

# Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing

Bin Wang<sup>1</sup>, Guangwu Guo<sup>2,3,4</sup>, Chao Wang<sup>1</sup>, Ying Lin<sup>1</sup>, Xiaoning Wang<sup>1</sup>, Mouming Zhao<sup>5</sup>, Yong Guo<sup>1</sup>, Minghui He<sup>1,4</sup>, Yong Zhang<sup>4,\*</sup> and Li Pan<sup>1,\*</sup>

<sup>1</sup>School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, Guangdong 510006, <sup>2</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, <sup>3</sup>Graduate University of Chinese Academy Sciences, Beijing 100049, <sup>4</sup>Beijing Genomics Institute at Shenzhen, Shenzhen 518000 and <sup>5</sup>College of Light Industry and Food Sciences, South China University of Technology, Guangzhou, Guangdong 510641, China

Received November 21, 2009; Revised March 16, 2010; Accepted March 26, 2010

## ABSTRACT

*Aspergillus oryzae*, an important filamentous fungus used in food fermentation and the enzyme industry, has been shown through genome sequencing and various other tools to have prominent features in its genomic composition. However, the functional complexity of the *A. oryzae* transcriptome has not yet been fully elucidated. Here, we applied direct high-throughput paired-end RNA-sequencing (RNA-Seq) to the transcriptome of *A. oryzae* under four different culture conditions. With the high resolution and sensitivity afforded by RNA-Seq, we were able to identify a substantial number of novel transcripts, new exons, untranslated regions, alternative upstream initiation codons and upstream open reading frames, which provide remarkable insight into the *A. oryzae* transcriptome. We were also able to assess the alternative mRNA isoforms in *A. oryzae* and found a large number of genes undergoing alternative splicing. Many genes and pathways that might be involved in higher levels of protein production in solid-state culture than in liquid culture were identified by comparing gene expression levels between different cultures. Our analysis indicated that the transcriptome of *A. oryzae* is much more complex than previously anticipated, and these results may provide a blueprint for further study of the *A. oryzae* transcriptome.

## INTRODUCTION

*Aspergillus oryzae* (koji mold) has been safely used for thousands of years in the manufacture of oriental fermented foods such as sauce, miso and sake, and its ability to produce large quantities of proteins has made the organism a suitable host for homologous and heterologous protein production (1). Although it is difficult to study *A. oryzae* by conventional genetic methods due to its multinucleate conidia and its lack of a sexual life cycle, the genome sequencing of *A. oryzae* (2) has established an alternative way to study it. In addition, the transcriptional regulation and metabolic modeling of *A. oryzae* under different culture conditions have been characterized by various methods such as expressed sequence tags (ESTs) (3,4) and oligonucleotide microarrays (5,6). However, because of the limited depth of coverage and sensitivity afforded by conventional sequencing and microarray methods, the functional complexity of the *A. oryzae* transcriptome has not yet been fully elucidated.

Recent research has shown that the high-throughput RNA-sequencing (RNA-Seq) technology is a powerful and cost-efficient tool for transcriptome analysis. It has been used to successfully interrogate transcriptomes of yeast (7,8), *Caenorhabditis elegans* (9), mice (10,11) and humans (12–14), and it has greatly accelerated our understanding of the complexity of gene expression, regulation and networks. Compared with the conventional approaches to transcriptome analysis, RNA-Seq offers several key advantages. First, RNA-Seq can accurately quantify gene expression levels with low, if any, background; second, RNA-Seq is very sensitive and can

\*To whom correspondence should be addressed. Tel: +86 20 39380616; Fax: +86 20 39380698; Email: btlipan@scut.edu.cn  
Correspondence may also be addressed to Yong Zhang. Tel/Fax: +86 755 25274032; Email: zhany@genomics.org.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

detect a larger dynamic range of gene expression levels in contrast to microarrays, which lack sensitivity for gene expression either at very low or very high levels and, therefore, have a much smaller range. Furthermore, RNA-Seq experiments are highly reproducible for both technical replicates and biological replicates (7). The most attractive advantage is that RNA-Seq can reveal much more precisely the boundaries of untranslated regions at single-nucleotide resolution and is useful for surveying complex transcriptome and sequence variations such as alternative splicing (AS) (14) and gene fusion (15). Taking these advantages into account, RNA-Seq is a revolutionary tool for transcriptome analysis.

To comprehensively understand the complex transcriptome of *A. oryzae* at whole genome level, we applied RNA-Seq to polyadenylated-enriched mRNA from four different culture conditions. Our analysis revealed numerous novel transcripts, newly identified exons and untranslated regions. The alternative upstream initiation codons (uATG) and upstream open reading frames (uORF) of many genes were also identified. In addition to significantly improving the current genome annotation of *A. oryzae*, we also assessed AS in the *A. oryzae* transcriptome and found that a far greater number of genes are undergoing AS than had previously been shown for *A. oryzae*. The functional analysis of differentially expressed genes involved in the protein production and secretion system indicated that filamentous *A. oryzae* grown under solid-state culture (SC) conditions exhibits superior protein production compared to organisms grown under liquid culture (LC) conditions. Our results substantially improved the global view of the *A. oryzae* transcriptome and paved the way for its further analysis.

## MATERIALS AND METHODS

### *Aspergillus oryzae* strain, culture conditions and DTT treatment

*Aspergillus oryzae* RIB40 was obtained from NITE Biological Resource Center (NBRC) in Japan. The CD medium used has been modified (16); it is composed of 0.3% NaNO<sub>3</sub>, 0.2% KCl, 0.05% MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.001% FeSO<sub>4</sub>·7H<sub>2</sub>O, 0.1% K<sub>2</sub>HPO<sub>4</sub> and 2.0% glucose at pH5.5. For transcriptome analysis, cultivation was conducted as follows.

For SC, 0.6 × 10<sup>7</sup> spores/plate were distributed over the CD medium-agar plate, which was covered by a layer of cellophane, and cultivated at 30°C until the mid-logarithmic phase (40 h). For ER (endoplasmic reticulum) stress treatment, the mycelia together with the cellophane were transferred to the surface of 2-cm thick filter paper soaked with 20 ml CD medium that contained 20 mmol/l dithiothreitol (DTT) for induction treatment. After 3 h, mycelia with or without DTT treatment were collected for the preparation of RNA-Seq samples.

For LC, 6 × 10<sup>5</sup> spores/ml medium were inoculated into 30 ml CD medium (total 1.8 × 10<sup>7</sup> spores) in a 100-ml flask and cultivated at 30°C at 200 r.p.m. until the mid-logarithmic phase (40 h). For induction treatment, DTT was added to the culture at a final concentration

of 20 mmol/l. Cultures without DTT treatment had an equivalent volume of distilled water added. After 2 h, mycelia were collected for RNA-Seq.

### Preparation of cDNA library for RNA-Seq

For Illumina sequencing, the total RNA of every sample was extracted using RNAiso™ Plus (TaKaRa) and then treated with RNase-free DNase I (TaKaRa) for 45 min according to the manufacturer's protocols. The integrity of total RNA was checked using Agilent Technologies 2100 Bioanalyzer, and all four samples had an RNA Integrity Number (RIN) value greater than eight.

The cDNA libraries were prepared according to the manufacturer's instructions (Illumina). The poly(A) containing mRNA molecules were purified using Sera-mag Magnetic Oligo(dT) Beads (Illumina) from 20 µg total RNA of each sample. Ten milli molar Tris-HCl was used to elute the mRNA from the magnetic beads. To avoid priming bias when synthesizing cDNA, the mRNA was first fragmented before cDNA synthesis. The mRNA was fragmented into small pieces using divalent cations at elevated temperature. The cleaved mRNA fragments were converted to double-stranded cDNA using SuperScript II, RNaseH and DNA Pol I, primed by random primers. The resulting cDNA was purified using the QIAquick PCR Purification Kit (Qiagen). Then, cDNA was subjected to end-repair and phosphorylation using T4 DNA polymerase, Klenow DNA polymerase and T4 PNK, and subsequent purification was performed using QIAquick PCR Purification Kit (Qiagen). These repaired cDNA fragments were 3'-adenylated using Klenow Exo- (Illumina) and purified using MinElute PCR Purification Kit (Qiagen), producing cDNA fragments with a single 'A' base overhung at their 3'-ends for subsequent adapter-ligation. Illumina PE adapters were ligated to the ends of these 3'-adenylated cDNA fragments, followed by purification using MinElute PCR Purification Kit (Qiagen). To select a size range of templates for downstream enrichment, the products of the ligation reaction were purified on a 2% TAE-agarose gel (Certified Low-Range Ultra Agarose, Biorad). A range of cDNA fragments (200 ± 25 bp) was excised from the gel and extracted using QIAquick Gel Extraction Kit (Qiagen). Fifteen rounds of PCR amplification were performed to enrich the adapter-modified cDNA library using primers complementary to the ends of the adapters [PCR Primer PE 1.0 and PCR Primer PE 2.0 (Illumina)]. The PCR products of size 200 ± 25 bp were purified using QIAquick Gel Extraction Kit (Qiagen) except that QIAquick spin columns were substituted with MinElute spin columns (Qiagen). Finally, after quantification on an Agilent Technologies 2100 Bioanalyzer using the Agilent DNA 1000 chip kit, the cDNA library products were sequenced using the 1G Illumina Genome Analyzer. Two biological replicates of every sample were analyzed. The raw Illumina sequencing data are deposited in GEO (<http://www.ncbi.nlm.nih.gov/geo/>) at NCBI with accession number GSE18851.

