Process Biochemistry 64 (2018) 51–62

Contents lists available at ScienceDirect

Process Biochemistry

journal homepage: www.elsevier.com/locate/procbio

Analyzing *Phanerochaete chrysosporium* gene expression patterns controlling the molecular fate of lignocellulose degrading enzymes

Ayyappa Kumar Sista Kameshwar, Wensheng Qin*

Department of Biology, Lakehead University, 955 Oliver Road, Thunder Bay, Ontario, P7B 5E1, Canada

ARTICLE INFO

Keywords:
Phanerochaete chrysosporium
Information
Storage and processing
Eukaryotic orthologous groups (KOG)
Transcription factors
Gene expression datasets
NCBI GEO

ABSTRACT

The outstanding degrading abilities of *Phanerochaete chrysosporium* is solely dependent on its lignocellulolytic, aromatic compound degrading and detoxifying enzymes. However, the gene expression and protein turnover of lignocellulolytic enzymes are controlled at cellular level by various genes involved in information storage and processing KOG group. Understanding the gene expression patterns and mechanisms involved in regulation of lignocellulose degrading enzymes will significantly help in strain improvement and developing recombinant strains. To study the common expression patterns, we have retrieved *P. chrysosporium* gene expression datasets from NCBI GEO and analyzed using GeneSpring® software based on the genome wide KOG annotations retrieved from JGI-MycoCosm database. Statistically significant genes obtained from our analysis were separated into replication, repair and recombination, chromatin structure and dynamics, transcription factors, RNA processing and modification and translation, ribosomal structure and biogenesis processes. We have observed various genes encoding for DNA damage, repair and recombination, mRNA splicing, amidases, polyadenylate binding factors, heat shock, helix loop helix, HMG-box, CCAAT (HAP5), CRE-B transcription factors, histone acetyl transferases (MYST, SAGA) commonly expressed among the datasets of natural plant biomass growth substrates. Further studies must be conducted to understand the role and involvement of these significant genes in plant biomass degradation by *P. chrysosporium*.

1. Introduction

In the past few years, genome and transcriptome of *Phanerochaete chrysosporium* were intensively studied because of its efficient lignocellulose degrading abilities and availability of highly annotated genome. These studies have mostly delineated the genes and mechanisms responsible for the process of lignocellulose degradation [1–7]. However, mechanisms involved in cellular regulation of genetic material coding for lignocellulolytic enzymes were not clearly explained till today. Recent genomic studies conducted by various research groups and development of fungal genome repositories such as JGI (Joint Genome Institute) MycoCosm [8] and 1000 fungal genome project have revealed various significant facts. The eukaryote specific classification of KOG (clusters of orthologous groups) is significantly used for finding the ortholog and paralog proteins [8]. All the sequenced genomes deposited in JGI-MycoCosm database are provided with their respective KOG classification or KOG ID. The present day JGI sequencing protocol predicts and classifies the sequenced genome into four major classes a) Cellular processes and signaling b) Information storage and processing c) Metabolism d) poorly characterized.

The KOG group information, storage and processing is further classified into five groups as RNA processing and modification (KOG-ID: A), chromatin structure and dynamics (KOG-ID: B), translation, ribosomal structure and biogenesis KOG-ID: J), transcription (KOG-ID: K) and replication, recombination and repair (KOG-ID: L). Present day KOG classification of *P. chrysosporium* harbors 1713 gene models coding for Information storage and processing group which were further divided into 489 (RNA processing and modification), 201 (chromatin structure and dynamics), 366 (translation, ribosomal structure and biogenesis), 415 (transcription) and 242 (replication, recombination and repair) gene models respectively. Large number of gene models were mostly present in single copies with few number of gene models occur in multiple copies (Fig. 1). Gene models involved in RNA processing and modification plays a crucial role in converting cellular genetic information from genes to proteins, thus determining the fate of cellular function and structure. However, RNA undergoes prior modifications and processing before performing the above functions. Majorly, RNA processing steps can be classified into three types a) trimming of RNA end segments resulting in a mature RNA form b) RNA splicing and c) sequence level modification of RNA segments.

* Corresponding author.
E-mail address: wqin@lakeheadu.ca (W. Qin).

http://dx.doi.org/10.1016/j.procbio.2017.10.007
Received 5 September 2017; Received in revised form 24 September 2017; Accepted 10 October 2017
Available online 12 October 2017
1359-5113/

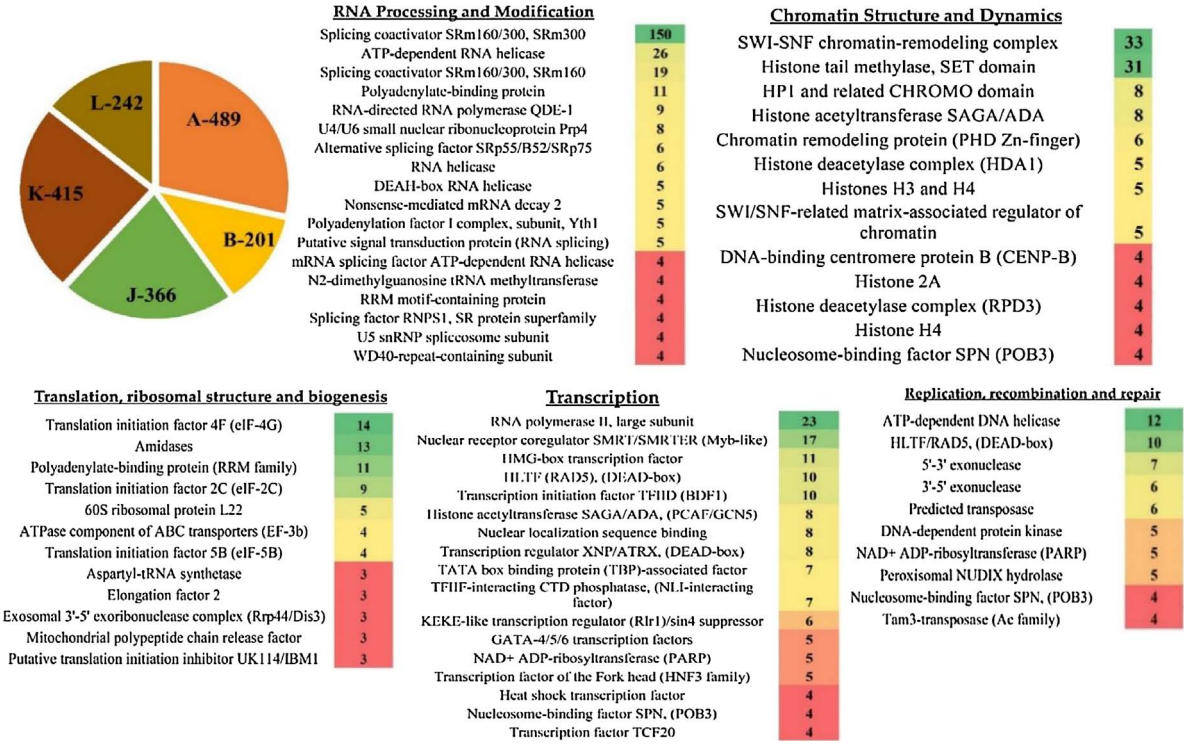


Fig. 1. Pie diagram showing the distribution of gene models A- RNA processing and modification, B-chromatin structure and dynamics, J- translation, ribosomal structure and biogenesis, K- transcription and L- replication, recombination and repair. And heatmaps showing the gene models occurring more than 3 copies in the *P. chrysosporium* genome.

Gene regulation is an important physiological process and was extensively studied in fungi. The process of gene regulation assures the up and down regulation of genes based on the growth conditions. According to Richard et al. (2014), in fungi, thirty-seven classes of gene regulators have been identified as zinc finger transcription factor proteins (C2H2 and binuclear zinc cluster protein (Zn2Cys6)), fungal specific transcription factors, bZIP, histone-like transcription, basic helix-loop-helix (HLH), heat shock factor (HSF), Myb-DNA binding, transcription enhancer factor (TEA) and GATA factors [9]. The production and secretion of extracellular fungal enzymes and their gene regulatory mechanisms especially cellulases and hemicellulases encoding genes were extensively studied in *Trichoderma* and *Aspergillus* species [10,11]. These transcription factors are required for coordination various cellular processes when cultured on different growth substrates. Occurrences of the gene sequences encoding for the above-mentioned transcription factors also might have been strongly relate to the fungal diversity. *P. chrysosporium* genome contains the gene sequences coding for the above mentioned crucial transcription factors.

The enormous genomic DNA of eukaryotes is well packed in the nucleus by the conserved histone proteins to form a complex ordered structure such as chromatin. After numerous stages of organization chromatin are further organized into chromosomes which allows precise cellular divisions [12]. Chromatin dynamics in eukaryotes can be introduced majorly by three ways a) ATP-dependent remodeling which might also lead to the interchanges in primary structures of histones [12–14] b) histones are also subjected to post translational modifications resulting in structural and functional outcomes [12,15] and c) DNA is also subjected to methylations [12,16]. All three processes chromatin modifications, chromatin remodeling and DNA methylation were found to be strongly interdependent. Steinfeld et al., have revealed the crucial role of chromatin modifiers in transcriptional regulation of common yeast, *Saccharomyces cerevisiae* [17].

Gene models classified under translation, ribosomal structure and biogenesis are required for the accurate conversion of genetic code to proteins. Fungal ribosomes (cellular nanomachines) majorly comprises two ribonucleoprotein subunits 40S (in turn contains 33 ribosomal proteins and a 18S ribosomal rRNA) and 60S subunits (consists of three ribosomal rRNA 25S, 5.8S, 5S and 46 ribosomal proteins) [18]. *P. chrysosporium* genome codes for 254 genes involved in ribosome structure, biogenesis and translation processes. DNA replication, repair and recombination events are complex and basic molecular processes in the living organisms. Process of DNA replication in eukaryotes is one of the highly studied molecular process and various genes and proteins involved in this key process have been clearly explained. The genome of *P. chrysosporium* harbors 155 unique gene models which includes various genes involved in the process of DNA replication, repair and recombination. The proteome, transcriptome and secretome studies conducted in the past have revealed the involvement of various oxidases, hydrolytic enzymes in degradation of plant biomass, however additional research must be conducted to reveal the regulation of the lignocellulolytic enzymes at molecular level [2,3,7,19].

In our present study, we have analyzed the gene expression patterns of *P. chrysosporium* to understand the common and significant expression of genes involved in RNA processing and modification, chromatin structure and dynamics, translation, ribosomal structure and biogenesis, transcription and replication, recombination and repair processes among *P. chrysosporium* gene expression datasets. To the best of our knowledge, this is the first comprehensive report on *P. chrysosporium* genes involved in information, storage and processing processes and their expression patterns.

