

Genomic and transcriptome analyses of a thermophilic bacterium *Geobacillus stearothermophilus* B5 isolated from compost reveal its enzymatic basis for lignocellulose degradation

Mengmeng Wang

Nanjing Agricultural University

Jiaxi Miao

Nanjing Agricultural University

Xuanqing Wang

Nanjing Agricultural University

Tuo Li

Nanjing Agricultural University

Han Zhu

Nanjing Agricultural University

Dongyang Liu (✉ liudongyang@njau.edu.cn)

Nanjing Agricultural University <https://orcid.org/0000-0001-7452-9423>

Qirong Shen

Nanjing Agricultural University

Research

Keywords: *Geobacillus stearothermophilus*, thermophilic composting, cellulose-degrading, comparative genomics, transcriptome

Posted Date: March 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-18489/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at *Microorganisms* on September 4th, 2020. See the published version at <https://doi.org/10.3390/microorganisms8091357>.

Abstract

Background Composting is a special global carbon cycle which sustains various microbes engendering the cellulose degradation. Studies have obtained substantial compost microbiomes, yet the expressions and functions of these lignocellulolytic enzymes remains obscure. Thus, the discovery of thermophilic microorganisms as considerable biochemical catalysts for biofuels is becoming more and more attractive.

Results A lignocellulose degrading strain isolated from thermophilic compost was identified as *Geobacillus stearothermophilus* B5, which could secrete considerable enzymes at the optimal temperature (60°C) and pH (7.5). One single contig of 3.37 Mbp was obtained from raw data and 3371 protein-coding genes were predicted, and the clusters of orthologous groups (COG) analysis revealed various genes with function of polymeric substrates degradation, especially for abundant CAZymes including glycoside hydrolases (GH, 29%) and glycosyl transferases (GT, 36%). Furthermore, the transcriptional responses of B5 at different temperature by using rice straw as sole carbon sources were also analyzed, based on which the mechanism of lignocellulose degradation at high temperature was revealed that B5 could resist the heat by up-regulating the heat shock proteins (HSP) and then secrete various Carbohydrate-Active enzymes (CAZymes) to realize energy balance.

Conclusions The comparative whole-genome along with transcriptome analysis indicated that *G. stearothermophilus* B5 owned the ability of lignocellulose degradation and could be considered as a potential inoculant in composting efficiency, thus are also valuable for the lignocellulosic bioenergy industry.

Background

The alternative production strategy has drawn more attention due to the increasing consciousness about the social sustainability, which lead to the development of renewable fuels. Lignocellulosic biomass which composed of cellulose, hemicellulose, and lignin are the most abundantly available carbon sources on the earth, thus considered as a potential feedstocks for the production of biofuels [1]. However, due to the complex and recalcitrant structure of plant cell walls, the process of regenerating these feedstocks are usually complex resulting the high cost. It is necessary and urgent to improve the efficiency of converting the feedstocks into useful biofuels. As the similar biochemical catalyzing system with biofuel production, composting was an economic and social adopted method, by which the organic solid wastes could be decomposed and transformed into stable and humus-like substances. Microorganisms especially fungi and bacteria can efficiently break down the linkage of organic matters, thus play key roles during the composting process [2]. In the recent years, composting itself has developed from an approach of agricultural wastes management [3] to a tool of mining novel microbes and enzymes for the conversion of lignocellulosic biomass to biofuels [4]. The discovery of efficient microbes and enzymes especially those with special functions at extreme conditions, such as heat, high-solid substrate and low moisture content, which will significantly reduce the production cost of advanced biofuels from lignocellulosic

biomass [5]. Though much researches have focused on fungi (*Aspergillus*, *Trichoderma*, *Penicillium*, et al.) earlier, the bacterium gradually attracted intensive attentions due to their resistance to the environmental extremes, faster growth, and expression of multienzyme complexes [6].

Thermophiles being active in thermal environment offer great potential for the discovery of industrially relevant enzymes, which degrade the complex organic substances by secreting abundant hydrolytic enzymes into the outside environment under self-heating, moist and aerobic environment [7]. The lignocellulolytic enzymes are mainly composed of cellulases and hemicellulases, and these catalytically enzymes can synergistically break down the cellulose and the hemicellulose molecules into its constituent monomers, separately [8]. In the past decades, a great deal of work has been done on isolating and engineering thermophilic microorganisms for the bioconversion of lignocellulosic substrates, however, the suitable saccharolytic organisms even from the most extreme and remote spots on earth are still difficult to meet the demands for the glucoamylase production [9]. Thus, various versatile bacterium with considerable thermo resistance has aroused great interests, especially for the genus of *Geobacillus*. The characteristics of simultaneously fermenting C5 and C6 sugars and thermostable enzymes by *Geobacillus* spp. is very attractive for their application in biofuel industry [10]. Xylanolytic activities were commonly detected in *Geobacillus* strains, and the responsible hemicellulose utilization locus associating with the hemicellulose metabolism have been emphasized in previous study [11]. Meanwhile, *Geobacillus* strains also exhibited considerable cellulolytic decomposition capacity [12]. Thus, many industrially important enzymes including glycoside hydrolases, oxidase and esterase could be excavated from *Geobacillus* spp. [13].

Despite of these advantages, the scientific research regarding this genus is still far from enough, and only limited whole genome sequence are available for public. The lack knowledge of *Geobacillus* species has definitely restrict the application of this genus as platform organisms, and more detailed genomic information, physiological and biochemical data are urgently needed. Here, we describe a complete genome sequence of *Geobacillus stearothermophilus* B5, along with its transcriptome response at different temperatures by using rice straw as sole carbon sources. The results will convey meaningful and unique information to the readers, and enlighten them about the importance and potential of *Geobacillus* spp. in the lignocellulosic bioenergy industry.

Results And Discussion

Identification and basic features of strain B5

The characterization of B5 strain was evaluated, and the colony was light yellow convex subtransparent with regular edges, sticky and 1-2 mm in diameter after incubation for 24 hours at 60 °C on LB plates (Fig. S1). The Congo red assay is a qualitative detection of reducing sugars, and commonly used to estimate cellulolytic activities. Here, clear zones were observed around the colonies on CMC-minimal media (Fig. 1A), showing remarkable cellulolytic activities. Under the microscope, it was shown that B5 was motile, Gram-positive, spore-forming, rod-shaped cell, and the oval spores were located terminally within a

swollen sporangium (Fig. S1). The size of the cell was from 4 to 6 μm in length, 0.5 to 1 μm in diameter with peritrichous flagella (Fig. 1B). 16S rRNA sequence of this strain indicated that B5 strain was closely related to members of *Geobacillus*. For *Geobacillus*, it has been demonstrated that *recN* gene is the most robust marker for assigning new bacterial strains at the species level [14], and homology search revealed that B5 strain was a member of genus *Geobacillus*, showing highest similarity (99.71%) to *G. stearothermophilus* DSM 458. Phylogenetic tree further indicated that B5 formed a distinct linkage with DSM 458 with 100% bootstrap support (Fig. S2). Combining the above results with physiological and biochemical characteristics (data not shown), B5 was identified as *G. stearothermophilus* B5 and the sequence data were stored in the NCBI SRA database (accession number CP034952).

The carbon source test results showed that B5 could utilize various carbon sources, including but not limited to sugars, amino acids, hexose acids, carboxylic acids, esters and fatty acids. The positive test included D-glucose, sucrose, D-mannose, D-cellobiose, D-mannitol, D-fructose, glycerol, D-turanose, inosine, L-pyroglutamic acid, pectin, glucuronamide, L-malic acid, L-lactic acid, and so on (Additional file 1 and Fig. S3). During composting process, microorganisms are responsible for transforming the organic matter into biomass, CO_2 , heat and humus-like end-products, and the broad carbon utilization illustrated that B5 could take advantage of kinds of carbon source during composting. B5 can grow at a broad temperature ranging from 40 $^{\circ}\text{C}$ to 73 $^{\circ}\text{C}$, and the optimal temperature is between 55 $^{\circ}\text{C}$ and 65 $^{\circ}\text{C}$ (Fig. 1C). When the temperature raised above 55 $^{\circ}\text{C}$, a mass of pathogens were killed, thus thermophilic composting is widely adopted for industrial application, and there is no doubt that exploring microorganisms with thermo-resistance is necessary and urgent. Here, the B5 strain just approached its optimal growth condition at 55 $^{\circ}\text{C}$, and then it could degrade fats, celluloses, hemicelluloses and some lignin efficiently. Generally, microbial activity even for most thermophiles declined rapidly at temperatures above 63 $^{\circ}\text{C}$, while B5 still maintained considerable activity even at 65 $^{\circ}\text{C}$, indicating that B5 might play an important role during thermophilic composting process. As one of the critical parameters, the pH values ranging from 6.7 to 9.0 are convenient for most composting microbes, and the optimum pH are between 5.5 and 8.0 [15]. Interestingly, B5 can survive with the pH values ranging from 4.0 to 9.5, and the optimum growing pH value (Fig. 1D) is similar with the pH value for best composting efficiency. Furthermore, B5 could survive with the NaCl concentration from 0 to 3.5% (Fig. 1E). The growth curve of B5 under the optimal conditions was shown in Fig. 1F, and the logarithmic growth period fell within 6 to 12 h.