## Mapping reads to reference genome and annotated genes

The *A. oryzae* RIB40 genome and gene information were downloaded from DOGAN (<http://www.bio.nite.go.jp/dogan/Top>). After removing reads containing sequencing adapters and reads of low quality (reads containing Ns > 5), the remaining reads were aligned to the *A. oryzae* RIB40 genome using SOAP (17), allowing up to 2 base mismatches. Reads that failed to be mapped were progressively trimmed off 1 base at a time from the 3'-end and mapped to the genome again until a match was found (unless the read had been trimmed < 27 bases). For paired-end reads, the insert between paired reads was set as 1 base ~5 kilobases, allowing them to span introns of varied sizes in the genome. The same strategy was performed for aligning paired-end reads to the non-redundant gene except that the insert was changed to 1 base ~1 kilobase.

## Normalized expression level of gene by RNA-Seq

The expression level of gene by RNA-Seq was normalized by the number of reads per kilobase of exon region per million mapped reads (RPKM) (11). The cutoff value for determining gene transcriptional activity was determined based on a 95% confidence interval for all RPKM values of each gene.

## Discovery of novel transcripts

A novel transcriptional active region (TAR) was determined in the intergenic regions by contiguous expression with each base supported by at least two uniquely mapped reads and length > 35 bp. Supported by the paired-end information, a novel transcript unit (TU) was constructed by the connectivity between novel TARs which were joined by at least one paired reads. Novel TUs with length < 150 bp or average expression of < 2 reads per base were excluded from further analysis. Finally, 1166 novel TUs were detected in the *A. oryzae* genome and were then screened by Augustus (18) using parameters trained from *Saccharomyces cerevisiae* known genes to predict putative protein-coding genes.

## UTR analysis by RNA-Seq data

To define the UTRs of annotated genes, we searched for a break in the transcribed region around genes, denoted by positions with reads count of zero, starting from either end of genes. Genes whose ends overlap with other genes were excluded from analysis. If no break was found, the UTR was discarded. The 5'- or 3'-UTRs with length > 1000 nt were discarded.

The uORF and uATG were screened in 5'-UTRs using computational methods. Any uORF of length > 150 bp or with the distance between the uORF and the gene start codon > 500 bp was excluded from the results. The length of uATG should be < 150 bp and uATGs must be in-frame to the gene's coding region and there should be no stop codons in the sequence of uATGs.

## Gene ontology analysis

We obtained the gene ontology (GO) terms of each *A. oryzae* gene by the software Blast2GO (version 2.3.4) (19) using the default parameters. Blast2GO was also used for GO functional enrichment analysis of certain genes, by performing Fisher's exact test with robust false discovery rate (FDR) correction to obtain an adjusted *P*-value between certain test gene groups and the whole annotation.

## Analysis of AS using RNA-Seq data

To identify all potential splice sites, we searched for the putative donor site ('GT' on the forward strand or 'AC' on the reverse strand) and the acceptor site ('AG' on the forward strand or 'CT' on the reverse strand) inside the intronic regions and set that a novel splice site was required to be supported by at least four uniquely mapped reads. We then enumerated all possible pairs of donor sites and acceptor sites and detected 108 000 potential splice junctions.

To determine the number of reads supporting each splice junction, all reads that could not match to the genome and that matched to the genome by trimming off several terminal bases were retrieved and aligned against the junction sequence with a tolerance of 2 bp mismatches. A nominated junction site was required to be supported by at least two unambiguous mapped reads with different match positions within the splice junction region and also with a minimum match of 5 bases on each side of the junction (such reads could be called 'trans-reads'). In total, 11 310 junction sites were identified.

To identify AS events in *A. oryzae* RIB40, the junction sequences identified were used to analyze seven types of AS including skipped exons (SE), retained introns (RI), alternative 5'-splice sites (A5SS), alternative 3'-splice sites (A3SS), mutually exclusive exons (MXE), alternative first exons (AFE) and alternative last exons (ALE) according to the method of E.T. Wang (14).

## Analysis of differential gene expression of *A. oryzae* grown under different culture conditions

Two biological replicates of every sample were used for differential gene expression analysis under different culture conditions. A 0.81–0.88 Pearson correlation coefficient was obtained for biological replicates. We applied R package DEGseq to identify differentially expressed genes, with Random Sampling model based on the reads count of each gene under different culture conditions (20). GO functional enrichment analysis was carried out using software Blast2GO mentioned above. KEGG pathway analyses of differentially expressed genes were performed using the Cytoscape software version 2.6.2 (<http://www.cytoscape.org/>) with the ClueGO plugin (<http://www.ici.upmc.fr/cluego/cluegoDownload.shtml>) (21).

## Validation of RNA-Seq analysis by RT-PCR and real-time quantitative PCR

PCR was performed to validate the RNA-Seq analysis including AS, novel transcripts, new candidate exons

and the expression status of the AF biosynthesis regulator genes (*aflR* and *aflJ*). The *A. oryzae* RIB40 samples for RT-PCR and qPCR (real-time quantitative PCR) were obtained just as that for Illumina sequencing. For AS, primers were designed inside exons or introns in order to detect different AS isoforms (Supplementary Table S1). For novel transcripts, new candidate exons, *aflR* and *aflJ*, primers were designed in exon regions (Supplementary Table S1). RNA was extracted using RNAiso<sup>TM</sup> Plus (TaKaRa) and then treated with RNase-free DNase I (TaKaRa) for 45 min according to the manufacturer's protocols. First-strand cDNA synthesis and subsequent PCR were performed using PrimeScript<sup>TM</sup> RT-PCR Kit (TaKaRa) and Fast SYBR<sup>®</sup> Green Master Mix (ABI) following its instructions. There was no PCR product in the negative control without reverse transcriptase.

## RESULTS AND DISCUSSION

### Summary of RNA-Seq data sets

To obtain a global view of the *A. oryzae* transcriptome at single-nucleotide resolution, poly(A)-enriched mRNA from SC and LC conditions, as well as the corresponding ER stress culture conditions (treatment with DTT), were subjected to high-throughput Illumina sequencing. We totally obtained 122 579 836 reads of an average length of 44 bp, representing ~145 *A. oryzae* genome lengths (Supplementary Table S2). For each of the four *A. oryzae* samples, >90% of all reads were uniquely mapped to the genome over 100% of their sequence with a tolerance of 2 bp mismatches. These unique reads covered a large part (~28.5 million bases) of *A. oryzae* genome, and resulted in an average sequence depth of 51–62 per base for each sample. Overall, 93.52% of the reads mapped to unique genomic regions, with 52.77% of the reads mapped to known exons and 40.75% located in intergenic and intronic regions (Figure 1A). The remaining reads were either of poor quality or multimapped to the genome (~6.48%).

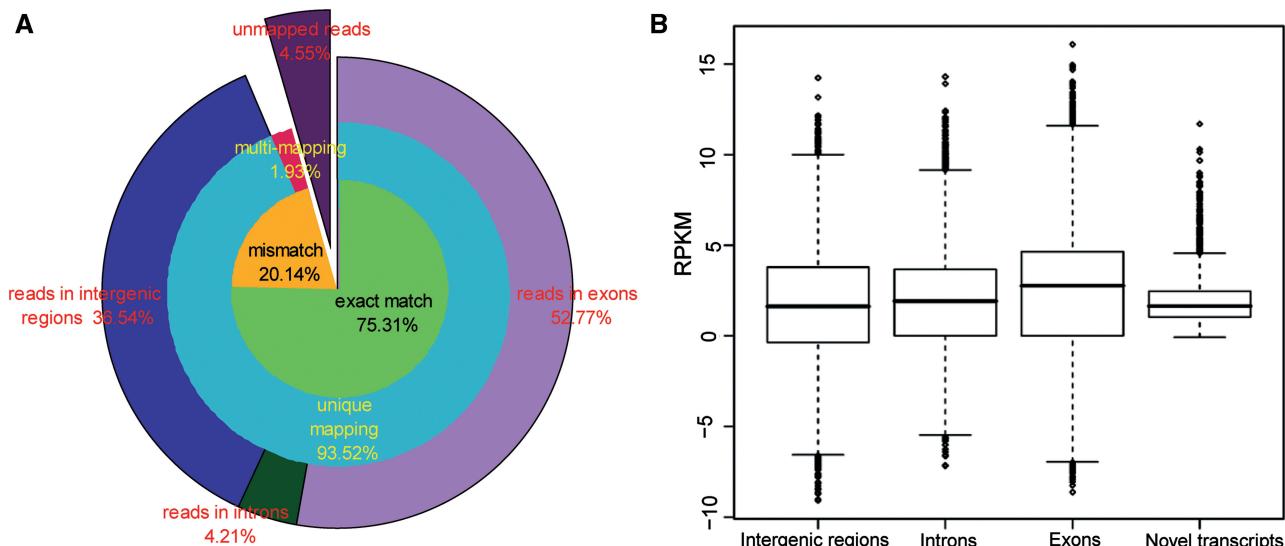
### Widespread transcription of the *A. oryzae* genome

RNA-Seq analysis revealed extensive expression of the whole *A. oryzae* genome (Supplementary Figure S1). Of the genome, 76.66% was expressed as RNA-Seq reads (Supplementary Table S2). We quantified the overall transcriptional activity of the genes in our data by calculating the number of reads per kilobase of exon region per million mapped reads (RPKM) (11) and found that 11 263 of the 12 074 protein-coding genes in the *A. oryzae* genome database (Dogan database) (2) showed expression (within a 95% confidence interval) in our four *A. oryzae* samples (Supplementary Table S3). In contrast, only 2953 (3) and 3320 (4) genes in the *A. oryzae* genome were validated by EST contigs. So, our RNA-Seq data was more sensitive and showed a more comprehensive landscape of the *A. oryzae* transcriptome. As expected, the expression level of the exons was higher than the intronic or intergenic regions (Figure 1B).

