RNA-SEQ AND PROTEOMICS BASED ANALYSIS OF REGULATORY RNA FEATURES AND GENE EXPRESSION IN *BACILLUS LICHENIFORMIS*

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

im Promotionsgrundprogramm Biologie
der Georg-August University School of Science (GAUSS)

vorgelegt von
Sandra Wiegand
aus Solingen

Göttingen 2013

Betreuungsausschuss

Prof. Dr. Rolf Daniel, Abteilung für Genomische und Angewandte Mikrobiologie und Laboratorium für Genomanalyse, Institut für Mikrobiologie und Genetik

Prof. em. Dr. Gerhard Gottschalk, Abteilung für Genomische und Angewandte Mikrobiologie und Laboratorium für Genomanalyse, Institut für Mikrobiologie und Genetik

Mitglieder der Prüfungskommission

Referent
Prof. Dr. Rolf Daniel, Abteilung für Genomische und Angewandte Mikrobiologie und Laboratorium für Genomanalyse, Institut für Mikrobiologie und Genetik

Korreferent
Prof. em. Dr. Gerhard Gottschalk, Abteilung für Genomische und Angewandte Mikrobiologie und Laboratorium für Genomanalyse, Institut für Mikrobiologie und Genetik

Weitere Mitglieder der Prüfungskommission

Prof. Dr. Stefanie Pöggeler, Abteilung Genetik eukaryotischer Mikroorganismen, Institut für Mikrobiologie und Genetik

PD Dr. Michael Hoppert, Abteilung für Allgemeine Mikrobiologie , Institut für Mikrobiologie und Genetik

PD Dr. Wilfried Kramer, Abteilung für Molekulare Genetik, Institut für Mikrobiologie und Genetik

Jun.-Prof. Dr. Kai Heimel, Abteilung für Molekulare Mikrobiologie und Genetik, Institut für Mikrobiologie und Genetik

Tag der mündlichen Prüfung: 25. September 2013

LIST OF PUBLICATIONS

Journal publications

RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation

Sandra Wiegand, Sascha Dietrich, Robert Hertel, Johannes Bongaerts, Stefan Evers, Sonja Volland, Rolf Daniel and Heiko Liesegang

BMC Genomics (2013), 14:667

Fermentation stage-dependent adaptations of *Bacillus licheniformis* during enzyme production

Sandra Wiegand, Birgit Voigt, Dirk Albrecht, Johannes Bongaerts, Stefan Evers, Michael Hecker, Rolf Daniel and Heiko Liesegang

Microbial Cell Factories (2013), 12:120

TraV: A Genome Context Sensitive Transcriptome Browser

Sascha Dietrich, **Sandra Wiegand** and Heiko Liesegang

PLoS One (2014), 9(4):e93677

Complete Genome Sequence of *Geobacillus* sp. GHH01, a Thermophilic Lipase-secreting Bacterium

Sandra Wiegand, Ulrich Rabausch, Jennifer Chow, Rolf Daniel, Wolfgang R. Streit and Heiko Liesegang

Genome Announcements (2013), Vol. 1, Issue 2, e00092-13

Poster presentations

Functional genome analysis of *Geobacillus sp*. HH01, an organism that secrets a thermostable lipase

Sandra Wiegand, Ulrich Rabausch, Wolfgang R. Streit and Heiko Liesegang

Annual meeting of the *Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM)* 2011 in Karlsruhe, Germany (2011)

Transcriptome deep-sequencing reveals regulatory processes in *Bacillus licheniformis*

Sandra Wiegand*, Sascha Dietrich* and Heiko Liesegang

* SW and SD contributed equally to this work

Gram+ 6th International Conference on Gram-positive Microorganisms in Montecatini Terme, Italy (2011)

Applicable features of *Geobacillus* sp. HH01

Sandra Wiegand, Ulrich Rabausch, Wolfgang R. Streit and Heiko Liesegang

ProkaGENOMICS 2011 - 5th European Conference on Prokaryotic and Fungal Genomics in Göttingen, Germany (2011)