Determination of various enzyme activities

Different enzyme activities were detected, and the CMCase activity kept increasing and reached the peak ($0.32 \pm 0.02 \text{ U ml}^{-1}$) on the 6th day (Fig. 2A) and then decrease until the end. The xylanase activity increased sharply and the peak ($0.14 \pm 0.01 \text{ U ml}^{-1}$) was obtained at the 3rd day (Fig. 2B). The tendency of α -amylase was similar with that of CMCase, and the highest value was obtained on the 6th day ($0.43 \pm 0.02 \text{ U ml}^{-1}$) (Fig. 2C). The protease activity increased sharply like that of xylanase until reaching to $0.62 \pm 0.03 \text{ U ml}^{-1}$ at the 3rd day, and then decreased gradually to the end (Fig. 2D). Factually, the

biodegradation of lignocellulosic biomass in the composting process required the synergism of various enzymes including cellulase, hemicellulase, urease and protease [16]. It is known that filamentous fungi such as *Aspergillus spp.*, *Trichoderma spp.* can secrete abundant extracellular hydrolytic enzymes with higher activities, while these mesophilic microbes and its enzyme activities tend to reduce or even completely lose at thermophilic phase [6, 17]. Interestingly, enzymes produced by thermophilic bacteria are usually more thermostable. Tai et al. [18] found that *Geobacillus* sp. could secrete CMCase retaining 90% activity after 1 h of incubation at 70 °C, and similar results were observed in another *Geobacillus* strain in which cellulase activity kept 100% stable after 24 h incubation at 60 °C [19]. As the member of *Geobacillus spp.*, B5 also could secrete various enzymes with considerable thermostability. Furthermore, cell-bound effect always existed in some enzymes [20], if so the actual organic matters degrading efficiency would be even higher than that measured in cell-free supernatant. A bottleneck for B5 strain is the low yields of thermophilic cells related to their growth, and the highest OD600 value of B5 was only around 1.2 even in the optimal conditions (Fig. 1H). In fact, this is also the bottleneck for most of the thermophiles, and some measures have been taken to improve the cell yields thus increasing enzyme production, such as medium composition, process configuration and special equipment [21]. For example, the cellulase production of some *Geobacillus* sp. was 2-fold increased by optimizing the culture conditions with additions of ammonium sulfate and yeast extract [22]. Thus, more research about B5 is necessary so as to increase the cell yield and its enzyme productions.

Genomic analysis of *G. stearothermophilus* B5 and comparison of COG categories

The *G. stearothermophilus* B5 genome analysis results revealed a GC content of 52.46%, and a single contig of total 3.37 Mbp with 3371 CDS, 32 rRNA, and 90 tRNA (Fig. 3A). The gene length/genome ratio was 85% and the intergenetic region length/genome ratio was 15%. Genes were then annotated with different databases as follows: COG (2472), GO (2363), NR (3342), Swiss-Prot (2617), KEGG (1797), and CAZy (100). B5 strain was compared to the other four *Geobacillus* strains which were previously reported to own considerable capacity of cellulase or hemicellulase productions, and the genome features of the five strains are presented in Table 1. The number of orthologous genes between B5 strain and other four *Geobacillus* strains were 2615 (HTA426), 2414 (NG80-2), 2417 (NBRC 101842) and 2602 (Y412MC52), and the core genome of the five strains consists of 2202 orthologous genes and the pan-genome pool consists of 6175 genes, among which 465 special genes were unique for B5 strain (Fig. 3B). Conserved genes and gene pools were always used to evaluate the variation of the genus families. Zhang et al. [23] compared the genomic of five *Bacillus* strains (four *Bacillus amyloliquefaciens* and one *Bacillus subtilis*), and the results showed that 73.9% conserved genes and 5643 genes pools were observed in five *Bacillus* strains, which indicated low variation of *Bacillus amyloliquefaciens*. In this study, 65.3% conserved genes and 6175 genes pools were observed when comparing to these five *Geobacillus* strains, which indicated abundant variations for *Geobacillus* strains, and the variations of this specie might be due to the thermophilic environment.

Besides, 2472 genes in B5 strain were annotated to 1581 COGs, and all available CDSs from five *Geobacillus* strains were assigned into 20 COGs functional categories (Fig. 3C), and there were fewer

differences in most of the cellular processes and signaling categories (D to V). These might be due to that those five strains belong to the same genus, and the major functional models were conserved. The major differences between these five strains were observed in metabolism categories (C to Q), especially in carbohydrate transport and metabolism (G). In order to deeply comprehend the potential roles of B5 strain during composting, specific COGs involved in carbon catabolic functions were also analyzed. Amino acid transport and metabolism, and carbohydrate transport and metabolism functions represented 10.76% and 7.24% of the COG categories, respectively (Additional file 2). With respect to amino acid transport and metabolism function, the top five abundant COGs were permeases of the major facilitator superfamily (COG0477), permeases of the drug/metabolite transporter superfamily (COG0697), aminotransferase (COG0436), deacetylase (COG0624), and lyase (COG0346); while the top five abundant COGs for carbohydrate transport and metabolism were permeases of the major facilitator superfamily (COG0477), permeases of the drug/metabolite transporter superfamily (COG0697), deacetylase (COG0726), glycosidase (COG0366), and phosphotransferase system IIC component (COG1263). COG0477 participated in encoding permeases of the major facilitator superfamily, which could catalyze the transport of kinds of substrates including carbohydrate, lipids, peptides, nucleotides, and some other molecules at the thermophilic condition [24]. COG0366 was versatile which could not only encode glycosidase responsible for the release of aromatic compounds [25], but also α -amylase which could destroy the alpha bonds between long-chain polysaccharides, like glycogen and starch [26]. Altogether, the deep investigation of COG categories indicated that B5 strain owned a good potential for degrading proteins and carbohydrates during composting. The extensive diversity of gene functions revealed considerable potential for *G. stearothermophilus* B5 in organic substance decomposition in the composting system.

CAZyme family analysis of B5 strain genome

CAZymes can cleave, build and rearrange oligo- and polysaccharides, which play important roles in bacteria and are vital for optimizing biomass degradation [27]. The degradation capacity of B5 strain during the composting process was revealed through the gene annotation against the CAZy database. B5 strain encoded 100 CAZymes which unevenly distributed between glycoside hydrolases (GH, 29.0%), glycosyl transferases (GT, 36.0%), carbohydrate esterases (CE, 20.0%), auxiliary activities (AA, 4.0%) and carbohydrate-binding modules (CBM, 11.0%) (Fig. S4). The GH and GT family members take up the largest share and fulfil vital functions in the cleavage of polymeric substrates [28].

