



OPEN ACCESS

EDITED BY
Sailesh Malla,
Chr. Hansen A/S, Denmark

REVIEWED BY
Michelle Meyer,
Boston College, United States
Marcus Lechner,
LOEWE Center for Synthetic
Microbiology (SYNMIKRO), Germany

*CORRESPONDENCE
Jan Gorodkin
gorodkin@rth.dk

SPECIALTY SECTION
This article was submitted to
Microbiotechnology,
a section of the journal
Frontiers in Microbiology

RECEIVED 31 March 2022
ACCEPTED 28 June 2022
PUBLISHED 04 August 2022

CITATION
Geissler AS, Poulsen LD, Doncheva NT,
Anthon C, Seemann SE,
González-Tortuero E, Breüner A,
Jensen LJ, Hjort C, Vinther J and
Gorodkin J (2022) The impact of PrsA
over-expression on the *Bacillus subtilis*
transcriptome during fed-batch
fermentation of alpha-amylase
production.
Front. Microbiol. 13:909493.
doi: 10.3389/fmicb.2022.909493

COPYRIGHT
© 2022 Geissler, Poulsen, Doncheva,
Anthon, Seemann, González-Tortuero,
Breüner, Jensen, Hjort, Vinther and
Gorodkin. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The impact of PrsA over-expression on the *Bacillus subtilis* transcriptome during fed-batch fermentation of alpha-amylase production

Adrian S. Geissler¹, Line D. Poulsen²,
Nadezhda T. Doncheva³, Christian Anthon¹,
Stefan E. Seemann¹, Enrique González-Tortuero¹,
Anne Breüner⁴, Lars J. Jensen³, Carsten Hjort⁴,
Jeppe Vinther² and Jan Gorodkin^{1*}

¹Department of Veterinary and Animal Sciences, Center for non-coding RNA in Technology and Health, University of Copenhagen, Copenhagen, Denmark, ²Department of Biology, University of Copenhagen, Copenhagen, Denmark, ³Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark, ⁴Novozymes A/S, Bagsværd, Denmark

The production of the alpha-amylase (AMY) enzyme in *Bacillus subtilis* at a high rate leads to the accumulation of unfolded AMY, which causes secretion stress. The over-expression of the PrsA chaperone aids enzyme folding and reduces stress. To identify affected pathways and potential mechanisms involved in the reduced growth, we analyzed the transcriptomic differences during fed-batch fermentation between a PrsA over-expressing strain and control in a time-series RNA-seq experiment. We observe transcription in 542 unannotated regions, of which 234 had significant changes in expression levels between the samples. Moreover, 1,791 protein-coding sequences, 80 non-coding genes, and 20 riboswitches overlapping UTR regions of coding genes had significant changes in expression. We identified putatively regulated biological processes *via* gene-set over-representation analysis of the differentially expressed genes; overall, the analysis suggests that the PrsA over-expression affects ATP biosynthesis activity, amino acid metabolism, and cell wall stability. The investigation of the protein interaction network points to a potential impact on cell motility signaling. We discuss the impact of these highlighted mechanisms for reducing secretion stress or detrimental aspects of PrsA over-expression during AMY production.

KEYWORDS

alpha-amylase, PrsA, *Bacillus subtilis*, RNA sequencing (RNA-seq), enzyme produced by microorganism

Introduction

Bacillus subtilis is a powerhouse for enzyme production in biotech industries (Schallmey et al., 2004; van Dijn and Hecker, 2013; Hohmann et al., 2016). Amylases are a specific class of enzymes that *B. subtilis* can produce commercially (Schallmey et al., 2004). The amylase enzyme, in particular the alpha-amylase (AMY), is a digestive enzyme (EC 3.2.1.1) that degrades starch molecules. Therefore, AMY is often an active component in laundry detergent for removing sticky stains from clothes. For successful AMY production and subsequent recovery, a host organism needs to both express and secrete AMY proteins in a biologically active form at a high rate (Spinnler, 2021). However, a major issue for commercial production is that the protein folding system of the cell is overwhelmed by the high rate of synthesis unless the strains used for production are genetically modified (Kontinen and Sarvas, 1993). The accumulation of unfolded AMY proteins causes stress that requires a bacterial cell to physiologically adapt to survive (Storz and Hengge, 2010). The Sec secretion pathway secretes AMY co-translationally (Fu et al., 2007). Therefore, unfolded AMY is extracellular, such that the corresponding stress signal triggers the heat shock response (Westers et al., 2004, 2006; Storz and Hengge, 2010; Lim and Gross, 2014; Yan and Wu, 2019). The simplified mechanism of this stress response has two components as follows (Westers et al., 2004, 2006; Storz and Hengge, 2010; Lim and Gross, 2014; Yan and Wu, 2019): First, the membrane-bound CssS receptor transduces the stress signal by phosphorylating CssR. Second, the phosphorylated CssR activates transcription of the two proteases, namely HrtA and HrtB, which degrade unfolded proteins and alleviate the stress condition. Furthermore, stress responses are intertwined with additional regulation in the core energy metabolism (Storz and Hengge, 2010), and such stress responses upregulate flagellar cell motility in order for a cell to physically escape the stress-causing location (Helmann et al., 1988; Márquez-Magaña et al., 1990; Yan and Wu, 2019). For instance, the level of cell motility is boosted by a low level of phosphorylated DegU (Kobayashi, 2007; Verhamme et al., 2007; Gupta and Rao, 2014), which is part of the core stress regulating DegU-DegS two-component system (Storz and Hengge, 2010; Laub, 2014). Nevertheless, these stress alleviating mechanisms can be opposed to the objective of achieving a high AMY yield: (i) the proteolytic degradation of AMY reduces yields, and (ii) a low phosphorylation level of DegU downregulates AMY expression (Gupta and Rao, 2014).

A state-of-the-art approach, which prevents the yield detrimental impact on the secretion of the stress response, is the over-expression of PrsA (Vitikainen et al., 2001; Quesada-Ganuza et al., 2019). Although the over-expression of PrsA reduces secretion stress by aiding AMY folding, it also has detrimental impacts such as hampered cell growth and even cell lysis (Vitikainen et al., 2001; Quesada-Ganuza et al., 2019). These detrimental phenotypes might be caused by

protein-protein interactions of specific PrsA protein domains with still unknown partner proteins (Quesada-Ganuza et al., 2019). Another unknown aspect of PrsA over-expression is its impact on the bacterial transcriptome, particularly during industrial fed-batch fermentation. The adaptation to glucose metabolism from maltose metabolism has a global impact on half of all transcriptional regulators even though both carbons are preferred by *B. subtilis* (Buescher et al., 2012). Thus, we would assume a substantially larger global impact on the transcriptome for the extreme secretion stress during PrsA over-expression (Quesada-Ganuza et al., 2019). We consider our assumption to be further supported by a large number of over a hundred proteins that require regulation to adapt bacterial motility (see above concerning stress) (Rajagopala et al., 2007). Furthermore, a pure protein-coding gene focus ignores the essential role regulatory small RNA (srRNA), RNA chaperones, and non-coding RNA (ncRNA) have in facilitating physiological changes impacting the entire cell during stress responses (Storz and Hengge, 2010). General stress regulatory mechanisms have been investigated in public datasets (Arrieta-Ortiz et al., 2015); however, metabolic and stress pathways undergo complex temporal adaptations (Hahne et al., 2010; Otto et al., 2010). Thus, both temporally resolved and condition-specific gene expression levels are needed to study stress pathways. Specifically for secretion stress during *B. subtilis* AMY fed-batch fermentation, no such dataset exists to our knowledge.

Here, we conducted fed-batch fermentation of two commercial *B. subtilis* strains. Both strains produce an AMY and are isogenic, except that one of them over-expresses PrsA. We studied the transcriptome during fermentation at six timepoints with RNA-seq and analyzed the expression levels of both known coding and non-coding annotations, and also of potential novel transcribed, yet unannotated regions. We complemented the differential expression analysis with a network analysis of known protein-protein interactions (PPI). This study found significant changes in gene expression levels between the studied strains for genes in the ATP biosynthesis and cell motility biological processes. Furthermore, the network analysis hints at mechanisms relating to competence transformation and cell motility that might be candidates for further tuning of AMY secretion yields.

Materials and Methods

Strains and fed-batch fermentation

The overall experimental setup is as previously described in Geissler et al. (2022). In summary, *B. subtilis* strain 168 $\Delta spoIIAC\Delta amyE\Delta apr\Delta nprE\Delta srfAC$ was maintained at 4°C on the LBGG medium. The AMY JE1 [sequence label *je1zyn* in Geissler et al. (2022)] was inserted by splicing by overlapping extension (SOE, inserted sequences are in

Supplementary Table 2) linear recombinant transformation, together with the commercial *sigA* promoter sequence P4199 and chloramphenicol marker in the *pel* locus. The PrsA over-expressing strain (referred to as the “+prsA” strain) had the insert by SOE of P4199, *prsA*, and spectinomycin marker in the *amyE* locus. A control strain without the *prsA* insert was included. After inoculation on SSB4 agar at 37°C, transfer on M-9 medium, sucrose 2M fed-batch fermentations were conducted in proprietary 2L tanks at 38°C. To avoid excessive overflow metabolite formation and to keep the culture in a sucrose metabolizing state, the fermentations were run as fed-batch fermentations without an initial batch phase. The feed medium consisted of sucrose (708 g/L), which was fed at a rate that increased linearly from 0.04 g/min at time = 0 h to 0.125 g/min at time = 8 h. The feed rate after 8 h of cultivation was kept constant at 0.125 g/min. Based on the dissolved oxygen tension data, the cultures entered a carbon-limited state after 9.4 h \pm 0.53 of fermentation. Fermentations were run in triplicates for 5 days. The selected replicate size allows detecting significant logFC in the expression of at least \pm 0.5 magnitude, as determined in benchmarks (Schurch et al., 2016). Samples were taken at six timepoints: 21, 26, 45, 71, 94, and 118 h after fermentation started. The samples were measured in cell density (OD650), and AMY activities were measured with an in-house assay. The assay (after 1/6,000 dilution) states the enzyme amount that breaks down 5.26 g starch per hour. This activity measure is proportional to the enzyme yield.

RNA-sequencing dataset

All samples were immediately mixed with 5 ml of 100% ethanol and stored on dry ice. The RNA extraction and purification method is the identical phenol-chloroform protocol of Geissler et al. (2022). RNA libraries and sequencing were conducted by BGI Hong Kong with DNBseq in single-ends of 50 bp length. RNA libraries were prepared with 3' adapter sequence AAGTCGGAGGCCAAGCGGTCTTAGGAAGACAA and the 5' adapter AAGTCGGATCGTAGCCATGTCGTTCTGTGAGC CAAGGAGTTG. The 36 samples (triplicates, two strains, and six timepoints) were sequenced in three batches with technical replicates for QC (**Supplementary Table 1**). The computational analyses were conducted in an adapted workflow of Geissler et al. (2022) (doi: 10.5281/zenodo.4534403), which provides a pipeline in a Snakemake framework nested in computational reproducible Anaconda environments (Koster and Rahmann, 2012). In concordance with the read quality assessment of FastQC (version 0.11.8) (Andrews, 2010), any adapter contaminations were removed with Trimmomatic (version 0.39) for up to two seed mismatches at a minimal 10 bp sequence overlap and 30 bp palindromic overlap (Bolger et al., 2014). In a sliding window of 4 bp, reads were clipped

for average Phred score quality below 20. From the 3' of reads, positions with quality below 3 were removed. Finally, a minimal length of 40 bp was required for filtered and cleaned reads. Reads were mapped against the respective +prsA and control genome sequence with Segemehl (version 0.3.4, default settings) (Hoffmann et al., 2009). The mapping and QC filtering statistics are given in **Supplementary Table 2**. Expression levels of coding and non-coding annotations (see below) in the respective strains were quantified for uniquely mapping reads with featureCounts (subread version 1.6.4, \geq 50% overlaps). Annotation coordinates in the respective strains were determined by liftOver (version 377) from the reference assembly (NC_000964.3) based on a pairwise alignment with LASTZ (version 1.0.4) (Harris, 2007; Liao et al., 2014; Haeussler et al., 2019).