Antisense Transcripts in *Bacillus licheniformis* DSM13 - Insights from RNA-Seq Data

Sascha Dietrich, **Sandra Wiegand**, Robert Hertel and Heiko Liesegang

ProkaGENOMICS 2011 - 5th European Conference on Prokaryotic and Fungal Genomics in Göttingen, Germany (2011)

TABLE OF CONTENTS

Abbreviations III

Chapter A Introduction 1
A.1 *Bacillus licheniformis* DSM13 1
A.2 The bacterial transcriptome 4
A.2.1 Post-transcriptional regulation 4
A.2.2 Regulatory RNAs localized in untranslated regions 5
A.2.2.1 *Trans*-encoded small RNAs 7
A.2.2.2 Cis-encoded antisense RNAs 9
A.2.3 Analysis of the bacterial transcriptome 11
A.3 Aim of the thesis 14
References 15

Chapter B RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation 23
Sandra Wiegand, Sascha Dietrich, Robert Hertel, Johannes Bongaerts, Stefan Evers, Sonja Volland, Rolf Daniel and Heiko Liesegang
Additional information 44

Chapter C Fermentation stage-dependent adaptations of *Bacillus licheniformis* during enzyme production 73
Sandra Wiegand, Birgit Voigt, Dirk Albrecht, Johannes Bongaerts, Stefan Evers, Michael Hecker, Rolf Daniel and Heiko Liesegang
Additional information 92

Chapter D TraV: A Genome Context Sensitive Transcriptome Browser 119
Sascha Dietrich, Sandra Wiegand and Heiko Liesegang

Chapter E Complete Genome Sequence of *Geobacillus* sp. GHH01, a Thermophilic Lipase-secreting Bacterium **129**
Sandra Wiegand, Ulrich Rabausch, Jennifer Chow, Rolf Daniel, Wolfgang R. Streit and Heiko Liesegang

Chapter F Discussion and outlook **133**
F.1 Transcriptome complexity 133
F.1.1 Untranslated regions 133
F.1.2 Non-coding antisense transcripts 138
F.1.3 The operon concept 140
F.2 Process monitoring and optimization 142
F.2.1 Process monitoring by RNA-Seq 142
F.2.2 Strategies for process monitoring and optimization 143
F.2.3 Putative optimization targets 145
References 150
Additional information 157

Summary **161**

Acknowledgements VII

ABBREVIATIONS

$(M+H)^+$	Protonated molecular ions
°C	Degrees Celsius
µg	Microgram
µL	Microliter
2D	Two dimensional
ABC	ATP-binding cassette
ANOVA	Analysis of variance
APP	Antarctic phosphatase
asRNA	Antisense RNA
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BAM	Binary version of a SAM file
BGSC	Bacillus Genetic Stock Center
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BPG	1,3-Biphosphoglycerate
C2	Organic molecule harboring two carbon atoms
cDNA	Complementary DNA
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CIGAR	Compact Idiosyncratic Gapped Alignment Report
CO_2	Carbon dioxide
CoA	Coenzyme A
ComA~P	Phosphorylated ComA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Da	Dalton
DegU~P	Phosphorylated DegU
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acid
DOOR	Database of prOkaryotic OpeRons
dRNA-Seq	Differential RNA sequencing
DTT	Dithiothreitol
e.g.	For example

EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory file format
EPS	Exopolymeric polysaccharides
EST	Expressed Sequence Tags
et al.	*Et alii*
F6P	Fructose 6-phosphate
FBP	Fructose 1,6-bisphosphate
FDR	False Discovery Rate
Fe	Iron
FMN	Flavin mononucleotide
FOM	Figure of merit
g	Gram
g	Gravitational constant
GADP	Glyceraldehyde 3-phosphate
G6P	Glucose 6-phosphate
GBK	GenBank file format
Gbyte	Gigabyte
GDH	Glutamate dehydrogenase
GFF3	Generic Feature Format Version 3
GO	Genome ontology
GOGAT	Glutamate synthase
GRAS	Generally recognized as safe
GS	Glutamine synthetase
h	Hours
H_2O_2	Hydrogen peroxide
HCl	Hydrogen chloride
i.e.	*Id est*
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
kVh	Kilovolt hour
L	Liter
ln	Natural logarithm
M	Molar
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
Mb	Megabase pairs
Mbyte	Megabyte
Mg	Magnesium
min	Minute
miRNA	MicroRNA
mL	Milliliter
MLST	Multi-locus sequence typing
mM	Millimolar

mm	Millimetre
MPSS	Massively Parallel Signature Sequencing
mRNA	Messenger RNA
NaCl	Sodium chloride
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NPKM	Nucleotide activity per kilobase of exon model per million mapped reads
nt	Nucleotides
N-terminal	Amino-terminal
OD	Optical density
ORF	Open reading frame
Oxo	Oxoglutarate
P	Phosphate
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PMSF	Phenylmethylsulfonyl fluoride
PNG	Portable Network Graphics
PNK	Polynucleotide kinase
Pyr	Pyruvate
RAM	Random-access memory
RBS	Ribosomal binding site
RK motif	Twin-arginine motif
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
rpm	Revolutions per minute
RR motif	Arginine-lysine motif
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcriptase quantitative PCR
s	Second
S/N	Signal-to-noise ratio
SAGE	Serial Analysis of Gene Expression
SAM	S-adenosyl methionine
SAM	Tab delimited text file that contains sequence alignment data
SD	Shine-Dalgarno sequence
SDS	Sodium dodecyl sulfate
Sec	Secretory
SNP	Single nucleotide polymorphism
sp.	Species
Spo0A~P	Phosphorylated Spo0A
SQL	Structured Query Language

sRNA	Small RNA
SRP	Signal recognition particle
SVG	Scalable Vector Graphics
TAP	Tobacco Acid Pyrophosphatase
Tat	Twin-arginine translocation
TAXI	TRAP-associated extracytoplasmic immunity protein
TCA	Tricarboxylic acid
TDS	Flat file data format to describe and save transcriptome mappings in TraV
TEX	Terminator™ 5´-Phosphate-Dependent Exonuclease
tmRNA	Transfer-messenger RNA
TOF	Time-of-flight mass spectrometer
TPP	Thiamine pyrophosphate
TRAP	Tripartite ATP-independent periplasmic dicarboxylate transporter
TraV	Transcriptome Viewer
tRNA	Transfer RNA
TSS	Transcription start sites
UTR	Untranslated region
V	Volt
v/v	Volume per Volume
Vh	Volt hour
W	Watt
w/v	Weight per volume
WWW	World wide web

Nucleobases

A	Adenine	C	Cytosine
G	Guanine	T	Thymine

Amino Acids

Ala/A	Alanine	Arg/R	Arginine
Asn/N	Asparagine	Asp/D	Aspartic acid
Cys/C	Cysteine	Gln/Q	Glutamine
Glu/E	Glutamic acid	Gly/G	Glycine
His/H	Histidine	Ile/I	Isoleucine
Leu/L	Leucine	Lys/K	Lysine
Met/M	Methionine	Phe/F	Phenylalanine
Pro/P	Proline	Ser/S	Serine
Thr/T	Threonine	Trp/W	Tryptophan
Tyr/Y	Tyrosine	Val/V	Valine
X	Any amino acid		