2. Methods

2.1. Data retrieval

The gene expression datasets of *P. chrysosporium* cultured on different growth substrates, were retrieved from NCBI-GEO (Gene expression omnibus database) (https://www.ncbi.nlm.nih.gov/geo/) using the term *P. chrysosporium*. Details about these gene expression datasets analyzed were reported in Table 1. All the relevant experimental metadata corresponding to the gene expression datasets were retrieved using the "Accession display" option of NCBI GEO website and

Table 1
List of the *P. chrysosporium* gene expression datasets retrieved from NCBI GEO repository:.

GEO- ID	Platform and Technology	Substrate	#Samples	Ref
GSE54542	NimbleGen *P. chrysosporium* arrays	Oak acetonic extractives	6	[63]
GSE27941	NimbleGen *P. chrysosporium* arrays	Ball milled aspen, Ball milled pine	6	[64]
GSE52922	NimbleGen *P. chrysosporium* arrays	P717 hybrid line, Transgenic line 82 Transgenic line 64	9	[65]
GSE14734	NimbleGen *P. chrysosporium* arrays	Cellulose, Glucose, Ball milled aspen	9	[3,4]
GSE14735	NimbleGen *P. chrysosporium* arrays	Replete, Carbon-limited Nitrogen-limited	9	[3,4]
GSE69008	NimbleGen *P. chrysosporium* arrays	Poplar wood substrates	24	[1]
GSE69461	Illumina HiSeq 2000	*Picea glauca* (spruce sapwood)	18	[66]

from the corresponding literature available. In NCBI GEO database, there are 8 gene expression datasets which are specifically studied on *P.* chrysosporium, out of which we have selected 7 (GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 and GSE69461) gene expression datasets for our current analysis (Table 1). These datasets were considered mainly because of the varied substrate conditions (from simple synthetic growth medium to complex plant biomass medium) used for the culture of *P. chrysosporium.* The experimental metadata accession ID's, gene expression platform details substrate used for the growth of *P. chrysosporium* and sample information were reported in Table 1.

2.2. Data analysis

All the datasets were analyzed using GeneSpring v14.8 (http://genespring-support.com/get-gs) software. The expression experiments were created using generic single-color workflow, by creating a prior generic single-color technology with the available supplementary information. Gene expression datasets with accession IDs GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 were log transformed with 2 log base, normalized using 75th percentile normalization, with a threshold value of raw signals set to 1.0 and baseline transformation to median of all samples. The experimental details of all the samples were retrieved from the corresponding literature and experimental metadata and later this data was used for grouping the samples. The probe sets were filtered using the option "Filter probe sets by expression" by selecting the raw data and filtered using the filter by percentile option (upper and lower percentiles set to 100 and 20 respectively).

Differentially expressed significant genes were obtained using the "Statistical analysis" option, and based on the experimental grouping and created interpretations T-test or one-way Anova was performed. However, for GSE69461 dataset, we have not performed any of the preprocessing steps such as normalization, log transformation or thresholding, as we have retrieved RPKM data for the individual samples. Based on the sample grouping T-test was performed with asymptotic p-value computation and Benjamini-Hochberg False discovery rate for multiple testing correction. Differentially expressed significant list of genes were retrieved from all the datasets and compared using Venny and Jvenn online softwares for obtaining common gene lists. We have retrieved the list of gene level annotations of *P. chrysosporium* RP78v2.2 encoding for information, storage and processing (KOG Group) which includes RNA processing and modification, chromatin structure and dynamics, translation, ribosomal structure and biogenesis, transcription and replication, recombination and repair processes, from JGI MycoCosm database.

3. Results

The extrinsic plant biomass degrading properties of *P. chrysosporium* is majorly credited to its lignocellulolytic CAZymes and wide range of aromatic compound degrading and detoxifying enzymes [6,20]. The expression of these enzymes is majorly regulated by a wide range of enzymes belonging to the information storage and processing group. Statistical analysis of GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 and GSE69461 gene expression dataset has resulted in 691, 583, 146, 320, 275, 865 and 1235 unique differentially expressed genes belonging to information storage and processing KOG group (Table 2). Several genes encoding for various lignocellulolytic CAZymes, aromatic compound degrading and large array of detoxifying enzymes were also found to be highly upregulated which were extensively discussed in our previous works [6,20].

We have observed a total of 8, 53, 3, 320, 58, 181 and 1119 unique genes encoding for information, storage and processing group were highly up regulated with higher fold change values (log2 FC > 2.0) among GSE14734, GSE14735, GSE27941, GSE54542, GSE52922, GSE69008 and GSE69461 datasets respectively (Table 2). The violin plots (sample and group level data on x-axis and normalized intensity values on y-axis) showing the distribution of normalized and baseline transformed samples, which briefly represents the differential expression of the normalized genes of all the samples at dataset level were shown (Fig. 2). Similarly, the volcano plots (log2 fold change values on x-axis and −log 10 corrected P-values on y-axis) and profile plots (experimental condition information on x-axis and normalized intensity values on y-axis) briefly representing the differentially expressed genes (with higher fold change values) among individual experimental conditions were shown (Fig. 3).

Table 2
The total number of differentially expressed and statistically significant genes (unique) belonging to information storage and processing group among the selected datasets:.

	Unique Genes after Statistical Analysis						(Log 2) Fold change ≥2.0					
Dataset	A	B	J	K	L	Total	A	B	J	K	L	Total
GSE14734	172	90	145	164	120	691	3	0	2	3	0	8
GSE14735	129	73	192	114	75	583	0	2	41	8	2	53
GSE27941	25	12	53	36	20	146	0	1	0	2	0	3
GSE54542	99	36	50	90	45	320	99	36	50	90	45	320
GSE52922	87	20	69	75	24	275	17	8	15	16	2	58
GSE69008	213	91	205	204	152	865	36	14	67	51	13	181
GSE69461	346	146	261	302	180	1235	310	127	248	274	160	1119

Note: A: RNA processing and modification, B: chromatin structure and dynamics, J: translation, ribosomal structure and biogenesis, K: transcription and L: replication, recombination and repair.

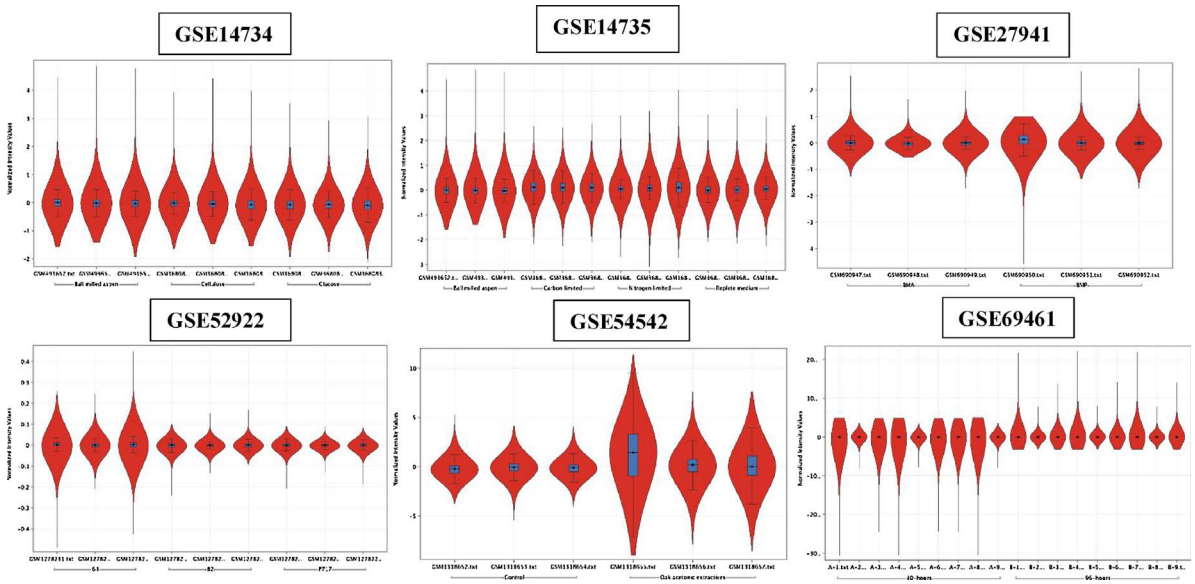


Fig. 2. Violin plots for the selected datasets GSE14734, GSE14735, GSE27941, GSE52922, GSE54542 (normalized and baselined) and GSE69461 (only baselined) briefly showing the distribution of samples.

3.1. RNA processing and modification (KOG ID: A)

The eukaryotic mRNA is substantially processed through 5′ capping, splicing and 3′ end processing before it gets exported, thus, these processes play a crucial role in determining the fate of transcripts [21]. Several studies conducted in the past have clearly explained that a wide range of enzymes coordinately function together to enhance the regulation information which affects in transcript export, localization and stability [21]. As mentioned earlier genome of *P. chrysosporium* encodes 489 genes involved in the RNA processing and modification class. Majorly, mRNA guanylyl transferases, methyltransferases are involved in 5'capping which happens immediately after early transcription by RNA polymerase II [22]. The 3'end processing involves mRNA cleavage and polyadenylation factors, polyadenylate polymerase, polyadenylate binding protein-II, polyadenylation factor complex, mRNA cleavage factor I and II [23,24]. Spliceosomes performs the complex splicing reactions on pre-mRNA in eukaryotes, a large set of proteins along with U1 to U6 small nuclear RNPs (snRNPs) [21,23,25]. In *P. chrysosporium* genome 19, 36 and 237 gene models were found to be involved in 5'end capping, 3' end processing and mRNA splicing processes.