Novel potentially transcribed regions

Reference annotations of coding, non-coding RNA (ncRNA), transcripts, untranslated regions (UTRs), and RNA structures were used from the BSGAtlas (version 1.1). The BSGAtlas uses separate annotation entries to specify which regions of an mRNA transcript are the coding, untranslated, or potential *cis*-regulatory RNA structure parts. Such a distinguishment to the UTR element has advantages since *cis*-regulatory RNA structures can overlap coding regions. The BSGAtlas unifies multiple databases and annotation resources, such that it also includes annotations for well-known ncRNA. Additional 141 putative ncRNA annotations from a tiling-array study were used (which are not part of the BSGAtlas) (Nicolas et al., 2012; Geissler et al., 2021). Relative to these reference annotations and all transcript and untranslated regions (UTRs) annotated in the BSGAtlas, we checked our RNA-seq data for transcription signals in 1,645 unannotated regions. The additional tiling-array annotations and un-annotated regions were determined with the R library plyranges (version 1.6.0) and GenomicRanges (version 1.38.0) combined with an overlap helper script from BSGAtlas' analysis code (doi: 10.5281/zenodo.4305872) in R (version 3.6.3) (R Development Core Team, 2008; Lawrence et al., 2013; Lee et al., 2018). Un-annotated regions shorter than 100 bp (the minimum length for >99% of the transcripts in the BSGAtlas) were excluded from any further expression analysis. The expression counts for all coding/non-coding sequences and *cis*-regulatory RNA structures were normalized with DESeq2's size-factor estimation (version 1.26.0) (Love et al., 2014). With respect to the downstream analysis of expression signals, we excluded the UTR annotations for improved interpretability, although we still retained all structured RNA *cis*-regulatory annotations. With the possible overlap between *cis*-regulatory RNAs and coding sequences, reads mapping within such overlaps can be counted twice during the quantification of expression. For a total of 542 unannotated regions, we observe

expression signals of normalized read counts relative to gap length of at least 4/50 bp (corresponds to four times average coverage) ([Supplementary Figure 1](#)). We chose not to narrow down the transcribed regions because we found that a read coverage-based approach (as suggested in the workflow used in the RNA-seq dataset, last section) resulted in fragmented results (see example in [Supplementary Figure 9](#)). These regions were assumed to be *novel potentially transcribed regions* (NPTRs) (see [Supplementary Table 3](#)); all other unannotated regions were excluded from the subsequent expression analysis. We used the open reading frame (ORF) predictions of prodigal (version 2.6.3, default settings) to check for potential not-yet annotated coding elements within NPTRs ([Hyatt et al., 2010](#)). We also verified the overall quality of the ORF predictions by checking for overlaps with all known coding gene annotations of the BSGAtlas. For each overlapping ORF-gene overlap (as detected by plyranges, see above), we computed the Jaccard similarity, which is the ratio of the length in the intersection of both annotations over their union.

Differential expression analysis

The expression levels of the coding/non-coding sequences, NPTRs, and *cis*-regulatory RNA structures were assessed for biological reproducibility in expression counts with scatter plots ([Supplementary Figure 2](#)). The scatter plots did not indicate visually striking patterns of batch effects according to the sequencing plan ([Supplementary Table 1](#)). The principal component analysis (PCA) inspection of the top 100 most variants that expressed annotations (without further differential expression analysis) confirmed the relevance of the experimental design in the latent structure of the expression data with the principal components corresponding to the strains and time aspect ([Supplementary Figure 3](#)). Differential expressions for pairwise comparisons between the strains at each of the six time points and within each strain along the time axis ([Figure 1C](#)) were assessed with the DESeq2's Wald test. Similar to the analysis presented in [Geissler et al. \(2022\)](#), the pairwise tests were weighted in a stage-wise procedure to guarantee an overall false discovery rate relative to the number of annotations: each annotation was screened for dynamic expression with a log-ratio test against a static expression model before confirming which of the pairwise tests had changes in expression. The screening and pairwise tests included a linear factor in the regression models to account for potential batch effects. The stage-wise weighting was conducted with stageR (version 1.8.0) ([Van den Berge et al., 2017](#)) and differential expression was called for adjusted *p*-values < 0.01. Overall, 2,127 annotations were detected as differentially expressed ([Table 1](#) and [Supplementary Table 4](#)). Based on the z-scaled log expected mean expression levels ([Supplementary Table 5](#)), expression profiles were grouped in 10 k-means clusters

(R implementation). The profiles per strain were clustered separately (one gene = two rows in the data matrix). The number of clusters was determined by the “elbow method” over the total within-cluster error curve ([Supplementary Figure 4; Thorndike, 1953](#)).

Regulated biological processes

We investigated the set of differentially expressed genes and their upregulation and downregulation for over-representation in biological processes as annotated in Gene Ontology (GO) terms, which are readily available for 78.3% of coding genes ([Caspi et al., 2014; Geissler et al., 2021](#)). For each pairwise differential expression test ([Figure 1C](#)), we inspected the set of upregulated genes (those with a positive logFC) and downregulated genes separately. The over-representation analysis was performed with topGO (version 2.37.0) ([Alexa and Rahnenführer, 2022](#)). Over-representation for the respective upregulated and downregulated genes was determined with a fisher test for the significance level of 0.01 relative to the background of all expressed genes, which were determined by the DESeq2's independent filtering procedure. This procedure discards the on average lowly expressed genes in order to maximize the number of differentially expressed genes (indicated by NA for *p*-values in [Supplementary Table 4; Love et al., 2014](#)). The minimal term size was set to 10, and the dependencies due to GO's hierarchy were de-correlated with topGO's “elim” algorithm. After filtering for a minimal observed/expected ratio of magnitude 2 (between the 80 and 85th percentile), *p*-values were adjusted for multiple testing with a false discovery rate (FDR). The over-represented processes and the associated differentially expressed genes are listed in [Supplementary Tables 6, 7](#) and [Supplementary Figure 6](#).

Protein–protein interaction network analysis

The protein–protein interaction network analysis was conducted in Cytoscape (version 3.8.2) ([Shannon, 2003](#)) for the differentially expressed protein-coding genes (both with and without significant logFC between strains). High-confidence protein associations (confidence score >0.8) were retrieved from the STRING v11 database using stringApp (version 1.6.0) for the *B. subtilis* strain 168 ([Doncheva et al., 2019; Szklarczyk et al., 2019](#)). The resulting network was clustered with the MCL algorithm (inflation value of 2.5, confidence scores as edge weights) implemented in the clusterMaker2 app (version 1.3.1) ([Enright, 2002; Morris et al., 2011](#)). The visualization of significant between strain logFCs on the network nodes was added with Omics Visualizer (version 1.3.0) ([Legeay et al., 2020](#)).

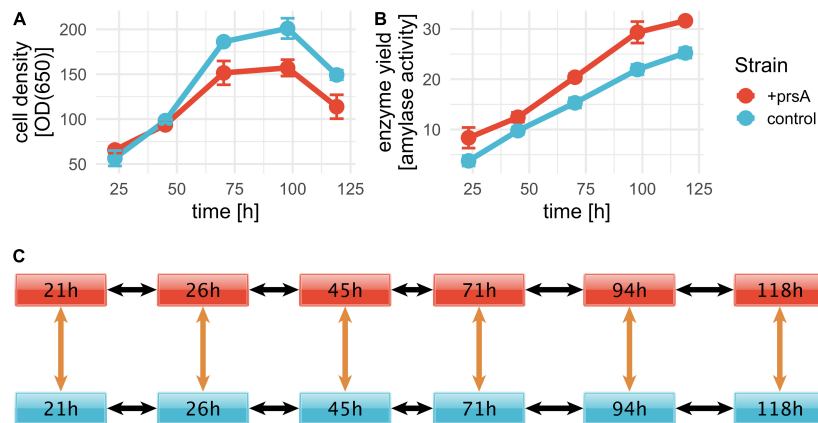


FIGURE 1

AMY fed-batch fermentations. Fed-batch fermentation was conducted in triplicates for a control strain (blue) and +prsA (red). RNA-seq samples were prepared at six timepoints: 21, 26, 45, 71, 94, and 118 h after fermentation start. Cell density and enzyme yield were measured for five timepoints: 23.2, 45, 70.2, 97.8, and 119 h. (A) The average cell density per strain over fermentation time was measured in optical density (OD) at 650 nm. The error bars indicate the standard deviation. (B) With a progressing fermentation, the yield increases. The shown yield is measured in enzyme activity (see section "Strains and fed-batch fermentation"). (C) For the differential expression, we investigated the significance of differential expression between the samples at six pairwise comparisons (orange arrows) and changes in expression over time in either strain for each pair (black arrows).

Global amino acid composition

In order to interpret the regulated biological processes (see above), we inspected the global amino acid compositions of all *B. subtilis* protein-coding genes. The nucleotide sequences of all coding sequences from the BSGAtlas were extracted with BSgenome (version 1.54.0) (Pagès, 2021). The corresponding amino acid sequences were determined according to the bacterial genetic code with Biostrings (version 2.54.0) (Pagès et al., 2019). Here, we used only the 99.3% of the coding genes that were completely relative to their corresponding amino acid sequences; that is, they used all codons encoded in their nucleotide sequences, correctly started with methionine, and ended with a stop codon. The composition in average proportion was determined for these complete sequences (Table 2).

Results

Novel potentially transcribed regions

Transcriptome analysis from RNA-seq data

To elucidate potential mechanisms of *B. subtilis* secretion stress during the production of the AMY enzyme JE1 (commercial name Natalase™) with a particular focus on PrsA over-expression, we conducted fed-batch fermentation in triplicates for two isogenic strain conditions: one control strain and one strain with PrsA over-expression (from here on referred to as +prsA). As expected from the reduced growth upon PrsA

over-expression (Vitikainen et al., 2001; Quesada-Ganuza et al., 2019), the +prsA strain has a lower cell density (Figure 1A) and higher AMY yield (Figure 1B). To capture the transcriptome dynamics during fermentation, we took out samples for RNA-seq analysis at six timepoints: 21, 26, 45, 71, 94, and 118 h after fermentation started. These timepoints correspond to sampling every 24 h (within a 3 h window) with one additional sample at the early phase of the fermentation.