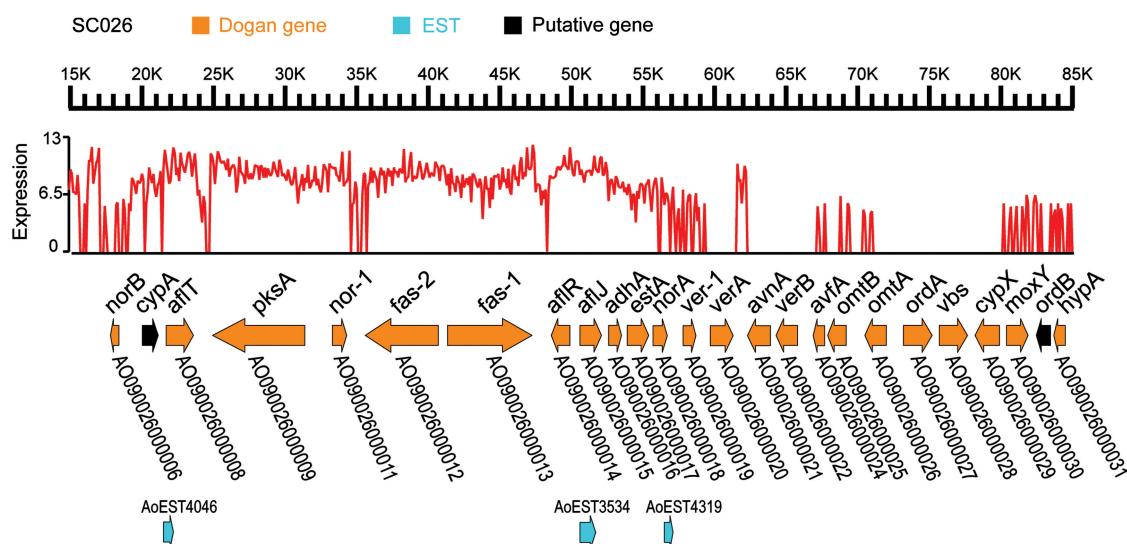
Furthermore, we also specifically focused on the transcriptional activity of the secondary metabolite genes in different samples. The *A. oryzae* RIB40 strain, used in the genome sequencing project and our study, belongs to the group I type of *A. oryzae* and contains the 25 genes of the entire aflatoxin (AF) biosynthesis pathway that are located in a 70-kb genomic region as a cluster. The transcriptional activity of the AF biosynthesis gene cluster was not detected by EST analysis except for *aflJ*, *aflT* and *norA* (3). With large numbers of reads representing a deep sampling of the *A. oryzae* transcriptome, we detected significant transcriptional activity for most of the AF biosynthesis genes (Figure 2). For example, *aflR*, the transcription factor gene required for transcriptional activation of most of the structural genes in the cluster, showed expression with a score of 1.54 RPKM in SC and 2.70 RPKM in LC, and upregulated to 4.06 RPKM and 4.09 RPKM in the corresponding cultures under ER stress (Supplementary Table S3). This is the first time the transcriptional activity of the *aflR* gene in *A. oryzae* has been detected. Moreover, expressions of *aflT*, *pksA*, *nor-1*, *fas-2*, *fas-1*, *aflJ*, *adhA*, *estA*, *norA*, *ver-1*, *omtB* and *hypA* were also detected in some of the four samples (Figure 2 and Supplementary Table S3). Expression levels of the two transcription factors, *aflR* and *aflJ*, were further confirmed by RT-PCR (Supplementary Figure S2). However, thus far, none of the known *A. oryzae* strains has produced AF under any culture conditions, and the reason for the lack of AF biosynthesis remains to be determined.

### Improvement of *A. oryzae* genome annotation by RNA-Seq

A total of 12 074 genes, encoding proteins longer than 100 amino acids, have already been predicted using gene-finding software tools (2). Surprisingly, ~40.75% of reads fell within the intergenic and intronic regions. These presumably were derived from as yet unrecognized transcripts and non-coding RNAs. Alternatively, they may correspond to novel exons of known genes. Extensive reads mapping and clustering revealed 1166 novel transcripts with significant expression levels above the surrounding intergenic region, of which 38.34% were longer than 500 bp (Supplementary Table S4 and Figure 3A and B). Based on the presence of ORF within the transcripts, 700 (60.03%) of the novel transcripts were likely to be protein-coding genes. To validate the novel transcripts determined by RNA-Seq, we randomly selected five of the novel transcripts and confirmed their transcriptional activity in all five cases by RT-PCR (Supplementary Figure S2). In addition to novel transcripts, 800 new candidate exons were identified in 513 annotated genes (Supplementary Table S5) by clustering successfully mapped reads, and these new exons were supported by paired-end reads. As illustrated in Figure 3C, a new candidate exon of 228 bp with average sequence depth of 11.69 was detected in the annotated gene AO090003000598, and this new candidate exon was further confirmed by RT-PCR (Supplementary Figure S2).



**Figure 1.** Genome-wide assessment of RNA-Seq reads. (A) Circle plot of the RNA-Seq reads mapping to the *A. oryzae* genome. The innermost and middle circles depict the percentage of RNA-Seq reads with exact/mis-match and unique/multi-mapping, respectively. The outermost circle represents the distribution of reads in exons, introns and intergenic regions according to the annotated gene set. The sector region penetrating all circles represents reads unmapped to the genome. (B) Box- and whisker plots of log<sub>2</sub>-transformed RPKM for intergenic regions, introns, exons and novel transcripts. The three horizontal lines in the boxes show the first quantile, median and third quantile, respectively. The other two horizontal lines show the inner boundaries and diamonds represent data outside of the inner boundaries.

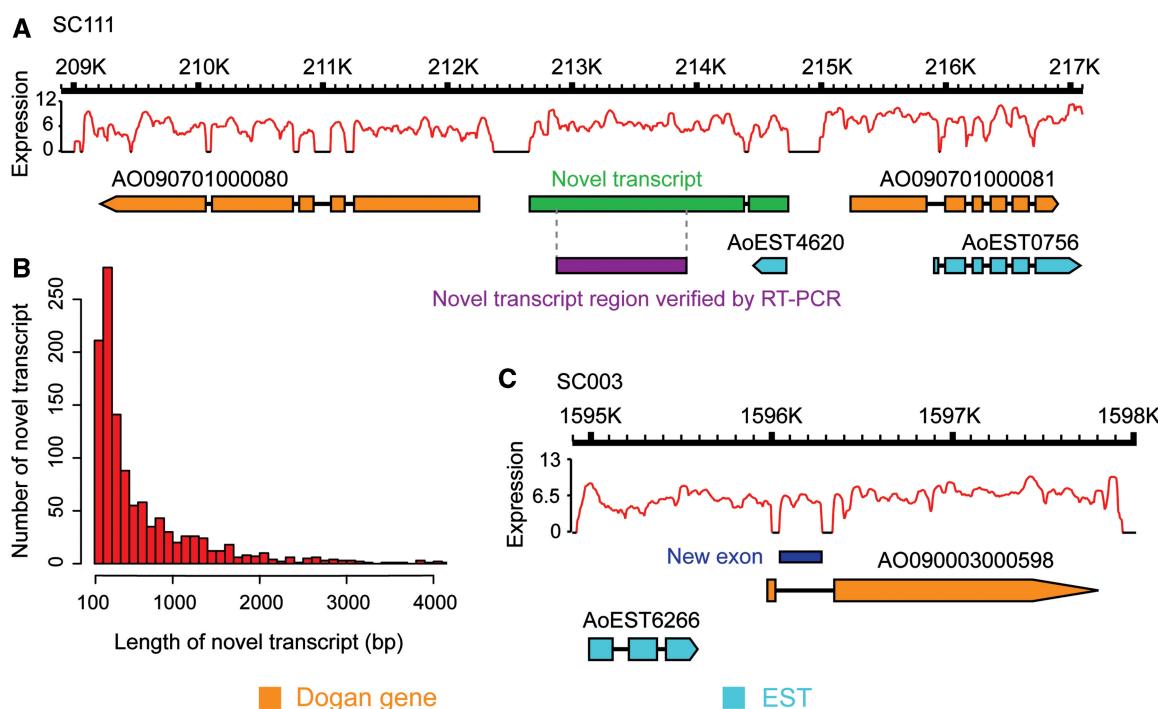


**Figure 2.** The transcriptional activity of the *A. oryzae* AF biosynthesis gene cluster detected by RNA-Seq. The red curve indicates the expression level (log<sub>2</sub> of reads count). Putative gene represents the hypothetical genes (*cypA* and *ordB*), and their gene structures were predicted using the software Augustus. Some elements of the AF biosynthesis gene cluster, such as *aftsT*, *aftsJ* and *noraA*, were also confirmed by EST contigs.