The GH family enzymes could hydrolyze the glycosidic bond between two carbohydrates or a carbohydrate and a non-carbohydrate moiety. Under thermophilic situation, GH members in B5 strain including cellulase (GH1, GH3, GH31), amylase (GH13), chitinase (GH18), together with some kinds of peptidoglycan hydrolase and oligosaccharide degrading enzymes were significantly up-regulated (Additional file 3). The genes encoding α -amylase (EC 3.2.1.1) were widely detected in B5 strain, which was considered as a crucial amylase. The biomass degradation pathways that GH families in B5 strain participated in included glycolysis (ko00010) and starch and sucrose metabolism (ko00500). The GH1 members encoded by B5 strain are 6-phospho- β -glycosidases, which were extremely thermostable and

almost lost no activity after incubation at 60 °C for 7 days, and they could utilize cellobiose [29]. One GH4 member encoding 6-phospho- α -glucosidase and one GH5 member encoding endo-1, 4- β -glucanase also had obvious effect in cellulose degradation. Besides, B5 strain contained a significantly higher number of ten GH13 family genes, which participated in starch hydrolyzation mainly [30]. There were also some CEs detected in B5 strain, which exhibited potential in de-acetylating xylan and xylooligosaccharide. A CE3 discovered in *T. reesei* previously could encode acetyl xylan esterase, thus enhanced the solubilization of xylans [31]. Moreover, a CE7 from *Thermoanaerobacterium* sp. was also validated to degrade xylan [32]. The existence of these genes may be important for B5 strain during the biodegradation process of cellulose and hemicellulose. Valuably, six CE4 members associated with the destruction of plant polysaccharides were also detected in B5 strain. These CE4 were acetyl xylan esterases which could catalyze the de-acylation of galactoglucomannan and acetylated manno-compounds. By the way, the CE4 also possessed peptidoglycan N-deacetylases with the ability of chitin degradation [33]. Three AA4 family members including vanillyl-alcohol oxidases (VAO) were detected, the function of which was to catalyze the conversion of multiple phenolic compounds bearing side chain at the para-position of aromatic rings [34]. As the vital components of CAZymes, glycosyl transferases could catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, and thirty-one GTs were detected in B5 dominated by GT4 and GT2. Actually, the combination of GT4 and GT2 family members make up approximately 50% of all glycosyl transferases, and they might be the originals from which other GT families evolved [35, 36]. The GT4 and GT2 families could catalyze various reactions including some key steps in N-glycosylation pathways. Besides, GT35 family genes were also detected, which could catalyze the phosphorolysis of specific glycosidic bonds within maltodextrins through removing the non-reducing glucosyl residues of linear oligosaccharides [37]. The CAZymes identified here indicated that B5 strain possessed a considerable potential of metabolizing some recalcitrant and readily degradable biomass. Our results also provided genetic evidence of both strong hydrolytic and transglycosylatic capabilities for B5 strain, which harbored various kinds of CAZymes genes.

Global analysis of transcriptome and DEGs

The raw data were processed according to the tophat2-cufflinks workflow [38], and finally get the expression of all unigenes represented by FPKM values. The FPKM distribution and relationship of different treatments were shown in Fig 4. Multi-dimensional scaling (MDS) method was applied to identify the sources of variability within the data. An obvious divergence was observed among different temperatures, and the repeatability of the three biological replicates are credible enough for DEG analysis (Fig. 4B). The results were further verified by the Spearman coefficient of correlation among the 9 data sets, which shared a value of more than 0.93 within the same treatment (Fig. 4C).

DEGs were detected to identify the temperature responding genes between each treatment, and a diagram was constructed showing DEGs to further understand the interaction of these treatments for DEGs (Fig. 4E). There were 980, 1019 and 601 DEGs between different treatments. Interestingly, we found 171 DEGs in all treatments, and most of their expression level decreased when temperature decrease or increase to

their limits (Additional file 4), which could be considered as the heat-response sensitive genes. Furthermore, the clustering analysis results showed that the treatments of T40 and T70 formed closer (Fig. 4F). It might be due to that the two treatments were all extreme environments, so many genes tended to exhibit in similar patterns. Briefly, all DEGs were divided into two groups: in group I, most genes expressed at a very low level in T60, and there were also many differences between T40 and T70; in group II, almost all the genes in T60 expressed substantively, and DEGs in T40 and T70 showed small differences.

Metabolism characteristic of B5 strain at mesophilic situation

It is well known that the metabolism systems of microbes are appropriately regulated, especially in non-optimal environment. The translation and synthesis of proteins are energy consuming process, which should be limited at some level when suffering extreme conditions. Compared with T60 (Additional file 5), the isopropylmalate synthase (gene2488), isopropylmalate dehydrogenase (gene2487), and aminotransferase (gene2492) genes were significantly up-regulated in T40 treatment. These were key enzymes associated with the process of leucine biosynthesis, and played a critical role in “valine, leucine and isoleucine biosynthesis” and “valine, leucine and isoleucine degradation” pathways. Besides, the expression of branched-chain amino acid transporters (gene3091) belonging to ABC transport system was also up-regulated in T40, which mainly participated in transportation or assimilation of branched-chain amino acids (leucine et al.) into the cell. Meanwhile, the genes involved in the degradation of branched-chain amino acids were down-regulated, such as branched-chain alpha-keto acid dehydrogenase (gene0976, 2204), which illustrated that B5 strain would enhance the biosynthesis of leucine and inhibit its degradation at mesophilic phase of composting. It was reported that the D-amino acids (D-Leu, D-Met, and D-Phe) could regulate the synthesis of peptidoglycan [39]. Always, many bacteria cell walls contained a large amount of peptidoglycan, especially for the gram-positive bacteria. Thus, the down-regulation of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase (gene1033) involved in the cell peptides synthesis might result in the inhibition of the cell walls synthesis in B5 strain. Actually, the synthesis of D-amino acids might be a common strategy for bacteria to adapt to the non-optimal environment. Here, the uptake and synthesis of leucine might also contribute to compensate the down-regulation of genes related to peptidoglycan synthesis.

Aromatic amino acids including tryptophan, phenylalanine and tyrosine were necessary to all microorganisms during primary metabolism, and these aromatic compounds were mainly produced via shikimate pathway [40]. In T40 treatment, the expression of chorismate mutase genes (gene2048, 2622) showed no difference by comparing to T60, while the anthranilate phosphoribosyl transferase (gene2045) and the pyridoxal phosphate dependent enzyme (gene1050) were up-regulated. These findings illustrated that the synthesis of phenylalanine and tyrosine might not be affected at composting mesophilic phase, while the synthesis of tryptophan was strengthened. The results above were further supported by the up-regulated expression of tryptophan-tRNA ligase (gene0160), which might be a signal of increasing demand for tryptophan. Furthermore, the expression of phosphoglycerate dehydrogenase (gene2090) and phosphoserine aminotransferase (gene0536) was up-regulated in T40. These two enzymes were the key members during serine biosynthesis process, while serine could be taken as a

precursor for the synthesis of tryptophan [41]. In addition, the up-regulated pyridoxal phosphate enzyme (gene1050) might also contribute to the conversion of serine to tryptophan. All above, the synthesis of tryptophan of B5 strain was up-regulated so as to get adapt to the low temperature and survive in the composting mesophilic phase. Similarly, the synthesis and uptake of tryptophan was also increased for *Saccharomyces cerevisiae* in cold stress, which enhanced the tolerance to low temperature [42].

The HSPs (heat shock proteins) and enrichment analyses of B5 strain at extreme thermophilic situation

B5 strain was isolated from the thermophilic composting, and it could survive at 70 °C or even higher temperature. Thus, it is certainly worth understanding its internal heat shock mechanism, which might be the prerequisite of B5 strain to secrete extracellular enzymes. Heat shock proteins performed chaperone functions by stabilizing proteins to ensure correct folding or helping refold denatured proteins so as to protect the cell from heat stress [43]. The expression changes of the heat shock proteins in T70 vs T60 treatment were displayed in Table 2, and their relative expression levels all increased at least 2 folds. Some HSPs were responsible for unfolding the insoluble protein aggregates, or serving as co-factor of Hsp70, such as *clpB* (gene0701), *clpX* (gene2482), *clpP* (gene1200) and *dnaJ* (gene2328). Hsp33 (gene0064) and *dnaK* (gene2329) belonged to class I heat shock proteins (chaperonin) with the functions of protein folding and unfolding, thus entrusted thermotolerance to cells exposed to extremely stressful conditions. Also, *grpE* (gene2330) was discovered in B5 strain, and it mainly serviced as co-factor of *dnaK*. It appeared that these genes played a pivotal role for B5 strain to survive in the thermal environment.

