



Contents lists available at ScienceDirect

Gene

journal homepage: [www.elsevier.com/locate/gene](http://www.elsevier.com/locate/gene)

Research paper

# The global transcriptional landscape of *Bacillus amyloliquefaciens* XH7 and high-throughput screening of strong promoters based on RNA-seq data

Yuling Liao<sup>a,b</sup>, Lianggang Huang<sup>a</sup>, Bin Wang<sup>a,\*</sup>, Feng Zhou<sup>a</sup>, Li Pan<sup>a,\*</sup><sup>a</sup> School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, Guangdong 510006, China<sup>b</sup> Star Lake Bioscience Co., Inc., Zhaoqing, Guangdong 526060, China

## ARTICLE INFO

## Article history:

Received 1 March 2015

Received in revised form 23 June 2015

Accepted 25 June 2015

Available online xxxxx

## Keywords:

*Bacillus amyloliquefaciens*

RNA sequencing

Untranslated regions

Transcription start site

Operon prediction

Promoter activity

## ABSTRACT

*Bacillus amyloliquefaciens* is an important industrial microbe for the production of many industrial enzymes and primary metabolites. Although the complete genome sequence of *B. amyloliquefaciens* has been now published, transcript structures of *B. amyloliquefaciens* remain poorly defined. In this study, high-throughput RNA sequencing (RNA-seq) technology was applied to dissect the transcriptome of *B. amyloliquefaciens* strain XH7. In total, 3936 out of a total of 4204 *B. amyloliquefaciens* genes (93.6%) were transcribed under the selected growth condition. Transcriptional start sites (TSS) of 1064 annotated genes and 749 operons were identified. To screen for strong promoters, a beta-galactoside reporter was fused to eight candidate promoters from 288 genes with higher expression levels (RPKM values) than the control gene *P<sub>43</sub>-bgab*. The results illustrated that the candidate promoter *P<sub>72</sub>* (promoter for the *sigW* gene) displayed the strongest beta-galactosidase specific activity during the post-log phase, suggesting that it could be used effectively for heterologous gene expression. The presented data will contribute to the further study of the *B. amyloliquefaciens* transcriptome by identifying useful promoters for industrial uses.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

*Bacillus amyloliquefaciens* is widely used in the agricultural, pharmaceutical, and food industries. Specifically, *B. amyloliquefaciens* is used as a plant growth-promoting rhizobacterium (PGPR) due to its ability to stimulate plant growth, as well as a biological control bacterium to suppress soil-borne pathogens due to its ability to produce antifungal and antibacterial metabolites (e.g., lipopeptides and polyketides) (Fan et al., 2012; Chen et al., 2007; Qiu et al., 2014). Additionally, *B. amyloliquefaciens* is used for the production of industrially important biochemicals, including many industrial enzymes (such as serine proteases and  $\alpha$ -amylases) and other products, such as purine nucleosides and riboflavin (Schallmeyer et al., 2004; Kröger et al., 2013; Zakataeva et al., 2010; Zhang et al., 2014; Yang et al., 2011).

**Abbreviations:** RNA-seq, RNA sequencing; TSSs, transcriptional start sites; RPKM, reads per kilobase of transcripts per million mapped reads; UTRs, untranslated regions; PGPR, plant growth-promoting rhizobacterium; CDSs, protein-coding sequences; CGMCC, China General Microbiological Culture Collection Center; LB, Luria-Bertani; Amp, ampicillin; Kan, kanamycin; cDNA, complementary DNA; GO, Gene Ontology; ORFs, Open reading frames; TFs, transcription factors; PCR, Polymerase Chain Reaction; RBS, ribosome binding site; PBS, phosphate-buffered saline;  $\beta$ -Gal,  $\beta$ -galactosidase; TESs, transcription end sites; ECF, extracytoplasmic function  $\alpha$ -factors.

\* Corresponding authors.

E-mail addresses: [btbinwang@scut.edu.cn](mailto:btbinwang@scut.edu.cn) (B. Wang), [btlipan@scut.edu.cn](mailto:btlipan@scut.edu.cn) (L. Pan).

RNA sequencing (RNA-seq) technology has been proven to be a powerful tool for transcriptome analysis. Compared with microarray methods, RNA-seq has several advantages. First, RNA-seq could detect a larger dynamic range of expression levels. Second, RNA-seq has very low background noise; thus, it can precisely quantify transcripts with low expression levels (Croucher and Thomson, 2010; Sorek and Cossart, 2010). Many transcriptome studies have been performed on *Bacillus* species, including *Bacillus anthracis* (Passalacqua et al., 2012; Martin et al., 2010), *Bacillus subtilis* (Rasmussen et al., 2009; Irnov et al., 2010; Nicolas et al., 2012), *Bacillus licheniformis* (Wiegand et al., 2013), and *Bacillus thuringiensis* (Wang et al., 2013a,b). However, much less transcriptomic information has been available for *B. amyloliquefaciens*. The genome of *B. amyloliquefaciens* strain XH7 has been sequenced by our group. The 3.9-Mb genome contains 4204 annotated protein-coding sequences (CDSs). On the basis of the obtained genome, a transcriptome analysis of *B. amyloliquefaciens* could contribute to a greater understanding of gene expression and regulation in *B. amyloliquefaciens*.

Promoters are important regulatory elements for high-level gene expression and recombinant protein production. There are several ways to search new promoters, such as using a promoter trap vector (Phan et al., 2010; Yang et al., 2013). Currently, mapping the transcription start site (TSS), combined with the high-throughput and unbiased RNA-seq technique, has been proven to be a novel and effective strategy for the

<http://dx.doi.org/10.1016/j.gene.2015.06.066>

0378-1119/© 2015 Elsevier B.V. All rights reserved.

Please cite this article as: Liao, Y., et al., The global transcriptional landscape of *Bacillus amyloliquefaciens* XH7 and high-throughput screening of strong promoters based on RNA-seq data, *Gene* (2015), <http://dx.doi.org/10.1016/j.gene.2015.06.066>

identification of active promoters. Wang et al. identified 1203 active promoter candidates in *B. thuringiensis* via a genome-wide TSS analysis based on transcriptome data (Wang et al., 2013a). Martin et al. determined the expression of 566 *B. anthracis* genes whose 5' untranslated regions (UTRs) have been identified. Their results showed that there was indeed a positive correlation between the gene expression level and the upstream promoter strength (Martin et al., 2010).

In this study, the transcriptome of XH7-*bgaB* carrying the reference gene *P*<sub>43</sub>-*bgaB* was defined using RNA-seq data. We investigated XH7 transcript structures, including TSSs, and operon structures on a genome-wide scale. Furthermore, the obtained TSSs provided candidates for exploring highly active promoters. In total, the obtained transcriptional map of *B. amyloliquefaciens* could contribute to understanding gene regulation and extending the useful promoter library for this species.

## 2. Materials and methods

### 2.1. Bacterial strains

Bacteria strains used in this study are listed in Supplementary Table s1.

XH7 was one of our lab collections. XH7 is a guanosine-producing strain obtained by traditional mutation breeding. The genome sequence of XH7 was deposited in GenBank under the accession number CP002927.1 (Yang et al., 2011). XH7-*bgaB* is a derivative of XH7, bearing the reporter gene *P*<sub>43</sub>-*bgaB* inserted in the location of *amyX* gene. The *P*<sub>43</sub> promoter is a strong promoter which was originated from *B. subtilis* strain 168 (Wang and Doi, 1984). The reporter gene, *bgaB* was originated from *Geobacillus kaustophilus*. *G. kaustophilus* (CGMCC 1.3655, ATCC 8005, IAM 1 1001) (Chen et al., 2008) was purchased from the China General Microbiological Culture Collection Center (CGMCC, Beijing, China). *B. subtilis* WB800 was obtained from our laboratory. *Escherichia coli* DH5 $\alpha$  was purchased from Takara (Dalian, China). *E. coli* JM110 was purchased from the Guangdong Institute of Microbiology.

### 2.2. Bacterial culture conditions

XH7-*bgaB* strains were activated from  $-80^{\circ}\text{C}$  glycerol stocks and grown on Luria-Bertani (LB) solid plates at  $37^{\circ}\text{C}$  for 20–24 h using a streak plate method. A single colony was inoculated into 10 mL of fresh LB medium supplemented with 1% glucose (rich LB medium) in 50-mL flasks. The cultures were grown aerobically at  $37^{\circ}\text{C}$  with shaking at 200 rpm for 18–20 h until the OD<sub>600</sub> reached 8.5–10. For RNA isolation, the cells were harvested by centrifugation at 12,000  $\times$ g for 2 min at  $4^{\circ}\text{C}$ . Plasmid-containing strains were cultured in LB broth containing appropriate antibiotic at  $37^{\circ}\text{C}$  with shaking at 200 rpm. *B. subtilis* WB800 (containing *bgaB* expression plasmid) was grown overnight at  $37^{\circ}\text{C}$  with shaking at 200 rpm in LB medium containing kanamycin, and 1% (vol/vol) of this pre-culture was used to be inoculated into 50 ml of the same medium in a 250 mL flask. 50  $\mu\text{g}/\text{mL}$  ampicillin (Amp), 25  $\mu\text{g}/\text{mL}$  kanamycin (Kan) were used for selecting transformants.

### 2.3. RNA isolation, library construction, and sequencing

A single colony of XH7-*bgaB* was cultured in rich LB medium at  $37^{\circ}\text{C}$  for 18–20 h as described. The cells were harvested, and the total RNA of above sample was isolated using the RNeasy Pure Cell/Bacteria Kit (TIANGEN, Beijing, China) following the manufacturer's instructions. Then, the total RNA (5  $\mu\text{g}$ ) was treated with RNase-free DNase I (Takara Bio, Otsu, Japan) for 30 min at  $37^{\circ}\text{C}$  to remove residual DNA. RNA quality was verified using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and was checked by RNase-free agarose gel electrophoresis, and the concentration of the total RNA was measured by a 2100 Bioanalyzer at 260 nm and 280 nm.