Genes encoding for ATP dependent helicases, DEAD-box superfamily, RNA binding proteins, mRNA guanylyl and methyl transferases

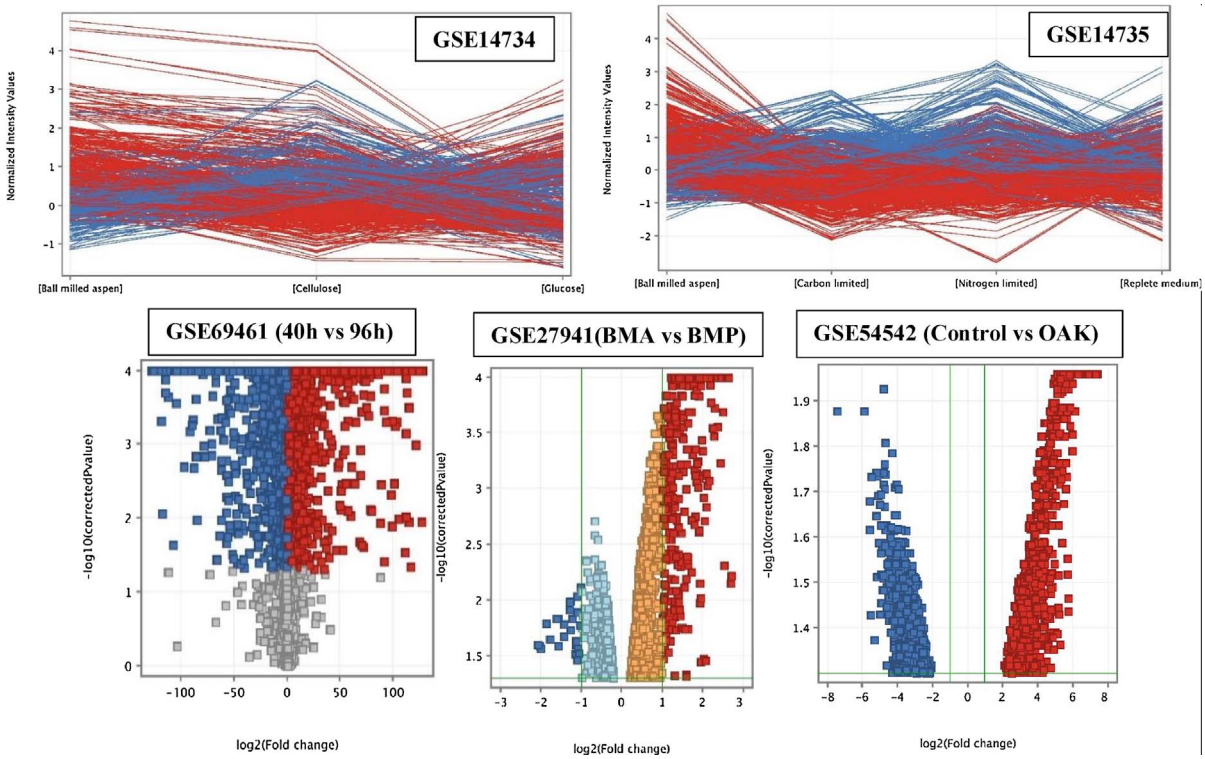


Fig. 3. Profile plots (GSE14734 and GSE14735) and volcano plots (GSE27941, GSE54542 and GSE69461) of the significant and differentially expressed genes with fold change cutoff > 2.0 among the selected datasets.

(5'end capping) and mRNA cleavage and polyadenylation factor, polyadenylation binding protein and other associated enzymes were found to be statistically significant and commonly expressed among all the gene expression datasets. However, in dataset GSE14734 and GSE14735 genes encoding for ATP-dependent RNA helicase (Phch1_5340), Fibrillarin and related nucleolar RNA-binding (Phch1_610), splicing factor hnRNP-F and related RNA-binding proteins (Phch1_4943) and ribosomal protein RPL1 (Phch1_128104) were found to be highly upregulated in cellulose, glucose, replete and nitrogen limited mediums and down regulated in ball milled aspen and carbon limited growth mediums. Whereas in GSE27941 dataset few genes encoding for RNA processing and modification were found to be statistically significant but were not highly expressed. In GSE54542, GSE52922, GSE69008 and GSE69461 gene expression datasets a large set of genes encoding for RNA processing and modification processes were found to be highly expressed. Total of 46 genes encoding for RNA processing and modification were found to be common among GSE69008 and GSE69461 datasets. In GSE54542 dataset 99 genes coding for RNA processing and modification were differentially expressed, transcripts encoding for spliceosomal proteins, U1, U4, U5 snRNP, WD40 repeats, ribosomal proteins (RPL1 and RPL2) and mRNA export factors and SWAP mRNA splicing regulator were found to be up regulated (with fold change ≥ 2.0) in oak acetonic extractive samples and down regulated in control samples. In GSE52922 dataset, 17 genes (ATP dependent RNA helicase, decapping enzyme complex, mRNA cleavage and polyadenylation factor and splicing coactivator and other splicing factors) were up regulated in 82 and 64 growth substrates. In GSE69008 dataset genes encoding for various RNA processing and modification processes were down regulated in A (high lignin and low glucose) growth substrates, were as found to be differentially up regulated in B (low lignin and high glucose) samples. Especially transcripts encoding for alternative splicing factor (SRp55/SRp75), mRNA cleavage and polyadenylation factor, mRNA deadenylase, RNA binding protein (p54nrb) up regulated in C (average lignin and average glucose) substrates. Finally, in GSE69461 dataset total of 310 genes encoding for RNA processing and modification were found to be expressed with fold change ≥ 2.0. We have found that 199 genes were highly up regulated in 96 h and 111 genes were highly upregulated in 40 h growth substrates. Out of which 40 genes encoding for DEAD-box family, decamping enzyme complex, histone H3 (Lys4) and mRNA methyl transferase, mRNA cleavage and polyadenylation factors, polyadenylating polymerase, RNA binding proteins, spliceosomes and splicing factors U1 to U6 snRNP were highly up regulated in 40 h growth substrate (Supplementary Information) (Fig. 4).

3.2. Chromatin structure and dynamics (KOG ID: B)

It is a well-known fact that vast eukaryotic genomes are efficiently packaged in to chromatin and further into stable chromosomes. Several state of the art reviews were already available on eukaryotic chromatin structure and dynamics and few were listed here [26–31]. As discussed earlier genome of *P. chrysosporium* codes 201 genes classified under chromatin structure and dynamics. Genes encoding for chromatin remodeling protein (PHD-Zn finger), chromosome condensation complex, histone acetyltransferase, and SWI-SNF chromatin remodeling complex proteins were found to be statistically significant and commonly expressed among GSE14734, GSE14735 and GSE27941 datasets. And in GSE54542, GSE52922, GSE69008 and GSE69012 datasets genes encoding for DNA binding centromere protein B, heterochromatin associated protein (HP1), histone acetyltransferase (SAGA) and type b catalytic unit, histones H3 and H4 and SWI-SNF chromatin remodeling complex proteins were significantly common. In GSE14734, GSE14735 and GSE2794 datasets total of 90, 73 and 12 genes were found to be statistically significant with a p-value < 0.05 after multiple testing correction. Total of 57 transcripts were found to be common between GSE14734 and GSE14735 datasets which majorly contain histone proteins (H1, H2A, H2B, H3, H4, H5), histone methyltransferases, histone acetyl transferases (MYST, SAGA/ADA, PCAF/SAGA), sirutin 5, structural maintenance of chromosome, SWI-SNF remodeling protein encoding genes. Chromatin remodeling complexes, heterochromatin associated protein, nucleosome assembly protein, histone deacetylase complexes, histone tail methylase, SWI-SNF (Snf5 subunit) chromatin remodeling, telomerase catalytic unit were up regulated in ball milled aspen samples. Only genes encoding for DNA-binding centromere protein B (Phchr1_5151) (in GSE14735), SWI-SNF chromatin remodeling complex protein (Phchr1_5319) (both in GSE14735 and GSE27941 datasets) were highly upregulated in ball milled aspen samples with fold change ≥ 2.0. However, no genes involved in chromatin structure and dynamics were found to be expressed with fold change < 2.0 (not very highly expressed).

In GSE54542 dataset, chromatin remodeling complex (Phchr_7694), heterochromatin associated protein (Phchr_1386), histone acetyltransferase (MYST) (Phchr_1138, 42216), histone acetyltransferase (SAGA) (Phchr_932), histone H3 methyltransferase complex (Phchr_136167), histone tail methylase (Phchr_136398) and telomerase length regulating protein kinase (Phchr_2237) were highly up regulated with fold change ≥ 2.0 in control samples. At the same time genes encoding for chromatin remodeling complex (Phchr_130373, 6773, 7315), DNA binding centromere (Phchr_5151), histone acetyl transferase (SAGA/ADA) (Phchr_2529, 3745 and 29427), histone deacetylase (Phchr_5894), histone H3 methyl transferase (Phchr_820, 3326), histone tail methylase (Phchr_5894) and SWI-SNF chromatin remodeling complex proteins were highly up regulated in oak acetonic extract samples. In GSE52922 dataset, genes encoding for DNA binding centromere protein, histone proteins (H2A, H2B, H3, H4) and SWI-SNF chromatin remodeling complex protein were highly upregulated in P717 growth samples and down regulated in 64 and 82 growth substrates. Similarly, in dataset GSE69008, genes encoding for heterochromatin associated protein (HP1), histone acetyl transferase (SAGA/ADA) and sirtuin 5 (SIR2 family) proteins were highly expressed in A (high lignin-low glucose) proteins. In B (low lignin −high glucose) growth substrate genes encoding for chromatin remodeling complex, histone acetyl transferase, histone (H3 and H4), nucleosome binding factor, sirtuin 5 (SIR2) and SWI-SNF chromatin remodeling complex proteins were highly up regulated. In dataset GSE69461, genes encoding for chromatin assembly factor, cell cycle regulated histone H1, chromatin remodeling complex, chromosome condensation complex, DNA topoisomerase, DNA binding centromere binding, heterochromatin associated protein, histone proteins (H1, H2A, H2B, H3, H4), histone acetyl transferases (SAGA/ADA, MYST), histone deacetylase complex (HDA1, RPD3, SIN3), nucleosome assembly and remodeling factors, histone tail methylase, sirtuin 4, 5, chromosome structural maintenance protein and SWI-SNF chromatin remodeling complex were highly up regulated in 96-h growth substrates (Supplementary Inforamtion-S1) (Fig. 5).

3.3. Translation ribosomal structure and biogenesis (KOG ID: J)

Translation is the central and crucial process for the conversion of nucleotide sequence to cellular expressing units. The process of translation is majorly dependent on ribosomes, which are highly conserved cellular nanomachines. Biogenesis of ribosomes happens in nucleolus of eukaryotic cells, where it synthesizes rRNA molecules (precursors) through DNA directed RNA polymerase I and III, these nascent rRNA molecules are further subjected to a set of RNA cleavage reactions and other chemical modifications [18,32]. A large set of ribosome biogenesis proteins are required for the proper rRNA folding and enzymatic processing to produce active ribosomal subunits [18,32]. Thus, a serious coordination of all the DNA dependent RNA polymerases is involved in synthesizing structurally active ribosomal subunits [32]. As mentioned above genome of *P. chrysosporium* encodes for 366 gene models classified to be involved in translation, ribosomal structure and