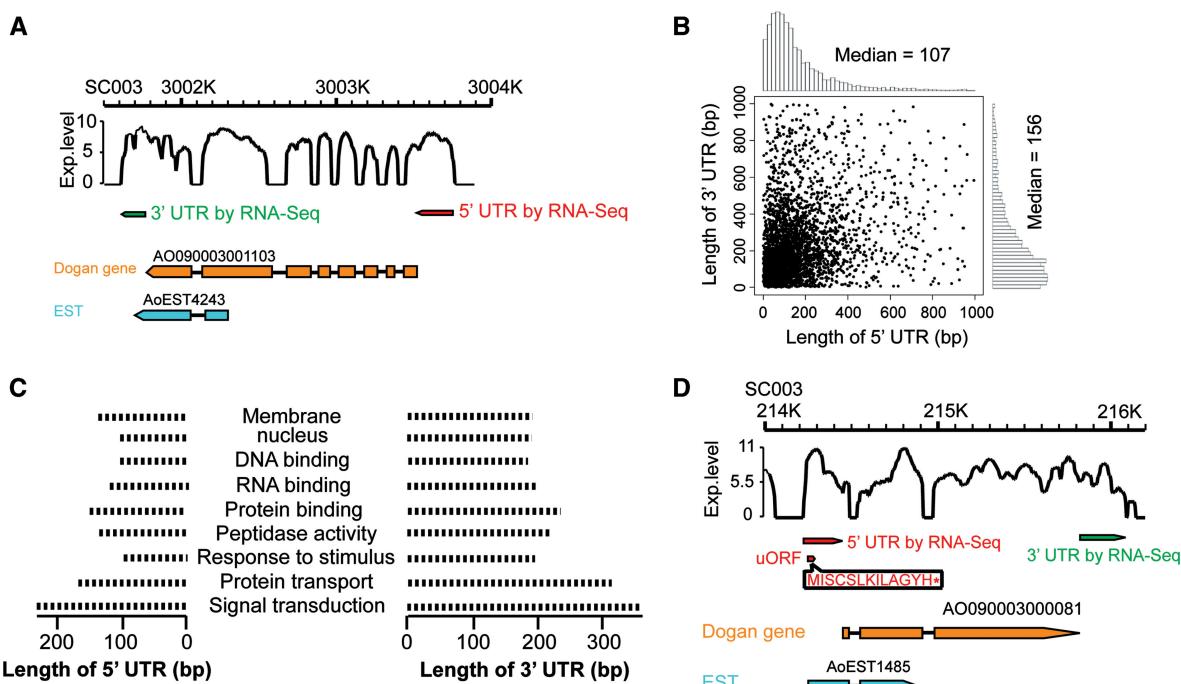
We also globally mapped the 5'- and 3'-boundaries of *A. oryzae* genes by searching for a sharp reduction of RNA-Seq reads signals at both ends of annotated genes. Genes whose 5'- or 3'-boundaries overlap with other genes were excluded from the analysis. The results defined or extended 5'-boundary regions for 4198 transcribed genes and 3'-boundary regions for 4357 transcribed genes in *A. oryzae* (Figure 4A and Supplementary Table S6). The median lengths of 5'- and 3'-UTR determined by RNA-Seq were 107 and 156 bp, respectively, similar to those of *S. pombe* (8) (Figure 4B). Furthermore, we compared the UTR length distributions for various GO

functional categories (Figure 4C). Obviously, transcripts related to signal transduction had remarkably longer 5'- and 3'-UTRs, which were the least stable, facilitating quick RNA turnover to amplify the cellular signal. In contrast, genes in GO categories of the nucleus, RNA binding and response to stimulus had significantly shorter UTRs. These results were similar to those in *S. pombe* (8), indicating that UTR length distributions may show some conservation in fungi.

Given that the identification and refinement of a large number of UTR boundaries could be useful for interpreting regulatory factors that mediate gene expression, we



**Figure 3.** Novel transcripts and new exons. (A) An example of transcripts previously unidentified in SC111. The green bar between the two annotated *A. oryzae* genes represents the novel transcript that was confirmed by RT-PCR (violet bar). (B) The length distribution of novel transcripts. (C) A new exon (blue bar) in the annotated intron of AO090003000598 was identified by RNA-Seq data.



**Figure 4.** UTRs determined by RNA-Seq. (A) 5'-UTR (red arrow) and 3'-UTR (green arrow) of AO090003001103 were defined by RNA-Seq. (B) Scatterplot and histograms showing the length distributions of 5'- and 3'-UTRs based on RNA-Seq data. (C) GO categories with significantly longer or shorter UTRs. (D) An example of uORFs (boxed and in red) in 5'-UTR of AO090003000081.

scanned the ORF upstream of the start codon (uORF), an important regulatory factor for gene expression in 5'-UTR (22–24). Our RNA-Seq data predicted uORFs upstream of the start codon for 1345 *A. oryzae* genes (11.14%)

(Figure 4D and Supplementary Table S7), about double the percentage in yeast genes (7), indicating that this transcriptional regulation mechanism might be important in *A. oryzae*. GO functional enrichment analysis revealed

that genes associated with histone deacetylase activity, protein deacetylase activity, peptidyl-amino acid modification, peptidyl-histidine modification, actin cytoskeleton, response regulator activity and signal transduction [FDR-adjusted  $P$ -value < 0.05] were significantly enriched in uORFs. In addition to uORFs, uATG, i.e. 5' and in-frame to the annotated ATG initiation codon, were detected in 272 genes (Supplementary Table S8), suggesting that these proteins are potentially longer than the annotated.

### AS in *A. oryzae*

AS is crucial for proteomic diversity and functional complexity in higher organisms (25–28). To assess the genome-wide extent of AS events in *A. oryzae*, we performed computational analysis to determine the known and putative splicing junctions and then identified sequence reads mapping to these regions using stringent criteria (see ‘Materials and methods’ section). We examined seven common types of AS events, including A5SS, A3SS, SE, RI, MXE, AFE and ALE (14), and found that 1032 *A. oryzae* genes underwent AS with 1375 AS events (Supplementary Table S9, Figure 5A and C). All types of AS events were detected except MXE. For our data, 76.98% of all *A. oryzae* genes contained two or more exons, and of these 11.10% produced two or more AS isoforms. Therefore, of the total number of *A. oryzae* genes, 8.55% were estimated to undergo AS.

Although a large number of genes undergoing AS was detected in *A. oryzae* in contrast to yeast, in which no evidence for the existence of AS has been identified by RNA-Seq (8), the number was far lower than observed in mammals. For example, ~86% of human genes were estimated to produce more than one distinct population of mRNA isoforms (12,14). Similarly, only 134 (1.6%) and 224 (3.6%) genes with AS events were discovered in *Magnaporthe grisea* and *Ustilago maydis*, respectively, by large-scale EST sequencing (27,29). In addition, as a part of genome annotation, AS forms were predicted for 277 genes (4.2% of the transcriptome) in *Cryptococcus neoformans* (26). *Aspergillus oryzae* genes (8.55%) with AS variants were detected through a very deep sampling of the transcriptome, suggesting that AS might play a major biological role among fungi albeit less than was seen in humans.