Except for the heat shock genes, the metabolism mechanism for B5 strain when confronting with the high temperature aroused more interests. To investigate how this strain responding to heat stress, the DEGs in T70 vs T60 treatment were taken to conduct the KEGG enrichment and protein-protein interaction analyses (Fig. 5 and Fig. 6). The most significant up-regulated pathway is ribosome (eco03010, q-value = 1.63×10^{-9}) covering 33 genes. The participated genes could encode various ribosomal proteins (Additional file 6) including 30S ribosomal protein S3 (gene0114) and 50S ribosomal protein L13 (gene0141). Protein S3 could encircle the mRNA along with protein S4 (gene2613) when entering the ribosome, and they also play an important role in mRNA helicase processivity [44]; while protein L13 was very important during the early stage of 50S assembly [45]. These results could be further supported by GO enrichment analysis, whose top three up-regulated enrichment terms were all related to ribosome synthesis (GO:1990904, 0005840, 0003735). All the results referred above suggested that normal protein synthesis and vigorous growth strongly increased so as to intensify the high heat resistance capacity, which were different from the performance of thermolabile strain *S. cerevisiae* with down-regulated ribosome protein when facing heat stress [46]. This might be due to the specialty of B5 strain that the increasing number of expressed proteins, especially some key enzymes, were related to cell survival under heat stress. Furthermore, the pantothenate and CoA biosynthesis (eco00770), arginine biosynthesis (eco00220), and pyrimidine metabolism (eco00240) pathways were significantly up-regulated as well. Unlike the ribosome pathway, the genes participated in carbon metabolism (eco01200), fatty acid degradation (eco00071), fructose and mannose metabolism (eco00051), starch and sucrose metabolism

(eco00500), biosynthesis of secondary metabolites (eco01110) and some biological process and cellular metabolic process related pathways were all down-regulated with various degrees. These results suggested that the catabolism of recalcitrant carbon sources could be synergistic with heat resistance of B5 strain because of the assistance from the high temperature which would destroy the lignocellulose thus providing more reducing sugar and polysaccharides.

Conclusion

An efficient lignocellulose degrading strain was isolated from thermophilic composting and identified as *G. stearothermophilus* B5. The whole-genome analysis of B5 along with comparative analysis had opened up a better understanding of its genomic information and highlighted the considerable ability of degrading lignocelluloses, thus indicated the potential application in agricultural wastes management fields. In addition, this study also explored the response of B5 in different composting phase based on the transcriptome analysis. The present work will certainly strengthen the genomic aspect of exploiting bacteria for efficient thermophilic composting.

Materials And Methods

Screening, isolation and identification of the cellulolytic strain

The compost samples were collected from a thermophilic composting heap at a local farm in Nanjing, China, mainly composed of rice chaff and chicken manure. For the screening of cellulolytic bacteria, ten grams of mixed samples were taken and then vortexed completely with 90 mL sterile water. The supernatant was serially diluted and then plated on screening medium described by López-Mondéjar [47] at 60 °C. The cellulose-degrading ability was preliminarily evaluated by qualitative assay following the workflow of Teather et al. [48] with cellulose-Congo red as the substrates. Morphological, physiological and biochemical characteristics of the selected strain were performed according to the protocol of Bergey's Manual of Determinative Bacteriology [49]. The morphological features of the cell was further observed by using a scanning electron microscope (SEM) after full preparation described by Abuga et al [50]. Meanwhile, the DNA of the selected strain was extracted and amplified by the universal 16S rRNA primers (27F 5'-AGAGTTTGATCMTGGCTCAG-3' and 1492R 5'-CGGTTACCTTGTTACGACTT-3') as well as *recN* gene primers (forward 5'- CGATTTGCGGCGACGATA -3' and reverse 5'-TACACCATGCAAAAACGGTTAC-3') [51]. The sequences were analyzed for similarities by BLAST against related sequences downloaded from NCBI database.

Optimal culture conditions and carbon source utilization for B5 strain

G. stearothermophilus B5 was first activated on LB solid plate and then transferred to relevant liquid media with an agitation speed of 170 rpm for 24 h. To determine the optimal culture conditions, growth tests were performed at different temperature, initial pH values and NaCl concentrations ranging from 35 to 75 °C, 4.0 to 9.5, and 0 to 4%, respectively. The growth curve was plotted under the optimal culture

condition at an interval of every one hour, and the growth rates were determined by evaluating the OD600 values. The carbon metabolic characterizations of B5 strain was determined by using GEN III MicroPlate (Biolog, USA), which contained 71 carbon sources along with a positive and negative control. Substrate utilization was evaluated colorimetrically alongside spectrophotometric cellular growth measurement.

Extracellular protein extraction and enzyme activity assays

B5 strain was cultivated in basal salt medium [52] supplemented with 1% (w/v) different carbon source including CMC sodium (Sigma, USA), birchwood xylan (Sigma, USA) and cassava starch (Sigma, USA) to determine different enzyme activities. Besides, B5 strain was grown in Tryptic Soy Broth medium (Hopebio, China) so as to detect the protease activity. After inoculation with 1×10^5 cfu·mL⁻¹ cells, the Erlenmeyer flask was inoculated at 60 °C in orbital shaker at 170 rpm for 7 days. The bacterial biomass and residual substrates were removed by centrifugation at 12,000 rpm for 10 min and then filtered through a 0.45 µm membrane, and the supernatant considered as crude enzymes was used in the subsequent experiments. CMCase and xylanase activities were determined by 3, 5-dinitrosalicylic acid (DNS) method according to Liu et al. [53] with CMC and xylan as the substrates, respectively, and α-amylase activity was measured according to P Dheeran, S Kumar, YK Jaiswal, DKJAm Adhikari and biotechnology [52] with cassava starch as the substrate. One unit of enzyme activity was defined as the amount of enzyme that released 1 µmol of reducing sugars per minute. Protease activity was measured according to Thebti et al. [54] with casein as substrate. One unit of protease activity was defined as the amount of enzyme that released 1 µmol of tyrosine per minute.

Genome sequencing, assembly and annotation

The genomic DNA was extracted by using E.Z.N.A.®MicroElute Genomic DNA Kit (Omega biotek, Norcross, GA, USA) according to the manufacturer's protocol and quantified by Nanodrop (Thermo Scientific). The purified DNA was sequenced by using PacBio Sequel Single Molecule Real-Time (SMRT) sequencing technology. After genome assembly, the coding sequences (CDSs) were predicted using prodigal software (v2.6.3). Common function annotation was performed by BLAST against the database of NCBI non-redundant (NR), Cluster of Orthologous Groups of proteins (COGs), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGGs) and Carbohydrate-Active enzymes (CAZymes).

RNA extraction and transcriptome sequencing

The strain was first grown in LB liquid medium at 60 °C, 160 rpm overnight. After removing the supernatant, cell concentration was washed with sterile distilled water and then transferred to basic medium with 1% rice straw as the carbon source for 10 hours. The culture temperature was set at 40°C (simulating the composting mesophilic phase), 60 °C (simulating thermophilic phase) and 70 °C (simulating extreme thermophilic phase). Samples collected from all the treatments were immediately frozen in liquid nitrogen for RNA extraction. The total RNA was extracted using the RNA Isolation Kit (Qiagen) and then sequenced for paired-end reads using the Illumina HiSeq2000 platform. After quality control, clean reads were retained and mapped to the genome of strain B5 using TopHat pipeline [55]. The

expressions of all unigenes were estimated by transforming the read density to fragments per kilo base of exon per million mapped reads values (FPKM). A threshold of FC > 2 (or < 0.5) and q-value < 0.01 were used to selected the significantly differential expression genes (DEGs).

Statistical analysis

All biochemical parameters here were measured in triplicates and analyzed by statistical analysis in R language (Version 3.6.1). The relationships between different treatments of the transcriptome data were analyzed using multi-dimensional scaling (MDS) method.

Abbreviations

Clusters of Orthologous Groups (COG); Kyoto Encyclopedia of Genes and Genomes (KEGG); Gene Ontology (GO); Glycoside Hydrolases (GH); Glycosyl Transferases (GT); Carbohydrate-Active enzymes (CAZymes); Carbohydrate Esterases (CE); Auxiliary Activities (AA); Carbohydrate-Binding Modules (CBM); Heat Shock Proteins (HSP); Differential Expression Genes (DEGs).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The materials and datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was financially supported by the National Key R & D Program of China (grant number 2018YFD0500201), the National Natural Science Foundation of China (grant number 31972513), the Independent Innovation Fund Project of Agricultural Science and Technology in Jiangsu Province (grant number CX (18)1005).

Author's contributions

Mengmeng Wang conceived the project and designed the analyses; Jiaxi Miao and Xuanqing Wang determined the enzyme activities; Tuo Li and Han Zhu extracted the RNA from different treatments; Mengmeng Wang wrote the manuscript; Dongyang Liu and Qirong Shen reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We thank members of the Liu laboratories for feedback and discussions.