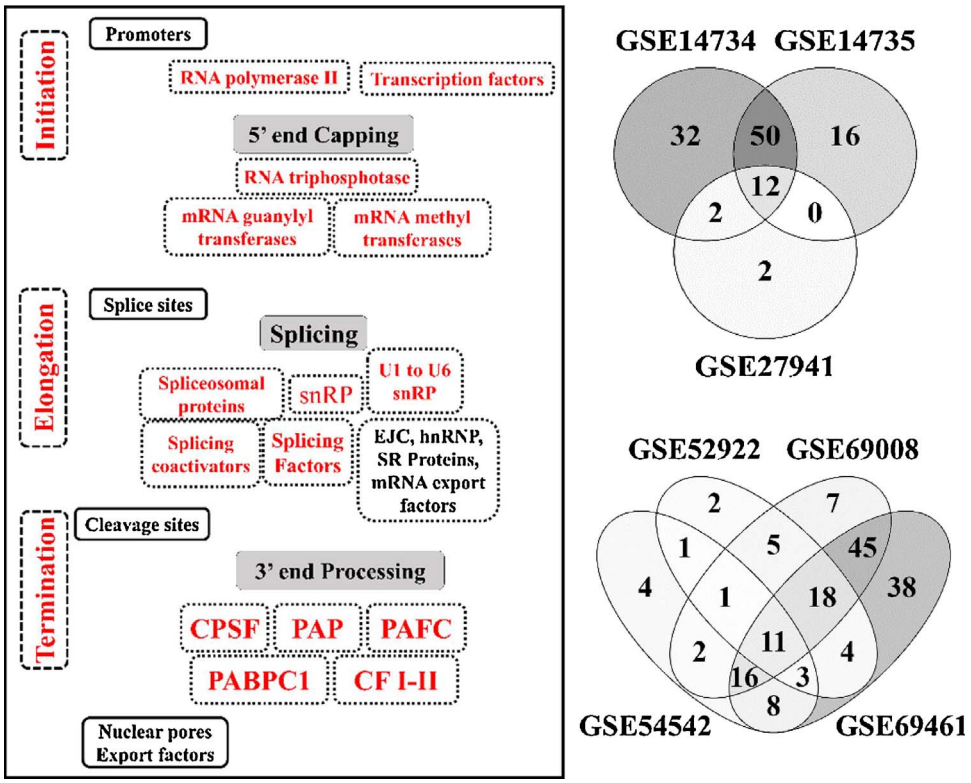


Fig. 4. List of commonly expressed and statistically significant genes among the selected gene expression datasets encoding for RNA processing and Modification group (KOG functional ID: A).

biogenesis. Our present metadata analysis has revealed that total of 15 and 8 commonly expressed statistically significant genes among GSE14734, GSE14735, GSE27941 and GSE54542, GSE52922, GSE69008 and GSE69461 datasets, with various other genes found to be common between the datasets (Fig. 3) (Supplementary information). Genes encoding for 60S acidic ribosomal protein P2 was up regulated in glucose substrate and translation initiation factor 4F (elF-4G) was highly upregulated in ball milled aspen and cellulose with fold change ≥2.0. In GSE14735 dataset, various genes encoding for translation ribosomal structure and biogenesis were significantly expressed with fold change values ≥2.0. Genes encoding for 40S ribosomal proteins (S2, S3, S3A, S4, S6, S8, SA, S12, S15, S26, SA/P40), 60S acidic ribosomal (P0, P2), 60S ribosomal protein (L3, L5, L6, L7, L7A, L10A, L11, L13, L14, L15/27, L18, L22, L24, L39), ABC transporters with ATPase domain, glutamyl tRNA synthetase, translation initiation factor (elF-2B, elF-4A and EF-1), ubiquitin/40S ribosomal protein S27A and ribosomal proteins S18 and S7 were highly up regulated in replete and ball milled aspen growth substrates. Several transcripts encoding for amidases were found to be highly up regulated in nitrogen limited followed by carbon-limited growth substrates, no genes were found to be expressed above fold change ≥2.0 in GSE27941 dataset.

In GSE54542 dataset, most of the genes involved in translation, ribosomal structure and biogenesis were found to be down regulated in control samples with genes encoding for amidases (Phchr_3719 and 6738), mitotic cell division protein, mitochondrial ribosomal protein L16, NMD protein affecting stability, RNaseP, translation initiation factor 6, tRNA splicing endonuclease and tRNA methyl transferase (GCD10). However, in oak acetonic extractives samples genes encoding for 40S (S14, S26, SA) 60S acidic protein P1, 60S (L13, L22) ribosomal proteins, amidases, tRNA synthetase (asparaginyl, aspartyl, glutamyl, glycyl, leucyl, lysyl), mitochondrial ribosomal proteins (L28, S14/S29, polyadenylate binding protein, RNA binding protein, translation initiation inhibitor, translation initiation factor (elF-2c, elF-3, elF4E, elF-4G, elF-6), translation repressor MUT5/PUF4 and tRNA methyl transferases were highly up regulated with fold change ≥2.0. In dataset GSE52922, genes encoding for 40S (S10, S3A, S7, SA), 60S (L15/L27,

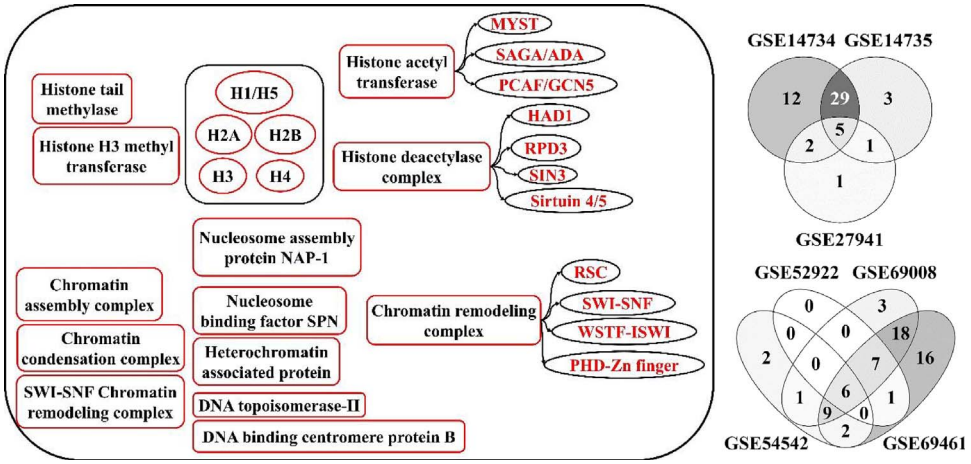


Fig. 5. List of commonly expressed and statistically significant genes among various datasets encoding for chromatin structure and dynamics group (KOG functional ID: B).

L3, L30, L39) ribosomal proteins, S4 and S7 ribosomal proteins were comparatively highly expressed in P717 growth substrate. Amidase, ABC transporters with ATPase domain, polyadenylate binding protein, translation initiation factor 4F and translation repressor MPT5/PUF4 were comparatively up regulated in 82 and 64 growth substrates (Fig. 3) (Supplementary Information).

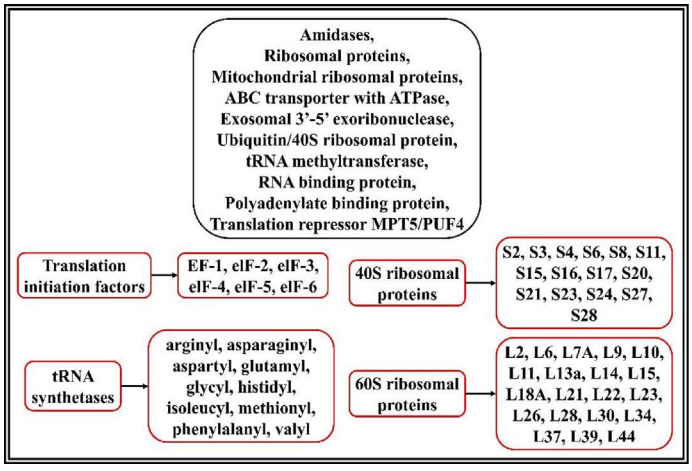
In GSE69008 dataset, most of the translation, ribosomal structure and biogenesis proteins were comparatively down regulated in A (high lignin-low glucose) and highly upregulated in C (average lignin-average glucose) samples. Genes encoding for 40S (S2, S3, S3A, S4, S6, S8, S15, S15/22, S23), 60S (L2/L8, L3, L5, L6, L7, L10, L10A, L13, L14, L15) and 60S acidic ribosomal protein P0, amidases, tRNA synthetase (asparaginyl, methionyl, threonyl), polyadenylate binding protein, ATP-dependent RNA helicase, translation initiation factor (EF-1, elF-1/SUI1, elF-2 alpha, beta, gamma, elf-4G, elf-5A and ubiquitin/40S ribosomal protein S27a were highly up regulated in C growth substrates. Although genes encoding for 40S (S2/S26, SA/P40), 60S (L15/L27, L19, L22, L44), 60S acidic ribosomal protein P2, tRNA synthetase (alanyl, aspartyl, glycyl, prolyl), ABC transporter with ATPase domain, Exosomal 3'-5' exoribonuclease complex, RNA binding protein (translation regulation), translation initiation factor (elF-2, elF-3a, EF-1, elF-2C, elF-3C, 3D, elF-4F) and tRNA isopentenyl pyrophosphate transferase were significantly up regulated in B (low lignin- high glucose) samples (Fig. 3) (Supplementary Information). Finally, in GSE69461 dataset a large set of genes encoding for translation, ribosomal structure and biogenesis were highly down regulated (with fold change ≥2.0) in 40-h samples. Several transcripts encoding for 40S and 60S ribosomal proteins, amidases, tRNA synthetases (asparaginyl, aspartyl, glutaminyl, glutamyl, isoleucyl, leucyl, lysyl, methionyl, phenylalanyl, tyrosyl), translation initiation factors, tRNA methyltransferases, ribosomal proteins, mitochondrial ribosomal proteins, ABC transporters, Exosomal 3'-5' exoribonuclease complex and WD40 nucleolar protein (translation) were highly up regulated in 96-h growth substrate (fold change ≥2.0). In 40-h growth substrates genes encoding for 40S (S2), 60S (P2, L10, L11, L30) ribosomal proteins, amidases, aspartyl, asparaginyl and cysteinyl tRNA synthetases and various translation initiation factors were found to be highly up regulated (Supplementary Information) (Fig. 6).

3.4. Transcription (KOG ID: K)

Previous studies have extensively determined that a wide range of transcription factors are involved in regulating the expression of lignocellulolytic enzymes [9,33–35]. Genome of *P. chrysosporium* encodes 415 gene models classified under the transcription group. Our present metadata analysis has revealed a large set of commonly expressed statistically significant genes coding for the essential transcription factors. We have observed that a total of 14 and 21 genes were found to be common among GSE14734, GSE14735, GSE27941 and GSE54542, GSE52922, GSE69008, GSE69461 datasets respectively. In GSE14734 dataset, genes encoding for CREB/ATF transcription factor, TATA box binding protein and transcription coactivator encoding genes were highly down regulated in both ball milled aspen and cellulose growth substrates, while comparatively highly expressed in glucose substrates. In dataset GSE14735, genes encoding for TATA box binding protein, HMG box transcription factors were found to be highly up regulated in carbon limited, ball milled aspen and nitrogen limited growth substrates, while genes encoding for GATA-4/5/6 transcription factors and elongation factor 1 were highly upregulated in cellulose containing growth substrates. And genes encoding for elongation factor 1 beta, protein arginine N-methyl transferases (PRMT1) related enzymes, RNA polymerase II were highly up regulated in replete growth substrates. In GSE27941 dataset, only genes encoding for TATA binding protein (RNA polymerase II) and component of TFIID and TFIIIB were found to be highly up regulated with fold change ≥2.0 in ball milled aspen samples, however genes encoding for a large set of transcription factors especially CREB/ATF, CCAAT (HAP5), class transcription repressor, HMG box, heat shock transcription factor were found to be statistically significant and highly up regulated in ball milled aspen with fold change ≥1.5 (Fig. 3) (Supplementary Information).