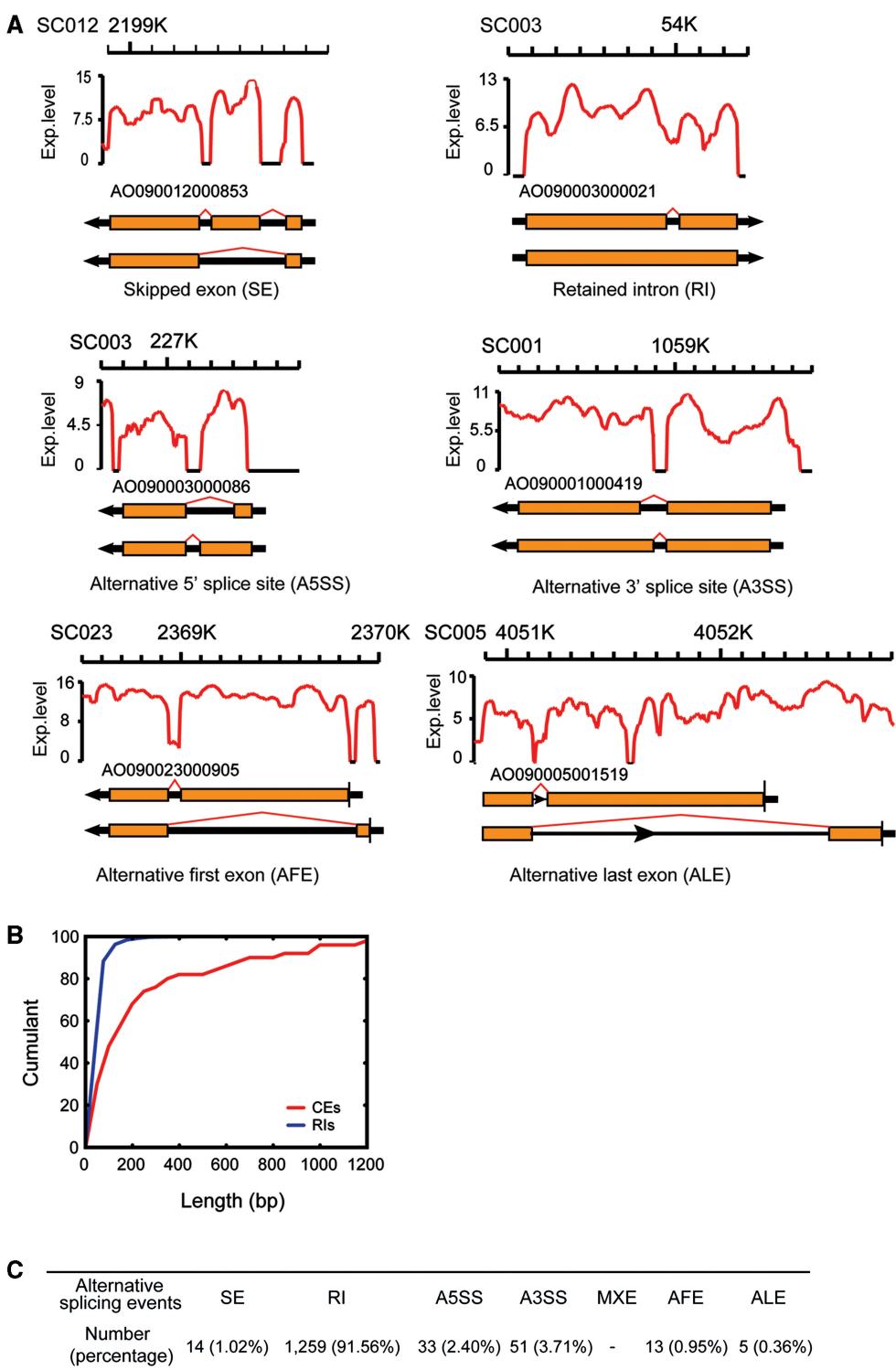
Like other fungi, such as *U. maydis* and *M. grisea*, the RI in *A. oryzae* is the predominant form of AS, accounting for 91.56% of all AS isoforms, while RI is a rare AS event in mammals (14,28). Contamination of the RNA sample with genomic DNA (gDNA) or the presence of unspliced transcripts might give rise to spurious ‘retained’ introns. We eliminated gDNA by selective RNA isolation and DNase I treatment in our sample preparation (see ‘Materials and methods’ section). To assess these measures, we randomly selected one case of RI events and performed PCR experiments. There were no PCR products for amplifying the genomic exon region using the same RNA samples for this study as templates (Supplementary Figure S2), suggesting that there was no remnant gDNA in the RNA samples. We also carried out

RT-PCR screening and qPCR screening for the same case, and it illustrated that the annotated exon and the RI had similar transcription levels while the spliced intron had a relatively low transcription level (Supplementary Figure S3). Therefore, the RI event was confirmed although there was a trace of unspliced mRNA in the RNA samples for RNA-Seq library construction. Overall, these results demonstrated that the RIs detected in our study represented bona fide AS events.

We found that cassette exons (CE, including SE, AFE, ALE and MXE) accounted for only a very small portion of the total AS events in *A. oryzae* and the CE fraction [CE/(RI+CE)] in *A. oryzae* was only 0.025, similar to many other fungi such as *A. flavus* (0.02) and *A. nidulans* (0.01) (28). In eukaryotes, it was proposed that there were two mechanisms of splice site recognition—introns definition (ID) and exon definition (ED)—in which the ID mechanism inconsistently identifies intron-exon boundaries to produce RIs and the ED mechanism results in CE by variable recognition of splice junctions (28,30–32). McGuire *et al.* (28) inferred that species with a high ratio of RIs to CEs (such as fungi and protists) primarily recognized splice junctions by the ID mechanism and showed no substantial length difference between RIs and constitutive introns, and that their CEs were shorter than constitutive exons. The length distributions of RIs and CEs observed in our data (Supplementary Figure 5B) were consistent with the above hypothesis: the median length of RIs (63 bp) is similar with that of *A. oryzae* constitutive introns (61 bp), and the median length of CEs (127 bp) is shorter than that of *A. oryzae* constitutive exons (226 bp). Therefore, *A. oryzae* might perform splice site recognition predominantly by the ID mechanism.

### Functional analysis of differential gene expression based on RNA-Seq data

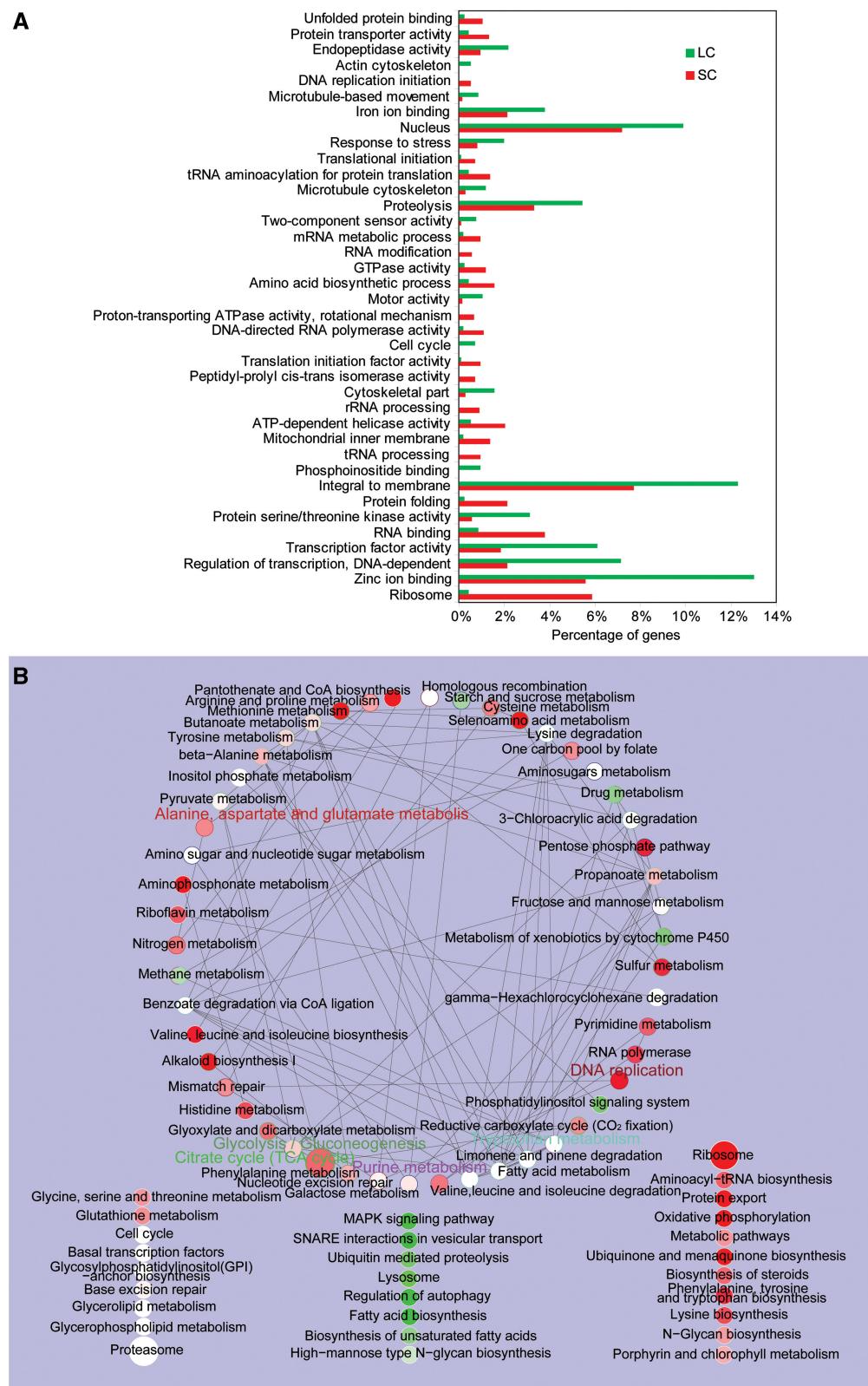
The molecular mechanisms resulting in higher levels of protein productivity in SC than in LC remain poorly understood, although a number of genes and metabolic pathways differently transcribed under various culture conditions have been illuminated (3,33–35). Here, to resolve this issue, we globally detected the differential gene expression of *A. oryzae* grown under SC and LC conditions based on RNA-Seq, an excellent quantitative method due to its very low background and deep sampling of the transcriptome. We found 4628 genes were differentially expressed between LC and SC, including 2355 and 2273 genes up- and downregulated on SC respectively ( $P$ -value < 0.001) (Supplementary Table S10). GO analysis revealed that genes upregulated on SC were mainly involved in RNA binding, ribosomes, ATP-dependent helicase activity, protein translation, protein folding, protein transporter activity and mitochondrial inner membrane (FDR < 0.05) (Figure 6A and Supplementary Table S10). In addition, the KEGG metabolic pathway analysis indicated that the genes upregulated on SC were specifically located in the pathways of ribosome, DNA replication, oxidative phosphorylation and the TCA cycle while genes



