular mechanisms and physiological basis of light-induced BF formation in *L. edodes*, as well as mushroom development and secondary metabolite production, which will facilitate the development of methods for promoting BF formation and improving the quality of the fruit body of mushrooms.

2. Materials and Methods

2.1. Cultivation of Fungal Mycelia. *L. edodes* strain 135 was obtained from the Agricultural Culture Collection of China (Beijing) (number ACCC50903). Fungal mycelia were grown at 22°C in cultivation bags containing 850 g of medium (80% dry sawdust and 20% dry corn bran) in the dark for 30 d, after which the substrate was fully colonized. Bags were then either exposed to a 12-h light/dark protocol (white light, 300 lx) for 50 d to induce BF formation on the mycelial surface (sample 313C) or maintained for 50 d in the dark (sample 313W, controls). Samples from the mycelial surface were then collected for protein preparation.

2.2. Protein Extraction. Protein extraction was performed using the trichloroacetic acid- (TCA-) acetone method as previously described [11], with some modifications. The samples were ground to a powder in liquid nitrogen, and 1 g of sample powder was extracted in 10 mL of 10% TCA in cold acetone at −20°C for 1 h. After centrifugation at 15,000 ×g for 15 min at 4°C, the deposits were washed three times with 10 mL of cold acetone at −20°C for 1 h and centrifuged at 15,000 ×g for 15 min at 4°C. The deposits were then collected and dried in a vacuum freeze dryer. The deposits were dissolved in a lysis solution containing 9 M urea, 4% CHAPS (w/v), 1% DTT (w/v), and 1% IPG buffer (v/v) (pH 3–10, GE Healthcare Bio-Science, Little Chalfont, UK) at 30°C for 1 h, and the solution was centrifuged at 15,000 ×g for 15 min at room temperature. The supernatants were collected and centrifuged again, and the concentration of the protein extracts was determined by the Bradford method [12]. The extracted protein solution was stored at −80°C until isoelectric focusing.

2.3. 2DE Image and Data Analysis. Samples containing 300 μg of proteins were brought to a total volume of 450 μL in fresh rehydration buffer (9 M urea, 4% CHAPS, 1% DTT, 1% IPG buffer, and a trace amount of bromophenol blue) and loaded on a DryStrip gel (GE Healthcare, 24 cm, pH = 3–10, NL). Isoelectric focusing was performed with an IPGphor system (GE Healthcare). The voltage was set to 50 V for 12 h, 500 V for 1 h, 1,000 V for 1 h, 10,000 V for 1 h, and 10,000 V for 8 h. The strips were subsequently incubated in equilibration buffer (6 M urea, 30% glycerol, 2% SDS, 50 mM Tris-HCl (pH 8.8), 1% DTT, and a trace amount of bromophenol blue) for 15 min followed by 15 min in 2.5% (w/v) iodoacetamide in equilibration buffer. Then, the IPG strips were loaded in a 12% (w/v) SDS-PAGE gel using the Ettan Dalt Six system (GE Healthcare). The gels were run at 100 V for 1 h and maintained at 200 V until the dye front reached the bottom of the gel. The gel was visualized by silver staining according to Yan's protocol [13] and scanned by Image Scanner software (GE Healthcare, USA) at a resolution of 300 dots per inch. The gel images were processed in three steps using PDquest 8.0 software (Bio-Rad): spot detection, volumetric quantification, and matching. A total of six 2DE gels resulting from three independent biological replicates of both the 313W and 313C sample were analyzed and the standard deviations were calculated from the three independent replicates. A threshold of $p \leq 0.05$ and a fold change of ≥2 or ≤0.5 were used to identify significant differentially expressed protein spots.

2.4. Protein Digestion. Gel spots were excised and destained with 15 mM $K_3Fe(CN)_6$ in 50 mM NaS_2O_3 at room temperature for 5 min. After removal of the destaining solution, the gels were washed twice with 200 μL of ddH_2O, and 50 μL of 25 mM NH_4HCO_3 and 100 μL of 50% ACN were then added, followed by 100 μL of 100% ACN. The gels were rehydrated in 5 μL of trypsin solution (Promega, Madison, WI, USA; 20 μg/mL in 25 mM NH_4HCO_3) for 30 min. Next, 20 μL of cover solution (25 mM NH_4HCO_3) was added, and the gels were digested for 16 h at 37°C. The supernatants were transferred to another tube, and the gels were extracted once with 50 μL of extraction buffer (67% ACN and 5% TFA). The peptide extracts and gel spot supernatants were combined and then completely dried.