Interestingly we have observed that in dataset GSE54542, a large set of genes encoding for various transcription factors were highly up regulated in oak acetonic extractive than control samples. Transcripts encoding for CAAT-binding factor, CREB/P300, CREB/ATF, calcium responsive transcription coactivator, chromatin remodeling complex, GATA-4/5/6, HMG-box, heat shock (HSF), helix loop helix (HLH), KEKE containing transcription regulator, Mlx related, nuclear receptor coregulator, transcription factor DATF1, RFX, TATA- binding factor, forkhead/HNF3, CCR4, upstream (L-myc-2) transcription factors were highly up regulated in oak acetonic extractive samples. However, genes encoding for CCAAT (HAP2), Cdk activating kinase, HMG box, leucine permease transcription regulator, TBP associated transcription factor, transcription factors (MEIS1, TCF20, Myb, CCAAT displacement CDP1), transcription repressor, Zn-finger transcription factors were highly up regulated in control samples with fold change ≥2.0. In GSE52922 dataset, genes encoding for GATA-4/5/6, helix-loop-helix (HLH), HMG-box, NAD + ADP-ribosyltransferase Parp, RFX, RNA polymerase II, TATA box binding protein, TFIIF interacting CTD phosphatase, transcription coactivator, upstream (L-myc-2) transcription factors and ubinuclein nuclear proteins were highly up regulated in 82 growth samples.

In GSE69008 dataset we have observed that genes classified under transcription were highly up regulated in B (low lignin-high glucose) followed by C (average lignin-average glucose) samples. Genes encoding for calcium responsive transcription coactivator, casein kinase



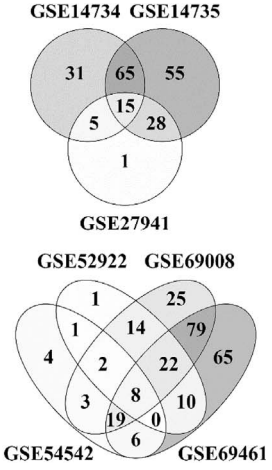


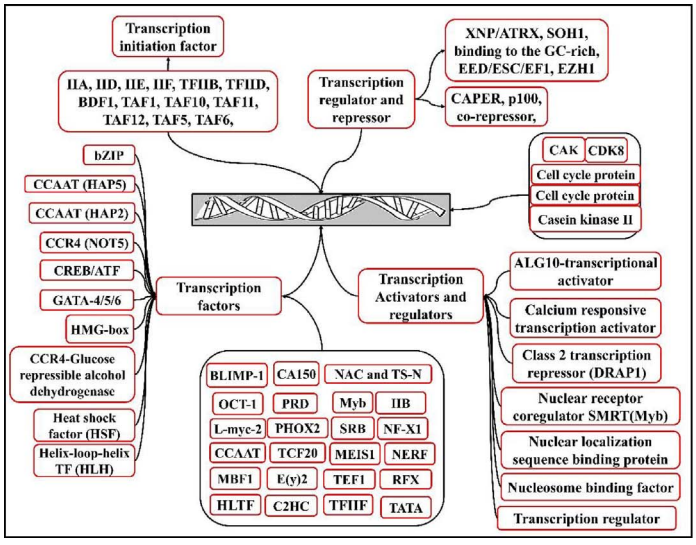
Fig. 6. List of commonly expressed and statistically significant genes among various datasets encoding for translation, ribosomal structure and biogenesis group (KOG functional ID: J).

II, cell cycle control protein, chromodomain helicase, dosage compensation regulatory complex, E3 ubiquitin ligase, GATA-4/5/6, glucose repressible alcohol dehydrogenase transcriptional effector (CCR4), heat shock (HSF), helix-loop-helix (HLH), transcription factors, negative regulation of transcription, nuclear receptor coregulator, transcription factors 5qNCA, RFX, OCT1, transcription coactivator(FOSB/c-Fos) and transcription initiation factor (TFIIB, TFIID) were comparatively highly up regulated in B (low lignin-high glucose). While genes encoding for HMG-box, KEKE-transcription regulatory, NAD + ADP-ribosyltransferase Parp, elongation factor 1, nuclear localization sequence binding protein, ubinuclein nuclear protein and upstream L-myc-2 transcription factors were down regulated in B samples. Total of 177 transcripts were highly regulated when cultured on *P. glauca* wood species of GSE69461 dataset, most of the genes encoding for transcription were found to be highly down regulated in 40-h samples and highly up regulated in 96- hour samples. Especially genes coding for transcription factors bZIP, CCAAT (HAP5), CCR4 (NOT5), GATA-4/5/6, leucine zipper transcription factors, Glucose-repressible alcohol dehydrogenase transcriptional effector, HSF, HLTF/DNA helicase, HMG-box, MADS-box, MOT2, RFX, TATA-box binding protein, BLIMP-1, CA150, E2F, MBF1, MEIS1, NERF, OCT-1, HNF3-forkhead, PHOX2, TCF20, Myb, CCAAT (CDP1) transcription factors and nuclear receptor coregulator (SMRT, Myb domains) were highly up regulated in 96-h samples. Simultaneously, wide range of transcripts encoding for alpha-1,2-glucosyltransferase, CCR4 (NOT5), cell cycle control protein, CREB/ATF, GATA-4/5/6, glucosyltransferase-Alg8p, HSF, HLH, HMG-box, homeobox, CCR4 associated factor, 5qNCA, FET5, RFX, TATA-box binding protein, MES1, NERF, NF-X1, HNF3-forkhead, PRD, TCF20, Myb, L-myc-2 transcription factors were highly up regulated in 40-h samples with fold change ≥2.0 (Fig. 3) (Supplementary Information) (Fig. 7).

3.5. Replication, repair and recombination (KOG ID: L)

DNA is the storehouse for the cellular genetic material from which other molecular components such as RNA and protein are derived. However, eukaryotic DNA experiences various DNA-damaging events every moment and the damaged DNA inhibits various necessary cellular mechanisms and pathways such as DNA replication and transcription, thus challenging the cellular fate [36]. Thus, eukaryotic cells have evolved with various repair mechanisms to reserve DNA from different types of damages. Eukaryotic cells employ five different types of repair mechanisms a) base excision repair, b) mismatch repair c) nucleotide excision repair d) homologous recombination and e) non-homologous end joining, state of the art review articles are already available on eukaryotic repair mechanisms [36]. As mentioned earlier *P. chrysosporium* genome encodes 242 genes classified under replication, repair and recombination group, based on the annotations these genes we have further tentatively grouped the genes based on their function as 60 (base excision repair), 20 (mismatch repair), 17 (nucleotide excision repair), 8 (homologous recombination) and 6 (non-homologous end joining). In GSE14734, GSE14735 and GSE27941 datasets, various genes encoding for replication, repair and recombination were expressed statistically significant, however no genes were expressed with fold change ≥2.0. Total of 52, 6, 5 genes were found to statistically significant and commonly expressed among GSE14734-GSE14735, GSE14734-GSE14735-GSE27941 and GSE14734 −GSE27941 datasets respectively (Supplementary Information). Most of the genes encoding for replication, repair and recombination were up regulated in cellulose grown samples followed by ball milled aspen. Transcripts encoding for 33 genes (5'-3' exonuclease (HKE1/RAT1, XRN1/KEM1/SEP1) ATP-dependent DNA helicase, DNA mismatch repair protein (MLH2, MutS), DNA polymerase, DNA repair protein RAD18, DNA licensing factor (MCM3, MCM4, MCM6, MCM7), DNA damage check point RHP9/CRB2/53BP1, Replication factor C, Signaling protein SWIFT and BRCT domain proteins, Single-stranded DNA-binding replication protein A) were found to be up regulated in glucose and down regulated in BMA and cellulose samples (Supplementary Information). Similarly, most of the replication, repair and recombination genes were up regulated in carbon and nitrogen limited growth conditions followed by replete medium (Supplementary Information).

In GSE54542 dataset, total of 45 genes were differentially expressed with a fold change ≥2.0, 28 genes (5'-3' exonuclease, ATP-dependent DNA helicase, Cdk activating kinase (CAK), DNA damage checkpoint protein RHP9/CRB2/53BP1, DNA mismatch repair protein − MLH1, DNA polymerase, Eukaryotic-type DNA primase, DNA repair protein XPA-interacting protein, Mismatch repair ATPase (MSH4, MSH5, MSH6), origin of recognition complex, DNA damage inducible protein, Replication factor C (RFC2, RFC4), SNF2 family DNA-dependent ATPase and Tam3-transpose) were highly up regulated in control. In oak acetonic extract samples, 17 genes were highly up regulated (3'-5' DNA helicase, 3'-5' exonuclease, 5'-3' exonuclease, ATPase related to holliday junction resolvase, damage specific DNA binding complex (DDB1), DNA polymerase, DNA repair and recombination protein (RAD52/RAD22, RHP57), DNA replication licensing factor (MCM2, MCM3, MCM4), endonuclease-III, HLTF/DNA helicase RAD5, NER factor (NEF2), origin recognition subunit1). In GSE52922 dataset, total of 24 replication, repair and recombination genes were found to be statistically significant out of which only 2 genes (DNA polymerase sigma, NAD+ ATP ribosyltransferase Parp) were significantly up



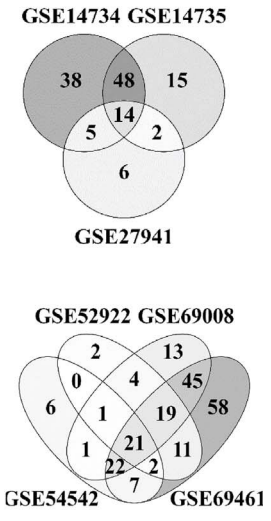


Fig. 7. List of commonly expressed and statistically significant genes among various datasets encoding for transcription group (KOG functional ID: K).