**Figure 5.** AS events in the *A. oryzae* transcriptome. (A) The schematic depicts the six primary types of AS events in the *A. oryzae* transcriptome. The red curve indicates the expression levels (log<sub>2</sub> of reads count). The orange bar denotes the exons and the red knuckle line highlights the linkage between two exons that were supported by at least two distinct junction reads. The vertical line shows the transcript start or end sites. MXE were not detected in our data. (B) A length comparison between RIs and CEs. Vertical coordinates represents the cumulative percentage of RIs or CEs of different lengths. (C) The numbers of each type of AS event in *A. oryzae* detected by RNA-Seq.

upregulated on LC were mainly associated with MAPK signaling pathway, SNARE interactions in vesicular transport, ubiquitin mediated proteolysis and lysosome (Figure 6B and Supplementary Table S10). Together,

these analyses indicated that *A. oryzae* drastically altered the manner of gene expression in response to different culture conditions and that the capacities for protein translation/modification and energy production were



**Figure 6.** Functional analysis of differentially expressed genes between SC and LC conditions. **(A)** GO functional enrichment analysis of differentially expressed genes in SC and LC. Fisher's exact tests with Benjamini and Hochberg FDR ( $<0.05$ ) were performed to get an adjusted  $P$ -value between test gene groups and the whole gene annotation. Crossbands in red represent upregulated genes in SC compared with LC, and crossbands in green represent upregulated genes in LC compared with SC. **(B)** KEGG pathway analysis of differentially expressed genes in SC and LC. Functional group networks were represented by nodes linked with each other based on their kappa score level ( $>0.3$ ). The node size represents the enrichment significance of the term and the most significant term per group is shown as the label. Terms with upregulated or downregulated genes under SC compared with LC conditions are shown in red or green, respectively, and the color gradient represents the extent of regulation. Terms with no significant difference between SC and LC are shown in white.

much more powerful in *A. oryzae* grown on SC compared to LC.

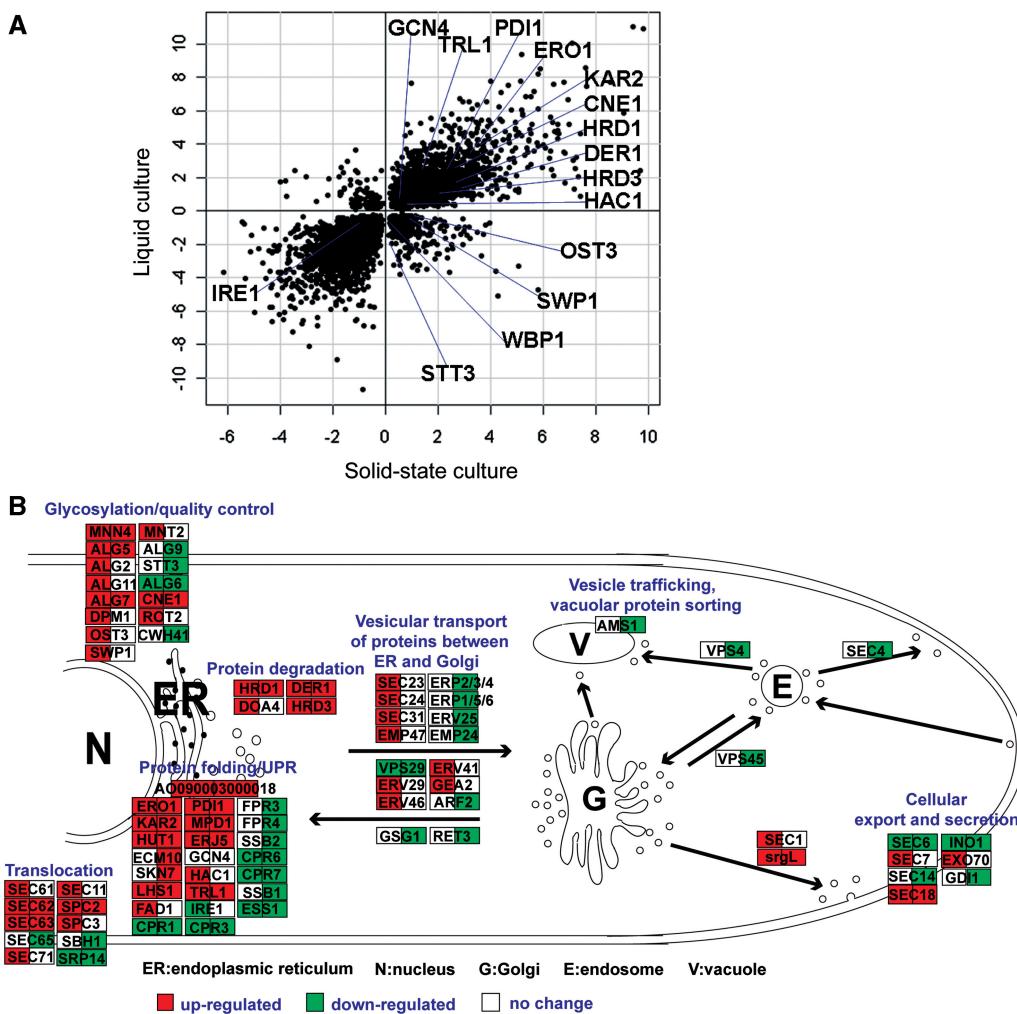
*Aspergillus oryzae* grown on SC resembles its natural growth status compared to LC, in terms of the diffusion of nutrients and gases, water activity and the continuity of medium distribution (3,36). For example, *A. oryzae* shows particular characteristics such as hyphal differentiation with asexual development and high-level production of enzymes in SC, which are absent in LC (36). Previously, many genes encoding extra- and intracellular enzymes and transport proteins, which are differentially expressed between SC and LC, have been identified by subtractive cloning of cDNA (36), cDNA microarrays (34), ESTs (3) and proteomic analysis (35). And it has been reported that the expression levels of genes associated with energy catabolism in SC were lower than those in LC, supporting a release of catabolite repression and consequently leading to high-level expression of the hydrolytic enzymes (34,35). However, in contrast, our results showed that *A. oryzae* genes associated with energy catabolism (such as the TCA cycle and oxidative phosphorylation) were upregulated on SC when compared with LC. For example, the key regulator genes of the TCA cycle, citrate synthase (AO090102000627), isocitrate dehydrogenase (AO090003000008, AO090005001404 and AO090012000629) and succinate dehydrogenase (AO090010000505, AO090020000415 and AO090020000596), were upregulated on SC (Supplementary Table S10), as well as some important genes of oxidative phosphorylation (AO090001000657, AO090005000617, AO090005000749, AO090005000888, AO090010000482 and AO090026000372). Our results indicated that the conversion of pyruvate to ethanol was repressed because *adhA* (AO090026000016, responsible for reducing acetaldehyde to ethanol) was expressed at a very low level and downregulated on SC (Supplementary Table S3). And it seemed that the conversion of pyruvate to acetate was also repressed because *alda* (AO090023000467, responsible for oxidizing acetaldehyde to acetate) was downregulated on SC (Supplementary Table S10). So, the flux of pyruvate in *A. oryzae* was targeted to the TCA cycle and oxidative phosphorylation on SC, and this may lead to the upregulation of genes associated with the TCA cycle and oxidative phosphorylation, consequently providing enough energy for the high-level expression of the hydrolytic enzymes. In addition, the metabolic fate of pyruvate suggests that *A. oryzae* utilizes glucose mainly via aerobic metabolism on SC.

It has been reported that *A. oryzae* ribosomal genes were highly expressed under glucose-depleted (AN) liquid culture conditions (34). Our results indicated that many *A. oryzae* genes associated with ribosome (90% of the total number of ribosome genes) were upregulated on SC (Supplementary Table S10). This might be because the low water activity in SC markedly reduced the diffusion rate of the glucose, forming glucose-limiting conditions similar to the glucose-depleted (AN) liquid culture conditions. In our previous study, we found that the expression levels of several protein folding related genes (such as *hacA*, *bipA*, *pdi* and *ppi*) were higher in SC than in LC (37).

Our RNA-Seq data also showed that 29 protein folding related genes were upregulated on SC compared to LC (Supplementary Table S10), suggesting that the protein folding is much more efficient under SC than under LC in *A. oryzae*. In addition, genes associated with ATP-dependent helicase activity, which are important in DNA replication and cell division, were highly transcribed in *A. oryzae* under SC (Supplementary Table S10). Pathway analysis also indicated that DNA replication of *A. oryzae* under SC conditions was much more active than under LC conditions (Figure 6B). This may be due to the hyphal differentiation with asexual development and the fast growth of *A. oryzae* on SC.