References

1. Adav SS, Ng CS, Arulmani M, Sze SKJJopr: **Quantitative iTRAQ secretome analysis of cellulolytic *Thermobifida fusca***. 2010, **9**(6):3016-3024.
2. Linton SM, Greenaway P: **Presence and properties of cellulase and hemicellulase enzymes of the gecarcinid land crabs *Gecarcoidea natalis* and *Discoplax hirtipes***. *Journal of experimental biology* 2004, **207**(23):4095-4104.
3. Bayer EA, Lamed R, Himmel MEJCoIB: **The potential of cellulases and cellulosomes for cellulosic waste management**. 2007, **18**(3):237-245.
4. Li L-L, Taghavi S, McCorkle SM, Zhang Y-B, Blewitt MG, Brunecky R, Adney WS, Himmel ME, Brumm P, Drinkwater CJBfb: **Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases**. 2011, **4**(1):23.
5. Reddy AP, Allgaier M, Singer SW, Hazen TC, Simmons BA, Hugenholtz P, VanderGheynst JSJB, bioengineering: **Bioenergy feedstock-specific enrichment of microbial populations during high-solids thermophilic deconstruction**. 2011, **108**(9):2088-2098.
6. Deswal D, Khasa YP, Kuhad RC: **Optimization of cellulase production by a brown rot fungus *Fomitopsis sp.* RCK2010 under solid state fermentation**. *Bioresource Technology* 2011, **102**(10):6065-6072.
7. Vargas-García M, Suárez-Estrella F, López M, Moreno J: **Microbial population dynamics and enzyme activities in composting processes with different starting materials**. *Waste management* 2010, **30**(5):771-778.
8. Miao J: **Effects of amino acids on the lignocellulose degradation by *Aspergillus fumigatus* Z5: insights into performance, transcriptional, and proteomic profiles**. 2019.
9. Blumer-Schuetz SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM: **Thermophilic lignocellulose deconstruction**. *FEMS Microbiology Reviews* 2014, **38**(3):393-448.
10. Hussein AH, Lisowska BK, Leak DJ: **The genus *Geobacillus* and their biotechnological potential**. In: *Advances in applied microbiology*. vol. 92: Elsevier; 2015: 1-48.
11. De Maayer P, Brumm PJ, Mead DA, Cowan DA: **Comparative analysis of the *Geobacillus hemicellulose* utilization locus reveals a highly variable target for improved hemicellulolysis**. *BMC*

Genomics 2014, **15**(1):836.

12. Rastogi G, Muppidi GL, Gurram RN, Adhikari A, Bischoff KM, Hughes SR, Apel WA, Bang SS, Dixon DJ, Sani RK, Joim *et al*: **Isolation and characterization of cellulose-degrading bacteria from the deep subsurface of the Homestake gold mine, Lead, South Dakota, USA.** 2009, **36**(4):585.
13. Bartosiak-Jentys J, Hussein AH, Lewis CJ, Leak DJ: **Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*.** *Microbiology* 2013, **159**(7):1267-1275.
14. Zeigler DR: **Application of a *recN* sequence similarity analysis to the identification of species within the bacterial genus *Geobacillus*.** *International Journal of Systematic and Evolutionary Microbiology* 2005, **55**(3):1171-1179.
15. Miller F: **Composting as a process based on the control of ecologically selective factors.** *Microbial Ecology* 1993:515-544.
16. Raut MP, Prince William SPM, Bhattacharyya JK, Chakrabarti T, Devotta S: **Microbial dynamics and enzyme activities during rapid composting of municipal solid waste – A compost maturity analysis perspective.** *Bioresource Technology* 2008, **99**(14):6512-6519.
17. Wang X-J, Bai J-G, Liang Y-X: **Optimization of multienzyme production by two mixed strains in solid-state fermentation.** *Applied Microbiology and Biotechnology* 2006, **73**(3):533-540.
18. Tai S-K, Lin H-PP, Kuo J, Liu J-KJE: **Isolation and characterization of a cellulolytic *Geobacillus thermoleovorans* T4 strain from sugar refinery wastewater.** 2004, **8**(5):345-349.
19. Rastogi G, Bhalla A, Adhikari A, Bischoff KM, Hughes SR, Christopher LP, Sani RK: **Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains.** *Bioresource Technology* 2010, **101**(22):8798-8806.
20. Lynd LR, Weimer PJ, Van Zyl WH, Pretorius ISJMMBR: **Microbial cellulose utilization: fundamentals and biotechnology.** 2002, **66**(3):506-577.
21. Turner P, Mamo G, Karlsson ENJMcf: **Potential and utilization of thermophiles and thermostable enzymes in biorefining.** 2007, **6**(1):9.
22. Abdel-Fattah YR, El-Helow ER, Ghanem KM, Lotfy WAJRjom: **Application of factorial designs for optimization of avicelase production by a thermophilic *Geobacillus* isolate.** 2007, **2**(1):13-23.
23. Zhang N, Yang D, Wang D, Miao Y, Shao J, Zhou X, Xu Z, Li Q, Feng H, Li S: **Whole transcriptomic analysis of the plant-beneficial rhizobacterium *Bacillus amyloliquefaciens* SQR9 during enhanced biofilm formation regulated by maize root exudates.** *BMC genomics* 2015, **16**(1):685.
24. Madej MG, Sun L, Yan N, Kaback HR: **Functional architecture of MFS D-glucose transporters.** *Proceedings of the National Academy of Sciences* 2014, **111**(7):E719-E727.
25. Canal-Llaubères RM: **4 - Enzymes and wine quality.** In: *Managing Wine Quality*. Edited by Reynolds AG: Woodhead Publishing; 2010: 93-132.
26. Janeček Š, Gabriško M: **Remarkable evolutionary relatedness among the enzymes and proteins from the α -amylase family.** *Cellular molecular life sciences* 2016, **73**(14):2707-2725.

27. Tamaru Y, López-Contreras AM: **Lignocellulosic biomass utilization toward biorefinery using meshophilic Clostridial species.** In: *Cellulose-Biomass Conversion*. InTech; 2013: 132-144.
28. Field R, Pergolizzi G, Kuhaudomlarp SS, Kalita E: **Glycan Phosphorylases in Multi-Enzyme Synthetic Processes.** *Protein and peptide letters* 2017, **24**.
29. Suzuki H, Okazaki F, Kondo A, Yoshida K-i: **Genome mining and motif modifications of glycoside hydrolase family 1 members encoded by *Geobacillus kaustophilus* HTA426 provide thermostable 6-phospho- β -glycosidase and β -fucosidase.** *Applied microbiology biotechnology* 2013, **97**(7):2929-2938.
30. Plaza-Vinuesa L, Hernandez-Hernandez O, Moreno FJ, de las Rivas B, Muñoz R: **Unravelling the diversity of glycoside hydrolase family 13 α -amylases from *Lactobacillus plantarum* WCFS1.** *Microbial Cell Factories* 2019, **18**(1):183.
31. Zhang J, Siika-aho M, Tenkanen M, Viikari L: **The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed.** *Biotechnology for Biofuels* 2011, **4**(1):60.
32. Shao W, Wiegel J: **Purification and characterization of two thermostable acetyl xylan esterases from *Thermoanaerobacterium* sp. strain JW/SL-YS485.** 1995, **61**(2):729-733.
33. Biely P: **Microbial carbohydrate esterases deacetylating plant polysaccharides.** *Biotechnology advances* 2012, **30**(6):1575-1588.
34. van den Heuvel RH, Fraaije MW, Mattevi A, van Berkel WJ: **Structure, function and redesign of vanillyl-alcohol oxidase.** In: *International Congress Series: 2002*. Elsevier: 13-24.
35. Magidovich H, Eichler JJFml: **Glycosyltransferases and oligosaccharyltransferases in Archaea: putative components of the N-glycosylation pathway in the third domain of life.** 2009, **300**(1):122-130.
36. Coutinho PM, Deleury E, Davies GJ, Henrissat B: **An Evolving Hierarchical Family Classification for Glycosyltransferases.** *Journal of Molecular Biology* 2003, **328**(2):307-317.
37. Geremia S, Campagnolo M, Schinzel R, Johnson LN: **Enzymatic Catalysis in Crystals of *Escherichia coli* Maltodextrin Phosphorylase.** *Journal of Molecular Biology* 2002, **322**(2):413-423.
38. Ghosh S, Chan C-KK: **Analysis of RNA-Seq data using TopHat and Cufflinks.** In: *Plant Bioinformatics*. Springer; 2016: 339-361.
39. Lam H, Oh D-C, Cava F, Takacs CN, Clardy J, de Pedro MA, Waldor MKJS: **D-amino acids govern stationary phase cell wall remodeling in bacteria.** 2009, **325**(5947):1552-1555.
40. Bentley R, Haslam EJCrib, biology m: **The shikimate pathway—a metabolic tree with many branche.** 1990, **25**(5):307-384.
41. Shen T, Liu Q, Xie X, Xu Q, Chen NJBRI: **Improved production of tryptophan in genetically engineered *Escherichia coli* with TktA and PpsA overexpression.** 2012, **2012**.
42. Vicent I, Navarro A, Mulet JM, Sharma S, Serrano R: **Uptake of inorganic phosphate is a limiting factor for *Saccharomyces cerevisiae* during growth at low temperatures.** *FEMS Yeast Res* 2015,

15(3).