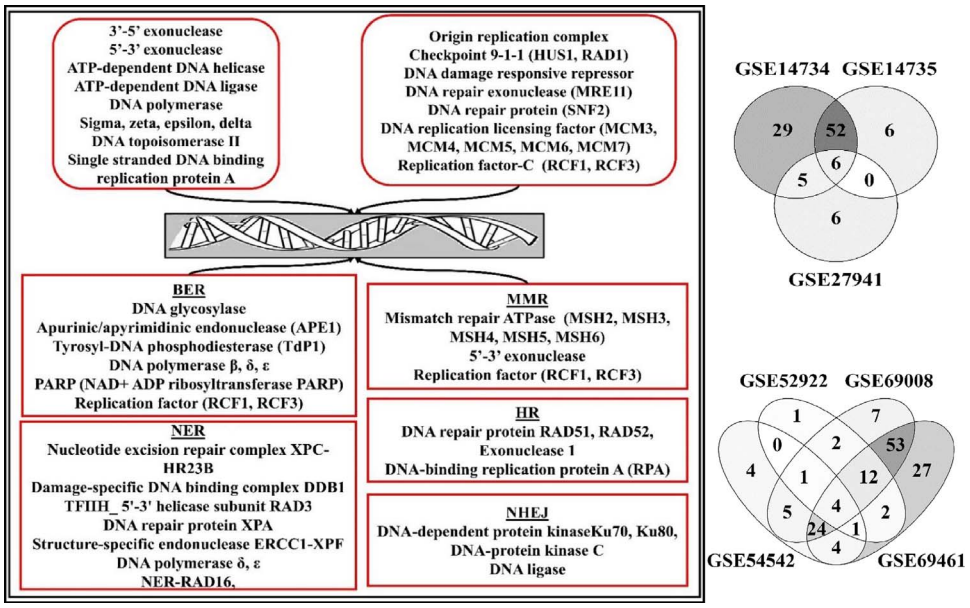


Fig. 8. List of commonly expressed and statistically significant genes (A), (B) among the selected gene expression datasets encoding for Replication, repair and recombination group (KOG functional ID: L).

regulated in 82 followed by P717 growth substrates (Supplementary Information). In GSE69008 dataset, genes encoding for 5'-3' exonuclease, transpose, ribonuclease HI, tam3-transposase (AC family) were significantly up regulated in A (high lignin-low glucose) and genes encoding for DNA topoisomerase I, nucleosome binding factor SPN, tam3-transposase were up regulated in B (low lignin-high glucose). In C (average lignin-average glucose) samples, 5'-3' exonuclease, NAD + ADP-ribosyltransferase Parp, nucleotide excision repair (NEF2), ribonuclease HI were up regulated and down regulated in A and B samples. Finally, in GSE69461 dataset total of 120 and 40 genes were found to be highly upregulated with a fold change ≥2.0. In 40-h samples genes encoding for 3-methyladenine DNA glycosidase, ATP-dependent helicase and ligase, damage specific DNA binding complex, DNA damage check point, damage inducible protein, DNA polymerase, topoisomerase I, G/T mismatch DNA glycosylase, DNA repair protein XPA, MSH2, NAD + ADP ribosyltransferase Parp were highly expressed. Genes involved in DNA replication and various repair mechanisms were found to be highly up regulated in 96-h samples (Supplementary Information) (Fig. 8).

4. Discussion

Extracellular digestion of the growth substrate by secreting highly oxidative and degradative enzymes, before the absorption of nutrients is one of the significant characteristics of fungi. White rot fungi such as *P. chrysosporium* secretes an array of oxidative and hydrolytic enzymes for the degradation of plant cell wall components, as a primary degrader of lignin. Thus, white rot fungi play a crucial role in maintaining the global carbon cycle [7]. However, the expression and turnover of *P. chrysosporium*'s lignocellulolytic enzymes are directly controlled by wide range of genes at DNA and RNA level. We have clearly listed some important set of genes encoding for different protein models classified under information, storage and processing in *P. chrysosporium's* genome (Fig. 9). The 30 million base pair genome of *P. chrysosporium* has revealed variation within complex gene families encoding for oxidases, peroxidases, cytochrome P450 monoxygenases and glycosyl hydrolases [7].

Maintaining the integrity and stability of eukaryotic genomic DNA is of primary importance to the cell, as the DNA damage compromises some of the essential cellular processes (transcription and replication) and sometimes it might even challenge cell's survival [36]. The genomic DNA experiences different types of damages such as endogenous or DNA replication related damage (DNA mismatches introduced at a rate of 10^{-4}–10^{-6}, chemically altered nucleotides ex: 8-oxo-dUTP and dGTP), exogenous or environmental damage (ultraviolet rays, ionizing radiations) and chemical damage (alkylating agents, heterocyclic amines, polycyclic aromatic hydrocarbons) [36]. Thus, eukaryotic cells have developed five different types of specific repair mechanisms (BER, MMR, NER, HR and NHEJ) [37]. Degradation of polyphenolic compounds (lignin) and other plant components present in the growth substrates (GSE54542, GSE52922, GSE69008 and GSE69461) might result in chemical damage to the DNA. Thus, resulting in higher expression of genes involved in DNA repair mechanisms. Various studies have confirmed the involvement of reactive oxygen species (ROS), hydroxy radicals (OH*) and hydrogen peroxide (H_2O_2) by ligninolytic peroxidases during the degradation of lignin [6,38–41]. Interaction of ROS, OH*, H_2O_2 and superoxide anion with DNA and other macromolecules results in various forms of DNA lesions challenging the cell survival by causing various detrimental effects [42]. According to Evan and Littlewood, multicellular eukaryotic organisms employ programmed cell death process known as apoptosis for the elimination of cells possessing irreparable damaged DNA [43]. However, unicellular organisms like Saccharomyces cerevisiae adapts the cells with irreparable damaged DNA and reenter into the cell cycle [44]. According to Glass et al., filamentous fungi might use apoptosis for removal of hyphal cells which have undergone anastomosis with a discordant pattern [45]. Recent studies conducted by Jung et al. on *Cryptococcus neoformans* (radiation resistant basidiomycetes fungus, usually found in radioactive habitats) have reported a special transcription factor Bdr1 containing basic leucine zipper domain which regulates the expression of genes coding for DNA repair genes [46]. However, the expression of Bdr1 is dependent on DNA damage response protein kinase (Rad53) [46]. Significantly common expression of genes encoding DNA repair protein (RAD1, RAD3, RAD5, RAD14, RAD16, RAD17, RAD18, RAD51 and RAD52) among the natural plant biomass growth substrates explains the occurrence of DNA lesions in *P. chrysosporium* (Supplementary information). Thus, DNA repair is a vital procedure for protecting the cellular genetic information.

DNA repair biochemical studies are usually performed using naked DNA substrates. However, this procedure is not physiologically applicable as the naked DNA is susceptible to nuclease digestion. Naked DNA will require more space and prone to other cellular insults until and unless it is organized [47]. Thus, eukaryotic naked DNA is structurally organized into nucleosomes (147 base pair DNA is packed

A		
	HMG-box	11
	HLTF	10
	GATA-4/5/6	5
	HST	4
	CCAAT(HAP5)	3
	CCR4-NOT5	3
	HLH	3
	C2HC (Zn finger)	3
	Myb	3
	C2H2-(Zn-finger)	1
	bZIP	1
	CCAAT (HAP3)	1
	CCAAT (HAP2)	1
	CREB	1
	CREB/ATF	1
	CCR4	1
	Alg8p	1
	HST	1
	HLH (BF/Olf-1)	1
	HMG-box (Capicua)	1
	TEA (r SPT4)	1
	TEA (TAT-SF1)	1
	TEA (TFIIS)	1
	Zn-finger TF	1

B		
	tRNA-Synthetase	42
	Translation Initiation factor	57
	Translation elongation factor	4
	Mitochondrial/chloroplast ribosomal proteins	28
	Mitochondrial ribosomal protein	10
	Exosomal 3'-5' exoribonuclease complex	12
	Elongation factor	6
	Amidases	13
	60S ribosomal protein subunits	51
	40S ribosomal protein subunits	31

C		
	Histone Acetyltransferases (HAT)	15
	Histone Deacetylases (HDAC)	16
	Histone Methyltransferases (HKMT)	5
	Histone Methylases	32
	Chromatin assembly complex (CAF)	4
	Chromatin remodeling complex (CRF)	14
	Chromosome condensation complex (CCC)	3

D		
	Base Excision Repair (BER)	60
	Mismatch Repair (MMR)	20
	Nucleotide Excision Repair (NER)	17
	Homologous Recombination (HR)	8
	Non-Homologous end Joining (NHEJ)	6

Fig. 9. Heat map (Color scales) listing the important groups of genes encoding for A) transcription factors B) translation ribosomal structure and biogenesis C) chromatin structure and dynamics D) DNA repair mechanisms in genome of *P. chrysosporium*.

around core histone proteins H2A, H2B, H3 and H4) which are further packed into chromatin. Subsequently, this chromatin condenses into chromosomes which enable accurate cellular division [48]. Chromatin's dynamic structure controls gene expression by controlling the nuclear processes such as DNA replication, transcription and DNA repair. During DNA replication, eukaryotic chromatin is exposed to proof reading and DNA repair mechanisms to ensure accurate transfer of genetic information. Present data analysis of the gene expression data revealed the genes involved in chromatin remodeling such as chromatin remodeling protein, histone acetyltransferase (MYST, SAGA/ADA, TRRAP), histone methyltransferases, chromatin remodeling complex (SWI-SNF, RSC, WSTF-ISWI and PHD Zn-finger), histone deacetylase complex (SIN3, RPD3). Along with these enzymes, genes coding for sirtuin 4, 5 (SIR2 family) structure maintenance of chromosome protein 4, nucleosome assembly protein and remodeling factors were common among the datasets. Based on our observations above-mentioned enzymes involved in chromatin structure, remodeling and dynamics were found to be highly expressed in natural plant biomass growth substrates (Supplementary information).

Gene regulation is of higher priority in physiology of all the organisms which ensures the up and down regulation of genes by responding to the growth conditions [9]. Studies have reported that 37 different classes of regulatory proteins were identified in fungi which control and coordinate fungal growth [9]. According to Benocci, Tiziano et al. [49], initial recognition of the growth substrate by fungi initiates the expression and secretion of plant biomass utilizing enzymes and the respective metabolic networks [49]. Plant cell wall components such as polysaccharides and lignin cannot act as inducers due to their large size as they cannot enter the fungal cell. Fungi identify the presence or absence of polymers by secreting polymer derived low molecular weight components (ex: mono (or) disaccharides). Majorly, the transcription factors regulating fungal plant cell wall degrading enzymes belong to zinc cluster family (contains Zn-fingers and cysteine/histidine residues). Studies have reported that most positive and negative regulators of fungal plant cell wall degrading enzymes belong to Zn2Cys6 and Cys2His2 classes [49]. Reports on commercially important fungi such as *Aspergillus nidulans, Trichoderma reesei, Neurospora crass, Fusarium sp, Colletotrichum gloeosporioides, Botrytis cinerea, Magnaportha oryzae, Aspergilus aculeatus* etc., have revealed about important transcription factors regulating the expression of plant cell wall polysaccharide utilizing enzymes (Table 3) [49].