2.5. Protein Identification by MALDI-TOF/TOF/MS and Database Search. Samples were resuspended in 5 μL of 0.1% TFA and mixed with a matrix comprising a saturated solution of α-cyano-4-hydroxy-trans-cinnamic acid (CHCA) in 50% ACN/0.1% TFA (1 : 1). Then, 1 μL of the mixture was spotted on a stainless steel sample target plate. Peptide MS and MS/MS were performed with an ABI 5800 MALDI-TOF/TOF Plus mass spectrometer (Applied Biosystems, Foster City, CA, USA). Data were acquired with a positive MS reflector using a CalMix5 standard to calibrate the instrument (ABI5800 Calibration Mixture). Both the MS and MS/MS data were integrated and processed by GPS Explorer V3.6 software (Applied Biosystems, USA) with default parameters. Based on the combined MS and MS/MS spectra, proteins were successfully identified based on the 95% or higher confidence interval of their scores in the MASCOT V2.3 search engine (Matrix Science Ltd., London, UK) with the following search parameters: a local database with peptide sequences created from the mycelial transcriptome (RNA-Seq) of *L. edodes* [10], trypsin as the digestion enzyme, one missed cleavage site, partial modifications of cysteine carbamidomethylation and methionine oxidization, no fixed modifications, and 200 ppm for precursor ion tolerance and 0.4 Da for fragment ion tolerance.

2.6. Quantitative Real-Time PCR (RT-qPCR) Validation of Proteomic Data. Eleven proteins exhibiting significant

upregulation in proteomic analysis (313C/313W) were selected for validation using RT-qPCR. cDNAs were synthesized according to the manufacturer's protocol (Takara, Dalian, China) and used as templates for real-time PCR using specific primers (Table S1 in Supplementary Material available online at http://dx.doi.org/10.1155/2016/5837293). Real-time PCR was performed using the FTC2000 System (Canada). Reaction mixtures (50 μL) contained 0.5 μL 20x SYBR green Mastermix, 1 μL of each primer (25 pmol/μL), 2 μL 10-fold diluted cDNA template, and 20.5 μL DEPC water. Amplification conditions were as follows: 94°C for 4 min, 35 cycles of 94°C for 20 s, 60°C for 30 s, and 72°C for 30 s. Each reaction was performed in triplicate using 18S rRNA as the internal control gene, and the relative gene expression levels were determined using the delta-delta Ct method [14].

3. Results

3.1. Growth of the L. edodes Mycelium. The cultivation substrate was fully colonized by white fungal mycelia after incubation of the test bags for 30 d in the dark. In the test bags that were subsequently exposed to a 12 h light/dark protocol, the mycelia gradually developed a BF (sample 313C; Figure 1(b)), whereas the mycelia in the control group remained white (sample 313W; Figure 1(a)).

3.2. 2DE Analysis of the Mycelial Protein Profile. To identify proteins that were differentially expressed during mycelial BF formation in response to light, a comparative proteomic analysis was performed on the mycelial surface components of the brown film and control groups. The extracted proteins were resolved and analyzed by 2DE with a pH range of 3 to 10, and the gels were visualized by silver staining and analyzed with PDquest software. Two representative images of the 2DE gel are shown in Figure 2; according to this figure, there were 789 ± 14 and 726 ± 30 examined protein spots for 313W and 313C, respectively, which demonstrates that more than 650 protein spots were detected, and most of the protein spots were distributed in the Mr range of 20.1 to 84.0 kDa. Of these spots, a total of 73 stained spots had significant and repeatable changes in abundance when the two treatments were compared (313C versus 313W).