43. Park C-J, Seo Y-S: **Heat shock proteins: a review of the molecular chaperones for plant immunity.** *The plant pathology journal* 2015, **31**(4):323.
44. Brauer D, Röming RJFI: **The primary structure of protein S3 from the small ribosomal subunit of *Escherichia coli*.** *FEBS LETTERS* 1979, **106**(2):352-357.
45. Herold M, Nierhaus KJJoBC: **Incorporation of six additional proteins to complete the assembly map of the 50 S subunit from *Escherichia coli* ribosomes.** 1987, **262**(18):8826-8833.
46. Ye Y, Zhu Y, Pan L, Li L, Wang X, Lin Y: **Gaining insight into the response logic of *Saccharomyces cerevisiae* to heat shock by combining expression profiles with metabolic pathways.** *Biochemical biophysical research communications* 2009, **385**(3):357-362.
47. López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P: **Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems.** *Scientific reports* 2016, **6**:25279.
48. Teather RM, Wood PJ: **Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen.** *Appl Environ Microbiol* 1982, **43**(4):777-780.
49. Holt JG, Krieg N, Sneath PH, Staley J, Williams SJBW, Wilkins: **Bergey's manual of determinative bacteriology. 9th.** 1994.
50. Abuga I, Sulaiman SF, Abdul Wahab R, Ooi KL, Abdull Rasad MSB: **In vitro antibacterial effect of the leaf extract of *Murraya koenigii* on cell membrane destruction against pathogenic bacteria and phenolic compounds identification.** *European Journal of Integrative Medicine* 2020, **33**:101010.
51. Canakci S, Inan K, Kacagan M, Belduz AOJJom, biotechnology: **Evaluation of arabinofuranosidase and xylanase activities of *Geobacillus* spp. isolated from some hot springs in Turkey.** 2007, **17**(8):1262.
52. Dheeran P, Kumar S, Jaiswal YK, Adhikari DKJAm, biotechnology: **Characterization of hyperthermostable α -amylase from *Geobacillus* sp. IIPN.** 2010, **86**(6):1857-1866.
53. Liu D: **Secretome diversity and quantitative analysis of cellulolytic *Aspergillus fumigatus* Z5 in the presence of different carbon sources.pdf.** 2013.
54. Thebti W, Riahi Y, Belhadj O: **Purification and Characterization of a New Thermostable, Haloalkaline, Solvent Stable, and Detergent Compatible Serine Protease from *Geobacillus toebii* Strain LBT 77.** *Biomed Res Int* 2016, **2016**:9178962.
55. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**(9):1105-1111.

Tables

Table 1 Genomic features of *G. stearothermophilus* B5 and comparison with four other reported cellulolytic *Geobacillus* spp. strains.

Features	B5	HTA426	NG80-2	Y412MC52	NBRC 101842
Size (bp)	3,390,870	3,592,666	3,608,012	3,673,940	3,539,687
GC content (%)	52.46	51.99	48.85	52.31	51.9
Contigs	1	2	2	2	164
Protein coding genes	3371	3546	3554	3596	3515
Mean gene length (bp)	851	861	853	869	849
Percent of coding region (%)	84.60	84.98	84.02	85.06	84.31
tRNA	90	87	88	87	82
rRNA	32	27	30	25	12

Table 2 The detailed information of heat shock protein-related genes

Gene ID	Gene	Details	Fold change
gene0064	-	Heat shock protein 33 (molecular chaperonin), a cytoplasmically localized protein	2.3
gene0701	<i>clpB</i>	Unfold insoluble protein aggregates, and co-factor of Hsp70/DnaK	2.8
gene1200	<i>clpP</i>	Clp protease	5.3
gene2328	<i>dnaJ</i>	Heat shock protein 40, co-factor of Hsp70	2.8
gene2329	<i>dnaK</i>	Folding and unfolding protein, providing thermotolerance on extreme environment	2.7
gene2330	<i>grpE</i>	Encoding protein GrpE, HSP-70 cofactor	4.5
gene2447	<i>mreB</i>	Rod shape-determining protein MreB	3.1
gene2482	<i>clpX</i>	ATP-dependent Clp protease	2.1

Additional Files

Supplementary figures.docx

Supplementary file 1: The utilization of different carbon sources for *G. stearothermophilus* B5;

Supplementary file 2: COG annotation of *G. stearothermophilus* B5;

Supplementary file 3: The gene expression of CAZymes in *G. stearothermophilus* B5;

Supplementary file 4: The heat-response sensitive genes in *G. stearothermophilus* B5;

Supplementary file 5: The significant differentially expressed genes of *G. stearothermophilus* B5 between T40 and T60;

Supplementary file 6: The up-regulated genes related to ribosome pathway.

Figures

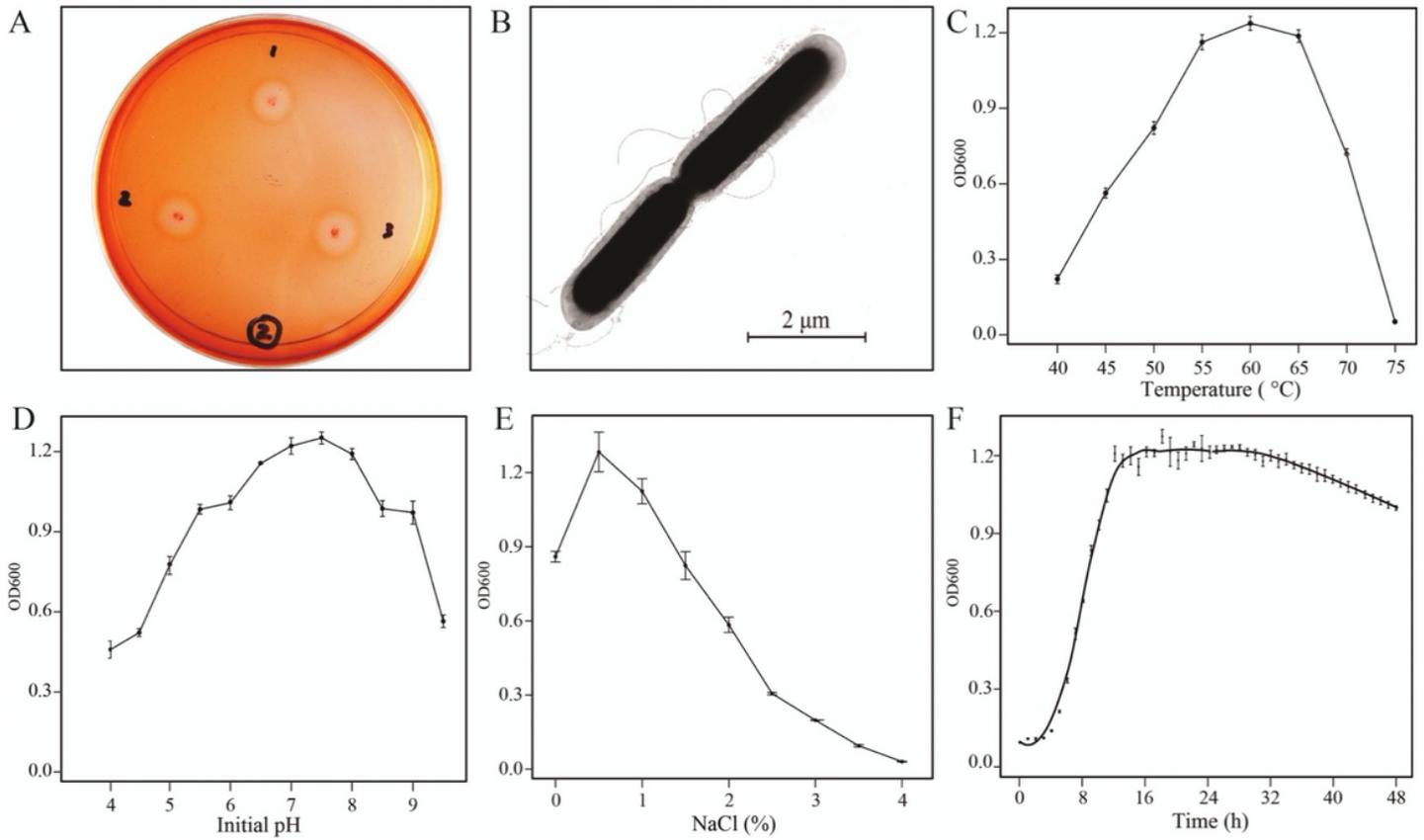


Figure 1

Basic features of *G. stearothermophilus* B5 and the growth curve within 48 hours. (A) Congo red zone clearing assay; (B) The microscopic observation of B5 strain under scanning electron microscope; (C-E) The growth condition at different temperature, pH values and salt concentration for B5 strain; (F) The growth curve under the optimal culture conditions.