Transcriptional activity for the reference annotations

In order to comprehensively investigate both the coding and non-coding RNA elements, we quantified the RNA-seq expression according to a recently developed transcript atlas for *B. subtilis* (Geissler et al., 2021). We included 141 additional annotations from a tiling-array study that was not included in the atlas due to unclear mechanism of transcription (annotations were ambiguous as to whether they are independent full RNA transcripts or only part thereof) (Nicolas et al., 2012; Geissler et al., 2021). In the following, we refer to these annotations, together with the less well-characterized RNA elements from the atlas, as putative ncRNA. These reference annotations combine gold standard curated information, computational RNA structure biology, and transcriptomic analysis of over 100 experimental conditions (Nicolas et al., 2012; Geissler et al., 2021). Additionally, these experimental conditions suggest that still 5% of remaining unannotated regions have evidence of expression activity (Geissler et al., 2021). Fed-batch fermentations were not part of the above-mentioned experimental conditions, such that there might be a larger potential to discover fed-batch-related regions

TABLE 1 Differentially expressed annotations.

Annotations	No. Annotations considered for analysis	Differentially expressed	Strain-specific expression
CDS	2,674	1,791 (67.0%)	1,026 (57.3%)
NPTRs	355	234 (65.9%)	123 (52.6%)
Putative ncRNA	107	68 (63.6%)	38 (55.9%)
Riboswitch	37	20 (54.1%)	10 (50.0%)
tRNA	22	9 (40.9%)	7 (77.8%)
sRNA	9	3 (33.3%)	2 (66.7%)
Synthetic PrsA	1	1 (100.0%)	1 (100.0%)
Synthetic AMY	1	1 (100.0%)	0
asRNA	1	0	0
SRP	1	0	0

For the differential expression analysis, multiple coding and non-coding annotations were considered (first column). The number of genes with minimal expression levels as determined by DESeq2's independent filtering, which were inspected for potential differential expression, is in the second column. The number of detected differentially expressed annotations in any of the pairwise comparisons (Figure 1C) is in the third column. The last column lists the number of annotations detected to have a significant difference in expression between the strains. The percentages provided in parenthesis are relative to the columns to the left (Only 355 of the 542 NPTRs passed the independent filtering).

TABLE 2 Amino acid composition.

Amino acid	All coding genes	Diff. expressed	Highly expressed	AMY	PrsA
Tryptophan	1.03% ± 0.99	1.06% ± 0.96	0.86% ± 0.81	4.12% (+3.1 SD)	0.35% (−0.7 SD)
Asparagine	4.07% ± 2.05	3.87% ± 1.75	3.86% ± 1.44	8.82% (+2.3 SD)	2.82% (−0.6 SD)
Histidine	2.30% ± 1.51	2.27% ± 1.33	2.12% ± 1.10	4.31% (+1.3 SD)	1.06% (−0.8 SD)
Tyrosine	3.57% ± 1.95	3.39% ± 1.56	3.13% ± 1.32	5.49% (+1.0 D)	3.17% (−0.2 SD)
Glycine	6.67% ± 2.74	6.95% ± 2.41	7.62% ± 2.08	8.82% (+0.8 SD)	6.34% (−0.1 SD)
Aspartic acid	5.11% ± 2.29	5.06% ± 2.10	5.17% ± 1.83	6.86% (+0.8 SD)	11.27% (+2.7 SD)
Threonine	5.32% ± 1.87	5.44% ± 1.75	5.51% ± 1.35	5.88% (+0.3 SD)	4.23% (−0.6 SD)
Arginine	4.19% ± 2.14	4.07% ± 1.91	4.24% ± 1.92	4.31% (+0.1 SD)	0.70% (−1.6 SD)
Proline	3.48% ± 1.71	3.55% ± 1.49	3.93% ± 1.24	3.53% (+0.0 SD)	0.35% (−1.8 SD)
Glutamine	3.87% ± 2.04	3.85% ± 1.82	3.65% ± 1.42	3.73% (−0.1 SD)	6.34% (+1.2 SD)
Alanine	7.36% ± 2.84	7.82% ± 2.65	8.25% ± 2.29	7.06% (−0.1 SD)	7.39% (+0.0 SD)
Phenylalanine	4.63% ± 2.40	4.53% ± 2.20	4.21% ± 1.94	4.31% (−0.1 SD)	1.76% (−1.2 SD)
Valine	6.75% ± 2.37	6.92% ± 2.11	7.38% ± 1.91	6.27% (−0.2 SD)	6.69% (−0.0 SD)
Methionine	2.46% ± 1.43	2.52% ± 1.31	2.50% ± 1.11	2.16% (−0.2 SD)	1.76% (−0.5 SD)
Serine	6.22% ± 2.26	6.28% ± 2.16	6.04% ± 2.04	4.51% (−0.8 SD)	4.93% (−0.6 SD)
Cysteine	0.91% ± 1.13	0.84% ± 0.91	0.70% ± 0.72	0.00% (−0.8 SD)	0.35% (−0.5 SD)
Lysine	7.50% ± 3.14	7.13% ± 2.78	6.79% ± 2.17	4.90% (−0.8 SD)	17.96% (+3.3 SD)
Glutamic acid	7.37% ± 3.29	7.20% ± 3.13	7.23% ± 2.77	4.31% (−0.9 SD)	9.15% (+0.5 SD)
Leucine	9.70% ± 3.06	9.72% ± 2.86	9.42% ± 2.35	6.67% (−1.0 SD)	8.45% (−0.4 SD)
Isoleucine	7.50% ± 2.65	7.53% ± 2.48	7.37% ± 2.15	3.92% (−1.4 SD)	4.93% (−1.0 SD)

The average amino acid compositions (rows) are shown for all *B. subtilis* endogenous (thus excl. AMY and PrsA) coding genes (second column), those that were detected as differentially expressed (third column), and the by average expression 10% most highly expressed genes (fourth column). The standard deviations are shown behind the “±” signs. The compositions of amino acids for the AMY enzyme (fifth) and the over-expressed PrsA (sixth) column are shown. The difference in standard deviations relative to the average for all genes are indicated in parenthesis. The bold font highlights amino acids with difference of more than two standard deviations.

from our RNA-seq data. Consequently, we investigated our RNA-seq data for expression in such unannotated regions.

Novel potentially transcribed regions

There are a total of 1,645 un-annotated contiguous stretches of the genome or gaps (stranded, meaning there can be antisense located annotations) between reference annotations of length

>100 bp (minimal length for 99.5% of transcripts in the atlas). We detect novel potentially transcribed regions (NPTRs) by inspecting the average RNA-seq read coverages over the entire unannotated gap region (read counts, DESeq2 size-factor normalized, relative to the lengths). Relative to the 50 bp sequencing lengths (see section “RNA-sequencing dataset”), 70% of atlas annotations were on average expressed by four reads

and 30% by one read. In contrast, only 20% (542) of unannotated regions were on average covered by four reads. This high coverage for these 542 NPTRs ([Supplementary Figure 1](#)) indicates that the NPTRs may have functional importance and that it would be relevant to include these in subsequent expression analysis (see [Supplementary Table 3](#)).

PrsA over-expression changes gene expression regulation of the global transcriptome

Differential expression

We assessed the impact of PrsA over-expression on the bacterial transcriptome by analyzing the expression levels of coding and non-coding sequences (see section “Transcriptional activity for the reference annotations” above), including the 141 additional annotations and the 542 NPTRs with DESeq2. For each region, we performed 16 pairwise differential expression tests: six tests between the two strains on each timepoint and 2×5 tests from one timepoint to the next in both strains ([Figure 1C](#)). Since each pairwise test corresponds to a separate hypothesis test, we used stage-wise testing to adjust for the overall false discovery rate (FDR) per annotation ([Love et al., 2014](#); [Van den Berge et al., 2017](#)). Compared to controlling the FDR per hypothesis, the overall FDR increases statistical power and guarantees the FDR relative to the gene/annotation number, independent of the number of hypotheses ([Van den Berge et al., 2017](#)). As part of the differential expression analysis, DESeq2’s independent filtering detected about half of all coding sequences and 355 of 542 NPTRs as expressed ([Love et al., 2014](#)). At an overall FDR p -adj. ≤ 0.01 , we detected differential expression for 1,793 coding sequences (67% of expressed genes), 234 NPTRs (66%), 68 putative ncRNAs (64%), 20 riboswitches (54%), 9 tRNAs (41%), and 3 sRNAs (33%) ([Table 1](#) and [Supplementary Table 4](#)). The differentially expressed coding genes include the AMY enzyme and the over-expressed PrsA. Between 50 and 78% of these biotypes had strain-specific expression patterns (significant difference for at least one of the six between strain tests). PrsA had strain-specific expression (as expected by not being inserted into the control strain’s genome). Notably, no strain-specific expression was detected for AMY.

To further assess the coding potential of the differentially expressed NPTRs, we leveraged a set of 4,226 ORF predictions ([Hyatt et al., 2010](#)). These ORF recall 4,085 of 4,332 known coding sequences perfectly (overlap with Jaccard similarity $> 90\%$), which corresponds to a recall of 94.3% with a precision of 96.7%. Only 141 ORF predictions do not recall coding genes. Furthermore, only 18 overlap unannotated regions (> 100 bp); for 3 ORF, the overlap is less than 5% of the ORF length ([Table 3](#)). Also, only two of the ORFs are fully located within an NPTR that has detected differential expression; in both cases, the ORFs antisense overlaps the 3’ ends of the coding genes:

The electron transfer flavoprotein *etfA* and the gene of unknown function *yobB*.

The regions with the highest expression changes

The strain-specific expression patterns of PrsA and the respective logFC between the two strains on all six timepoints were the most extreme observed in this study with logFC values up to a factor of 20 at each timepoint. Other extreme logFC values were observed for genes from operons encoding a variety of biological functions ([Table 4](#)). The NAD biosynthesis genes of the *nadABC* operon ([Rodionov et al., 2008](#)) also have extreme logFC, but they undergo both extreme upregulation and downregulation in the control strain with *nadA* and *nadB* being downregulated from timepoint 21 to 26 h (both logFCs < -6 , adj. $p < 0.004$) and subsequently upregulated from 26 to 45 h (both logFCs $\sim +7$, adj. $p < 3e-10$). Due to the secretion stress, the production strains attempt to sporulate despite being unable to do so ([Geissler et al., 2022](#)). Consistently, the two sporulation genes, namely *safA* and *coxA*, were among the most extremely regulated (logFC > 6 , adj. $p < 2.3e-5$). Other extreme changes in expression (logFC < -5) were observed for the spore killing factors *skfA* and *skfB* ([González-Pastor, 2011](#)), the sporulation controlling factor *spoIIIGA* ([Ramos-Silva et al., 2019](#)), the bacitracin resistance genes *bceA* and *bceB* ([Ohki et al., 2003](#)), the for NADH during fermentation essential lactate dehydrogenase *ldh* ([Cruz Ramos et al., 2000](#); [Larsson et al., 2005](#)), and an NPTR antisense to the gene of unknown function *yta* ([Asai et al., 2007](#)).

Biological processes and differentially expressed genes are mutually associated

The investigation of the overall expression profiles from a k-means clustering ([Figure 2A](#)) on the average expected expression at each timepoint ([Supplementary Table 5](#)) shows marked differences in the expression dynamics between the strains ([Figure 2C](#)). Also, all profiles indicate a substantial shift in dynamics between timepoints 45–71 h, during which the cell population increased the most ([Figure 1A](#)): For instance, profiles 4 and 5 drop in expression levels at that timepoint but recover and even exceed the starting expression level whereas profiles 7 and 8 have drastically downregulated expression at that timepoint and do not recover ([Figure 2B](#)). Genes and other biotypes with strain-specific expression patterns had predominately different expression profiles between the strains, whereas those without strain-specific expression had the same ([Supplementary Figure 5](#)). Therefore, *B. subtilis* regulates gene expression both timepoint- and strain-specifically.