The extracted RNA sample was used for the generation of all cDNA libraries construction. RNA samples were subjected to rRNA depletion using the Ribo-Zero (Gram-positive bacteria) kit (Epicentre, Madison, WI, USA) according to the manufacturer's instructions. All mRNAs were broken into short (200-nt) fragments by adding fragmentation buffer. First-strand cDNA was generated using random hexamer-primed reverse transcription, followed by the synthesis of the second-strand cDNA using RNase H and DNA polymerase I. The cDNA fragments were purified using the QIA quick PCR extraction kit (Qiagen, Valencia, CA, USA). These purified fragments were then washed with EB buffer for end reparation and poly (A) addition and ligated to sequencing adapters. Following agarose gel electrophoresis and extraction of cDNA from gels, the cDNA fragments (200 bp  $\pm$  25 bp) were purified and enriched by PCR to construct the final cDNA library. The cDNA library was sequenced on the Illumina sequencing platform (Illumina HiSeq™ 2000, Illumina, San Diego, CA, USA) using the paired-end technology in a single run (Guangzhou, China). The processing of original images to sequences, base-calling, and quality value calculations were performed using the Illumina GA Pipeline (version 1.6), in which 100-bp paired-end reads were obtained. A Perl program was written to select clean reads by removing low-quality sequences (in which more than 50% of the bases had a quality of less than 20 in one sequence), reads with more than 5% of N bases (unknown bases) and reads containing adaptor sequences. Then the clean reads were assembled using Trinity to construct unique consensus sequences. The raw sequencing data have been deposited in the Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra/>) at Nation Center for Biotechnology Information (NCBI) under the accession number SRP055848.

### 2.4. RNA-Seq data analysis

The obtained clean reads were mapped to the reference sequence of XH7 (cp002927.1). Genome coverage was calculated using BEDTools (Quinlan and Hall, 2010). The gene expression level was normalized by the reads per kb per million reads method (RPKM) (Mortazavi et al., 2008).

Gene Ontology (GO) terms of *B. amyloliquefaciens* genes were obtained by Blast2GO (Conesa et al., 2005) and visualized by WEGO (Ye et al., 2006) using the default parameters. Blast2GO was used to obtain the annotation results of unigenes in GO (<http://www.geneontology.org/>). Subsequently, WEGO (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>), was used to classify the annotation results of the unigenes in the GO database.

For each predicted translational start, we searched for the break in the transcribed region around genes using the default parameters as Vijayan et al. described (Vijayan et al., 2011; Passalacqua et al., 2009). Open reading frames (ORFs) that shared the same 5' TSS were defined as being in the same operon. DBTBS (<http://dbtbs.hgc.jp/>), a database of transcriptional regulation in *Bacillus subtilis*, was used to analyze putative binding sites for  $\sigma$ -factors and transcription factors (TFs) (Sierro et al., 2008). Operons were predicted from the XH7 genome download from DOOR (Database of prokaryotic Operons), (<http://csbl.bmb.uga.edu/DOOR/>) (Mao et al., 2009).

### 2.5. Validation of RNA-seq analysis by RT-PCR

To search for genes that were highly expressed compared with the reporter gene *P*<sub>43</sub>-*bgaB*, an independent experiment was conducted out by RT-PCR. Cells were harvested under the same condition mentioned above, and 1  $\mu\text{g}$  isolated RNA was used to synthesize cDNA using the PrimeScript™ RT reagent Kit (Takara, Dalian, China) according to the manufacturer's instructions. Primers designed for each gene are given in Supplementary Table s2. RT-PCR was performed using an Applied Biosystems 7500 using SYBR® Premix Ex Taq™ (Tli RNaseH Plus), ROX plus (Takara, Dalian, China) according to the manufacturer's protocol. The PCR cycles were as follows: 1 cycle of 30s at  $95^{\circ}\text{C}$ ,

followed by 40 cycles at 95 °C for 5 s, 54 °C for 30 s and 72 °C for 34 s. After amplification, fluorescent data were converted to threshold cycle (Ct) values for analysis, *bgab* and *gapA* genes were used as the internal reference gene.

## 2.6. Construction of plasmids

Recombinant plasmids used in this study are listed in Supplementary Table s1.

All primers used in this study are listed in Supplementary Table s2.

### 2.6.1. Construction of an integrative plasmid

To introduce the reporter gene  $P_{43}$ -*bgab* into the *B. amyloliquefaciens*, the integrative plasmid pKS2-*amyX* was constructed as follows (Supplementary Fig. s1). Using primers *amyX*-F/R the homogeneous *amyX*-truncated fragment was amplified using XH7 chromosomal DNA as a template. The *amyX*-truncated fragment contained additional *Afl*III and *Eco*RI sites at its 5' and 3' ends, respectively. The PCR product was digested and ligated to the *Afl*III and *Eco*RI sites of the expression plasmid pBEP<sub>43</sub>-*bgab* to yield pBE-*amyX*. The expression plasmid pBEP<sub>43</sub>-*bgab* was our lab stock, and the *bgab* gene encodes a thermostable  $\beta$ -galactosidase from *G. kaustophilus*. Using the primers pK2-*amyX*-F/R, the fragment *amyX*-truncated- $P_{43}$ -*bgab*, which was flanked by *Xho*I and *Pst*I at its 5' and 3' ends, was amplified from plasmid pBE-*amyX*. After digestion with *Xho*I and *Pst*I, the fragment was cloned into the corresponding sites of pKS2, yielding pKS2-*amyX*, which was subsequently introduced into *E. coli* strain JM110 (*dam*<sup>-</sup>, *dcm*<sup>-</sup>) that was deficient in adenine and cytosine methylation, in preparation for electrotransformation into the XH7 strain.

### 2.6.2. Construction of expression plasmids

All promoter-5' UTR DNA region complexes were designated as  $P_{r1}$  with the corresponding serial numbers. The translational fusion plasmid pBEP<sub>r1</sub>-*bgab* was constructed as follows (Supplementary Fig. s2). First, a fragment with the p43 ribosome binding site (RBS) sequence was generated by annealing two artificially synthesized rbs-F/R oligonucleotides, and then cloned into the pMD™18-T vector. The fragment was flanked by *Afl*III and *Sal*I sites at its 5' and 3' ends, respectively. Subsequently, the fragment was digested and cloned into the corresponding sites of pBEP<sub>43</sub>-*bgab*, yielding pBE-rbs. Second, the promoter-5' UTR DNA region complex of  $P_{r1}$  was amplified from the genomic DNA of XH7 using the primers r1-F/R that was flanked by the *Eco*RI and *Kpn*I sites at its 5' and 3' ends, respectively. The PCR-amplified fragment was digested and ligated into the *Eco*RI and *Kpn*I sites of the pBE-rbs to construct the expression plasmid pBEP<sub>r1</sub>-*bgab*.

## 2.7. Introduction of plasmids into XH7 and *B. subtilis* WB800

The plasmids used for electrotransformation were prepared from the *E. coli* JM110 (*dam*<sup>-</sup>, *dcm*<sup>-</sup>) strain and introduced into XH7 and *B. subtilis* WB800 by electroporation as previously described (Zakataeva et al., 2010). The cells were exposed to a single electrical pulse using a field strength of 12.5 KV/cm for 4.5–6 ms.

## 2.8. Determination of $\beta$ -galactosidase activity

According to a previously described method (Hirata et al., 1985), the  $\beta$ -galactosidase ( $\beta$ -Gal) activities were measured as follows. 250- $\mu$ L cultures were collected via centrifugation at 10,000  $\times g$  for 1 min. The supernatant was discarded, and the cells were washed and resuspended in phosphate-buffered saline (PBS) and then treated with 2 mg/mL lysozyme at 37 °C for 10 min. After lysozyme treatment, the solution was placed into liquid nitrogen for 5 min and then sonicated at 500 W at 40 °C for 5 min. The  $\beta$ -Gal activity of the pretreatment sample supernatant was measured.  $\beta$ -Gal specific activities were converted to Miller

units (Miller U/mL). Data are averages of three independent experiments; error bars indicate standard errors of mean values.

## 2.9. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of expressed proteins

SDS-PAGE analysis of  $\beta$ -Gal was performed. The protein fractions were electrophoresed on 5% (w/v) stacking gels and 12% (w/v) separating gels, and they were stained with Coomassie Blue R-250.

## 3. Results

### 3.1. Growth kinetics of strain XH7-*bgab*

XH7-*bgab* carrying the reporter gene  $P_{43}$ -*bgab* was grown in LB broth for RNA sample preparation. According to the results ( $\beta$ -Gal, Fig. 1A), strain XH7 and XH7-*bgab* reached the late exponential growth phase after approximately 16–18 h, at which time XH7-*bgab* achieved the highest  $\beta$ -Gal expression level. Therefore, we chose the 18-h growth sample for RNA isolation. Fig. 1B shows that the  $\beta$ -Gal activity of XH7-*bgab* was subject to catabolite repression by glucose with LB medium as a control. Because the growth rate of XH7-*bgab* in LB was slower than in rich LB medium, to maintain rapid growth and high  $\beta$ -Gal activity, we chose rich LB medium for RNA-seq sample preparation.