Since the protein secretory behavior of *A. oryzae* on SC has become a model for many filamentous fungi, we globally monitored the transcriptome of *A. oryzae* under ER stress on both SC and LC to further investigate the protein secretory pathway. ER stress could result in the accumulation of unfolded proteins in the ER to activate the unfolded protein response (UPR), in which the unconventional splicing of *HAC1* mRNA by kinase/RNase Ire1p and tRNA ligase Rlg1p produces Hac1p (a bZIP-type transcription factor) and Hac1p further activates the transcription of UPR target genes in the protein secretory pathway (38). We induced UPR of *A. oryzae* grown on SC or LC by treating *A. oryzae* cells with the strong reducing agent DTT, which prevents disulfide bond formation and disrupts protein folding, resulting in the accumulation of unfolded proteins in ER. Control experiments were *A. oryzae* grown on SC or LC without DTT treatment. Our results depicted a comprehensive transcription profile of ER-stress target genes that participate in the process of UPR, ER-associated degradation (ERAD), protein folding, glycosylation and quality control of *N*-glycosylated folded proteins (Figure 7A and Supplementary Table S11). As anticipated, UPR target genes were induced under both stress conditions. However, genes for protein glycosylation had the opposite transcriptional regulation under ER stress between SC and LC. The regulation model of the protein secretory pathway under ER stress on SC and LC is illustrated in Figure 7B. The results showed that much more UPR target genes in the protein secretory pathway were upregulated under ER stress on SC than on LC and far fewer UPR target genes were downregulated under ER stress on SC than LC, suggesting that the regulation capacity of UPR in protein production and secretion of *A. oryzae* on SC was much more efficient than on LC. This study constitutes the first global analysis of ER stress in *A. oryzae*, especially under SC. Previous analyses of ER stress in *A. oryzae* had not reached such a high resolution and there has been no comprehensive interpretation regarding SC conditions (37,39). As the ER stress response and protein secretion of *A. oryzae* on SC are known to resemble the natural status, our global analysis of ER stress in *A. oryzae* on SC is significant.

Our previous study showed that the expression levels of the UPR target genes (*hacA*, *bipA*, *pdi* and *ppi*) were higher on SC than on LC under control conditions (37). Herein, we confirmed that many more UPR target genes were



**Figure 7.** Differentially expressed genes between DTT stress conditions and non-stress (control) conditions for both SC and LC. (A) Scatterplot depicts the ratio (log<sub>2</sub>-transformed) of the gene expression level under DTT stress conditions to the gene expression level under non-stress conditions for both SC and LC. Some of the differentially expressed genes are labeled with the names of their homologs in *S. cerevisiae*, including genes related to UPR (HAC1, IRE1, TRL1 and GCN4), ERAD (DER1, HRD1 and HRD3), protein folding (PDI1, ERO1 and KAR2) as well as glycosylation and related quality control (OST3, WBP1, SWP1, STT3 and CNE1). (B) The regulation model of genes in the *A. oryzae* protein secretory pathway under DTT stress conditions compared with non-stress conditions for both SC and LC. Boxes represent differentially expressed genes ( $P$ -value  $<0.001$  and  $>2$ -fold change) between DTT stress conditions and non-stress conditions for SC or LC, designated by the names of their homologs in *S. cerevisiae*. Each box is separated into two parts, representing SC (left part) and LC (right part) respectively. Red, up-regulated; green, down-regulated; white, no significant change. N, nucleus; ER, endoplasmic reticulum; E, endosome; V, vacuole.

upregulated on SC than on LC under ER stress conditions (Figure 7B). Taken together, we infer that the high-level expression of UPR target genes was due to the SC conditions and the regulatory ability of UPR was much more efficient on SC than on LC in *A. oryzae* under either induction or non-induction conditions (Figure 7A and B; Supplementary Tables S10 and S11). Therefore, the UPR in *A. oryzae* is a culture-condition dependent biological process. In addition, there was a dissimilarity in glycosylated modification during the production of Pre-S2 antigen in *A. oryzae* between submerged culture conditions and SC conditions (40), and our results also indicated that the capacity for protein glycosylation of *A. oryzae* under SC was greater than under LC, suggesting that the protein glycosylation modification in *A. oryzae* also depends on the culture conditions.

It has been reported that the control of UPR was far from a simple binary switch based on the HAC1 signal mechanism and exhibited complex fine-tuning via a signaling network (41,42). For example, GCN4, which regulates amino acid biosynthesis in *S. cerevisiae*, is an important regulator that activates a large subset of UPR target genes and is activated at the translation level (41,42), which was also confirmed in our study under ER stress conditions (GCN4 homologs gene AO090009000459 in *A. oryzae*, Figure 7A and Supplementary Table S11) and in *Trichoderma reesei* under ER stress conditions (GCN4 homologs gene CPC1) (43). Repression under secretion stress (RESS) is another important regulation mechanism under ER stress conditions. It has been reported that RESS repressed the transcription of secretory protein genes under ER stress conditions in *T. reesei* (44),

*Aspergillus niger* (45) and *A. nidulans* (46). Our results also confirmed the existence of the RESS mechanism in *A. oryzae* because the transcription of major secretory protein genes [Taka-amylase A (AO090023000944, AO090120000196), oryzin (AO090003001036), pepsinogen (AO090003000693), alkaline phosphatase (AO090012000725) and maltase glucoamylase (AO090038000471)] were clearly repressed under ER stress (Supplementary Tables S3 and S12). However, it is still unknown whether RESS is directly involved in the regulation of UPR or independent of UPR (44). In addition, DTT treatment resulted in translational repression of a large number of genes associated with ribosome and translation initiation factor activity (Supplementary Table S12), supporting the hypothesis of a translational repression as a mechanism of reducing ER throughput.

In this study, we report a full-scale transcriptome dataset of the filamentous fungi *A. oryzae* for the first time, using RNA-Seq technology combined with the paired-end method, which allows precise detection of read position and linkage. We subsequently performed elaborate analyses of this dataset to improve the genome annotation of *A. oryzae*, including evidence for many novel transcripts and new exons, the definition and extension of untranslated regions, and the prediction of uORFs and uATGs in 5'-UTRs.

We also assessed AS events in order to further illustrate the complexity of the *A. oryzae* transcriptome, with 1375 AS events identified in 1032 *A. oryzae* genes. That is to say, 8.55% of the total number of *A. oryzae* genes was estimated to undergo AS. This is the highest number reported among fungi up to now, but it is far lower than the numbers reported in higher eukaryotes, especially in humans where ~86% of all genes were estimated to produce two or more mRNA isoforms. In addition, the high ratio of RIs to CEs suggested that AS events in *A. oryzae* were predominantly controlled by the ID mechanism.

Our RNA-Seq data, due to its digital characteristics, provided detailed and accurate information about gene expression levels, allowing us to detect differential gene expression among various culture conditions. This helped us to elucidate many differences between *A. oryzae* grown on SC and on LC with regard to protein translation/modification, energy catabolism and DNA replication. Based on the differential gene expression between ER stress conditions and control conditions, we established a regulation model of the protein secretory pathway for both SC and LC. This is the first global analysis of ER stress in *A. oryzae* under SC, and will be valuable for further studying a series of ER-associated biological processes such as UPR, ERAD, protein folding, glycosylation and RESS.

In conclusion, with high resolution and unprecedented throughput, RNA-Seq allowed us to survey the global transcriptome of *A. oryzae* much more precisely and accurately than previous transcriptional analyses by EST or microarrays. It provided us with a global view of the protein-coding potential of the genome to improve genome annotation and enabled us to study protein secretion and many other ER-related biological processes.

These results will undoubtedly help us to comprehensively understand the complexity of the *A. oryzae* transcriptome and pave the way for further research in fungi.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Natural Science Foundation of China (30970045); Guangdong Technological Foundation of China (2008A010900002, 2007A010900001); National Key Technology R&D Program of China (2007BAK36B03); National High Technology R&D Program of China (2006AA020304). Funding for open access charge: National Natural Science Foundation of China (30970045).

*Conflict of interest statement.* None declared.