We assessed which biological processes [annotated in Gene Ontology (GO), terms ([Ashburner et al., 2000](#))] are over-represented among the differentially expressed genes in each time and strain pairwise comparison ([Figure 1C](#)). We compared the numbers of respective upregulated or

TABLE 3 ORF overlapping un-annotated regions.

Overlapping gap	ID	Gap length	ORF start	ORF end	ORF strand	ORF length	% ORF within gap
Low expression	gap-865	18994	164501	164923	–	18994	100
Low expression	gap-964	382	635149	635262	–	382	100
Low expression	gap-1109	277	1233396	1233599	–	277	100
Low expression	gap-1244	15761	1784706	1784837	–	15761	100
Low expression	gap-1402	1357	2747665	2747853	–	1357	100
Low expression	gap-643	8279	3341033	3341254	+	8279	100
NPTR [†]	gap-1322	7336	2221838	2221951	–	7336	100
NPTR with diff. expr.	gap-1293	743	2050930	2051037	–	743	100
NPTR with diff. expr.	gap-538	7687	2915229	2915390	+	7687	100
Low expression	gap-1495	1272	3419414	3419656	–	1272	96.7
Low expression	gap-1367	140	2560095	2560259	–	140	84.8
NPTR [†]	gap-113	1795	718436	718597	+	1795	83.3
Low expression	gap-1166	197	1453322	1453525	–	197	79.4
NPTR with diff. expr.	gap-1354	1038	2460278	2460625	–	1038	26.7
NPTR without diff. expr.	gap-1296	581	2056650	2057114	–	581	23.2
NPTR with diff. expr.	gap-1058	2975	1033458	1034003	–	2975	4.8
Low expression	gap-449	2395	2555468	2555845	+	2395	2.1
Low expression	gap-90	2286	608879	609391	+	2286	1.4

Shown are each overlap between ORFs (that did not recall a coding gene) and un-annotated regions (> 100 bp). Some of these un-annotated regions with sufficient expression signal were considered as NPTR and further processed for differential expression analysis (column 1–3). Bold texts highlights gaps with detected differential expression. Based on the coordinates of the predicted ORF (column 4–7), the overlap of the ORF with the un-annotated gap is expressed relative to the length of the ORF (column 8).

[†]Un-annotated region with sufficient expression signal to be considered as NPTR, but which was excluded from further analysis by the independent filtering procedure of the differential expression analysis.

downregulated genes relative to the number of expressed genes (see section “Materials and methods”). A total of 24 processes had significant over-representation (Fisher’s exact test, FDR p -adj. ≤ 0.01). We inspected the list of differentially expressed genes per process (Supplementary Table 7) in combination with meta-information available in the BSGAtlas, particularly KEGG pathway annotations (Kanehisa and Goto, 2000; Geissler et al., 2021). Notably, the detected over-represented processes annotate genes with differentially expressed logFC predominately above the background logFC distribution of genes without detected differential expression (Supplementary Figure 8). Furthermore, some of the top 10 most extremely upregulated and downregulated genes (Table 4) were annotated by the detected processes (Supplementary Table 7), namely cell wall macromolecule catabolic process (*safA* and *skfA*), response to stress (*nadC* and *nadE*), and ATP biosynthetic process (*ldh*). We further inspected the detected biological processes (Figure 3) for their relevance with respect to fed-batch fermentation, as described in the sections later.

Nucleotide biosynthesis

It is well established that an ample supply of nucleotides is needed for efficient AMY protein expression (Hosoda et al., 1959) and thus also the nucleotide precursors, such as UMP and IMP, are of regulatory interest (Peifer et al., 2012; Hohmann et al., 2016). Consistently, the over-representation investigation indicates an upregulation of UMP (GO:0006222)

and IMP (GO:0006189) biosynthesis in the +*prsA* strain from timepoint 26 to 45 h and 95 to 118 h, respectively. The monosaccharide catabolic genes (GO:0046365), especially the genes involved in the ribose synthesis *via* the pentose phosphate pathway (Supplementary Table 7), are upregulated in the control strain from timepoint 45 to 71 h. The pteridine-containing compound metabolic process (GO:0042558) was over-represented by genes upregulated from the first to the second timepoint in both strains. These specific genes are also part of the folate biosynthesis pathway, which is essential for both purine and pyrimidine synthesis (Kilstrup et al., 2005) and therefore quintessential for AMY production (Hohmann et al., 2016; Hosseini et al., 2018).

PrsA over-expression affects genes involved in energy metabolism

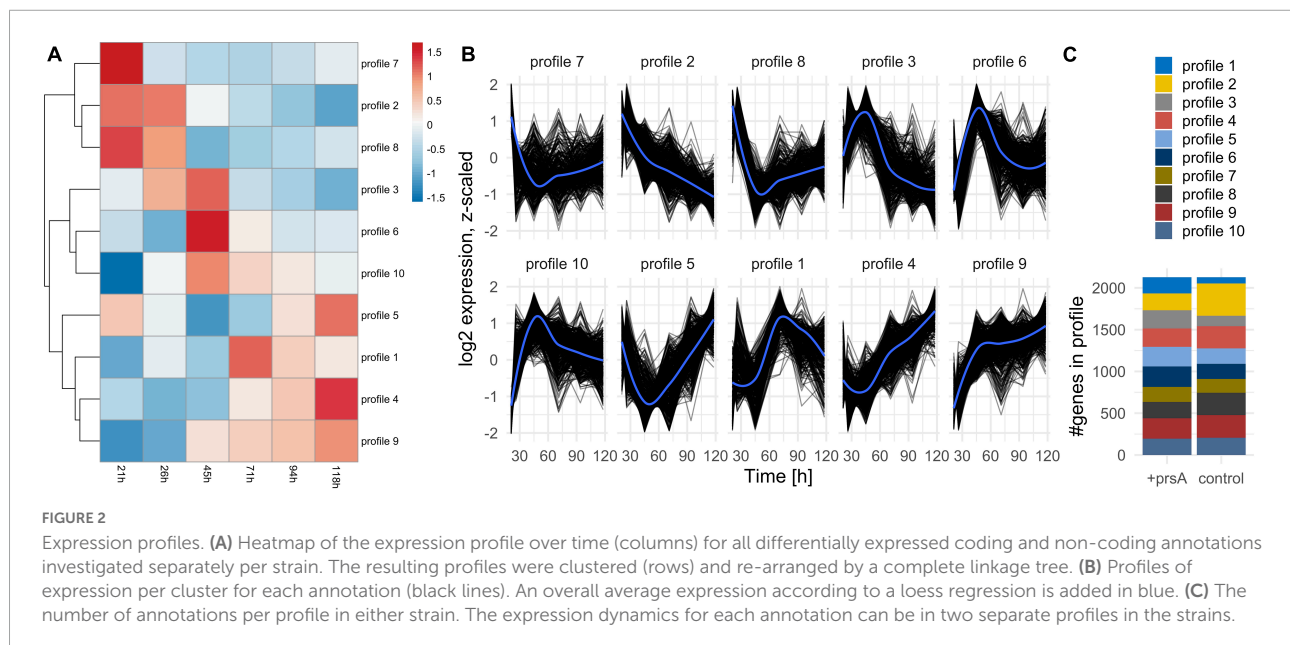
ATP biosynthesis

The ATP biosynthetic process (GO:0006754) was significantly downregulated in +*prsA* compared to the control strain on the first timepoint of the fermentation. Furthermore, the data suggest that the energy derivation by oxidation of organic compounds (GO:0015980) was further downregulated in +*prsA* from the first to the second timepoint within the first day of fermentation. The differentially expressed genes associated with both processes comprise a long list

TABLE 4 Most extreme observed logFCs.

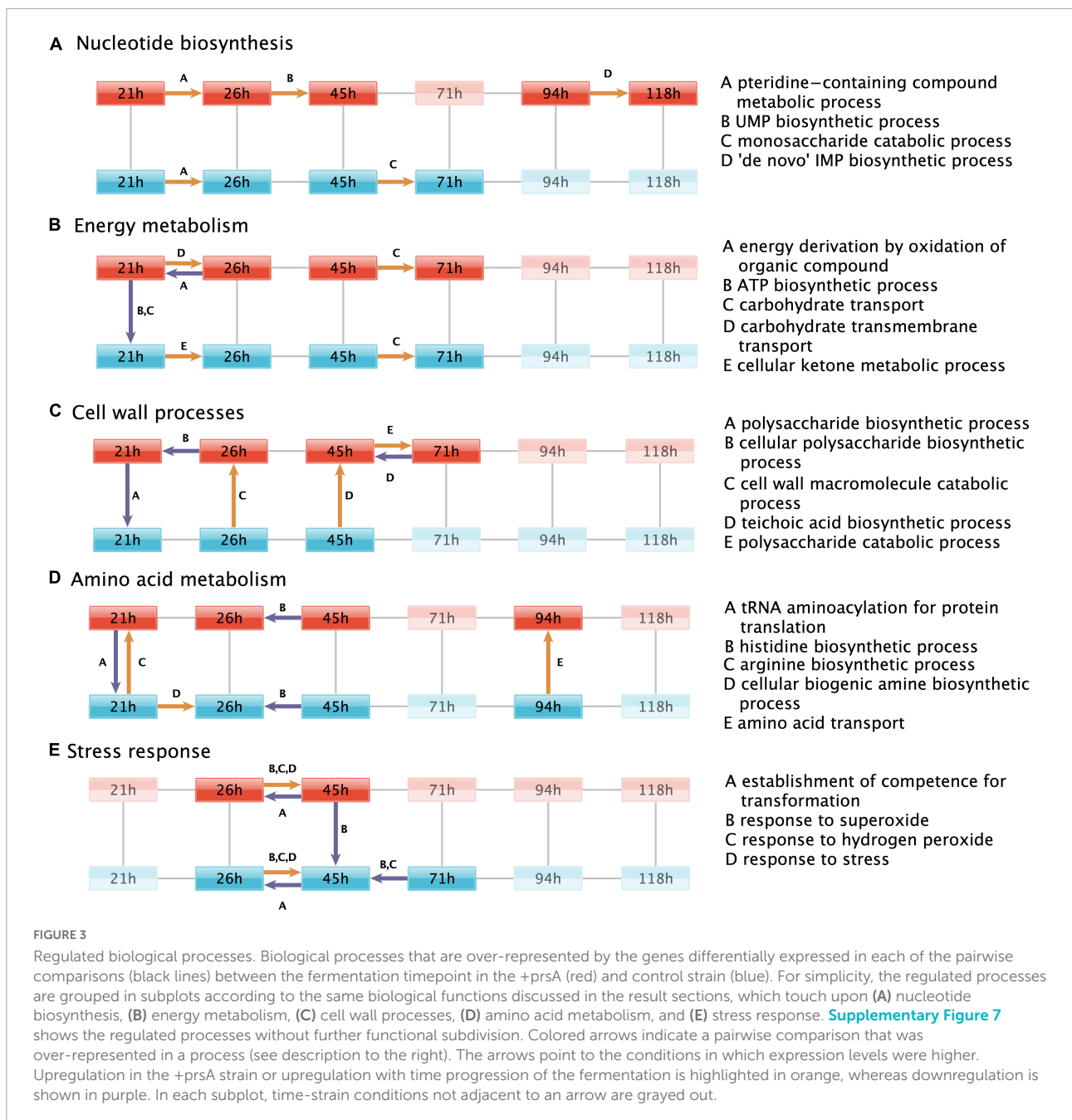
Name	Type	Test	logFC	Adj. P	Location
nadB	Coding	Between strains, 26 h	7.4	1.07E-10	2847871 <- 2849466
nadB	Coding	Control, 26->45 h	7.2	2.65E-10	2847871 <- 2849466
nadC	Coding	Control, 26->45 h	7.0	2.89E-09	2847048 <- 2847917
nadC	Coding	Between strains, 26 h	6.9	5.26E-09	2847048 <- 2847917
nadA	Coding	Control, 26->45 h	6.8	7.82E-11	2845955 <- 2847061
nadA	Coding	Between strains, 26 h	6.8	1.33E-10	2845955 <- 2847061
safA	Coding	Control, 26->45 h	6.4	4.05E-09	2844675 <- 2845838
safA	Coding	Between strains, 26 h	6.2	8.98E-09	2844675 <- 2845838
coxA	Coding	Between strains, 26 h	6.2	1.82E-05	2843931 <- 2844527
coxA	Coding	Control, 26->45 h	6.1	2.28E-05	2843931 <- 2844527
spoIIIGA	Coding	Control, 26->45 h	-5.2	8.47E-67	1603779->1604708
skfB	Coding	Control, 26->45 h	-5.4	2.86E-60	214175->215407
skfA	Coding	+prsA, 26->45 h	-5.4	7.31E-09	213941->214108
skfA	Coding	Control, 26->45 h	-5.7	4.39E-24	213941->214108
bceA	Coding	Control, 26->45 h	-6.2	5.76E-107	3111327 <- 3112088
nadA	Coding	Control, 21->26 h	-6.2	0.002258927	2845955 <- 2847061
bceB	Coding	Control, 26->45 h	-6.5	1.62E-127	3109397 <- 3111337
nadB	Coding	Control, 21->26 h	-6.5	0.003566971	2847871 <- 2849466
ldh	Coding	+prsA, 21->26 h	-6.6	2.42E-13	329774->330739
gap-1449	NPTR	control, 26->45 h	-6.9	1.85E-48	3108525 <- 3109352