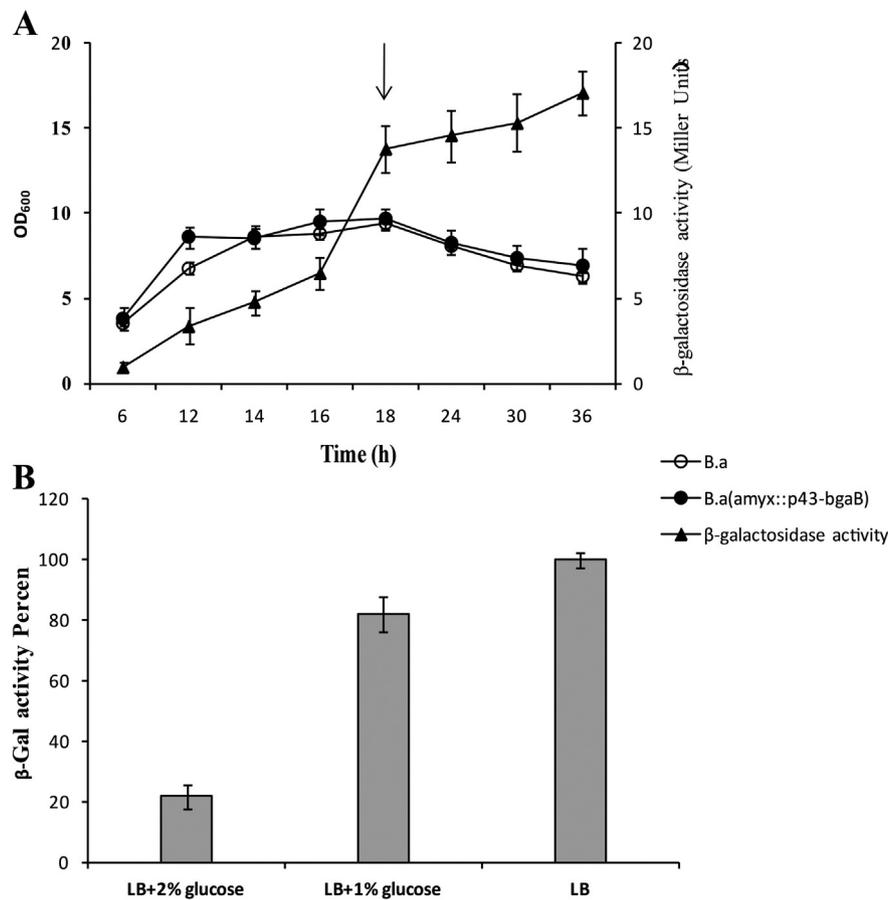
### 3.2. Summary of RNA-seq data

To obtain a global view of the *B. amyloliquefaciens* transcriptome, isolated mRNAs, after the removal of ribosomal mRNAs, were subjected to high-throughput Illumina sequencing. We obtained a total of 17.16 million reads with an average length of 100 bp. 9.54 million reads (55.61%) were uniquely mapped to the genome with a tolerance of a 2-bp mismatch, which represented a 240-fold coverage of the XH7 genome. Overall, 98.32% of all reads were mapped to genome regions. The remaining reads were either of low quality or only adapters (~1.68%). Approximately 76% of the expressed genes had a sequencing coverage of >90%. For the coding transcriptome, the number of reads mapped to each CDS was calculated and normalized for gene length and library depth to generate RPKM values. Based on the RPKM values, we quantified the expression of 3936 genes (93.63%) in the XH7 genome database (Fig. 2 and Supplementary Table s3).

Differentially expressed genes were subjected to GO functional enrichment analysis. 3936 genes were assigned to 38 functional groups. WEGO analysis illustrated that in the *B. amyloliquefaciens* genome, the majority of WEGO terms were categorized as metabolic processes, cellular processes, cell components, cell, catalytic activity, binding (14.5, 12.4, 11.5, 11.5, 16.7, and 12.6%, respectively) (Fig. 3A). GO annotation of genes within the top 10% confidence interval for all RPKM values (Fig. 3B) showed that GO terms of reproduction, macromolecular complexes, electron carrier activity, structural molecule activity, and translational regulator activity contained more highly expressed genes. The remainder of the highly expressed genes were involved in cellular component biogenesis, localization, and pigmentation (7.7, 7.9, and 4.7%, respectively).

### 3.3. Validation of RNA-seq analysis

To validate the RNA-seq results, an independent experiment was performed to confirm gene transcription by RT-PCR. An RNA sample was isolated as previously described, reverse-transcribed, and analyzed. The threshold cycle (Ct) values were determined at the threshold fluorescence value of 0.2. Seven genes were randomly chosen including downstream genes of  $P_{r2}$ ,  $P_{r4}$ ,  $P_{r6}$ ,  $P_{r7}$ ,  $P_{r8}$  (*sigW*, *yuaF*, *yqeZ*, *yceC*, *acoA*) and *yvyD*, *kata*. RT-PCR results showed that *sigW*, *yvyD*, *yuaF*, and *kata* were highly expressed comparable to *bgab*, and *acoA*, *yqeZ* and *yceC*



**Fig. 1.** Cell growth and  $\beta$ -Gal activity change curve of XH7-*bgaB* in rich LB medium. (A) The growth curve of strain XH7 (open circle) and XH7-*bgaB* (solid circle), and the  $\beta$ -Gal activity change curve (solid triangle) were demonstrated. (B) The  $\beta$ -Gal activity of XH7-*bgaB* was detected in different mediums. Data points represent means of three replicated studies of each sample with the standard errors of means.

were closed to the expression of *bgaB* (Supplementary Table s4 and Fig. s3), which was consistent with RNA-seq data.

### 3.4. Transcription start site (TSS) and transcriptional termination site determination

To understand the transcription of *B. amyloliquefaciens* genes, we globally defined the precise location of UTRs using RNA-seq data. After searching for a break in the transcribed region around genes, those genes whose ends overlap with other genes were excluded, and the remainder of genes were used to define the UTRs. TSSs were determined using a computer algorithm as previously described (Vijayan et al., 2011). Following this principle, we identified 1064 TSSs in the XH7 genome, among which 476 are located within operons. Based on the TSS map, the length distribution of the 5' untranslated upstream regions for this set of 1064 transcripts is shown in Fig. 4A and Supplementary Table s5. More than 50% of 5' UTRs varied between 10 bp and 50 bp in length according to the statistical analysis. 15.3% of these 5' UTRs had lengths shorter than 10 bp, while 16.3% had long 5' UTRs (ranging from 110 bp to 750 bp). Fifty-seven genes had 5' UTRs shorter than 5 bp (leaderless transcription). In contrast, *BAXH7\_01306* and *ynnC*, encoding an acyltransferase and integral inner membrane protein, respectively, had 5' UTRs longer than 700 bp.

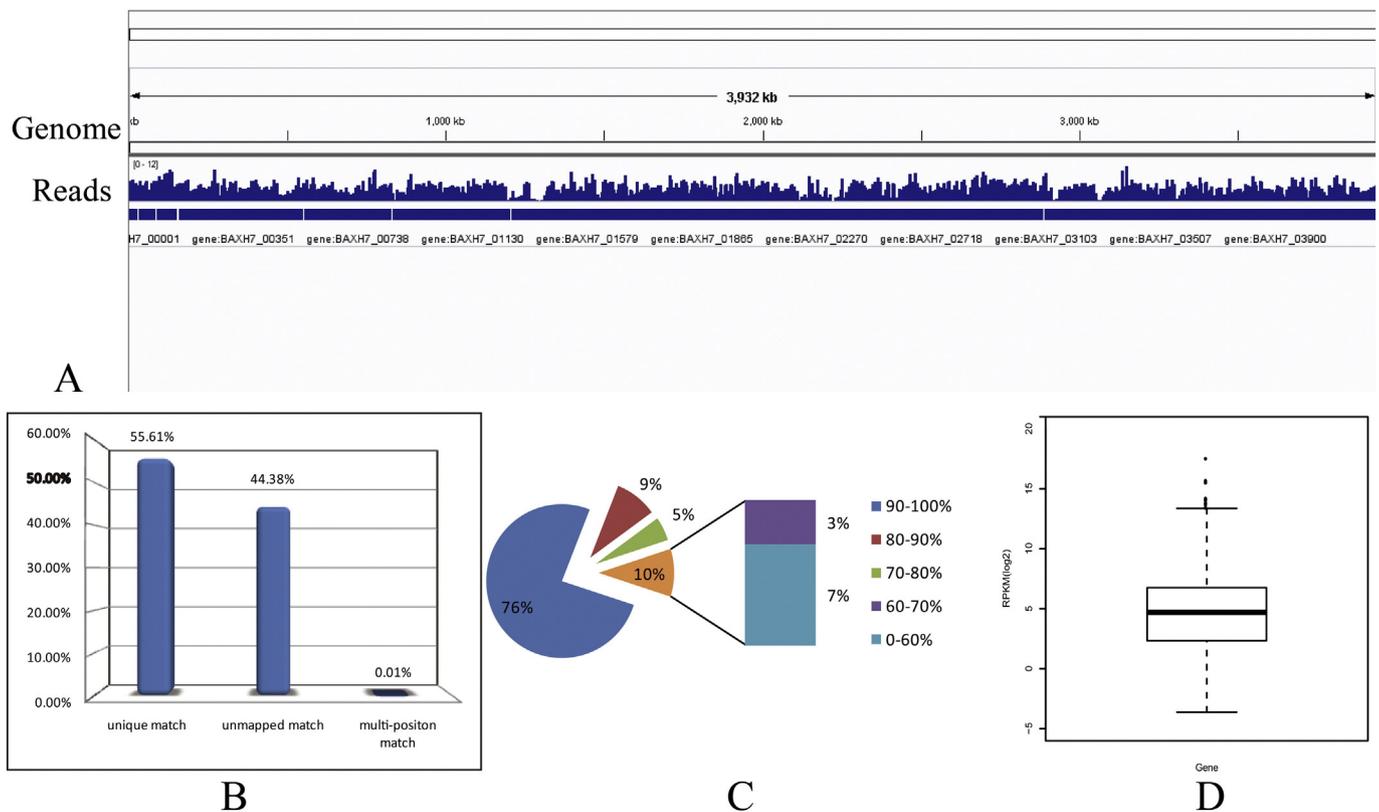
Additionally, 635 transcription end sites (TESs) were detected based on RNA-seq data, 302 of which belong to operons. The medium length of 3' UTRs was 18 bp (Fig. 4B), and 65.66% of the 3' UTRs were between 10 bp and 30 bp in length. 15.1% of them were longer than 110 bp, and

3.6% exceeded 500 bp. For 420 genes possessing both 5' and 3' UTRs, 194 belonged to multi-gene operons.

The GO functional analysis illustrated that genes belonging to the GO terms of organelle, cellular component organization and biogenesis, molecular transducer activity, and multicellular organismal process had significantly increased UTR lengths (Fig. 4C). Genes belonging to the GO terms of antioxidant activity and virion had particularly short UTRs.