Genes encoding for various transcription factors were statistically significant and commonly expressed among *P. chrysosporium* gene expression datasets. Genes encoding for transcription factors CREB/ATF, HMG-box, MADS-box, CCAAT (HAP5, HAP2), CCR4, calcium-responsive transcription coactivator, GATA/4/5/6 and L-myc2 were statistically significant and commonly expressed among all the datasets. Previous reports suggest that carbon catabolite repression (CCR) regulates the expression and secretion of plant cell wall utilizing enzymes when cultured on certain carbon sources [10,34,50–52]. Foreman et al. have revealed that genes encoding for plant cell wall utilizing enzymes were coregulated by XYR1, ACE2, HAP-2/3/5 complex (positive regulators) with repressors being ACE1 and CRE1 [53]. The cis acting element CCAAT motif is present on promoter and enhancer regions in most of the eukaryotic genes. In filamentous fungi, all the CCAAT box binding proteins were mostly classified under HAP-like factors [52,53]. Previous reports suggest that HAP-2/3/5 complex is involved in generation of open chromatin structure which subsequently required for the total transcriptional activation of cellulases [54,55]. Studies conducted on *Aspergillus* sp. have reported that CRE-A, CRE-B and CRE-C

Table 3
Transcription factors involved in regulation of plant cell wall utilizing enzymes studied in various commercially important fungi:.

Plant biomass component	Transcription factors	Enzymes
Cellulose	CLR-1, CLR-2, ClrA/1 (Cellulase regulators-Zn2Cys6-class) XlnR (Xylanolytic transcriptional activator-Zn2Cys6) ACE2, ACE3 (Activators of cellulase expression-Zn2Cys6) ClbR (Cellobiose response regulator-Zn2Cys6) McmA (MADS-box protein-MADS-box) HAP2, HAP3, HAP5 (Multimeric protein complex) ACE1, CreA/1, BglR/COL-26 (Cellulase enzyme repressors) CRE1, CREA, CREB, CREC	Cellulases
Hemicellulose	XlnR/XYR1 (Xylanolytic transcriptional activator-Zn2Cys6) AraR, ARA1 (Arabinose responsive regulators Zn2Cys6) GalR (GalX) (Galactose responsive regulator-Zn2Cys6)	Xylanases, Arabinases, Galactases
Starch, Inulin	InuR (Inulinolytic regulation-Zn2Cys6) AmyR, MalR (Amylolytic regulation-Zn2Cys6)	Inulases, Amylases, Maltases,
Pectin	RhaR, GaaR(GaaX) (Pectinolytic regulation-Zn2Cys6)	Rhamnosidases, Pectinases

are involved in carbon catabolite repression regulatory mechanisms [56–60]. According to de Vries, Ronald P, and Jaap Visser, expression of genes encoding for cellulase, xylanase, arabinase, and various endoxylanase, xylosidase, feruloyl esterase and few pectinase are affected by CRE-A mediated repression [50]. Significant expression of heat shock, HMG-box, helicase like, RFX, Myb, L-myc2 transcription factors only among the natural plant biomass growth substrates might reveal the increased cellular stress on *P. chrysosporium* during degradation of lignin and other plant extractives. According to Gettemy, Jessica. M et al. [61], expression of *P. chrysosporium* genes encoding for manganese peroxidase (MnP) is strongly regulated by manganese (Mn), heat shock and hydrogen peroxide (H_2O_2) when cultured on nitrogen limited growth substrates [19,61].

Before the process of translation in the cellular cytoplasm, messenger RNA is subjected to 5′ capping, mRNA-splicing, 3'end processing and mRNA export, these cellular events are strongly interdependent and influence cellular fate of a transcript [21]. Genes encoding for ATP-dependent RNA helicase, DEAH-box RNA helicase, fibrillarin related nucleolar RNA binding proteins, polyadenylation factor complex, splicing coactivator SRM 160/300, splicing factor RNPS1, hnRNP-F, 3b-subunit 4, mRNA splicing factor were statistically significant and commonly expressed among all the datasets (Supplementary Information). Serine-arginine nuclear matrix protein (SRM 160/300 protein) play crucial role in regulating alternative splicing, possibly by localizing splicing machinery components to the transcription active site. SR proteins are regulated by phosphorylation and dephosphorylation reactions as these are phosphoproteins in nature [62]. Higher expression of genes encoding for splicing factors in the natural plant biomass growth substrates might be due to its involvement in regulation of lignocellulolytic enzymes. Besides the abovementioned common enzymes genes encoding for various 40S and 60S ribosomal proteins, amidases, translation initiation factor 2C, 4F, PUF3, were expressed commonly among GSE14734, GSE14735 and GSE27941 datasets. In natural plant biomass substrates (GSE54542, GSE52922, GSE69008 and GSE69461) Exosomal 3'-5' exoribonuclease, polyadenylate binding protein, translation initiation inhibitor UK114, initiation factor 4F and repressor MPT5 proteins were commonly expressed. Studies have confirmed the involvement of the enzymes in regulation and maintenance of the protein turnover, however the exact involvement and role of the abovementioned enzymes in expression and regulation of lignocellulolytic and detoxification enzymes in *P. chrysosporium* is not explained till today.

In this study, we have revealed the common gene expression patterns of *P. chrysosporium* involved in regulation of protein expression and turnover of lignocellulolytic enzymes. We have extensively reported about various genes involved in information storage and processing of *P. chrysosporium.* Higher and significant expression of various genes encoding for information storage and processing especially DNA damage, repair and recombination mechanisms, mRNA splicing, histone acetyltransferases by *P. chrysosporium* were common and highly expressed among datasets cultured on natural plant biomass growth substrate. Thus, degradation of natural plant biomass containing lignin and other plant extractives along with plant polysaccharides causes various DNA level lesions. These results also convey that expression of lignocellulolytic genes internally depends on expression or repression of various genes involved in information storage and processing. However, further investigations must be performed to understand and analyze the exact involvement of the abovementioned genes in expression and regulation of *P. chrysosporium* lignocellulolytic enzymes. Application of next generation sequencing techniques such as ChIP (chromatin immunoprecipitation), RNA sequencing and mRNA splicing studies might reveal the functional involvement of the reported genes which would significantly help in strain improvement and production of recombinant strains.

Competing interests

The authors have declared that no competing interest exists.

Acknowledgement

This work was supported by Natural Sciences and Engineering Research Council of Canada Funding (RGPIN-2017-05366) to Wensheng Qin and Ontario Trillium Scholarship (OTS) to Ayyappa Kumar Sista Kameshwar. We would also like to thank Agilent for providing the trial license of GeneSpring v14.8 software.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.procbio.2017.10.007.

References

[1] O. Skyba, D. Cullen, C.J. Douglas, S.D. Mansfield, Gene expression patterns of wood decay fungi Postia placenta and Phanerochaete chrysosporium are influenced by wood substrate composition during degradation, Appl. Environ. Microbiol. (2016) 16–00134.
[2] A.V. Wymelenberg, J. Gaskell, M. Mozuch, S.S. BonDurant, G. Sabat, J. Ralph, O. Skyba, S.D. Mansfield, R.A. Blanchette, I.V. Grigoriev, Gene expression of wood decay fungi Postia placenta and Phanerochaete chrysosporium is significantly altered by plant species, Appl. Environ. Microbiol. (2011) 00508–00511.
[3] A.V. Wymelenberg, J. Gaskell, M. Mozuch, P. Kersten, G. Sabat, D. Martinez, D. Cullen, Transcriptome and secretome analyses of Phanerochaete chrysosporium reveal complex patterns of gene expression, Appl. Environ. Microbiol. 75 (12) (2009) 4058–4068.
[4] A.V. Wymelenberg, J. Gaskell, M. Mozuch, G. Sabat, J. Ralph, O. Skyba, S.D. Mansfield, R.A. Blanchette, D. Martinez, I. Grigoriev, Comparative transcriptome and secretome analysis of wood decay fungi Postia placenta and