The table lists the top 10 most extreme upregulated and downregulated genomic elements according to their logFC of differential expression (fourth column). prsA is excluded since it was upregulated with an approximate logFC of 20 between strains at all-time points. For each genomic element, the locations (last column) are relative to the reference genome (see section "Materials and methods"). The pairwise tests (third column) refer to the conducted differential expression analysis (Figure 1C), and the corresponding adjusted *P*-values are listed in the fifth column.



(> 50, see [Supplementary Table 7](#)) of core energy metabolic enzymes from the citrate cycle, oxidative phosphorylation, and glycolysis. Nevertheless, the list also overlaps with the starch and sucrose metabolism pathway, particularly with glycogen biosynthesis (*glgA*, *glgB*, *glgC*, *glgD*, and *glgP*) (Kiel et al., 1994).

Consistent with these observations, the carbohydrate transport (GO:0008643) was also downregulated in +prsA on the first timepoint. In contrast, the cellular ketone metabolic process (GO:0042180) was upregulated in the control strain from the first to the second timepoint. Ketones are essential for the



biosynthesis of menaquinone (Lu et al., 2008). Menaquinone is *B. subtilis*' respiration coenzyme, similar in function to ubiquinone in human mitochondria (Lemma et al., 1990). Nevertheless, the ATP biosynthetic process (GO:0006754) was not detected significantly over-represented by the regulated genes at the other fermentation timepoints.

Altering carbohydrate transport during fermentation

The over-representation analysis also suggests that both strains have an upregulated carbohydrate transport

(GO:0008643, GO:0034219) from 45 to 71 h. The transport might also be upregulated in the +prsA strain from the first to the second timepoint.

PrsA over-expression affects genes involved in cell wall destabilizing processes

Low PrsA protein abundances and increased concentrations of teichoic acid can reduce cell growth and cell wall disruption

(Driessen et al., 1998; Hyryläinen et al., 2000). For instance, the inhibition of the *dlt* operon—which is key to teichoic acid synthesis—increases AMY yields (Hyryläinen et al., 2000; Yan and Wu, 2017). However, our data suggest that not only *dltB* expression is upregulated in +*prsA* on timepoint 45 h (logFC = 0.86, adj. $p < 2.11e-5$) but also the entire teichoic acid biosynthetic process (GO:0019350). Additional processes relating to cell wall molecules and polysaccharide biosynthetic (GO:0033692, GO:0000271) were observed as downregulated in +*prsA*. Nevertheless, not only does our data suggest that the biosynthesis is downregulated, the corresponding catabolic processes (GO:0016998, GO:0000272) might be upregulated.

Upregulation of amino acid metabolism during PrsA over-expression

Regulated amino acid metabolism

Genes of the arginine biosynthetic process (GO:0006526) are over-represented among the genes upregulated in the +*prsA* strain on the first timepoint and for the amino acid transport (GO:0006865) at timepoint 94 h after fermentation started. The histidine biosynthetic process (GO:0000105) was detected as downregulated from timepoint 26 h to the timepoint 45 h in both strains. The data suggest also that the tRNA aminoacylation for protein translation (GO:0006418) is downregulated in +*prsA* on the first timepoint, and that the cellular biogenic amine biosynthetic process (GO:0042401) is upregulated in the control strain from the first to the second timepoint.

Expected changes in amino acid metabolism

Given the observed potential regulation in amino acid metabolism above, we investigated to which extent these might be the result of the peptide sequence of the secreted AMY. The inspection of the codon composition of all coding genes suggests that the AMY and the over-expressed PrsA contain substantially more tryptophan, asparagine, aspartic acid, and lysine (more than 2 standard deviations from the average proportion, Table 2). Tryptophan was the strongest over-represented amino acid in AMY (+3.1 standard deviations). In comparison, the subset of neither differentially expressed endogenous (excl. AMY and PrsA) coding genes nor 10% of most highly expressed endogenous genes have changes in the overall composition (within 1 standard deviation). The PrsA was extremely over-expressed in the +*prsA* strain (logFC > 19, adj. $p \leq 5.27e-40$). By average expression, AMY was the 5th and PrsA the 34th highest expressed gene (see Supplementary Table 6). Thus, the enrichment of these four amino acids in AMY and PrsA should have physiological relevance: given the high energetic cost of tryptophan biosynthesis (Akashi and Gojobori, 2002), the evolutionarily adapted amino acid metabolism will be affected (Smith and Chapman, 2010).

Protein–protein interactions of stress response and competence transformation

Stress response turning point

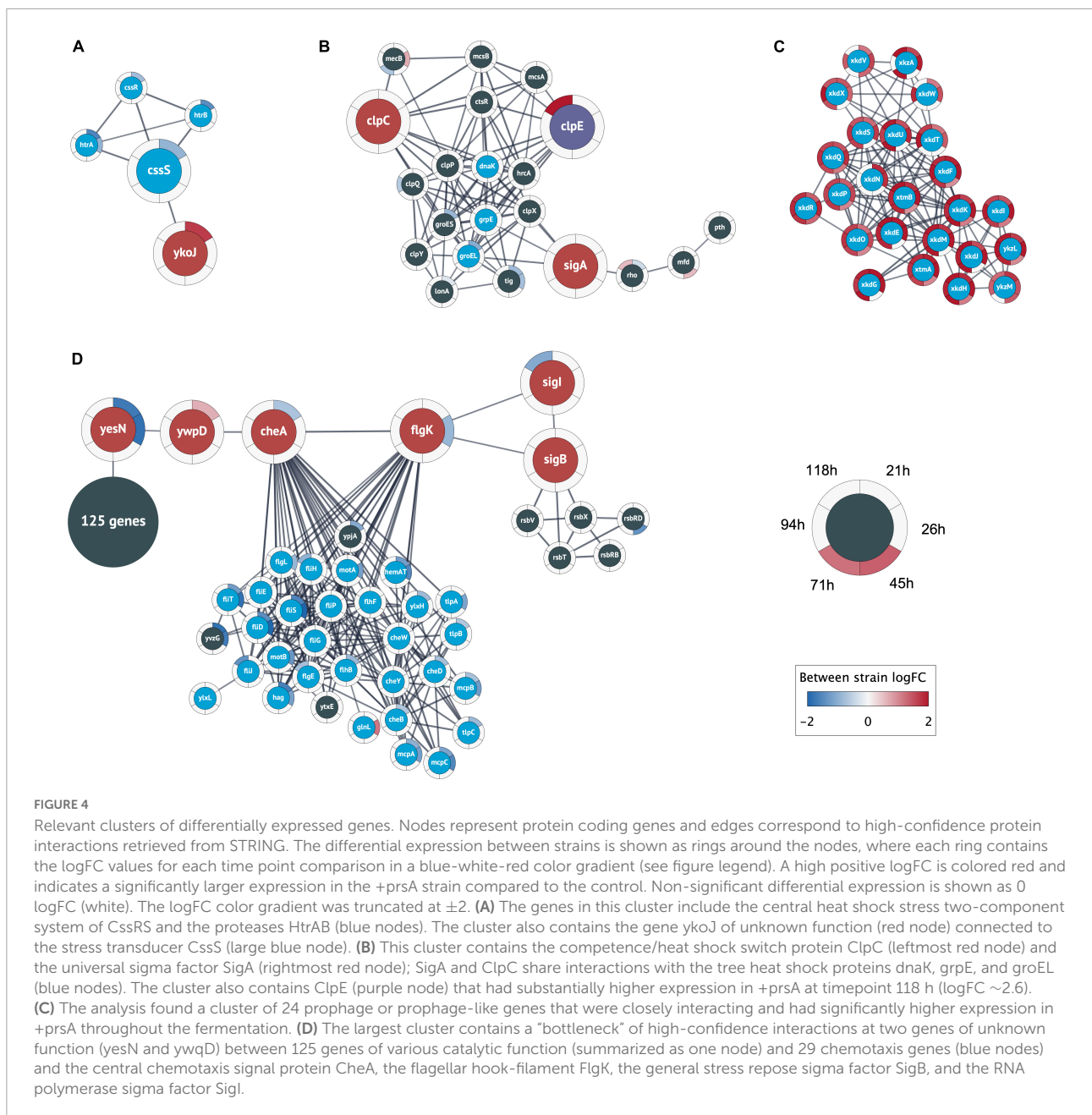
The over-representation investigation reveals that both strains upregulate parts of their stress response concerning the reactive oxygen species (ROS) response (GO:0006950 and the two children terms GO:0042542, GO:0000303) from timepoint 26 h to 45 h. Simultaneously, the strains downregulate the establishment of competence for transformation (GO:0030420). The protein ClpC is the key switch between heat shock (including secretion stress) and competence regulation (Turgay et al., 1997). During stress, a three-protein complex of ClpC, MecA, and ComK is formed (Turgay et al., 1997). The bound central competence regulator ComK can no longer act as a transcription regulator, which prevents the establishment of competence (Turgay et al., 1997). According to our results, *clpC* undergoes significant differential expression during fermentation in both strains, but neither *comK* nor *mecA* had significant expression changes though both were expressed (Supplementary Table 4). Given that the molecular mechanism of the ClpC switch (i) is post-translational, (ii) does not directly impact the transcription levels of the involved genes, and (iii) involves a third factor, the analysis by pairwise comparison of expression levels cannot detect that specific interaction. Therefore, we complemented the expression analysis with protein-protein interaction (PPI) network analysis.