### 3.5. Operon identification

The RNA-seq data, as well as the TSS and TES data, enabled us to identify operon structures in *B. amyloliquefaciens*. We compared operons from the RNA-seq data with operons from the XH7 genome using DOOR (Database of prokaryotic Operons) (Supplementary Table s6). We obtained the operon map based on the genome prediction from the DOOR analysis to assign 2681 (62.6%) of all XH7 genes to 883 operons. Operon predictions based on RNA-seq data resulted in 2467 (57.6%) genes located in 749 operons, 367 of which were completely identical with the genome prediction results from DOOR (Fig. 5A). The distribution of operons from two datasets was shown in Fig. 5B. Most operons are bicistronic (55.5% in the genome; 49.6% in the RNA-seq data), and 12 operons encompass more than 10 genes. The longest operon, which encoded flagellar biosynthesis proteins, contained 29 genes (*BAXH7\_01813* to *BAXH7\_01841*) according to the genome prediction and 30 genes (from *BAXH7\_01815* to *BAXH7\_01844*)



**Fig. 2.** The genome-scale transcriptional profile of XH7-*bgaB*. (A) RNA-Seq reads mapping to the XH7-*bgaB* genome. (B) The column represents the percentage of RNA-seq reads with unique match, unmapped match and multi-position match, respectively. (C) The pie chart depicted the distribution of genes' coverage according to the annotated gene set. (D) Box-and Whisker plots of log<sub>2</sub>-transformed RPKM for all transcriptional activity genes. The three horizontal lines in graph show the first quartile, median and third quartiles, respectively. The vertical lines in the box stand for the inner boundaries, the diamonds stand for the data outside the inner boundaries.

according to the RNA-seq data, Based on the RNA-seq data, we determined multi-gene operons (an example was shown in Fig. 5C).

### 3.6. Screening of highly active promoters based on RNA-seq data

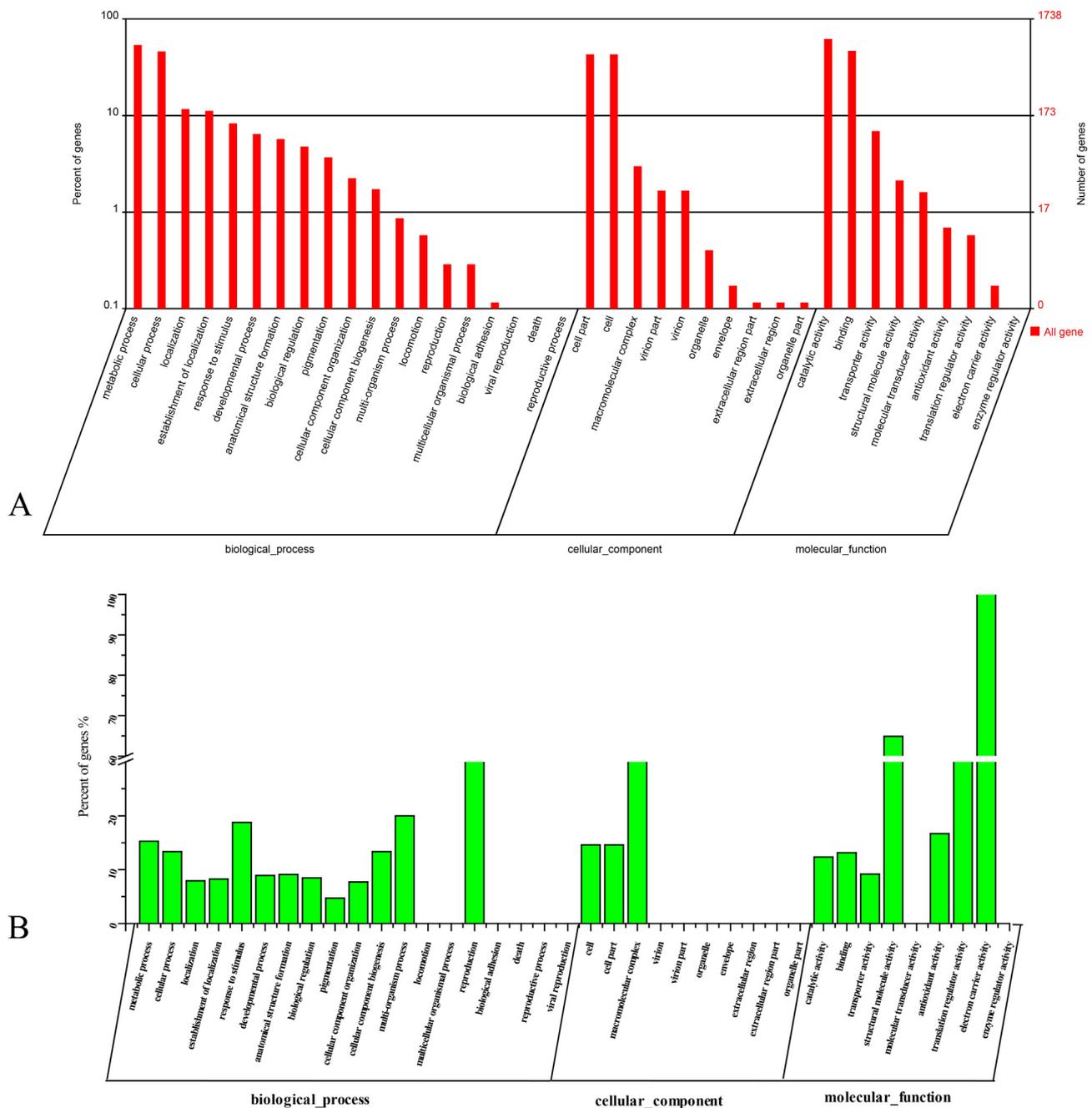
Compared with the marker gene *P*<sub>43</sub>-*bgaB*, 288 genes demonstrated high expression. After removing multi-gene operons, 198 nucleotide sequences upstream of the start codons of the 288 genes were analyzed for putative binding sites for  $\sigma$ -factors and transcription factors (TFs). Among the 198 active promoter candidates, 72 were possibly controlled by a single  $\sigma$ -factor and 65 were possibly controlled by multiple  $\sigma$ -factors (Supplementary Table s7). Using the DBTBS database, we identified 137 active promoters, including  $\sigma^A$ -dependent promoters (19, 13.9%),  $\sigma^B$ -dependent promoters (6, 4.4%),  $\sigma^D$ -dependent promoters (6, 4.4%),  $\sigma^E$ -,  $\sigma^F$ -,  $\sigma^G$ -, and  $\sigma^K$ -dependent promoters, (sporulation-specific factors, 13, 9.5%), (Haldenwang, 1995),  $\sigma^H$ -dependent promoters (11, 8.0%),  $\sigma^I$ -dependent promoters (3, 2.2%),  $\sigma^W$ - and  $\sigma^X$ -dependent promoters (14, 10.2%) and multiple  $\sigma$ -factor-dependent promoters (65, 47.4%). The box plots (Fig. 6A) showed the RPKM values of the 137 active promoters controlled by different  $\sigma$ -factors. The results demonstrated that  $\sigma^W$ - and  $\sigma^X$ -dependent promoters exhibited the highest activities, followed by  $\sigma^L$ -dependent promoters and multiple  $\sigma$ -factor-dependent promoters.  $\sigma^E$ -,  $\sigma^F$ -,  $\sigma^G$ - and  $\sigma^K$ -dependent promoters displayed the lowest activity among the active promoters. The downstream genes of these active promoters were used for the GO functional enrichment analysis (Supplementary Fig. s4). The results demonstrated that high activity genes were involved in cell, cell constituents, binding, catalytic, cellular processes, and metabolic processes.

We identified 11 promoters that have overlapping  $\sigma$ -factors binding sites, and the structural connection with the different  $\sigma$ -factors was

shown in Supplementary Table S8. The *yuaJ* gene has only a 3-bp overlap for  $\sigma^W$  and  $\sigma^H$  binding sites, and the *yxjG* gene has a 24-bp overlap for the binding sites of the  $\sigma^G$  and  $\sigma^A$  factors. In some cases, different  $\sigma$ -factors recognized very similar promoter consensus sequences, such as those in the *msmX*, *BAXH7\_04036*, *fenE*, and *clpC* genes.

There were 33 different TF binding sites found in the 137 active promoters, and four classes, from global to local, were defined according to the hierarchical architecture of the *B. subtilis* genetic regulatory network: CcpA, ResD, AbrB, and Fur (Goelzer et al., 2008), including class1: CcpA (41); class2: ResD (21) and SpoOA (19); class3: AbrB (21), CodY (66), DegU (60), FNR (9), PurR (18), PerR (23), SpoIIID (12), and TnrA (26); and class4: Fur (18) and so on (Supplementary Table s9).

We chose candidates with the highest RPKM values (from  $P_{r1}$  to  $P_{r8}$ ) according to the RNA-seq data to identify highly active promoters (Supplementary Table s10).  $P_{r1}$  was located within the coding region of the *BAXH7\_03271* gene;  $P_{r2}$  was located within the coding region of the *sigW* gene and regulated by *sigW-ybbM* operon;  $P_{r3}$  was located within the coding region of the *rapA* gene and regulated by *phrA-rapA* operon;  $P_{r4}$  was located within the coding region of the *yuaF* gene and regulated by *yuaFGI* operon;  $P_{r5}$  was located within the coding region of the *ydH* gene.  $P_{r6}$  was located within the coding region of the *yqeZ* gene and regulated by *yqeZ-yqfA-yqfB* operon;  $P_{r7}$  was located within the coding region of the *yceC* gene and regulated by *yceCDEFGH* operon;  $P_{r8}$  was located within the coding region of the *acoA* gene and regulated by *acoABCL* operon. The putative binding sites for  $\sigma$ -factors and TFs of eight candidates were predicted in DBTBS (Supplementary Table s11). However, the putative  $\sigma$ -factor binding sites of  $P_{r1}$ ,  $P_{r3}$ , and  $P_{r5}$  could not be predicted. Additionally, we identified and defined TF families for the candidates (Moreno-Campuzano et al., 2006). The results



**Fig. 3.** GO functional enrichment analysis of XH7-*bgaB* genes from RNA-seq. (A) WEGO illustration of all expression genes from RNA-seq (red column). (B) GO annotation (green column) of genes within top 10% confidence interval for all RPKM values.