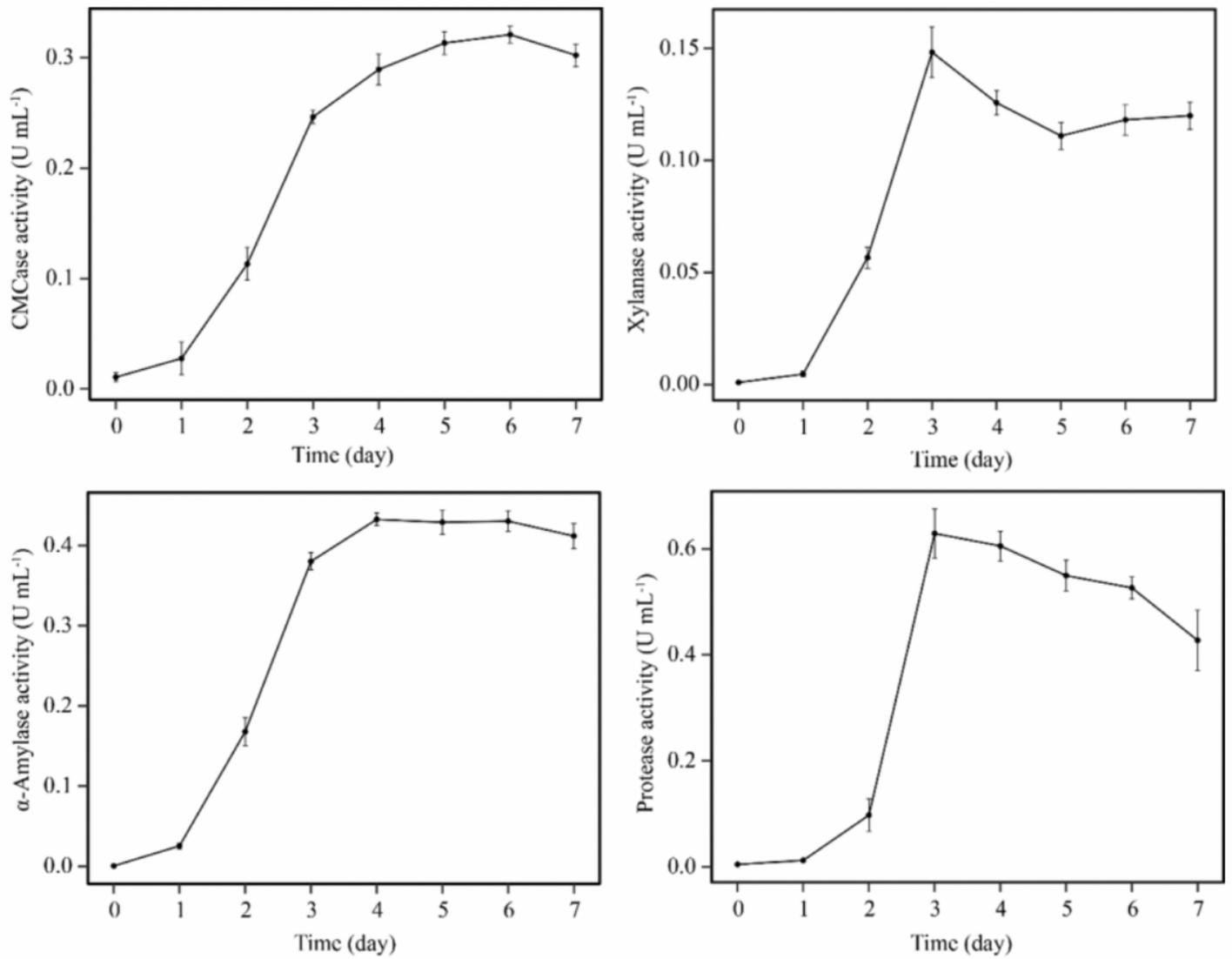


Figure 2

Various enzyme activities of B5 strain determined by spectrophotometric assays. The time course profile of different enzyme activities for CMCase (A), xylanase (B), α -amylase (C) and protease (D), respectively. The results are replicated three times, and bar indicate the standard error of replicates.



Figure 3

Genomic feature of *G. stearothermophilus* B5 and the comparison with other four *Geobacillus* spp. strains. (A) Circular map of *G. stearothermophilus* B5 genome features. Circles from the innermost to the outmost are the following features: GC skew, GC content, ncRNA (red represents rRNA, blue represents tRNA and green represents sRNA), COG on backward chains, and COG on forward chains; (B) Venn diagram for the uniqueness and intersection of encoding genes between *G. stearothermophilus* B5 and

other four *Geobacillus* spp. strains; (C) Comparison of COG function categories of *G. stearothermophilus* B5 and other *Geobacillus* spp. strains.

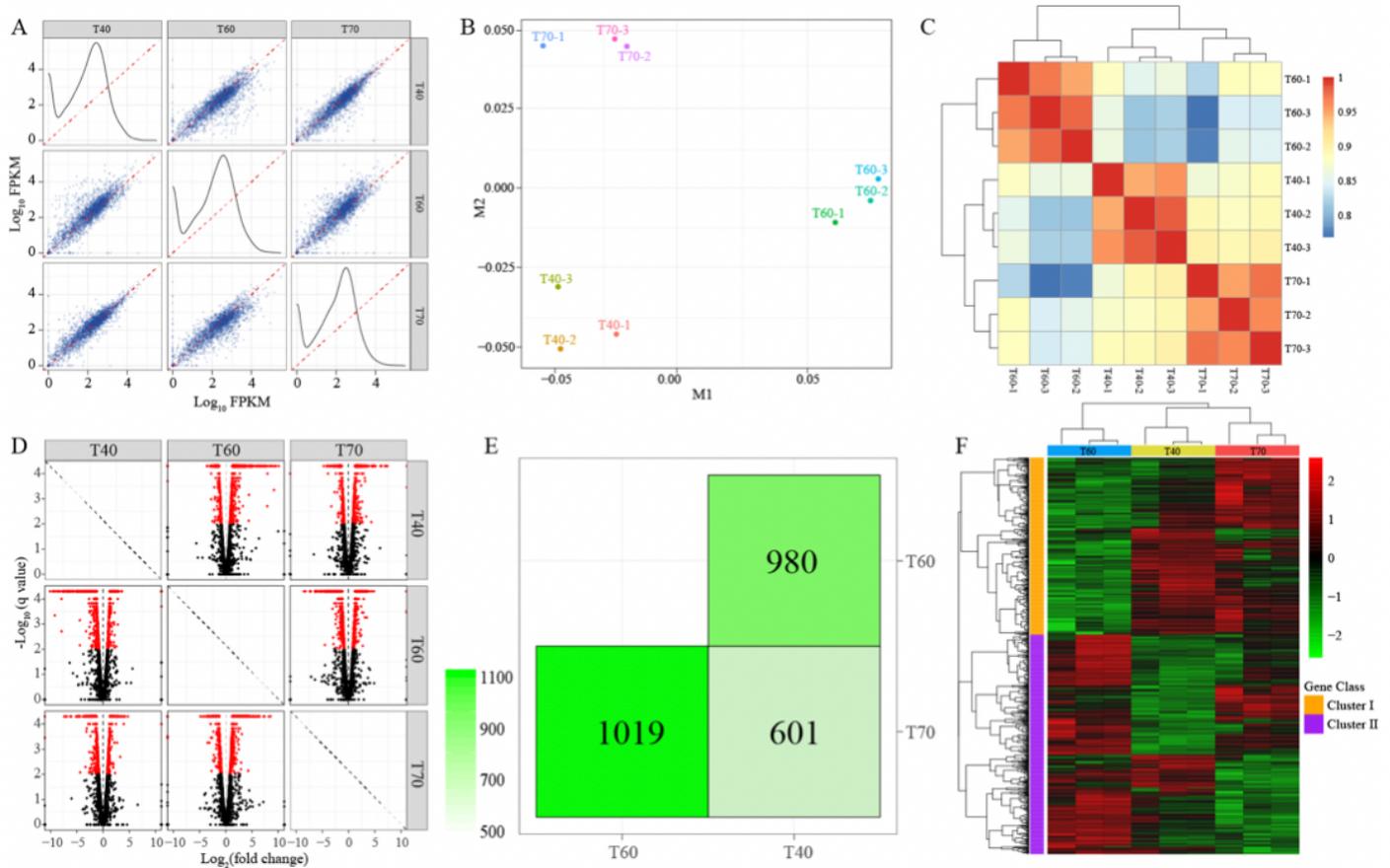


Figure 4

Global analysis of the transcriptome data among different treatments. (A-C) The distribution and correlations of the FPKM values of different treatments shown by scatterplots, MDS and heatmap, respectively; (D-E) Different expression genes among different treatments; (F) hierarchical clustering analysis of gene expression profiles with all the DEGs.