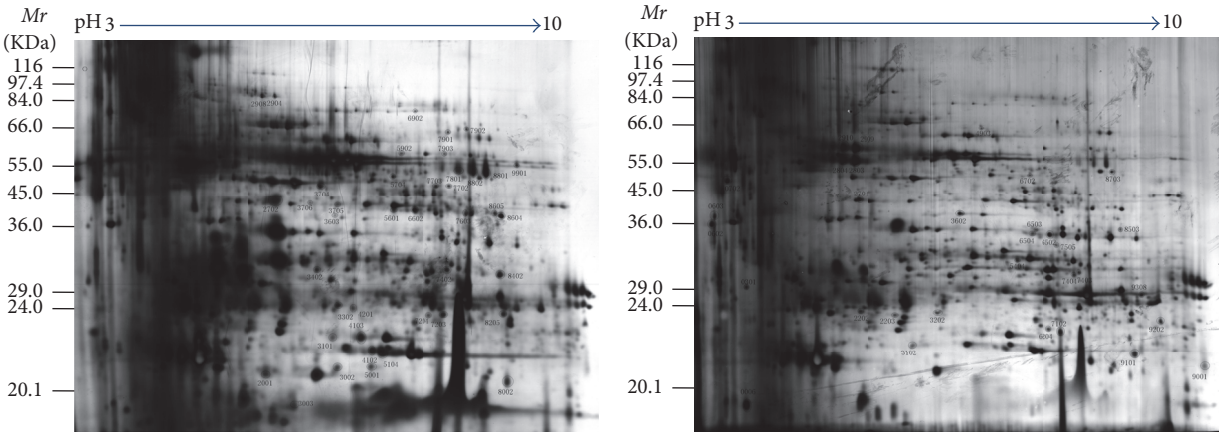
3.3. Identification of Differentially Expressed Proteins (DEPs). Among these 71 protein spots, 52 were successfully identified by MALDI-TOF/TOF, and a local protein database was constructed from the recently sequenced mycelium transcriptome [10]. Among the 52 proteins, 23 were upregulated and 29 were downregulated when 313C was compared to 313W. Moreover, 23 proteins were upregulated (Table S2); for example, small heat shock proteins such as heat shock cognate 70, spot 7404, which was induced by light treatment during BF formation, was upregulated compared to the control. Meanwhile, two proteins (20S proteasome subunit, spot 6504, and proteasome component pts1, spot7505) involved in proteasome were also upregulated in 313C. Short-chain dehydrogenase/reductase (SDR, spot 0301) was 3.61-fold upregulated in 313C compared to 313W, and manganese superoxide dismutase (spot 2202) was 2.39-fold upregulated in 313C compared to 313W. Four nucleoside diphosphate kinase proteins (spot 0006, spot 7102, spot 6204, and spot 9202) were all ≥2-fold upregulated in 313C compared to 313W, while for 29 proteins their expression was significantly downregulated in 313C, including NAD-aldehyde dehydrogenase (spot 6602), protein disulfide isomerase (spot 8402), succinate-semialdehyde dehydrogenase (spot 7702), heat shock protein (spot 5704), glycoside hydrolase family 13 protein (spot 5902), proliferation-associated 2G4 (spot 9901), tripeptidyl peptidase A (spot 8801), heat shock cognate 70 (spot 3704), hsp70-like protein (spot 6902), small ubiquitin-related modifier (spot 7203), and FG-GAP repeat-containing protein (spot 2904). Carbohydrate-binding module family 13 protein (spot 8002) was more than 9-fold downregulated in 313C. Some proteins were distributed across multiple spots at different positions on the same gel, including adenosine kinase (spots 2909 and 2910), heat shock cognate 70 (spots 3704, 3705, and 4201), and nucleoside diphosphate kinase (spots 0006, 7102, 6204, and 9202). The multiple spots suggest that these proteins may be subjected to posttranslational modifications, such as phosphorylation, N-acetylation, and glycosylation.

3.4. Functional Classification of the DEPs. In order to understand the roles of proteins associated with mycelial BF formation in this mushroom, the differentially expressed proteins were sorted into different categories based on their functions (Table S2, Figure 3(a)). As shown in Figure 3, in terms of quantitative changes in BF formation, proteins involved in small molecule metabolic processes, response to oxidative stress, and organic substance catabolic processes accounted for the majority of differentially expressed proteins in 313C. In addition, as is shown in Figure 3(a), the most represented groups of proteins were proteins involved in small molecule metabolic processes (39%), response to oxidative stress (5%), and organic substance catabolic processes (5%), followed by oxidation-reduction processes (3%), single-organism catabolic processes (3%), positive regulation of protein complex assembly (3%), and protein metabolic processes (3%). This finding implies that proteins involved in small molecule metabolic processes may play major roles in 313C under light conditions.