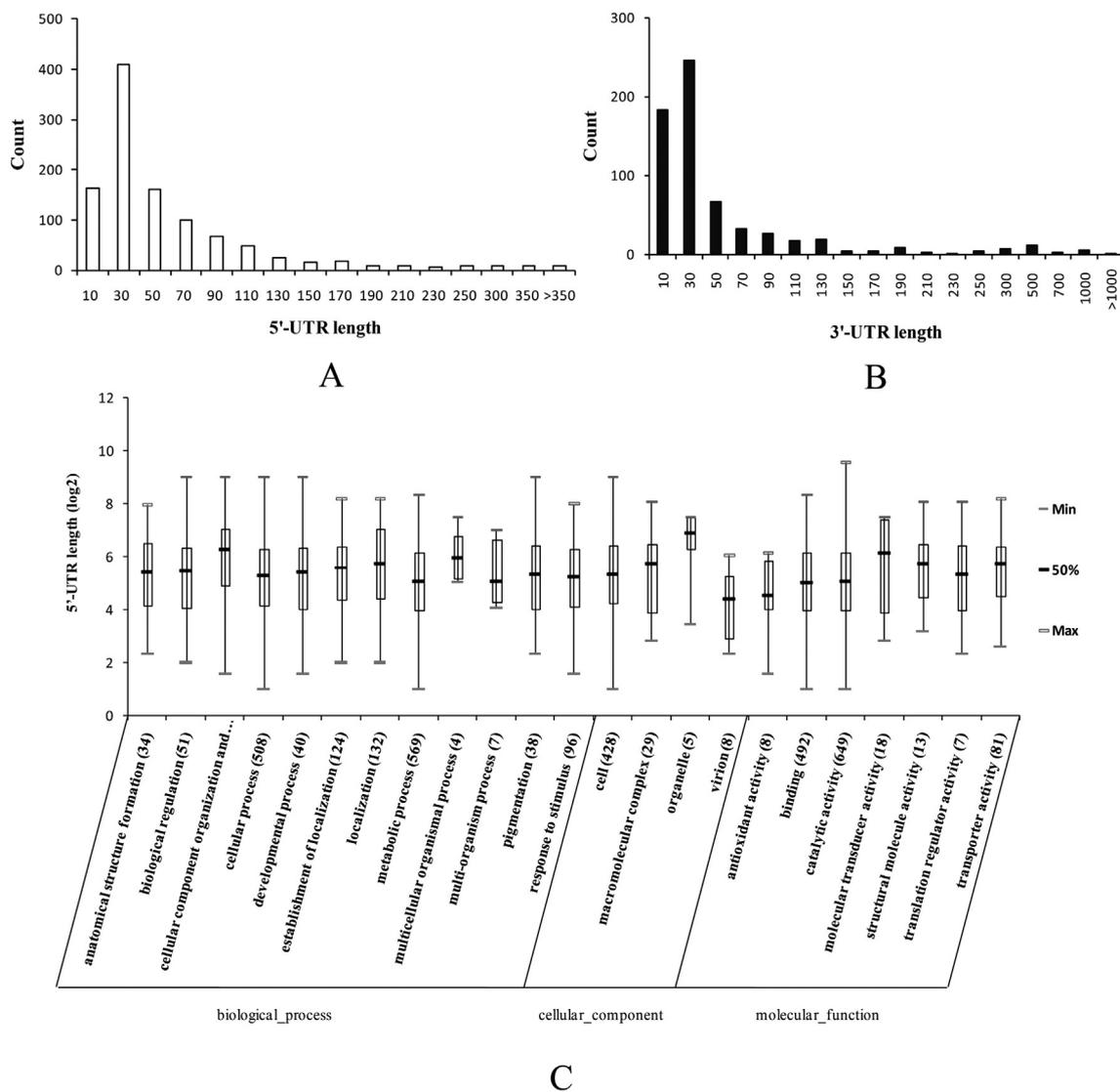
showed that 16 of 29 (55.2%) TFs were predicted to be repressor proteins, seven (24.1%) were activator and six (20.7%) were dual regulator. This shows that more promoters were repressed than activated.

$\beta$ -Gal activity analysis showed that  $P_{r2}$  displayed the strongest activity during post-log phase, followed by  $P_{r6}$ ,  $P_{r7}$ , and  $P_{r3}$ . Moreover, the maximal  $\beta$ -Gal specific activity for  $P_{r2}$  reached 5300 Miller units in LB medium ( $P_{r43}$  reached 5720 Miller units). Unfortunately, the  $\beta$ -Gal activities of  $P_{r5}$ ,  $P_{r4}$ ,  $P_{r1}$ , and  $P_{r8}$  were less than 1000 Miller units (Fig. 6B). Moreover,  $\beta$ -Gal expression plasmids were transformed into *B. subtilis* WB800. SDS-PAGE analyses of  $\beta$ -Gal from *B. subtilis* WB800 harbouring pBEP<sub>r2</sub>-*bgaB* and pBEP<sub>r43</sub>-*bgaB* were performed (Fig. 7). The results illustrated that  $\beta$ -Gal was expressed controlled by  $P_{r2}$  promoter at the postexponential phases of growth. And the  $\beta$ -Gal expression level controlled by  $P_{r2}$  promoter was comparable to that controlled by  $P_{r43}$

promoter, a commonly used strong promoter in *B. subtilis*. SDS-PAGE analyses were consistent with the enzyme activity assay of  $\beta$ -Gal.

#### 4. Discussion

Here we presented the global transcriptional landscape of XH7 using RNA-seq approach. In earlier studies, Fan et al. identified 302 genes involved in interactions between *B. amyloliquefaciens* FZB42 with plants using a two-color DNA microarray system (Fan et al., 2012). In this study, 9.54 million reads (55.61%) were generated and represented a 240-fold coverage of the XH7 genome. A total of 3936 genes (91.8% of 4286 total genes in the genome) were detected under our experimental condition. To validate the RNA-seq results, independent experiment was performed to confirm gene transcription by RT-PCR. In summary,



**Fig. 4.** Classification and distribution of UTRs. (A) Length distribution of 5' UTRs. (B) Length distribution of 3' UTRs. (C) Categories of GO functional enrichment analysis of 5' UTRs length box plots. Y axis stands for  $\log_2$  values of 5' UTRs length. X axis was GO categories.

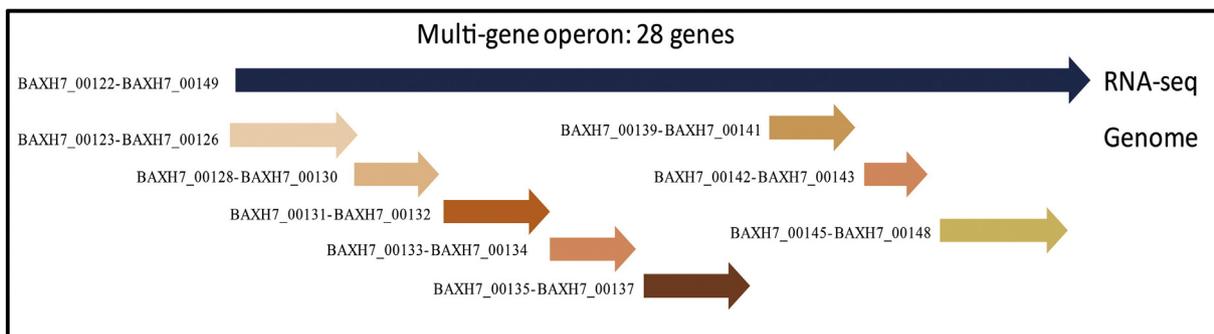
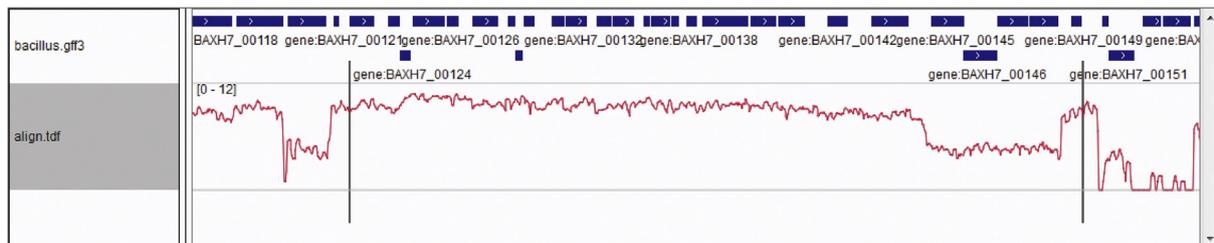
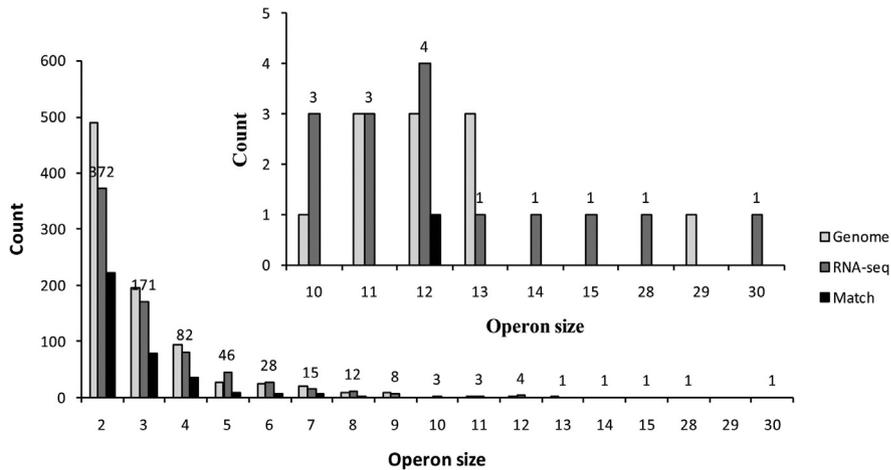
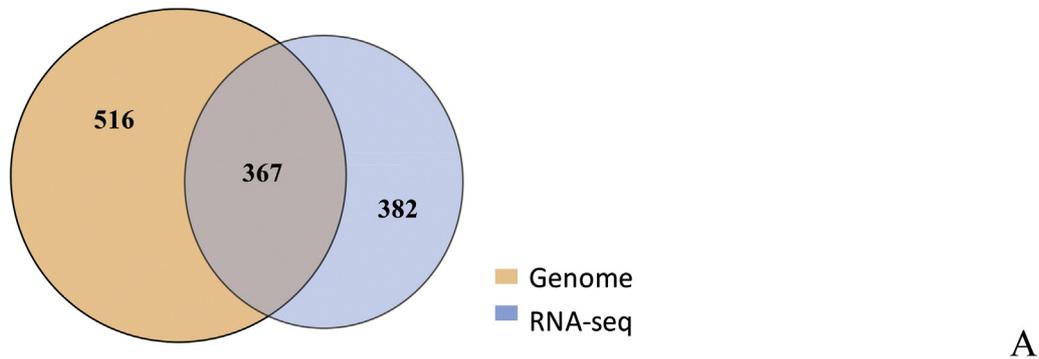
RNA-seq analysis revealed that sufficient sequencing depth was obtained to define the gene transcripts (Wang et al., 2011a; Robles et al., 2012). And the RT-PCR analysis suggested that the RNA-seq data were reliable. Based on the complete genome sequence of XH7, the transcriptional map of XH7 advanced our understanding of its gene expression and genome content.