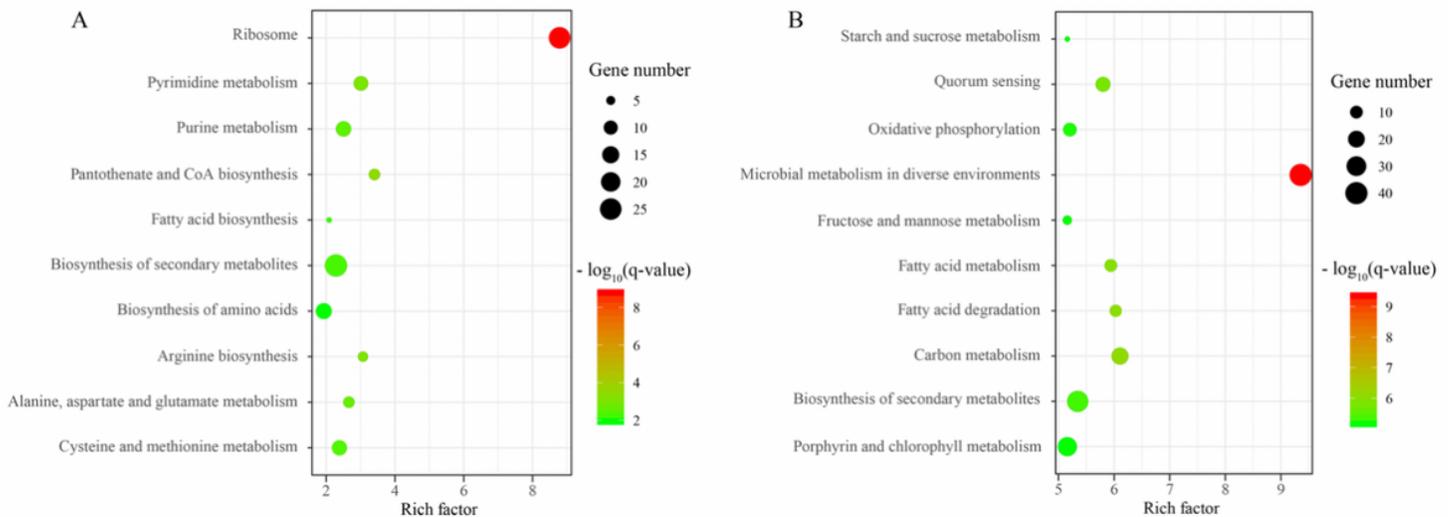


Figure 5

KEGG pathway enrichment of DEGs in T70 vs T60 treatment. A is the enrichment analysis of up-regulated genes (T70 vs T60); B is the enrichment analysis of down-regulated genes (T70 vs T60). The x axis represents the rich factor, and the y axis shows the name of the KEGG pathway; dot size represents the number of associated genes and the color indicates the $-\log_{10}(q\text{-value})$. The top 10 enriched GO are shown in the figure.

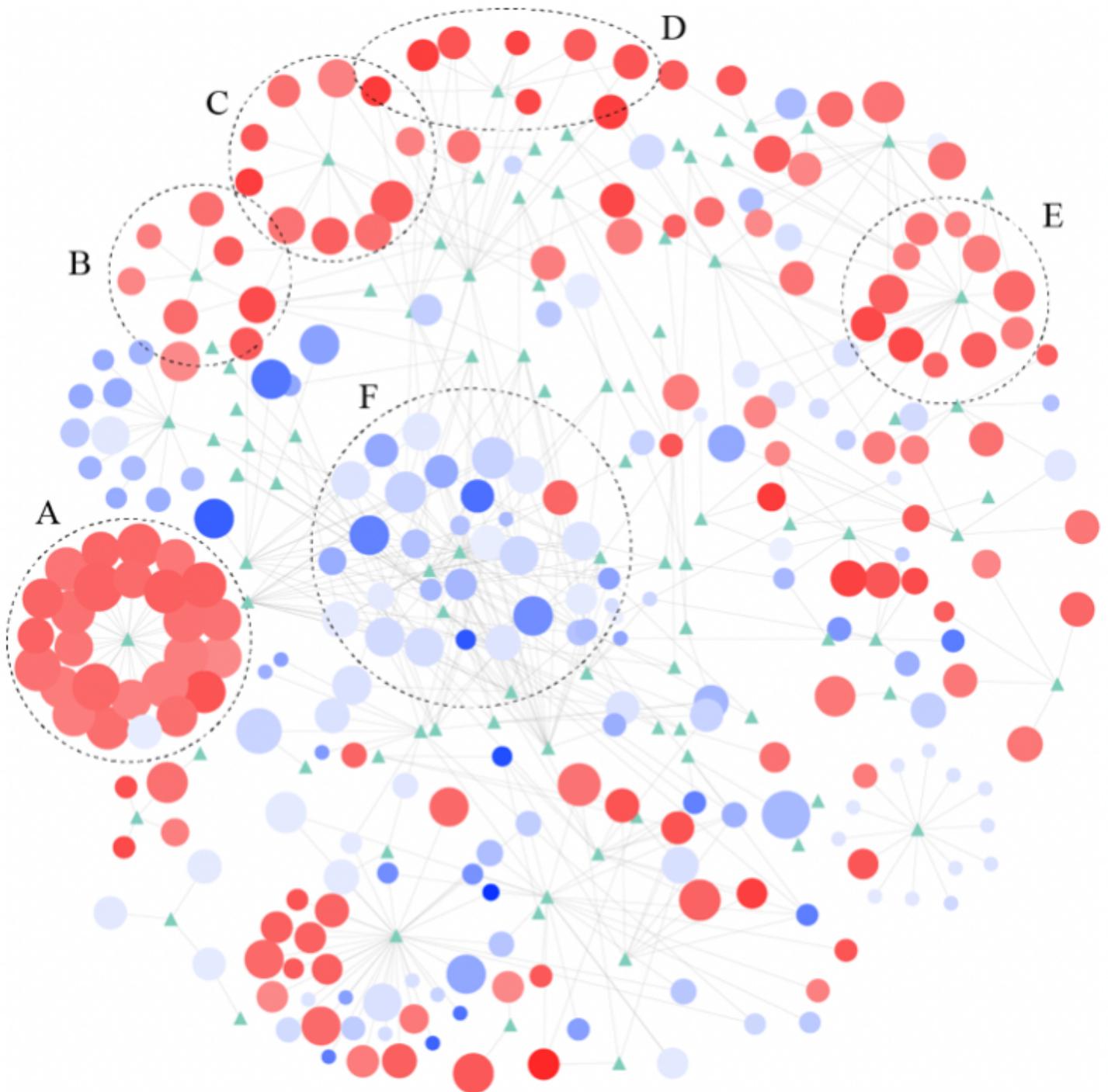


Figure 6

Protein-protein interaction based on the KEGG pathway analysis of differentially expressed genes in T70 vs T60 treatment. Nodes are proteins (circle) and KEGG categories (green diamonds); edges are protein interactions defined by KEGG database. Node sizes show the expression level in T70 treatment, and the colors indicate the fold change (red means up-regulated, blue means down-regulated) of expression values between T70 vs T60 treatments. The black dotted circles A is the ribosome pathway, B is pantothenate and Coa biosynthesis, C is cysteine and methionine metabolism, D is arginine biosynthesis, E is pyrimidine metabolism, F are metabolisms related to carbon source.

Supplementary Files

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

- [Additionalfile5.xlsx](#)
- [Additionalfile6.xlsx](#)
- [Additionalfile1.xlsx](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile4.xlsx](#)
- [SupplementaryFigures.docx](#)