3.5. Gene Ontology (GO) Analysis of Differentially Expressed Proteins. To gain further knowledge of the biological functionality of the differentially expressed proteins between 313C and 313W, these proteins were used to perform GO analysis and were assigned to three GO vocabularies: biological process (GO-BP), cellular component (GO-CC), and molecular function (GO-MF). As is shown from the results of GO-BP analysis (Figure 3(a)), DEPs were involved in diverse biological processes. The most highly enriched GO-BP category was small molecule metabolic processes (39%), demonstrating that these processes are of functional importance for brown film formation in *L. edodes*. The second highly enriched GO-BP categories were in response to oxidative stress (5%) and organic substance catabolic processes (5%). As for the GO-MF categories (Figure 3(c)), the top three

(a) (b)

Figure 1: Comparison of the surface of the mycelium grown under different illumination conditions. (a) 50 d of 24 h of darkness; (b) 50 d of a 12 h dark/light regimen (sample 313C).



(a) (b)

Figure 2: Representative 2DE protein patterns from *Lentinula edodes* mycelium sample 313W that did not produce a brown film (a) and sample 313C (b) that produced a brown film. 313W indicates samples that did not form a brown film when cultivated in the dark, while 313C denotes samples that formed a brown film when cultivated under light/dark conditions. Proteins were loaded on a 24 cm IPG strip with a nonlinear immobilized pH gradient ranging from 3 to 10 for isoelectric focusing, followed by electrophoresis in a 12% SDS-PAGE gel and silver staining.

GO-MFs were oxidoreductase activity acting on CH-OH group of donors (13%), unfolded protein binding (10%), and peptidase activity (10%) (Figure 3(c)). These were followed by nucleoside phosphate binding (8%) and identical protein binding (5%). There were other categories also involved in oxidoreductase activity acting on the aldehyde or oxo group of donors (5%) and oxidoreducatse activity acting on superoxide radicals as acceptors (3%). These molecular functions of oxidoreductase activity had a higher proportion (21%) in BF processes, suggesting that the molecular function category was of functional importance for light-induced BF formation.

3.6. Analysis on Selected Transcripts of Differentially Accumulated Proteins. The proteome results were further confirmed at the transcription level. Eleven proteins, which exhibited upregulated expression, were selected and investigated by RT-qPCR analysis. The genes analyzed included nucleoside diphosphate kinase, proteasome component pts1, manganese superoxide dismutase, and adenosine kinase. The results