The transcript structures of bacteria represent a critical step in understanding the function and regulatory architecture of their genomes (Albrecht et al., 2011; Vijayan et al., 2011; Perkins et al., 2009; Martin et al., 2010; Pfeifer-Sancar et al., 2013; Mendoza-Vargas et al., 2009; Cortes et al., 2013; Kröger et al., 2012; McGrath et al., 2007). Based on the *B. amyloliquefaciens* transcriptome, 1064 TSSs and 749 operon structures throughout the genome were determined. We found that more than 50% of the 5' UTRs were shorter than 30 bp, and that 37.4% of the 3' UTRs were about 30 bp in length. Wang et al. demonstrated that the length of most (52%) 5' UTRs varied between 10 and 50 bp in *B. thuringiensis* (Wang et al., 2013a). They showed that different species exhibited significant differences in the lengths of their 5' UTRs. These results are consistent with a conclusion from Wang et al. Additionally, the relationship between UTR length and gene function was studied. The results showed that genes classified by the GO terms organelle, cellular component organization and biogenesis, molecular transducer activity, and multicellular organismal process had significantly longer UTRs,

while those classified by the GO terms of antioxidant activity and virion had particularly short UTRs. Sharma et al. found that genes involved in the regulation of cellular processes, such as cell division, pathogenesis, or transformation, had significantly longer UTRs, while plasmid- and transposon-related genes had particularly short UTRs (Sharma et al., 2010).

Gene organization in operons is crucial to understand the regulatory networks of prokaryotic genomes (Wang et al., 2011b; Passalacqua et al., 2009; Brouwer et al., 2008; Fortino et al., 2014; Charaniya et al., 2007; Santini et al., 2013; Mendoza-Vargas et al., 2009; Koide et al., 2009; Ramachandran et al., 2012; Choi et al., 2014). We presented operon maps of *B. amyloliquefaciens* based on RNA-seq data, and compared it with operons from the XH7 genome using DOOR (Database of prokaryotic OpeRons). The results confirmed the operons identified by RNA-seq data, and determined the extent of new operon structures. Martin et al. found that OperonDB (a database of predicted operons in microbial genomes) predictions with confidences ranging from 50% to 75% were much less conclusive with the *B. anthracis* transcript map predicted from raw RNA-seq reads (Martin et al., 2010). Accordingly, the high-throughput RNA-seq technology exhibited a high efficiency in identifying operon maps.

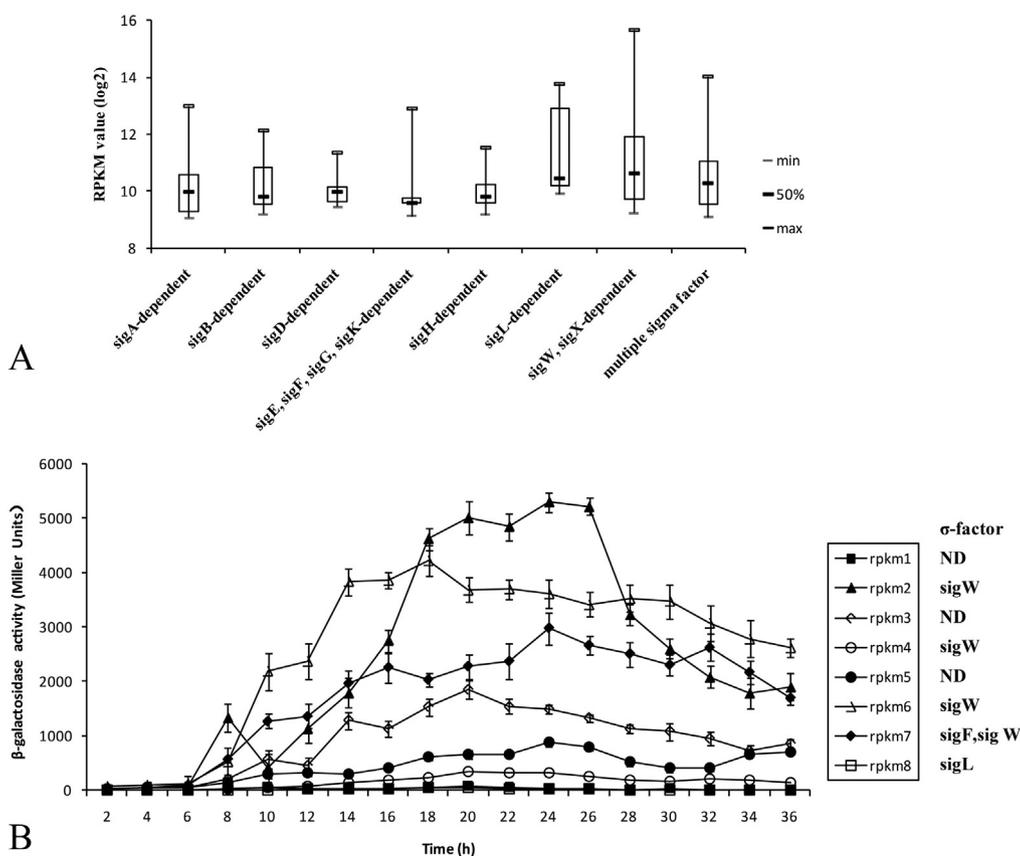
Promoters have very important regulatory effects on gene expression and enzyme production (Wang et al., 2013a,b; Schumann, 2007).



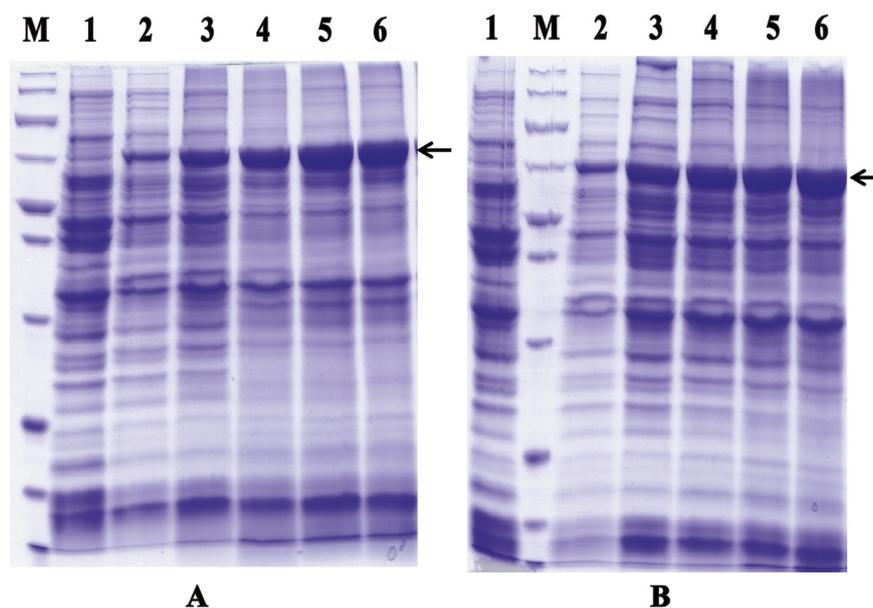
**Fig. 5.** Operon identification. (A) Venn diagram showing overlap among operon identification. Number 367 means operons of RNA-seq were completely identical with genome prediction results from Door. Number 883 means operons prediction from Door analysis (orange). Number 749 means operons identified from RNA-seq (blue). (B) Distribution of operons of XH7-*bgaB* from RNA-seq was demonstrated according to the operon encompass the number of genes. (C) Top: Combined cDNA reads (read wave line) mapped the XH7-*bgaB* chromosome region from BAXH7\_00122 to BAXH7\_00149. Bottom: an example of scheme of operon structure identification from RNA-seq and prediction from genome.