Phanerochaete chrysosporium, Appl. Environ. Microbiol. 76 (11) (2010) 3599–3610.
[5] A.K.S. Kameshwar, W. Qin, Metadata analysis of phanerochaete chrysosporium gene expression data identified common CAZymes encoding gene expression profiles involved in cellulose and hemicellulose degradation, Int. J. Biol. Sci. 13 (1) (2017) 85–99.
[6] A.K.S. Kameshwar, W. Qin, Gene expression metadata analysis reveals molecular mechanisms employed by Phanerochaete chrysosporium during lignin degradation and detoxification of plant extractives, Curr. Genet. (2017) 1–18.
[7] D. Martinez, L.F. Larrondo, N. Putnam, M.D.S. Gelpke, K. Huang, J. Chapman, K.G. Helfenbein, P. Ramaiya, J.C. Detter, F. Larimer, Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78, Nat. Biotechnol. 22 (6) (2004) 695–700.
[8] I.V. Grigoriev, R. Nikitin, S. Haridas, A. Kuo, R. Ohm, R. Otillar, R. Riley, A. Salamov, X. Zhao, F. Korzeniewski, MycoCosm portal: gearing up for 1000 fungal genomes, Nucleic Acids Res. (2013) gkt1183.
[9] R.B. Todd, M. Zhou, R.A. Ohm, H.A. Leeggangers, L. Visser, R.P. De Vries, Prevalence of transcription factors in ascomycete and basidiomycete fungi, BMC Genomics 15 (1) (2014) 214.
[10] A. Amore, S. Giacobbe, V. Faraco, Regulation of cellulase and hemicellulase gene expression in fungi, Curr. Genomics 14 (4) (2013) 230–249.
[11] P. Hortschansky, H. Haas, E.M. Huber, M. Groll, A.A. Brakhage, The CCAAT-binding complex (CBC) in Aspergillus species, Biochimica et Biophysica Acta (BBA)-Gene Regul. Mech. 1860 (5) (2017) 560–570.
[12] G. Brosch, P. Loidl, S. Graessle, Histone modifications and chromatin dynamics: a focus on filamentous fungi, FEMS Microbiol. Rev. 32 (3) (2008) 409–439.
[13] A. Lusser, J.T. Kadonaga, Chromatin remodeling by ATP-dependent molecular machines, Bioessays 25 (12) (2003) 1192–1200.
[14] A. Eberharter, P.B. Becker, ATP-dependent nucleosome remodelling: factors and functions, J. Cell Sci. 117 (17) (2004) 3707–3711.
[15] B.M. Turner, Defining an epigenetic code, Nat. Cell Biol. 9 (1) (2007) 2–6.
[16] J. Dobosy, E. Selker, Emerging connections between DNA methylation and histone acetylation, Cell. Mol. Life Sci. 58 (5) (2001) 721–727.
[17] I. Steinfeld, R. Shamir, M. Kupiec, A genome-wide analysis in Saccharomyces cerevisiae demonstrates the influence of chromatin modifiers on transcription, Nat. Genet. 39 (3) (2007) 303–309.
[18] J.L. Woolford, S.J. Baserga, Ribosome biogenesis in the yeast Saccharomyces cerevisiae, Genetics 195 (3) (2013) 643–681.
[19] M.H. Gold, M. Alic, Molecular biology of the lignin-degrading basidiomycete Phanerochaete chrysosporium, Microbiol. Rev. 57 (3) (1993) 605–622.
[20] A.K.S. Kameshwar, W. Qin, Metadata Analysis of Phanerochaete chrysosporium gene expression data identified common CAZymes encoding gene expression profiles involved in cellulose and hemicellulose degradation, Int. J. Biol. Sci. 13 (1) (2017) 85.
[21] S. Hocine, R.H. Singer, D. Grünwald, RNA processing and export, Cold Spring Harbor perspectives in biology 2 (12) (2010) a000752.
[22] S. Shuman, Structure, mechanism, and evolution of the mRNA capping apparatus, Prog. Nucleic Acid Res. Mol. Biol. 66 (2000) 1–40.
[23] H. Busch, R. Reddy, L. Rothblum, Y. Choi, SnRNAs, SnRNPs, and RNA processing, Annu. Rev. Biochem 51 (1) (1982) 617–654.
[24] E. Wahle, U. Rüegsegger, 3′-End processing of pre-mRNA in eukaryotes, FEMS Microbiol. Rev. 23 (3) (1999) 277–295.
[25] H. Stark, R. Lührmann, Cryo-electron microscopy of spliceosomal components, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 435–457.
[26] E.M. Bradbury, Chromatin structure and dynamics: state-of-the-art, Mol. Cell 10 (1) (2002) 13–19.
[27] J. Zlatanova, S.H. Leuba, Chromatin Structure and Dynamics: State-of-the-art, Gulf Professional Publishing, 2004.
[28] T. Kouzarides, Chromatin modifications and their function, Cell 128 (4) (2007) 693–705.
[29] J. Widom, Structure, dynamics, and function of chromatin in vitro, Annu. Rev. Biophys. Biomol. Struct. 27 (1) (1998) 285–327.
[30] P. Tessarz, T. Kouzarides, Histone core modifications regulating nucleosome structure and dynamics, Nat. Rev. Mol. Cell Biol. 15 (11) (2014) 703–708.
[31] J. Strauss, Y. Reyes-Dominguez, Regulation of secondary metabolism by chromatin structure and epigenetic codes, Fungal Genet. Biol. 48 (1) (2011) 62–69.
[32] S.J. Brown, M.D. Cole, A.J. Erives, Evolution of the holozoan ribosome biogenesis regulon, BMC Genomics 9 (1) (2008) 442.
[33] M. Ilmen, C. Thrane, M. Penttilä, The glucose repressor genecre1 ofTrichoderma: isolation and expression of a full-length and a truncated mutant form, Mol. Gen. Genet. 251 (4) (1996) 451–460.
[34] G.J. Ruijter, J. Visser, Carbon repression in Aspergilli, FEMS Microbiol. Lett. 151 (2) (1997) 103–114.
[35] G.P. Borin, C.C. Sanchez, E.S. Santana, G.K. Zanini, R.A.C. Santos, A.O. Pontes, A.T. Souza, Dal R.M.M.T.S, D.M. Riaño-Pachón, G.H. Goldman, Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by Aspergillus niger and Trichoderma reesei, BMC Genomics 18 (1) (2017) 501.
[36] T.S. Dexheimer, DNA Repair Pathways and Mechanisms, DNA Repair of Cancer Stem Cells, Springer, 2013, pp. 19–32.
[37] F. Altieri, C. Grillo, M. Maceroni, S. Chichiarelli, DNA damage and repair: from molecular mechanisms to health implications, Antioxid. Redox Signaling 10 (5) (2008) 891–938.
[38] P.A. Belinky, N. Flikshtein, S. Lechenko, S. Gepstein, C.G. Dosoretz, Reactive oxygen species and induction of lignin peroxidase in Phanerochaete chrysosporium, Appl. Environ. Microbiol. 69 (11) (2003) 6500–6506.
[39] M. Morel, A.A. Ngadin, J.-P. Jacquot, E. Gelhaye, Reactive oxygen species in Phanerochaete chrysosporium relationship between extracellular oxidative and intracellular antioxidant systems, Adv. Bot. Res. 52 (2009) 153–186.
[40] P.A. Belinky, N. Flikshtein, C.G. Dosoretz, Induction of lignin peroxidase via reactive oxygen species in manganese-deficient cultures of Phanerochaete chrysosporium, Enzyme Microb. Technol. 39 (2) (2006) 222–228.
[41] L. Zacchi, J.M. Palmer, P.J. Harvey, Respiratory pathways and oxygen toxicity in Phanerochaete chrysosporium, FEMS Microbiol. Lett. 183 (1) (2000) 153–157.
[42] E.I. Azzam, J.-P. Jay-Gerin, D. Pain, Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury, Cancer Lett. 327 (1) (2012) 48–60.
[43] G. Evan, T. Littlewood, A matter of life and cell death, Science 281 (5381) (1998) 1317–1322.
[44] D.P. Toczyski, D.J. Galgoczy, L.H. Hartwell, CDC5 and CKII control adaptation to the yeast DNA damage checkpoint, Cell 90 (6) (1997) 1097–1106.
[45] N.L. Glass, D.J. Jacobson, P.K. Shiu, The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi, Annu. Rev. Genet. 34 (1) (2000) 165–186.
[46] K.-W. Jung, D.-H. Yang, M.-K. Kim, H.S. Seo, S. Lim, Y.-S. Bahn, Unraveling fungal radiation resistance regulatory networks through the genome-wide transcriptome and genetic analyses of cryptococcus neoformans, mBio 7 (6) (2016) (e01483-16).
[47] J.A. Downs, M.C. Nussenzweig, A. Nussenzweig, Chromatin dynamics and the preservation of genetic information, Nature 447 (7147) (2007) 951.
[48] K. Mekhail, D. Moazed, The nuclear envelope in genome organization, expression and stability, Nature reviews, Mol. Cell Biol. 11 (5) (2010) 317.
[49] T. Benocci, M.V. Aguilar-Pontes, M. Zhou, B. Seiboth, R.P. Vries, Regulators of plant biomass degradation in ascomycetous fungi, Biotechnol. Biofuels 10 (1) (2017) 152.
[50] R.P. de Vries, J. Visser, Aspergillus enzymes involved in degradation of plant cell wall polysaccharides, Microbiol. Mol. Biol. Rev. 65 (4) (2001) 497–522.
[51] N. Aro, T. Pakula, M. Penttilä, Transcriptional regulation of plant cell wall degradation by filamentous fungi, FEMS Microbiol. Rev. 29 (4) (2005) 719–739.
[52] C.P. Kubicek, M. Mikus, A. Schuster, M. Schmoll, B. Seiboth, Metabolic engineering strategies for the improvement of cellulase production by Hypocrea jecorina, Biotechnol. Biofuels 2 (1) (2009) 19.
[53] P.K. Foreman, D. Brown, L. Dankmeyer, R. Dean, S. Diener, N.S. Dunn-Coleman, F. Goedegebuur, T.D. Houfek, G.J. England, A.S. Kelley, Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei, J. Biol. Chem. 278 (34) (2003) 31988–31997.
[54] S. Zeilinger, M. Schmoll, M. Pail, R.L. Mach, C.P. Kubicek, Nucleosome transactions on the Hypocrea jecorina (Trichoderma reesei) cellulase promoter cbh2 associated with cellulase induction, Mol. Genet. Genomics 270 (1) (2003) 46–55.
[55] M. Bolotin-Fukuhara, Thirty years of the HAP2/3/4/5 complex, Biochimica et Biophysica Acta (BBA)-Gene Regul. Mech. 1860 (5) (2017) 543–559.
[56] H.N. Arst, D.J. Cove, Nitrogen metabolite repression in Aspergillus nidulans, Mol. Gen. Genet. 126 (2) (1973) 111–141.
[57] C. Dowzer, J.M. Kelly, Analysis of the creA gene, a regulator of carbon catabolite repression in Aspergillus nidulans, Mol. Cell. Biol. 11 (11) (1991) 5701–5709.
[58] M. Hynes, J.M. Kelly, Pleiotropic mutants ofAspergillus nidulans altered in carbon metabolism, Mol. Gen. Genet. 150 (2) (1977) 193–204.
[59] R.A. Lockington, J.M. Kelly, Carbon catabolite repression in Aspergillus nidulans involves deubiquitination, Mol. Microbiol. 40 (6) (2001) 1311–1321.
[60] R. Todd, R. Lockington, J. Kelly, The Aspergillus nidulans creC gene involved in carbon catabolite repression encodes a WD40 repeat protein, Mol. Gen. Genet. 263 (4) (2000) 561–570.
[61] J.M. Gettemy, D. Li, M. Alic, M.H. Gold, Truncated-gene reporter system for studying the regulation of manganese peroxidase expression, Curr. Genet. 31 (6) (1997) 519–524.
[62] A. Eldridge, R. Issner, Y. Li, E. Reifenberg, P.A. Sharp, B.J. Blencowe, Role of the SRm160/300 splicing coactivator in exon-enhancer function, Biochem. Cell Biol. 77 (4) (1999) 393.
[63] A. Thuillier, K. Chibani, G. Belli, E. Herrero, S. Dumarçay, P. Gérardin, A. Kohler, A. Deroy, T. Dhalleine, R. Bchini, Transcriptomic responses of Phanerochaete chrysosporium to oak acetonic extracts: focus on a new glutathione transferase, Appl. Environ. Microbiol. 80 (20) (2014) 6316–6327.
[64] A.V. Wymelenberg, J. Gaskell, M. Mozuch, S.S. BonDurant, G. Sabat, J. Ralph, O. Skyba, S.D. Mansfield, R.A. Blanchette, I.V. Grigoriev, Significant alteration of gene expression in wood decay fungi Postia placenta and Phanerochaete chrysosporium by plant species, Appl. Environ. Microbiol. 77 (13) (2011) 4499–4507.
[65] J. Gaskell, A. Marty, M. Mozuch, P.J. Kersten, S.S. BonDurant, G. Sabat, A. Azarpira, J. Ralph, O. Skyba, S.D. Mansfield, Influence of Populus genotype on gene expression by the wood decay fungus Phanerochaete chrysosporium, Appl. Environ. Microbiol. 80 (18) (2014) 5828–5835.
[66] P. Korripally, C.G. Hunt, C.J. Houtman, D.C. Jones, P.J. Kitin, D. Cullen, K.E. Hammel, Regulation of gene expression during the onset of ligninolytic oxidation by phanerochaete chrysosporium on spruce wood, Appl. Environ. Microbiol. 81 (22) (2015) 7802–7812.