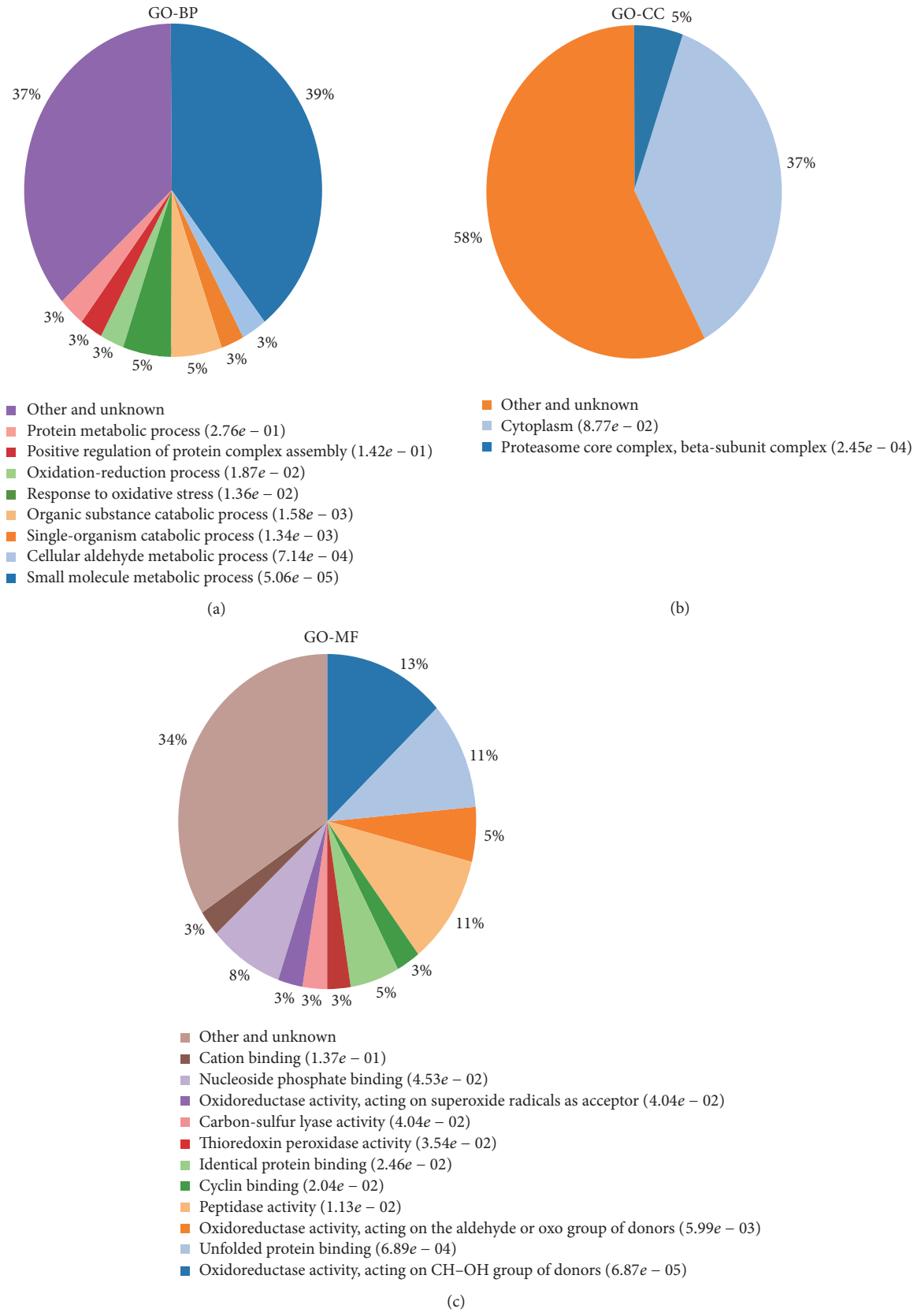


Figure 3: Functional category distribution of the differentially expressed proteins. Pie charts representing the distribution of the 52 identified proteins according to biological function are shown.

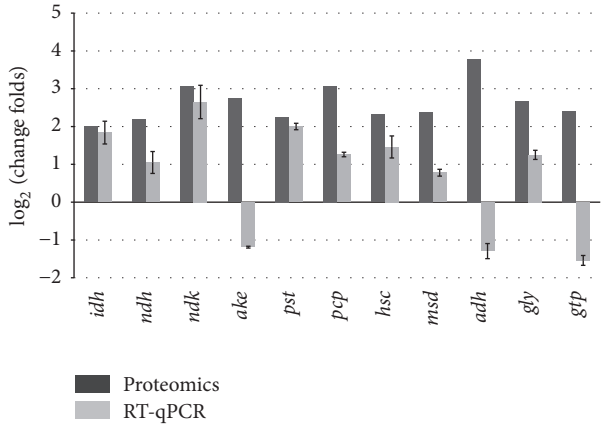


Figure 4: Validation of proteomic data by RT-qPCR. Open and solid bars indicate fold changes obtained from proteomic data and by RT-qPCR, respectively. Error bars represent the mean ± SD of triplicate experiments. The full name of the gene evaluated by qPCR and their fold changes were 3-isopropylmalate dehydrogenase (*idh*, 1.841), NAD-aldehyde dehydrogenase (*ndh*, 1.054), nucleoside diphosphate kinase (*ndk*, 2.645), adenosine kinase (*ake*, –1.189), 20S proteasome subunit (*pst*, 1.998), proteasome component pts1 (*pcp*, 1.26), heat shock cognate 70 (*hsc*, 1.457), manganese superoxide dismutase (*msd*, 1.146), alcohol dehydrogenase (*adh*, –1.293), glyoxalase (*gly*, 1.247), and glutathione S-transferase-like protein (*gtp*, –1.549). The reference gene of qPCR was 18S rRNA.

showed that most proteins (Figure 4) exhibited upregulated expression at the mRNA level comparing 313C to 313W, which correlated to the proteomic data. However, the mRNA level of three genes was different from the mRNA level identified with proteome analysis; the three genes corresponded to adenosine kinase, alcohol dehydrogenase, and glutathione *S*-transferase-like proteins.