In this study, we scanned bacterial transcriptional data and TSSs to search for highly expressed promoters. We selected 288 genes with higher expression levels (RPKM values) than that of the marker gene

*P<sub>43</sub>-bgaB*. Among the 288 genes, we identified 137 active promoters and analyzed putative binding sites for  $\sigma$ -factors and TFs. Based on 137 active promoters, the candidate promoter *P<sub>12</sub>* displayed the



**Fig. 6.** Candidates identification for searching highly active promoters. (A) 137 active promoters were classified as  $\sigma^A$ -dependent promoters,  $\sigma^B$ -dependent promoters,  $\sigma^D$ -dependent promoters,  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$ -dependent promoters,  $\sigma^H$ -dependent promoters,  $\sigma^I$ -dependent promoters,  $\sigma^W$ ,  $\sigma^X$ -dependent promoters and multiple  $\sigma$ -factor promoters. And Y axis stands for  $\log_2$  values of different sigma factors promoters. X axis indicated different  $\sigma$ -factors.  $\sigma^E$ ,  $\sigma^F$ ,  $\sigma^G$ , and  $\sigma^K$ -dependent promoters (13) are sporulation-specific factors,  $\sigma^E$ -dependent promoters (3),  $\sigma^F$ -dependent promoters (6),  $\sigma^G$ -dependent promoters (2) and  $\sigma^K$ -dependent promoters (2) together.  $\sigma^W$  and  $\sigma^X$ -dependent promoters (14) were extracytoplasmic function  $\sigma$ -factors (ECF) which combined  $\sigma^W$  dependent promoters (13) and  $\sigma^X$ -dependent promoters (1) together. The multiple sigma factor promoters were recognized by more than two kinds of sigma factor. (B)  $\beta$ -Gal activities of the candidate  $P_{r1}$ – $P_{r8}$  acting throughout the life cycle.  $\sigma$ -factor of candidates located on the right side of the legend. Data points represent means of three replicated studies of each sample with the standard errors of the means.



**Fig. 7.** SDS-PAGE analysis of  $\beta$ -Gal in *B. subtilis* WB800 at different times. (A) Lane 1, molecular mass markers (top to bottom: 250, 150, 100, 70, 50, 40, 30, 20, 15, 10 and 5 kDa); Lane 2, intracellular proteins of *B. subtilis* WB800 (pBEP<sub>43</sub>-SmyQ) as a control; and Lanes 3–6, intracellular proteins of *B. subtilis* WB800 (pBEP<sub>43</sub>-bgaB) at 8 h, 16 h, 24 h, 32 h, 40 h. (B) Lane 1, intracellular proteins of *B. subtilis* WB800 (pBEP<sub>43</sub>-SmyQ) as a control; Lane 2, molecular mass markers (top to bottom: 250, 150, 100, 70, 50, 40, 30, 20, 15, 10 and 5 kDa); and Lanes 3–6, intracellular proteins of *B. subtilis* WB800 (pBEP<sub>43</sub>-bgaB) at 8 h, 16 h, 24 h, 32 h, 40 h. The cells were harvested by centrifugation and the crude proteins were extracted by boiling (20  $\mu$ L sample volumes). The protein bands of  $\beta$ -Gal are marked by arrows.

strongest  $\beta$ -Gal specific activity during post-log phase. Interestingly,  $\sigma^W$ -dependent promoters ( $P_{r2}$ ) had relatively higher activity than  $\sigma^A$ - or  $\sigma^B$ -dependent promoters. Additionally, multiple  $\sigma$ -factor-dependent promoters that demonstrated high activities, such as those of the *yceE*, *yvlB*, and *yoeB* genes, were also recognized by  $\sigma^W$ .  $\sigma^W$  and  $\sigma^X$  belong to extracytoplasmic function  $\sigma$ -factors (ECF), which constitute the largest and most diverse family of alternative  $\sigma$ -factors responding to cell envelope stress, iron levels, and oxidation states (Haldenwang, 1995; Pátek and Nešvera, 2011). Additionally,  $\sigma^W$  and  $\sigma^X$ , whose regulons overlap, might have overlapping promoter selectivity (Sharma et al., 2010; Huang et al., 1998; Ho and Ellermeier, 2012). The results would contribute to the understanding of the *B. amyloliquefaciens* regulatory network controlled by  $\sigma^W$ -dependent promoters. Martin et al. showed that there was a positive correlation between gene expression and upstream promoter strength by analyzing gene expression in *B. anthracis* (Martin et al., 2010). Our data support the concept of using highly expressed genes to discover strong promoters based on RNA-seq data. However, the regulation of gene expression is much more complicated, because promoters contain different TF binding sites (Moreno-Campuzano et al., 2006). Therefore, to improve the efficiency of promoter  $P_{r2}$ , further studies will be necessary to characterize the DNA element of  $P_{r2}$  promoter.

In conclusion, using RNA-seq technology, we characterized the transcriptome of *B. amyloliquefaciens* on a genome scale. Our data provided further understanding of the transcriptional organization and gene structure of *B. amyloliquefaciens*. These results were a substantial contribution to our knowledge of *B. amyloliquefaciens* gene regulation, and the identified TSSs provided substantial new information that will aid the search for industrially useful promoters.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2015.06.066>.

## Acknowledgments

We wish to acknowledge the financial support by the 863 program (National High Technology Research and Development Program of China) under contract NO. 2014AA021304, the Natural Science Foundation of Guangdong Province (S2012030006235), and the Science and Technology Planning Project of Guangdong Province (2012B010900028, and 2013B010404007).

## References

- Albrecht, M., Sharma, C.M., Dittrich, M.T., Müller, T., Reinhardt, R., Vogel, J., Rudel, T., 2011. The transcriptional landscape of *Chlamydia pneumoniae*. *Genome Biol.* 12, R98. <http://dx.doi.org/10.1186/gb-2011-12-10-r98>.
- Brouwer, R.W.W., Kuipers, O.P., Van Hijum, S.A.F.T., 2008. The relative value of operon predictions. *Brief. Bioinform.* 9, 367–375. <http://dx.doi.org/10.1093/bib/bbn019>.
- Charaniya, S., Mehra, S., Lian, W., Jayalak, K.P., Karypis, G., Hu, W.S., 2007. Transcriptome dynamics-based operon prediction and verification in *Streptomyces coelicolor*. *Nucleic Acids Res.* 35, 7222–7236. <http://dx.doi.org/10.1093/nar/gkm501>.
- Chen, X.H., Koumoutsis, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W.R., Reva, O., Junge, H., Voigt, B., Jungblut, P.R., Vater, J., Süßmuth, R., Liesegang, H., Strittmatter, A., Gottschalk, G., Borriss, R., 2007. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nat. Biotechnol.* 25, 1007–1014. <http://dx.doi.org/10.1038/nbt1325>.
- Chen, W., Chen, H., Xia, Y., Zhao, J., Tian, F., Zhang, H., 2008. Production, purification, and characterization of a potential thermostable galactosidase for milk lactose hydrolysis from *Bacillus stearothermophilus*. *J. Dairy Sci.* 91 (5), 1751–1758. <http://dx.doi.org/10.3168/jds.2007-617>.
- Choi, S., Jung, J., Jeon, C.O., Park, W., 2014. Comparative genomic and transcriptomic analyses of NaCl-tolerant *Staphylococcus* sp. OJ82 isolated from fermented seafood. *Appl. Microbiol. Biotechnol.* 98, 807–822. <http://dx.doi.org/10.1007/s00253-013-5436-2>.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Jalón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. <http://dx.doi.org/10.1093/bioinformatics/bti610>.
- Cortes, T., Schubert, O.T., Rose, G., Arnvig, K.B., Comas, I., Aebersold, R., Young, D.B., 2013. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* 5, 1121–1131. <http://dx.doi.org/10.1016/j.celrep.2013.10.031>.