Protein–protein interaction network analysis

We retrieved PPIs from the STRING database for the *B. subtilis* strain 168. STRING provides a list of functional associations from multiple evidence channels, such as curated knowledge from known metabolic pathways and protein complexes, physical PPIs from lab experiments (e.g., pull-down assays), predicted interactions from text mining of the biomedical literature, or associations based on co-expression analysis (Szklarczyk et al., 2019). The resulting network of 4,774 high-confidence associations (confidence score >0.8) among 1,770 of the 1,791 differentially expressed protein-coding sequences was clustered into 201 protein clusters using MCL (Enright, 2002; Morris et al., 2011; Doncheva et al., 2019). In combination with the significant logFCs between the +*prsA* and control strains (Legeay et al., 2020), we manually inspected four clusters with interesting patterns regarding this study's outset (Figure 4). These are described in the following sections below.

Two-component system

The first PPI cluster consists of the C_{ss}RS two-component system, including the involved proteases (see Introduction, Figure 4A). However, the cluster contains an additional association between the stress signal transducer C_{ss}S and YkoJ of unknown function. The *ykoJ* expression during secretion of



a vaccine compound (beta-toxoid) positively depends on CssS (Nijland et al., 2007). In contrast, the expression during AMY might have a negative dependency with *cssS* being significantly lower expressed in +prsA on timepoint 21 h after fermentation start (logFC = -0.9 , adjusted $p = 8.5e-10$) and *ykoJ* significantly higher (logFC = 1.7 , adj. $p = 1.2e-7$). To our knowledge, the association YkoJ-CssS has not been characterized in the context of AMY production.

Competence switch

The second cluster (Figure 4B) contains the above-described heat shock/competence protein switch ClpC

(Turgay et al., 1997). The cluster also contains ClpC’ repressor CtsR (Derré et al., 1999) and the universal sigma factor SigA. Furthermore, SigA and ClpC share associations with the three heat shock proteins DnaK, RrpE, and GroEL. Although *mecA* was not detected as differentially expressed, the paralog *mecB* was, and it is part of this second cluster (Persuh et al., 2002). *B. subtilis*’ other two Clp-proteins ClpP and ClpE are also part of this cluster. ClpE had a significantly higher expression on timepoint 118 h in +prsA (logFC = 2.6 , adj. $p = 0.0005$), which is relevant because ClpE destabilizes the functionality of the repressor CtsR (Miethke et al., 2006).

Prophage genes

A third cluster (Figure 4C) contains a set of tightly associated 24 PBSX prophage and prophage-like genes that were all significantly higher expressed in +prsA compared to control at various timepoints during the entire duration of the fermentation ($\log_{FC} \in [0.9, 2.4]$, adj. $p \in [2.6e-23, 9.2e-3]$). PBSX, a defective *B. subtilis* prophage (Wood et al., 1990a), is known to be potentially heat-induced (Wood et al., 1990b), and they have a potential association with the level of lytic stress resistance (Buxton, 1980).

Potential cell motility regulation

Finally, the fourth cluster has an interesting pattern of associations involving many chemotaxis genes (Figure 4D). This cluster is structured into two separate interconnected components: On the one side, there are 29 chemotaxis proteins and on the other 125 protein-coding genes with various catalytic functions [116 of 125 (92.8%) genes are annotated in the general catalytic activity term GO:0003824]; however, both parts are connected by a backbone of associated genes. This backbone includes the central flagella motion frequency regulator CheA, the flagellar hook-filament FlgK, the general stress sigma factor SigB, the heat-shock protein sigma factor SigI, and the two partially characterized signal transducers YesN and YwspD (Fabret et al., 1999; Petersohn et al., 2001; Zuber et al., 2001; Asai et al., 2007; Mukherjee and Kearns, 2014). Interacting with SigB are five stress regulatory proteins induced by SigB (according to STRING annotations). Both YesN and YwspD are described as histidine kinases, although the corresponding response regulator remains unknown (Fabret et al., 1999; Caspi et al., 2014; Zhu and Stülke, 2018; Geissler et al., 2021). Even if the regulators are unknown, the backbone has an interesting pattern of antagonistic logFC: (i) YesN is significantly lower expressed in +prsA on timepoint 21 h ($\log_{FC} = -1.7$, adj. $p = 1.4e-6$) and 26 h ($\log_{FC} = -1.84$, adj. $p = 8.5e-5$), (ii) YwspD is higher expressed in +prsA on 21 h ($\log_{FC} = 0.6$, adj. $p = 0.0037$), and (iii) CheA lower again on 21 h ($\log_{FC} = -0.7$, adj. $p = 0.0025$). The bottom-line is that the PPI analysis elucidates the tight associations between heat shock, competence transformation, cell motility, general stress response, and translation.

Discussion

In this study, we investigated how PrsA over-expression in *B. subtilis* impacts the transcriptome during fed-batch alpha-amylase (AMY) fermentation. We carried out a temporally resolved RNA-seq study to analyze expression levels and regulation of biological processes with respect to secretion stress. We inspected a comprehensive set of coding and non-coding reference annotations and 542 novel potentially transcribed regions (NPTRs). The fermentation process strongly affects gene expression and we observe a large number of differentially expressed genes both between the strange and

overtime: a total of 1,793 coding genes (67% of expressed genes), 234 NPTRs (66%), 68 putative ncRNAs (64%), 20 riboswitches (54%), 9 tRNAs (41%), and 3 sRNAs (33%) were differentially expressed. The PrsA over-expressing strain, which is consistent with prior descriptions had increased yield and reduced growth (Quesada-Ganuza et al., 2019), was observed to have a significant strain-specific differential expression for more than half of the transcribed genes. Subsequent in-depth analysis of regulated biological processes (Figure 3) and the PPI network of differentially expressed coding genes (Figure 4) shed light on the complex intertwined processes of stress pathways, core energy metabolism, and cell motility (Helmann et al., 1988; Márquez-Magaña et al., 1990; Storz and Hengge, 2010; Yan and Wu, 2019).

Concerning the NPTR, we assessed their potential to contain ORF relative to predictions that recalled 94.3% of known genes with high precision of 96.7%. A marginal fraction of these ORF overlap unannotated regions (Table 3). Therefore, our data do not suggest the presence of ORF in the NPTR, including those with detected differential expression in this dataset. Future investigation for potential conservation of RNA—let alone assessment of their biological function—requires RNA structure alignments that can have average sequence identities below 40% (Yao et al., 2006, 2007; Weinberg et al., 2010). We predicted the NPTR relative to a reference transcript annotation that integrates a comprehensive set of annotation databases and resources (Nicolas et al., 2012; Caspi et al., 2014; Geissler et al., 2021; Pedreira et al., 2022). Among these resources is SubtiWiki, an active community effort that comprehensively collects previously identified coding and non-coding genes (Zhu and Stülke, 2018). Therefore, we consider the NPTR to extend beyond known transcribed regions.

Amino acid and energy metabolism

The observation of the potentially downregulated ATP biosynthesis in the +prsA strain surprised us: (i) The AMY hypersecretion is stressful and energy-intensive for the cells (Song et al., 2015). (ii) It has been hypothesized that ATP might be required for PrsA chaperone activity (Yan and Wu, 2017). (iii) The reduction of ATP levels can also increase the general stress response of *B. subtilis* (Haldenwang, 1995; Petersohn et al., 2001; Yan and Wu, 2019). The potential downregulation of ATP biosynthesis in the +prsA strain seems counterintuitive because the strain has both lower stress and higher yield than the control (Quesada-Ganuza et al., 2019). However, the reduced ATP biosynthesis might be due to the impact of the hypersecreted AMY and over-expressed PrsA on amino acid metabolism. Contrary to the evolutionary energetic adaption of the amino acid composition for secreted proteins (Smith and Chapman, 2010), the four amino acids tryptophan, asparagine, aspartic acid, and lysine are over-represented in the AMY and PrsA proteins (Table 2). Although the specific metabolism

processes for these four amino acids were not detected as significantly regulated during fermentation (Figure 3), more general amino acid processes (e.g., transport) or biosynthetic processes for other amino acids (arginine and histidine) were significantly over-represented by regulated genes. On the one hand, the upregulation of arginine synthesis and related transport mechanisms improves osmotic stress resistance (Du et al., 2011; Zaprasi et al., 2015), which in turn is beneficial to AMY production in *B. subtilis* (Zhao et al., 2018). On the other hand, the over-represented amino acids might explain the reduced ATP biosynthesis. (i) Tryptophan is the amino acid with the highest biosynthetic cost in *B. subtilis*, with a 42.9% higher cost than the second most costly amino acid (phenylalanine) (Akashi and Gojobori, 2002). (ii) The biosynthesis, in particular for costly amino acids, diverges intermediate metabolites from ATP biosynthesis (Akashi and Gojobori, 2002). In the case of tryptophan, the intermediate metabolites have already diverged from glycolysis, which also impacts the downstream citrate cycle (Kanehisa and Goto, 2000; Akashi and Gojobori, 2002). However, a more definite inspection to confirm the regulation of the amino acids and ATP metabolism would require an investigation of concentrations of the individual metabolites with for instance metabolomics.

Cell wall destabilizing processes

The over-expression of PrsA is known to lead to reduced cell growth and cell lysis (Quesada-Ganuza et al., 2019). It was suggested that protein-protein interactions of specific PrsA protein domains are causal for these phenotypes (Quesada-Ganuza et al., 2019). Our data suggest that, on a transcription regulatory level, the PrsA-over-expressing strain has both increased polysaccharide catabolism and reduced polysaccharide biosynthesis. We hypothesize that this strongly contributes to cell wall breakdown, which leads to detrimental phenotypes. Therefore, investigating the associated differentially expressed genes could potentially be the outset to trace back the causality chain of why their regulation changes, and as a path forward to finding candidates that stabilize cell walls and increase yields. Furthermore, the PPI network analysis highlighted 24 tightly associated PBSX prophage and prophage-like genes (Figure 4C) that might be decisive in unraveling the PrsA over-expression lysis phenomena (Buxton, 1980; Quesada-Ganuza et al., 2019), particularly due to the heat-induced (and thus secretion stress-related) expression of the PBSX genes (Wood et al., 1990b).

Stress and cell motility

The protein-protein interaction network analysis resulted in four clusters of proteins that we found to be relevant to

this study's outset (Figure 4). These were the genes of the CsrRS two-component secretion stress response in one cluster (Figure 4A), while the known ClpC regulatory switch and its associations with secretion stress, competence transformation, and associations with the universal sigma factor SigA belong to another cluster (Figure 3B; Turgay et al., 1997). Furthermore, the analysis provided a large cluster (Figure 4D) of cell motility-related genes, which is consistent with the large number of proteins involved in regulating bacterial motility (Rajagopala et al., 2007). A closer inspection of the latter cluster suggests that the proteins YesN and YwsqD might have a signaling role in balancing between cell motility and 125 genes that are annotated to have various metabolic catalytic functions, e.g., the phosphogluconate dehydrogenase. To our knowledge, the potential relationship between cell motility and AMY fermentation has not been elucidated so far, although a potential hypothesis could be that the signaling facilitates the regulation of flagellar cell motility to escape from the stress region (Helmann et al., 1988; Márquez-Magaña et al., 1990; Yan and Wu, 2019). Nevertheless, a follow-up study is needed to verify cell motility regulation during AMY production.