4. Discussion

As the genome sequences of a variety of organisms are gradually being completed, proteomics will become an increasingly important form of analysis to identify functionally important proteins based on differences in protein expression.

NAD-aldehyde dehydrogenase is an energy metabolism-related enzyme that is differentially expressed during neurosporaxanthin biosynthesis, which is stimulated and regulated by light [15]. The results of our differential proteomic analysis suggest that aldehyde dehydrogenase participates in critical reaction steps in pigment biosynthesis during the process of mycelial BF formation in *L. edodes* and might be regulated by light. In the BF sample (313C), nucleoside diphosphate kinase (NDPK) was upregulated. NDPK regulates various biological processes and signal transduction pathways [16] and can interact with other proteins, including phytochrome [17], catalase [18], and MAP kinase [19]; these proteins function as substrates or regulators in signal transduction pathways. In fungi, light is one of the most important factors for inducing the shape of *Neurospora crassa*, and nucleoside diphosphate kinase-1 (NDK-1), a 15-kDa candidate light signal transduction protein [20, 21], is rapidly phosphorylated after blue light irradiation and may be under the control of WC proteins. In yeast, NDK regulates various signal transduction pathways, such as spore growth and photomorphogenesis [22]. In this analysis, four protein spots corresponding to NDPK were upregulated in 313C, indicating that this enzyme plays an important role in light-induced BF formation. Differential expression of proteins in pathways similar to those of oxidative stress and the light response in plants and other fungi was also observed. Polyketide synthase (PKS) is a large biosynthetic enzyme that contains a variety of active domains that synthesize many complex chemical molecules, some of which have therapeutic and pharmaceutical applications [23], and in some fungi others play important roles in pigment biosynthesis [24–26]. Ketoreductase was significantly upregulated in the BF sample, confirming the involvement of this and related enzymes in complex biological processes and its correlation with light-induced secondary metabolism, which suggests that the pigments in BF likely are comprised of polyketide compounds. We also observed significant upregulation of glutathione S-transferase (GST) in the BF mycelium. GST participates in many reactions involved in the response to biotic and abiotic stresses. Moreover, GSTs can be induced by different types of light in some plants, which is the result of UV light-dependent chalcone synthase signaling pathways [27].

In the present study, differential expression of the 20S proteasome subunit, which plays an essential role in the catalytic activity of the 26S proteasome, was also observed. The proteasome is involved in plant photomorphogenesis [28] and regulates flower development processes [29–31]. Other proteins identified in this study included superoxide dismutase (SOD), a protein whose major role is to transform superoxide anion into hydrogen peroxide and molecular oxygen for defense against cell damage by reactive oxygen species (ROS) [28], and manganese superoxide dismutase (MnSOD), a ubiquitous metalloenzyme also involved in the defense against ROS. Another important DEP was alcohol dehydrogenase (ADH), which plays an important role in regulating alcohol metabolism, anaerobic respiration, and physiological resilience [32]. Anaerobic respiration based on ethanol fermentation is also a major energy supply system under oxidative stress, and ADH activity is significantly increased during hypoxia and in waterlogged environments. Consistent with the results of previous studies, our results suggest that the *L. edodes* mycelium requires more oxygen during BF formation and produces more water, which often leads to submergence and hypoxic stress and, consequently, a need for additional energy supplements, suggesting that the process of BF formation in the mycelium of this mushroom is similar to hypoxia-resistant mechanisms in other crops that involve ethanol fermentation.

In this study, 52 different proteins involved in light-induced BF formation in the *L. edodes* mycelium were identified, providing insights into the mechanism of light-induced BF formation at the proteome level. During light-induced BF formation, a series of physiological and biochemical responses are activated which involve proteins that participate in small molecule metabolic processes and proteins with oxidoreductase activity.

Competing Interests

The authors declare that no financial competing interests exist.

Acknowledgments

The authors thank Dr. Lie-bo Shu from Shanghai Boyuan Biotechnology, for technical assistance in MS analysis. This research was supported by the Shanghai Municipal Agricultural Commission, China, 2015 (Hu Nong Qing Zi no. 2015 1-9 and no. 1-10), the National Natural Science Foundation of China (Grant no. 31170098), and the Science and Technology Commission of the Shanghai Municipality (Grant no. 13391912202).