- Croucher, N.J., Thomson, N.R., 2010. Studying bacterial transcriptomes using RNA-seq. *Curr. Opin. Microbiol.* 13, 619–624. <http://dx.doi.org/10.1016/j.mib.2010.09.009>.
- Fan, B., Carvalhais, L.C., Becker, A., Fedoseyenko, D., Von Wirén, N., Borriss, R., 2012. Transcriptomic profiling of *Bacillus amyloliquefaciens* FZB42 in response to maize root exudates. *BMC Microbiol.* 12, 116. <http://dx.doi.org/10.1186/1471-2180-12-116>.
- Fortino, V., Smolander, O.P., Auvinen, P., Tagliaferri, R., Greco, D., 2014. Transcriptome dynamics-based operon prediction in prokaryotes. *BMC Bioinform.* 15, 145. <http://dx.doi.org/10.1186/1471-2105-15-145>.
- Goelzer, A., Brikci, F.B., Martin-Verstraete, I., Noirot, P., Bessières, P., Aymerich, S., Fromion, V., 2008. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst. Biol.* 2, 20. <http://dx.doi.org/10.1186/1752-0509-2-20>.
- Haldenwang, W.G., 1995. The sigma factors of *Bacillus subtilis*. *Microbiol. Rev.* 59, 1–30.
- Hirata, H., Negoro, S., Okada, H., 1985. High production of thermostable  $\beta$ -galactosidase of *Bacillus stearothermophilus* in *Bacillus subtilis*. *Appl. Environ. Microbiol.* 49, 1547–1549.
- Ho, T.D., Ellermeier, C.D., 2012. Extra cytoplasmic functionofactor activation. *Curr. Opin. Microbiol.* 15, 182–188. <http://dx.doi.org/10.1016/j.mib.2012.01.001>.
- Huang, X., Fredrick, K.L., Helmann, J.D., 1998. Promoter recognition by *Bacillus subtilis*  $\sigma^W$ : autoregulation and partial overlap with the  $\sigma^X$  regulon. *J. Bacteriol.* 180, 3765–3770.
- Imov, I., Sharma, C.M., Vogel, J., Winkler, W.C., 2010. Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Res.* 38, 6637–6651. <http://dx.doi.org/10.1093/nar/gkq454>.
- Koide, T., Reiss, D.J., Bare, J.C., Pang, W.L., Facciotti, M.T., Schmid, A.K., Pan, M., Marzolf, B., Van, P.T., Lo, F.Y., Pratap, A., Deutsch, E.W., Peterson, A., Martin, D., Baliga, N., 2009. Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol. Syst. Biol.* 5, 285. <http://dx.doi.org/10.1038/msb.2009.42>.
- Kröger, C., Dillon, S.C., Cameron, A.D., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G.C., Lohan, A.J., Loftus, B., Lucchini, S., Ussery, D.W., Dorman, C.J., Thomson, N.R., Vogel, J., Hinton, J.C., 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *PANS.* 109, E1277–E1286. <http://dx.doi.org/10.1073/pnas.1201061109>.
- Kröger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankarana, S.K., Liu, L., Liu, Y., Shin, H., Chen, R.R., Wang, N.S., Li, J., Du, G., Chen, J., 2013. Developing *Bacillus* spp. as a cell factory for production of microbial enzymes and industrially important biochemicals in the context of systems and synthetic biology. *Appl. Microbiol. Biotechnol.* 97, 6113–6127. <http://dx.doi.org/10.1007/s00253-013-4960-4>.
- Mao, F., Dam, P., Chou, J., Olman, V., Xu, Y., 2009. DOOR: a database for prokaryotic operons. *Nucleic Acids Res.* 37, D459–D463. <http://dx.doi.org/10.1093/nar/gkn757>.
- Martin, J., Zhu, W., Passalacqua, K.D., Bergman, N., Borodovsky, M., 2010. *Bacillus anthracis* genome organization in light of whole transcriptome sequencing. *BMC Bioinform.* 11, S10. <http://dx.doi.org/10.1186/1471-2105-11-S3-S10>.
- McGrath, P.T., Lee, H., Zhang, L., Iniesta, A.A., Hottes, A.K., Tan, M.H., Hillson, N.J., Hu, P., Shapiro, L., McAdams, H.H., 2007. High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nat. Biotechnol.* 25, 584–592. <http://dx.doi.org/10.1038/nbt1294>.
- Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juárez, K., Contreras-Moreira, B., Huerta, A.M., Collado-Vides, J., Moret, E., 2009. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* 4 (10), e7526. <http://dx.doi.org/10.1371/journal.pone.0007526>.
- Moreno-Campuzano, S., Janga, S.C., Pérez-Rueda, E., 2006. Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes-agenomic approach. *BMC Genomics* 7, 147. <http://dx.doi.org/10.1186/1471-2164-7-147>.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. <http://dx.doi.org/10.1038/nmeth.1226>.
- Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeker, M., Aymerich, S., Becher, D., Bisicchia, P., Botella, E., Delumeau, O., Doherty, G., Denham, E.L., Fogg, M.J., Fromion, V., Goelzer, A., Hansen, A., Härtig, E., Harwood, C.R., Homuth, G., Jarmer, H., Jules, M., Klipp, E., Chat, L.L., Lecoqte, F., Lewis, P., Liebermeister, W., March, A., Mars, R.A.T., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rügheimer, F., Sappa, P.K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stülke, J., Wiegert, T., Devine, K.M., Wilkinson, A.J., Van Dijk, J.M., Hecker, M., Völker, U., Bessières, P., Noirot, P., 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335, 1103–1106. <http://dx.doi.org/10.1126/science.1206848>.
- Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., Bergman, N.H., 2009. Structure and complexity of a bacterial transcriptome. *J. Bacteriol.* 191, 3203–3211. <http://dx.doi.org/10.1128/JB.00122-09>.
- Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., Bergman, N.H., 2012. Strand-specific RNA-Seq reveals ordered patterns of sense and antisense transcription in *Bacillus anthracis*. *PLoS ONE* 7 (8), e43350. <http://dx.doi.org/10.1371/journal.pone.0043350>.
- Pátek, M., Nešvera, J., 2011. Sigma factors and promoters in *Corynebacterium glutamicum*. *J. Biotechnol.* 154, 101–113. <http://dx.doi.org/10.1016/j.jbiotec.2011.01.017>.
- Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assef, S.A., He, M., Croucher, N.J., Pickard, D.J., Maskell, D.J., Parkhill, J., Choudhary, J., Thomson, N.R., Dougan, G., 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid *Bacillus Salmonella* Typhi. *PLoS Genet.* 5, e1000569. <http://dx.doi.org/10.1371/journal.pgen.1000569>.
- Pfeifer-Sancar, K., Mentz, A., Rückert, C., Kalinowski, J., 2013. Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics* 14, 888. <http://dx.doi.org/10.1186/1471-2164-14-888>.

- Phan, T.T.P., Nguyen, H.D., Schumann, W., 2010. Establishment of a simple and rapid method to screen for strong promoters in *Bacillus subtilis*. *Protein Expr. Purif.* 71, 174–178. <http://dx.doi.org/10.1016/j.pep.2009.11.010>.
- Qiu, M., Xu, Z., Li, X., Li, Q., Zhang, N., Shen, Q., Zhang, R., 2014. Comparative proteomics analysis of *Bacillus amyloliquefaciens* SQR9 revealed the key proteins involved in in situ root colonization. *J. Proteome Res.* 13, 5581–5591. <http://dx.doi.org/10.1021/pr500565m>.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic feature. *Bioinformatics.* 26, 841–842. <http://dx.doi.org/10.1093/bioinformatics/btq033>.
- Ramachandran, V.K., Shearer, N., Jacob, J.J., Sharma, C.M., Thompson, A., 2012. The architecture and ppGpp-dependent expression of the primary transcriptome of *Salmonella typhimurium* during invasion gene expression. *BMC Genomics* 13, 25. <http://dx.doi.org/10.1186/1471-2164-13-25>.
- Rasmussen, S., Nielsen, H.B., Jarmer, H., 2009. The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol. Microbiol.* 73, 1043–1057. <http://dx.doi.org/10.1111/j.1365-2958.2009.06830.x>.
- Robles, J.A., Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J., Taylor, J.M., 2012. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* 13, 384. <http://dx.doi.org/10.1186/1471-2164-13-484>.
- Santini, T., Turchi, L., Ceccarelli, G., Franco, C.D., Beccari, E., 2013. Transcriptional analysis of *ftsZ* within the *dcw* cluster in *Bacillus mycoides*. *BMC Microbiol.* 13, 27. <http://dx.doi.org/10.1186/1471-2180-13-27>.
- Schallmeyer, M., Singh, A., Ward, O.P., 2004. Developments in the use of *Bacillus* species for industrial production. *Can. J. Microbiol.* 50, 1–17. <http://dx.doi.org/10.1139/W03-076>.
- Schumann, W., 2007. Production of recombinant proteins in *Bacillus subtilis*. *Adv. Appl. Microbiol.* 62, 137–189. [http://dx.doi.org/10.1016/S0065-2164\(07\)62006-1](http://dx.doi.org/10.1016/S0065-2164(07)62006-1).
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findei, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P.F., Vogel, J., 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464, 250–255. <http://dx.doi.org/10.1038/nature08756>.
- Sierro, N., Makita, Y., De Hoon, M., Nakai, K., 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res.* 36, D93–D96. <http://dx.doi.org/10.1093/nar/gkm910>.
- Sorek, R., Cossart, P., 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* 11, 9–16. <http://dx.doi.org/10.1038/nrg2695>.
- Vijayan, V., Jain, I.H., O'Shea, E.K., 2011. A high resolution map of a cyanobacterial transcriptome. *Genome Biol.* 12, R47. <http://dx.doi.org/10.1186/gb-2011-12-5-r47>.
- Wang, P.Z., Doi, R.H., 1984. Overlapping promoters transcribed by *Bacillus subtilis* sigma 55 and sigma 37 RNA polymerase holoenzymes during growth and stationary phases. *J. Biol. Chem.* 259 (13), 8619–8625.
- Wang, Y., Ghaffari, N., Johnson, C.D., Braga-Neto, U.M., Wang, H., Chen, R., Zhou, H., 2011a. Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. *BMC Bioinforma.* 12 (Suppl. 10), S5. <http://dx.doi.org/10.1186/1471-2105-12-S10-S5>.
- Wang, Y., Li, X., Mao, Y., Blaschek, H.P., 2011b. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *BMC Genomics* 12, 479. <http://dx.doi.org/10.1186/1471-2164-12-479>.
- Wang, J., Ai, X., Mei, H., Fu, Y., Chen, B., Yu, Z., He, J., 2013a. High-throughput identification of promoters and screening of highly active promoter-59-UTR DNA region with different characteristics from *Bacillus thuringiensis*. *PLoS ONE* 8 (5), e62960. <http://dx.doi.org/10.1371/journal.pone.0062960>.
- Wang, J., Mei, H., Zheng, C., Qian, H., Cui, C., Fu, Y., Su, J., Liu, Z., Yu, Z., He, J., 2013b. The metabolic regulation of sporulation and parasporal crystal formation in *Bacillus thuringiensis* revealed by transcriptomics and proteomics. *Mol. Cell. Proteomics* 12, 1363–1376. <http://dx.doi.org/10.1074/mcp.M112.023986>.
- Wiegand, S., Voigt, B., Albrecht, D., Bongaerts, J., Evers, S., Hecker, M., Daniel, R., Liesegang, H., 2013. Fermentation stage-dependent adaptations of *Bacillus licheniformis* during enzyme production. *Microb. Cell Factories* 12, 20.
- Yang, H., Liao, Y., Wang, B., Lin, Y., Pan, L., 2011. Complete genome sequence of *Bacillus amyloliquefaciens* XH7, which exhibits production of purine nucleosides. *J. Bacteriol.* 193, 5593–5594. <http://dx.doi.org/10.1128/JB.05880-11>.
- Yang, M., Zhang, W., Ji, S., Cao, P., Chen, Y., Zhao, X., 2013. Generation of an artificial double promoter for protein expression in *Bacillus subtilis* through a Promoter Trap System. *PLoS ONE* 8 (2), e56321. <http://dx.doi.org/10.1371/journal.pone.0056321>.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Lars Bolund, L., Wang, J., 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297. <http://dx.doi.org/10.1093/nar/gkl031>.
- Zakataeva, N.P., Nikitina, O.V., Gronskiy, S.V., Romanenkov, D.V., Livshits, V.A., 2010. A simple method to introduce marker-free genetic modifications into the chromosome of naturally nontransformable *Bacillus amyloliquefaciens* strains. *Appl. Microbiol. Biotechnol.* 85, 1201–1209. <http://dx.doi.org/10.1007/s00253-009-2276-1>.
- Zhang, W., Gao, W., Feng, J., Zhang, C., He, Y., Cao, M., Li, Q., Sun, Y., Yang, C., Song, C., Wang, S., 2014. A markerless gene replacement method for *B. amyloliquefaciens* LL3 and its use in genome reduction and improvement of poly- $\gamma$ -glutamic acid production. *Appl. Microbiol. Biotechnol.* 98, 8963–8973. <http://dx.doi.org/10.1007/s00253-014-5824-2>.