Conclusion

In conclusion, our transcriptome study highlights the expression dynamics of secretion stress during fed-batch AMY fermentation. The comparison of expression levels in a PrsA over-expressing strain to a control strain showed differential expression for nearly half of the transcribed genes. A wide variety of upregulated and downregulated biological processes is related to energy and amino acid metabolism. Also, the data shows potential associations of the cell lysis phenomenon of PrsA over-expression with the stress response and cell motility. Overall, these results identify genes and biological processes, which are affected during fermentation and by the overexpression of PrsA and provide a starting point for future genetic modification of *B. subtilis* for improved yield.

Data availability statement

The genomic sequences and RNA-seq data were deposited in the GEO database (GSE189556). The expression coverages are presented as a browser for interactive investigation (<http://rth.dk/resources/bsg/prsa>). The annotations of the BSGAtlas are accessible at <https://rth.dk/resources/bsgatlas/>. The additional putative ncRNA annotations are part of the supplementary information of Nicolas et al. (2012). The RNA-seq data were processed with a reproducible pipeline located at [doi: 10.5281/zenodo.4534403](https://doi.org/10.5281/zenodo.4534403).

Author contributions

AG conducted the entire computational analysis and wrote the manuscript. LP extracted the RNA. ND contributed to the analysis and methodology design of the PPI network. CA contributed to the discussion of the expression analysis. EG-T contributed to the writing in the early stage. AB prepared the bacterial strains. LJ contributed to the discussion of gene clustering, enrichment analysis, and PPI network analysis. SES, CH, JV, and JG supervised the work. JG and AG made the study design. JG was the main project coordinator. All authors read and approved the manuscript.

Funding

This work was supported by the Innovation Fund Denmark (5163-00010B) and Novo Nordisk Foundation (NNF14CC0001).

Acknowledgments

We thank Annaleigh Ohrt Fehler for pivoting the samples for sequencing and both Bjarke Krysel Christensen and Thomas B. Kallehauge for support in conducting the fermentation and sampling.

References

- Akashi, H., and Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *PNAS* 99, 3695–3700. doi: 10.1073/pnas.062526999
- Alexa, A., and Rahnenführer, J. (2022). *topGO: Enrichment Analysis for Gene Ontology. R package Version 2.48.0*. Available online at: <https://git.bioconductor.org/packages/topGO> (accessed March 4, 2021).
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed July 18, 2017).
- Arrieta-Ortiz, M. L., Hafemeister, C., Bate, A. R., Chu, T., Greenfield, A., Shuster, B., et al. (2015). An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol. Syst. Biol.* 11:839. doi: 10.15252/msb.20156236
- Asai, K., Ootsuji, T., Obata, K., Matsumoto, T., Fujita, Y., and Sadaie, Y. (2007). Regulatory role of RsgI in sigI expression in *Bacillus subtilis*. *Microbiology* 153, 92–101. doi: 10.1099/mic.0.29239-0
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Buescher, J. M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., et al. (2012). Global Network Reorganization During Dynamic Adaptations of *Bacillus subtilis* Metabolism. *Science* 335, 1099–1103. doi: 10.1126/science.1206871
- Buxton, R. S. (1980). Selection of *Bacillus subtilis* 168 Mutants with Deletions of the PBX Prophage. *J. Gene. Virol.* 46, 427–437. doi: 10.1099/0022-1317-46-2-427
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, D459–D471. doi: 10.1093/nar/gkt1103
- Cruz Ramos, H., Hoffmann, T., Marino, M., Nedjari, H., Presecan-Siedel, E., Dreesen, O., et al. (2000). Fermentative Metabolism of *Bacillus subtilis*: physiology and Regulation of Gene Expression. *J. Bacteriol.* 182, 3072–3080. doi: 10.1128/JB.182.11.3072-3080.2000
- Derré, I., Rapoport, G., and Msadek, T. (1999). CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. *Mol. Microbiol.* 31, 117–131. doi: 10.1046/j.1365-2958.1999.01152.x
- Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2019). Cytoscape StringApp: network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* 18, 623–632. doi: 10.1021/acs.jproteome.8b00702
- Driessen, A. J., Fekkes, P., and van der Wolk, J. P. (1998). The Sec system. *Curr. Opin. Microbiol.* 1, 216–222. doi: 10.1016/S1369-5274(98)80014-3
- Du, Y., Shi, W. W., He, Y. X., Yang, Y. H., Zhou, C. Z., and Chen, Y. (2011). Structures of the substrate-binding protein provide insights into the multiple compatible solute binding specificities of the *Bacillus subtilis* ABC transporter OpuC. *Biochem. J.* 436, 283–289. doi: 10.1042/BJ20102097
- Enright, A. J. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575
- Fabret, C., Feher, V. A., and Hoch, J. A. (1999). Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. *J. Bacteriol.* 181, 1975–1983. doi: 10.1128/JB.181.7.1975-1983.1999
- Fu, L. L., Xu, Z. R., Li, W. F., Shuai, J. B., Lu, P., and Hu, C. X. (2007). Protein secretion pathways in *Bacillus subtilis*: implication for optimization

Conflict of interest

AB and CH were employed by the Novozymes A/S.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.909493/full#supplementary-material>

- of heterologous protein secretion. *Biotechnol. Adv.* 25, 1–12. doi: 10.1016/j.biotechadv.2006.08.002
- Geissler, A. S., Anthon, C., Alkan, F., González-Tortuero, E., Poulsen, L. D., Kallehauge, T. B., et al. (2021). BSAtlas: a unified *Bacillus subtilis* genome and transcriptome annotation atlas with enhanced information access. *Microb. Genom.* 7:000524. doi: 10.1099/mgen.0.000524
- Geissler, A. S., Fehler, A. O., Poulsen, L. D., González-Tortuero, E., Kallehauge, T. B., Alkan, F., et al. (2022). CRISPRi screen for enhancing heterologous α -amylase yield in *Bacillus subtilis*. *bioRxiv* [Preprint]. doi: 10.1101/2022.03.30.486407
- González-Pastor, J. E. (2011). Cannibalism: a social behavior in sporulating *Bacillus subtilis*. *FEMS Microbiol. Rev.* 35, 415–424. doi: 10.1111/j.1574-6976.2010.00253.x
- Gupta, M., and Rao, K. K. (2014). Phosphorylation of DegU is essential for activation of amyE expression in *Bacillus subtilis*. *J. Biosci.* 39, 747–752. doi: 10.1007/s12038-014-9481-5
- Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., et al. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 47, D853–D858. doi: 10.1093/nar/gky1095
- Hahne, H., Mäder, U., Otto, A., Bonn, F., Steil, L., Bremer, E., et al. (2010). A Comprehensive Proteomics and Transcriptomics Analysis of *Bacillus subtilis* Salt Stress Adaptation. *J. Bacteriol.* 192, 870–882. doi: 10.1128/JB.01106-09
- Haldenwang, W. G. (1995). The sigma factors of *Bacillus subtilis*. *Microbiol. Rev.* 59, 1–30.
- Harris, R. S. (2007). *Improved Pairwise Alignment of Genomic DNA*. Ph.D thesis. University Park, PA: Pennsylvania State University.
- Helmann, J. D., Marquez, L. M., and Chamberlin, M. J. (1988). Cloning, Sequencing, and Disruption of the *Bacillus subtilis* c28 Gene. *J. Bacteriol.* 170:7.
- Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., et al. (2009). Fast Mapping of Short Sequences with Mismatches, Insertions and Deletions Using Index Structures. *PLoS Comput. Biol.* 5:e1000502. doi: 10.1371/journal.pcbi.1000502
- Hohmann, H. P., van Dijk, J. M., Krishnappa, L., and Prágai, Z. (2016). “Host Organisms: *Bacillus subtilis*,” in *Industrial Biotechnology*, eds C. Wittmann and J. C. Liao (Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA), 221–297. doi: 10.1002/9783527807796.ch7
- Hosoda, J., Kohiyama, M., and Nomura, M. (1959). Studies on amylase formation by *Bacillus subtilis*: vii. Effect of purine, pyrimidine and their analogues on exoenzyme formation by uracil- and adenine-requiring mutants. *J. Biochem.* 46, 857–864. doi: 10.1093/oxfordjournals.jbchem.a126976
- Hosseini, S., Curilovs, A., and Cutting, S. M. (2018). Biological Containment of Genetically Modified *Bacillus subtilis*. *Appl. Environ. Microbiol.* 84, e02334–17. doi: 10.1128/AEM.02334-17
- Hyatt, D., Chen, G. L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Hyyryläinen, H.-L., Vitikainen, M., Thwaite, J., Wu, H., Sarvas, M., Harwood, C. R., et al. (2000). d-Alanine Substitution of Teichoic Acids as a Modulator of Protein Folding and Stability at the Cytoplasmic Membrane/Cell Wall Interface of *Bacillus subtilis*. *J. Biol. Chem.* 275, 26696–26703. doi: 10.1016/S0021-9258(19)61432-8
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kiel, J. A. K. W., Boels, J. M., Beldman, G., and Venema, G. (1994). Glycogen in *Bacillus subtilis*: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol. Microbiol.* 11, 203–218. doi: 10.1111/j.1365-2958.1994.tb00301.x
- Kilstrup, M., Hammer, K., Ruhdal Jensen, P., and Martinussen, J. (2005). Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol. Rev.* 29, 555–590. doi: 10.1016/j.fmr.2005.04.006
- Kobayashi, K. (2007). Gradual activation of the response regulator DegU controls serial expression of genes for flagellum formation and biofilm formation in *Bacillus subtilis*. *Mol. Microbiol.* 66, 395–409. doi: 10.1111/j.1365-2958.2007.05923.x
- Kontinen, V. P., and Sarvas, M. (1993). The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion. *Mol. Microbiol.* 8, 727–737. doi: 10.1111/j.1365-2958.1993.tb01616.x
- Koster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522. doi: 10.1093/bioinformatics/bts480
- Larsson, J. T., Rogstam, A., and von Wachenfeldt, C. (2005). Coordinated patterns of cytochrome bd and lactate dehydrogenase expression in *Bacillus subtilis*. *Microbiology* 151, 3323–3335. doi: 10.1099/mic.0.28124-0
- Laub, M. T. (2014). “The Role of Two-Component Signal Transduction Systems in Bacterial Stress Responses,” in *Bacterial Stress Responses*, eds G. Storz and R. Hengge (Washington, DC, USA: ASM Press), 45–58. doi: 10.1128/9781555816841.ch4
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118
- Lee, S., Lawrence, M., and Cook, D. (2018). *plyranges: A Fluent Interface for Manipulating Genomic Ranges*. Available online at: <https://github.com/sa-lee/plyranges> (accessed July 9, 2017).
- Legeay, M., Doncheva, N. T., Morris, J. H., and Jensen, L. J. (2020). Visualize omics data on networks with Omics Visualizer, a Cytoscape App. *F1000 Res.* 9:157. doi: 10.12688/f1000research.22280.2
- Lemma, E., Uden, G., and Kröger, A. (1990). Menaquinone is an obligatory component of the chain catalyzing succinate respiration in *Bacillus subtilis*. *Arch. Microbiol.* 155, 62–67. doi: 10.1007/BF00291276
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Lim, B., and Gross, C. A. (2014). “Cellular Response to Heat Shock and Cold Shock” in *Bacterial Stress Responses*, eds G. Storz and R. Hengge (Washington, DC, USA: ASM Press), 91–114. doi: 10.1128/9781555816841.ch7
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Lu, X., Zhang, H., Tonge, P. J., and Tan, D. S. (2008). Mechanism-based inhibitors of MenE, an acyl-CoA synthetase involved in bacterial menaquinone biosynthesis. *Bioorg. Med. Chem. Lett.* 18, 5963–5966. doi: 10.1016/j.bmcl.2008.07.130
- Márquez-Magaña, L. M., Helemann, J. D., Ferrari, E., Helen, Parker HM, Ordal, G. W., and Chamberlin, M. J. (1990). Studies of σ -D-Dependent Functions in *Bacillus subtilis*. *J. Bacteriol.* 172, 3435–3443.
- Miethke, M., Hecker, M., and Gerth, U. (2006). Involvement of *Bacillus subtilis* ClpE in CtsR Degradation and Protein Quality Control. *J. Bacteriol.* 188, 4610–4619. doi: 10.1128/JB.00287-06
- Morris, J. H., Apeltsin, L., Newman, A. M., Baumbach, J., Wittkop, T., Su, G., et al. (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* 12:436. doi: 10.1186/1471-2105-12-436
- Mukherjee, S., and Kearns, D. B. (2014). The Structure and Regulation of Flagella in *Bacillus subtilis*. *Annu. Rev. Genet.* 48, 319–340. doi: 10.1146/annurev-genet-120213-092406
- Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., et al. (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. *Science* 335, 1103–1106. doi: 10.1126/science.1206848
- Nijland, R., Heerlien, R., Hamoen, L. W., and Kuipers, O. P. (2007). Changing a Single Amino Acid in Clostridium perfringens β -Toxin Affects the Efficiency of Heterologous Secretion by *Bacillus subtilis*. *AEM* 73, 1586–1593. doi: 10.1128/AEM.02356-06
- Ohki, R., Giyanto, N., Tateno, K., Masuyama, W., Moriya, S., Kobayashi, K., et al. (2003). The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis*. *Mol. Microbiol.* 49, 1135–1144. doi: 10.1046/j.1365-2958.2003.03653.x
- Otto, A., Bernhardt, J., Meyer, H., Schaffer, M., Herbst, F.-A., Siebourg, J., et al. (2010). Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis*. *Nat. Commun.* 1:137. doi: 10.1038/ncomms1137
- Pagès, H. (2021). *BSgenome: Software Infrastructure for Efficient Representation of Full Genomes and their SNPsR Package Version 1.64.0*. Available online at: <https://bioconductor.org/packages/BSgenome> (accessed November 22, 2021).
- Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2019). *Biostrings: Efficient Manipulation of Biological Strings*. Available online at: <https://bioconductor.org/packages/Biostrings> (accessed July 9, 2019).
- Pedreira, T., Elfmann, C., and Stülke, J. (2022). The current state of Subti Wiki, the database for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* 50, D875–D882. doi: 10.1093/nar/gkab943
- Peifer, S., Barduhn, T., Zimmet, S., Volmer, D. A., Heinzle, E., and Schneider, K. (2012). Metabolic engineering of the purine biosynthetic pathway in