References

[1] S. Lee, E. J. Lee, E. J. Yang et al., "Proteomic of identification of annexins, calcium-dependent membrane binding proteins that mediate osmotic stress and abscisic acid signal transduction in *Arabidopsis*," *Plant Cell*, vol. 16, no. 6, pp. 1378–1391, 2004.

[2] N. Imin, T. Kerim, J. J. Weinman, and B. G. Rolfe, "Low temperature treatment at the young microspore stage induces protein changes in rice anthers," *Molecular and Cellular Proteomics*, vol. 5, no. 2, pp. 274–292, 2006.

[3] A. K. Mukherjee, M.-J. Carp, R. Zuchman, T. Ziv, B. A. Horwitz, and S. Gepstein, "Proteomics of the response of *Arabidopsis thaliana* to infection with *Alternaria brassicicola*," *Journal of Proteomics*, vol. 73, no. 4, pp. 709–720, 2010.

[4] Y. Kim, M. P. Nandakumar, and M. R. Marten, "Proteomics of filamentous fungi," *Trends in Biotechnology*, vol. 25, no. 9, pp. 395–400, 2007.

[5] V. Gergely, K. M. Kubachka, S. Mounicou, P. Fodor, and J. A. Caruso, "Selenium speciation in *Agaricus bisporus* and *Lentinula edodes* mushroom proteins using multi-dimensional chromatography coupled to inductively coupled plasma mass spectrometry," *Journal of Chromatography A*, vol. 1101, no. 1-2, pp. 94–102, 2006.

[6] Y. Sakamoto, A. Ando, Y. Tamai, K. Miura, and T. Yajima, "Protein expressions during fruit body induction of *Flammulina velutipes* under reduced temperature," *Mycological Research*, vol. 106, no. 2, pp. 222–227, 2002.

[7] Y. Sakamoto, A. Ando, Y. Tamai, and T. Yajima, "Pileus differentiation and pileus-specific protein expression in *Flammulina velutipes*," *Fungal Genetics and Biology*, vol. 44, no. 1, pp. 14–24, 2007.

[8] Y. Sakamoto, A. Ando, Y. Tamai, K. Miura, and T. Yajima, "Differential protein expression in the fruiting dikaryon and the non-fruiting monokaryon of *Flammulina velutipes*," *Mycological Research*, vol. 105, no. 2, pp. 177–182, 2001.

[9] K. Horie, R. Rakwal, M. Hirano et al., "Proteomics of two cultivated mushrooms *Sparassis crispa* and *Hericium erinaceum* provides insight into their numerous functional protein components and diversity," *Journal of Proteome Research*, vol. 7, no. 5, pp. 1819–1835, 2008.

[10] L.-H. Tang, H.-H. Jian, C.-Y. Song et al., "Transcriptome analysis of candidate genes and signaling pathways associated with light-induced brown film formation in Lentinula edodes," *Applied Microbiology and Biotechnology*, vol. 97, no. 11, pp. 4977–4989, 2013.

[11] S. C. Carpentier, E. Witters, K. Laukens, P. Deckers, R. Swennen, and B. Panis, "Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis," *Proteomics*, vol. 5, no. 10, pp. 2497–2507, 2005.

[12] M. M. Bradford, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding," *Analytical Biochemistry*, vol. 72, no. 1-2, pp. 248–254, 1976.

[13] J. X. Yan, R. Wait, T. Berkelman et al., "A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry," *Electrophoresis*, vol. 21, no. 17, pp. 3666–3672, 2000.

[14] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method," *Methods*, vol. 25, no. 4, pp. 402–408, 2001.

[15] É. Gráczer, A. Merli, R. K. Singh et al., "Atomic level description of the domain closure in a dimeric enzyme: *Thermus thermophilus* 3-isopropylmalate dehydrogenase," *Molecular BioSystems*, vol. 7, no. 5, pp. 1646–1659, 2011.

[16] B. Lee, Y. Yoshida, and K. Hasunuma, "Photomorphogenetic characteristics are severely affected in nucleoside diphosphate kinase-1 (ndk-1)-disrupted mutants in *Neurospora crassa*," *Molecular Genetics and Genomics*, vol. 275, no. 1, pp. 9–17, 2006.