## REFERENCES

1. Nevalainen,K.M., Te'o,V.S. and Bergquist,P.L. (2005) Heterologous protein expression in filamentous fungi. *Trends Biotechnol.*, **23**, 468–474.
2. Machida,M., Asai,K., Sano,M., Tanaka,T., Kumagai,T., Terai,G., Kusumoto,K., Arima,T., Akita,O., Kashiwagi,Y. et al. (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, **438**, 1157–1161.
3. Akao,T., Sano,M., Yamada,O., Akeno,T., Fujii,K., Goto,K., Ohashi-Kunihiro,S., Takase,K., Yasukawa-Watanabe,M., Yamaguchi,K. et al. (2007) Analysis of expressed sequence tags from the fungus *Aspergillus oryzae* cultured under different conditions. *DNA Res.*, **14**, 47–57.
4. Vongsangnak,W., Olsen,P., Hansen,K., Krogsbaard,S. and Nielsen,J. (2008) Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genomics*, **9**, 245.
5. Andersen,M.R., Vongsangnak,W., Panagiotou,G., Salazar,M.P., Lehmann,L. and Nielsen,J. (2008) A trispecies *Aspergillus* microarray: comparative transcriptomics of three *Aspergillus* species. *Proc. Natl Acad. Sci. USA*, **105**, 4387–4392.
6. Tamano,K., Sano,M., Yamane,N., Terabayashi,Y., Toda,T., Sunagawa,M., Koike,H., Hatamoto,O., Umitsuki,G., Takahashi,T. et al. (2008) Transcriptional regulation of genes on the non-syntenic blocks of *Aspergillus oryzae* and its functional relationship to solid-state cultivation. *Fungal Genet. Biol.*, **45**, 139–151.
7. Nagalakshmi,U., Wang,Z., Waern,K., Shou,C., Raha,D., Gerstein,M. and Snyder,M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, **320**, 1344–1349.
8. Wilhelm,B.T., Marguerat,S., Watt,S., Schubert,F., Wood,V., Goodhead,I., Penkett,C.J., Rogers,J. and Bahler,J. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, **453**, 1239–1243.
9. Hillier,L.W., Reinke,V., Green,P., Hirst,M., Marra,M.A. and Waterston,R.H. (2009) Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res.*, **19**, 657–666.
10. Cloonan,N., Forrest,A.R.R., Kolle,G., Gardiner,B.B.A., Faulkner,G.J., Brown,M.K., Taylor,D.F., Steptoe,A.L., Wani,S., Bethel,G. et al. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods*, **5**, 613–619.
11. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
12. Pan,Q., Shai,O., Lee,L.J., Frey,J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human

- transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
13. Sultan,M., Schulz,M.H., Richard,H., Magen,A., Klingenhoff,A., Scherf,M., Seifert,M., Borodina,T., Soldatov,A., Parkhomchuk,D. et al. (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, **321**, 956–960.
  14. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
  15. Maher,C.A., Kumar-Sinha,C., Cao,X.H., Kalyana-Sundaram,S., Han,B., Jing,X.J., Sam,L., Barrette,T., Palanisamy,N. and Chinnaian,A.M. (2009) Transcriptome sequencing to detect gene fusions in cancer. *Nature*, **458**, 97–101.
  16. Ishida,H., Matsumura,K., Hata,Y., Kawato,A., Sugiyama,K., Abe,Y., Imayasu,S. and Ichishima,E. (2001) Establishment of a hyper-protein production system in submerged *Aspergillus oryzae* culture under tyrosinase-encoding gene (*melO*) promoter control. *Appl. Microbiol. Biotechnol.*, **57**, 131–137.
  17. Li,R., Li,Y., Kristiansen,K. and Wang,J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713–714.
  18. Stanke,M., Diekhans,M., Baertsch,R. and Haussler,D. (2008) Using native and synteny-mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*, **24**, 637–644.
  19. Conesa,A., Gotz,S., Garcia-Gomez,J.M., Terol,J., Talon,M. and Robles,M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.
  20. Wang,L., Feng,Z., Wang,X., Wang,X. and Zhang,X. (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, **26**, 136–138.
  21. Bindea,G., Mlecnik,B., Hackl,H., Charoentong,P., Tosoloni,M., Kirilovsky,A., Fridman,W.H., Pages,F., Trajanoski,Z. and Galon,J. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **25**, 1091–1093.
  22. Ruiz-Echevarria,M.J. and Peltz,S.W. (2000) The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. *Cell*, **101**, 741–751.
  23. Vilela,C. and McCarthy,J.E. (2003) Regulation of fungal gene expression via short open reading frames in the mRNA 5' untranslated region. *Mol. Microbiol.*, **49**, 859–867.
  24. Hinnebusch,A.G. (2005) Translational regulation of *GCN4* and the general amino acid control of yeast. *Annu. Rev. Microbiol.*, **59**, 407–450.
  25. Modrek,B. and Lee,C. (2002) A genomic view of alternative splicing. *Nat. Genet.*, **30**, 13–19.
  26. Loftus,B.J., Fung,E., Roncaglia,P., Rowley,D., Amedeo,P., Bruno,D., Vamathevan,J., Miranda,M., Anderson,I.J., Fraser,J.A. et al. (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*, **307**, 1321–1324.
  27. Ho,E.C., Cahill,M.J. and Saville,B.J. (2007) Gene discovery and transcript analyses in the corn smut pathogen *Ustilago maydis*: expressed sequence tag and genome sequence comparison. *BMC Genomics*, **8**, 334.
  28. McGuire,A.M., Pearson,M.D., Neafsey,D.E. and Galagan,J.E. (2008) Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol.*, **9**, R50.
  29. Ebbole,D.J., Jin,Y., Thon,M., Pan,H., Bhattacharai,E., Thomas,T. and Dean,R. (2004) Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: analysis of expressed sequence tags. *Mol. Plant Microbe Interact.*, **17**, 1337–1347.
  30. Berget,S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
  31. Romfo,C.M., Alvarez,C.J., van Heeckeren,W.J., Webb,C.J. and Wise,J.A. (2000) Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.*, **20**, 7955–7970.
  32. Ast,G. (2004) How did alternative splicing evolve? *Nat. Rev. Genet.*, **5**, 773–782.
  33. Hata,Y., Ishida,H., Ichikawa,E., Kawato,A., Sugiyama,K. and Imayasu,S. (1998) Nucleotide sequence of an alternative glucoamylase-encoding gene (*glab*) expressed in solid-state culture of *Aspergillus oryzae*. *Gene*, **207**, 127–134.
  34. Maeda,H., Sano,M., Maruyama,Y., Tanno,T., Akao,T., Totsuka,Y., Endo,M., Sakurada,R., Yamagata,Y., Machida,M. et al. (2004) Transcriptional analysis of genes for energy catabolism and hydrolytic enzymes in the filamentous fungus *Aspergillus oryzae* using cDNA microarrays and expressed sequence tags. *Appl. Microbiol. Biotechnol.*, **65**, 74–83.
  35. Oda,K., Kakizono,D., Yamada,O., Iefuji,H., Akita,O. and Iwashita,K. (2006) Proteomic analysis of extracellular proteins from *Aspergillus oryzae* grown under submerged and solid-state culture conditions. *Appl. Environ. Microbiol.*, **72**, 3448–3457.
  36. Akao,T., Gomi,K., Goto,K., Okazaki,N. and Akita,O. (2002) Subtractive cloning of cDNA from *Aspergillus oryzae* differentially regulated between solid-state culture and liquid (submerged) culture. *Curr. Genet.*, **41**, 275–281.
  37. Ye,L. and Pan,L. (2008) A comparison of the unfolded protein response in solid-state with submerged cultures of *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.*, **72**, 2998–3001.
  38. Ma,Y.J. and Hendershot,L.M. (2001) The unfolding tale of the unfolded protein response. *Cell*, **107**, 827–830.
  39. Kasuya,T., Nakajima,H. and Kitamoto,K. (1999) Cloning and characterization of the *bipA* gene encoding ER chaperone BiP from *Aspergillus oryzae*. *J. Biosci. Bioeng.*, **88**, 472–478.
  40. Maruyama,J.I., Ohnuma,H., Yoshikawa,A., Kadokura,H., Nakajima,H. and Kitamoto,K. (2000) Production and product quality assessment of human hepatitis B virus pre-S2 antigen in submerged and solid-state cultures of *Aspergillus oryzae*. *J. Biosci. Bioeng.*, **90**, 118–120.
  41. Leber,J.H., Bernales,S. and Walter,P. (2004) *IRE1*-independent gain control of the unfolded protein response. *PLoS Biol.*, **2**, 1197–1207.
  42. Patil,C.K., Li,H. and Walter,P. (2004) Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. *PLoS Biol.*, **2**, 1208–1223.
  43. Arvas,M., Pakula,T., Lanthaler,K., Saloheimo,M., Valkonen,M., Suortti,T., Robson,G. and Penttila,M. (2006) Common features and interesting differences in transcriptional responses to secretion stress in the fungi *Trichoderma reesei* and *Saccharomyces cerevisiae*. *BMC Genomics*, **7**, 32.
  44. Pakula,T.M., Laxell,M., Huuskonen,A., Uusitalo,J., Saloheimo,M. and Penttila,M. (2003) The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei* - Evidence for down-regulation of genes that encode secreted proteins in the stressed cells. *J. Biol. Chem.*, **278**, 45011–45020.
  45. Al-Sheikh,H., Watson,A.J., Lacey,G.A., Punt,P.J., MacKenzie,D.A., Jeenes,D.J., Pakula,T., Penttila,M., Alcocer,M.J.C. and Archer,D.B. (2004) Endoplasmic reticulum stress leads to the selective transcriptional downregulation of the glucoamylase gene in *Aspergillus niger*. *Mol. Microbiol.*, **53**, 1731–1742.
  46. Sims,A.H., Gent,M.E., Lanthaler,K., Dunn-Coleman,N.S., Oliver,S.G. and Robson,G.D. (2005) Transcriptome analysis of recombinant protein secretion by *Aspergillus nidulans* and the unfolded-protein response *in vivo*. *Appl. Environ. Microbiol.*, **71**, 2737–2747.