- Corynebacterium glutamicum* results in increased intracellular pool sizes of IMP and hypoxanthine. *Microb. Cell Factories* 11:138. doi: 10.1186/1475-2859-11-138
- Persuh, M., Mandic-Mulec, I., and Dubnau, D. (2002). A MecA Paralog, YpbH, Binds ClpC, Affecting both Competence and Sporulation. *J. Bacteriol.* 184, 2310–2313. doi: 10.1128/JB.184.8.2310-2313.2002
- Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J. D., Völker, U., and Hecker, M. (2001). Global Analysis of the General Stress Response of *Bacillus subtilis*. *J. Bacteriol.* 183, 5617–5631. doi: 10.1128/JB.183.19.5617-5631.2001
- Quesada-Ganuza, A., Antelo-Varela, M., Mouritzen, J. C., Bartel, J., Becher, D., Gjermansen, M., et al. (2019). Identification and optimization of PrsA in *Bacillus subtilis* for improved yield of amylase. *Microb. Cell Factories* 18:158. doi: 10.1186/s12934-019-1203-0
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing. R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rajagopala, S. V., Titz, B., Goll, J., Parrish, J. R., Wohlbold, K., McKeivitt, M. T., et al. (2007). The protein network of bacterial motility. *Mol. Syst. Biol.* 3:128. doi: 10.1038/msb4100166
- Ramos-Silva, P., Serrano, M., and Henriques, A. O. (2019). From Root to Tips: sporulation Evolution and Specialization in *Bacillus subtilis* and the Intestinal Pathogen *Clostridioides difficile*. *Mol. Biol. Evol.* 36, 2714–2736. doi: 10.1093/molbev/msz175
- Rodionov, D. A., Li, X., Rodionova, I. A., Yang, C., Sorci, L., Dervyn, E., et al. (2008). Transcriptional regulation of NAD metabolism in bacteria: genomic reconstruction of NiaR (YrxA) regulon. *Nucleic Acids Res.* 36, 2032–2046. doi: 10.1093/nar/gkn046
- Schallmey, M., Singh, A., and Ward, O. P. (2004). Developments in the use of *Bacillus* species for industrial production. *Can. J. Microbiol.* 50, 1–17. doi: 10.1139/w03-076
- Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22, 839–851. doi: 10.1261/rna.053959.115
- Shannon, P. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Smith, D. R., and Chapman, M. R. (2010). Economical Evolution: microbes Reduce the Synthetic Cost of Extracellular Proteins. *mBio* 1, e00131–10. doi: 10.1128/mBio.00131-10
- Song, Y., Nikoloff, J. M., and Zhang, D. (2015). Improving Protein Production on the Level of Regulation of both Expression and Secretion Pathways in *Bacillus subtilis*. *J. Microbiol. Biotechnol.* 25, 963–977. doi: 10.4014/jmb.1501.01028
- Spinnler, H. E. (2021). “Production of enzymes: Fermentation and genetic engineering,” in *Enzymes Novel Biotechnological Approaches for the Food Industry*, eds S. Kermasha and N. A. Michael Eskin (Amsterdam: Elsevier), 45–58. doi: 10.1016/B978-0-12-800217-9.00003-4
- Storz, G., and Hengge, R. (eds) (2010). *Bacterial Stress Responses*. Washington, DC, USA: ASM Press, doi: 10.1128/9781555816841
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika* 18, 267–276. doi: 10.1007/BF02289263
- Turgay, K., Hamoen, L. W., Venema, G., and Dubnau, D. (1997). Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev.* 11, 119–128. doi: 10.1101/gad.11.1.119
- Van den Berge, K., Soneson, C., Robinson, M. D., and Clement, L. (2017). stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biol.* 18:151. doi: 10.1186/s13059-017-1277-0
- van Dijl, J., and Hecker, M. (2013). *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microb. Cell Factories* 12:3. doi: 10.1186/1475-2859-12-3
- Verhamme, D. T., Kiley, T. B., and Stanley-Wall, N. R. (2007). DegU coordinates multicellular behaviour exhibited by *Bacillus subtilis*. *Mol. Microbiol.* 65, 554–568. doi: 10.1111/j.1365-2958.2007.05810.x
- Vitikainen, M., Pummi, T., Airaksinen, U., Wahlstrom, E., Wu, H., Sarvas, M., et al. (2001). Quantitation of the Capacity of the Secretion Apparatus and Requirement for PrsA in Growth and Secretion of alpha-Amylase in *Bacillus subtilis*. *J. Bacteriol.* 183, 1881–1890. doi: 10.1128/JB.183.6.1881-1890.2001
- Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H., et al. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* 11:R31. doi: 10.1186/gb-2010-11-3-r31
- Westers, H., Darmon, E., Zanen, G., and van Dijl, J. M. (2004). The *Bacillus* secretion stress response is an indicator for alpha-amylase production levels. *Letts. Appl. Microbiol.* 39, 65–73. doi: 10.1111/j.1472-765X.2004.01539.x
- Westers, H., Westers, L., Darmon, E., van Dijl, J. M., Quax, W. J., and Zanen, G. (2006). The CssRS two-component regulatory system controls a general secretion stress response in *Bacillus subtilis*. *FEBS J.* 273, 3816–3827. doi: 10.1111/j.1742-4658.2006.05389.x
- Wood, H. E., Dawson, M. T., Devine, K. M., and McConnell, D. J. (1990a). Characterization of PBSX, a defective prophage of *Bacillus subtilis*. *J. Bacteriol.* 172, 2667–2674. doi: 10.1128/jb.172.5.2667-2674.1990
- Wood, H. E., Devine, K. M., and McConnell, D. J. (1990b). Characterisation of a repressor gene (xre) and a temperature-sensitive allele from the *Bacillus subtilis* prophage. *PBSX Gene* 96, 83–88. doi: 10.1016/0378-1119(90)90344-Q
- Yan, S., and Wu, G. (2017). Bottleneck in secretion of alpha-amylase in *Bacillus subtilis*. *Microb. Cell Factories* 16:124. doi: 10.1186/s12934-017-0738-1
- Yan, S., and Wu, G. (2019). Proteases HtrA and HtrB for alpha-amylase secreted from *Bacillus subtilis* in secretion stress. *Cell Stress Chaperones* 24, 493–502. doi: 10.1007/s12192-019-00985-1
- Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M., et al. (2007). A Computational Pipeline for High- Throughput Discovery of cis-Regulatory Noncoding RNA in Prokaryotes. *PLoS Comput. Biol.* 3:e126. doi: 10.1371/journal.pcbi.0030126
- Yao, Z., Weinberg, Z., and Ruzzo, W. L. (2006). CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 22, 445–452. doi: 10.1093/bioinformatics/btk008
- Zapras, A., Bleisteiner, M., Kerres, A., Hoffmann, T., and Bremer, E. (2015). Uptake of Amino Acids and Their Metabolic Conversion into the Compatible Solute Proline Confers Osmoprotection to *Bacillus subtilis*. *Appl. Environ. Microbiol.* 81, 250–259. doi: 10.1128/AEM.02797-14
- Zhao, L., Ye, J., Fu, J., and Chen, G.-Q. (2018). Engineering peptidoglycan degradation related genes of *Bacillus subtilis* for better fermentation processes. *Bioresour. Technol.* 248, 238–247. doi: 10.1016/j.biortech.2017.05.134
- Zhu, B., and Stülke, J. (2018). SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.* 46, D743–D748. doi: 10.1093/nar/gkx908
- Zuber, U., Drzewiecki, K., and Hecker, M. (2001). Putative Sigma Factor SigI (YkoZ) of *Bacillus subtilis* Is Induced by Heat Shock. *J. Bacteriol.* 183, 1472–1475. doi: 10.1128/JB.183.4.1472-1475.2001