[17] G. Choi, H. Yi, J. Lee et al., "Phytochrome signalling is mediated through nucleoside diphosphate kinase 2," *Nature*, vol. 401, no. 6753, pp. 610–613, 1999.

[18] Y. Fukamatsu, N. Yabe, and K. Hasunuma, "*Arabidopsis* NDK1 is a component of ROS signaling by interacting with three catalases," *Plant and Cell Physiology*, vol. 44, no. 10, pp. 982–989, 2003.

[19] H. Moon, B. Lee, G. Choi et al., "NDP kinase 2 interacts with two oxidative stress-activated MAPKs to regulate cellular redox state and enhances multiple stress tolerance in transgenic plants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 1, pp. 358–363, 2003.

[20] K. Oda and K. Hasunuma, "Genetic analysis of signal transduction through light-induced protein phosphorylation in *Neurospora crassa* perithecia," *Molecular and General Genetics*, vol. 256, no. 6, pp. 593–601, 1997.

[21] Y. Ogura, Y. Yoshida, N. Yabe, and K. Hasunuma, "A point mutation in nucleoside diphosphate kinase results in a deficient light response for perithecial polarity in *Neurospora crassa*," *Journal of Biological Chemistry*, vol. 276, no. 24, pp. 21228–21234, 2001.

[22] H. Izumiya and M. Yamamoto, "Cloning and functional analysis of the *ndk1* gene encoding nucleoside-diphosphate kinase in *Schizosaccharomyces pombe*," *The Journal of Biological Chemistry*, vol. 270, no. 46, pp. 27859–27864, 1995.

[23] G. Yadav, R. S. Gokhale, and D. Mohanty, "Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases," *Journal of Molecular Biology*, vol. 328, no. 2, pp. 335–363, 2003.

[24] Y. Chen, P. Feng, Y. Shang, Y.-J. Xu, and C. Wang, "Biosynthesis of non-melanin pigment by a divergent polyketide synthase in *Metarhizium robertsii*," *Fungal Genetics and Biology*, vol. 81, pp. 142–149, 2015.

[25] J. W. Cary, P. Y. Harris-Coward, K. C. Ehrlich et al., "Functional characterization of a veA-dependent polyketide synthase gene in *Aspergillus flavus* necessary for the synthesis of asparasone,

a sclerotium-specific pigment," *Fungal Genetics and Biology*, vol. 64, pp. 25–35, 2014.

[26] H. O. Akamatsu, M. I. Chilvers, J. E. Stewart, and T. L. Peever, "Identification and function of a polyketide synthase gene responsible for 1,8-dihydroxynaphthalene-melanin pigment biosynthesis in *Ascochyta rabiei*," *Current Genetics*, vol. 56, no. 4, pp. 349–360, 2010.

[27] L. Loyall, K. Uchida, S. Braun, M. Furuya, and H. Frohnmeyer, "Glutathione and a UV light-induced glutathione *S*-transferase are involved in signaling to chalcone synthase in cell cultures," *Plant Cell*, vol. 12, no. 10, pp. 1939–1950, 2000.

[28] C. S. Hardtke and X.-W. Deng, "The cell biology of the COP/DET/FUS proteins. Regulating proteolysis in photomorphogenesis and beyond?" *Plant Physiology*, vol. 124, no. 4, pp. 1548–1557, 2000.

[29] A. Samach, J. E. Klenz, S. E. Kohalmi, E. Risseeuw, G. W. Haughn, and W. L. Crosby, "The UNUSUAL FLORAL ORGANS gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem," *Plant Journal*, vol. 20, no. 4, pp. 433–445, 1999.

[30] D. Zhao, M. Yang, J. J. Solava, and H. Ma, "The *ASK1* gene regulates development and interacts with the *UFO* gene to control floral organ identity in *Arabidopsis*," *Developmental Genetics*, vol. 25, no. 3, pp. 209–223, 1999.

[31] I. Fridovich, "Superoxide radical and superoxide dismutases," *Annual Review of Biochemistry*, vol. 64, pp. 97–112, 1995.

[32] H. Kato-Noguchi, "Low temperature acclimation mediated by ethanol production is essential for chilling tolerance in rice roots," *Plant Signaling and Behavior*, vol. 3, no. 3, pp. 202–203, 2008.