CHAPTER A

INTRODUCTION

A.1 *Bacillus licheniformis* DSM13

Members of the genus *Bacillus* have been exploited for biotechnological purposes since the production of natto from soy beans using *Bacillus subtilis* was developed in Japan over thousand years ago (Schallmey et al., 2004). As enzymes deriving from bacterial fermentation processes have become accessible for technical applications in the 1960s, they have been produced in large scale employing *B. licheniformis* and *B. amyloliquefaciens* (Maurer, 2004). It is estimated, that 50% of the total enzyme market is nowadays constituted by enzymes obtained from *Bacillus* species, especially from *B. subtilis*, *B. clausii*, and *B. licheniformis* (Schallmey et al., 2004). Main reasons for this development are the species' high growth rates combined with their large capacities to secrete enzymes into the extracellular space. In contrast, in Gram-negative bacteria like *Escherichia coli,* proteins tend to accumulate in the periplasm or cytoplasm, leading to incorrect protein folding or the formation of insoluble inclusion bodies (Hintz, 2003; Schallmey et al., 2004). The enzymes produced by *Bacilli* are mainly employed by detergent, food and textile industries, and comprise large shares of amylases, pullanases, and glucose isomerases (Kirk, 2002; Schallmey et al., 2004). However, approximately 40% of the total enzyme sales account for proteases; alkaline proteases such as Subtilisin Carlsberg, Subtilisin BPN' and Savinase utilized in household detergents constitute the largest proportion among this group (Gupta et al., 2002a, 2002b).

The genus *Bacillus* currently comprises more than 250 described species, after being subject to numerous taxonomic changes during recent years (Logan and de Vos, 2008). For example, several members of this genus were transferred to the newly defined genus *Geobacillus* (Nazina et al., 2001). Two major species groups, the *B. subtilis* and the *B. cereus* group, were identified within the genus *Bacillus* (Ash et al., 1991; Wang et al., 2007). According to this

definition, *B. licheniformis* is a member of the *B. subtilis* group, which also includes *B. pumilus* and *B. amyloliquefaciens* (Rooney et al., 2009). An multi-locus sequence typing (MLST)-based phylogenetic analysis of the *B. licheniformis* species itself revealed that the type strain *B. licheniformis* DSM13 shares a high degree of similarity with *B. licheniformis* ATCC 9945A (Madslien et al., 2012), which is frequently used for laboratory purposes (Hoffmann et al., 2010; Rachinger, 2010).

B. licheniformis was first described as *Clostridium licheniforme* in 1898, referring to the formation of lichen-shaped colonies (Logan and de Vos, 2008). The organism is facultative anaerobic, Gram-positive, forms motile rods and develops endospores (Logan and de Vos, 2008; Slepecky and Hemphill, 2006). *B. licheniformis* has been reported to show growth in slightly acidic environments and in the presence of up to 7% NaCl at temperatures ranging from 15 °C to 55 °C (Logan and de Vos, 2008; Slepecky and Hemphill, 2006). Like most other members of the genus *Bacillus*, *B. licheniformis* is a widely distributed saprophyte and has mainly been isolated from soil. Nevertheless, isolates have also been derived from other sources, including inner plant tissue, feathers, milk, marine sponges and clinical specimens (Logan and de Vos, 2008; Sayem et al., 2011). The species has occasionally been reported as opportunistic pathogen and members have been isolated in cases of bacteremia, peritoneum inflammation, food poisoning and eye infection from immunocompromised patients (Agerholm et al., 1997; Haydushka et al., 2012; Logan and de Vos, 2008). Also, there seem to be cases in which infection with *B. licheniformis* has led to bacteremia in immunocompetent patients (Galanos et al., 2003; Haydushka et al., 2012; Sugar and McCloskey, 1977). However, the U.S. Environmental Protection Agency (1997) stated that infections occur most probably when an individual is exposed to atypically high bacterial cell counts. In this context, enzymes with applications as food additives produced in *B. licheniformis* have been *generally recognized as safe* (GRAS) by the U.S. Food and Drug Administration (1999-2013).

The genome sequence of *B. licheniformis* DSM13 (4.22 Mb), which was also deposited as *B. licheniformis* ATCC 14580 at the American Type Culture Collection (Euzéby, 1997), is available since 2004 and was determined independently in the course of two separate studies (Rey et al., 2004; Veith et al., 2004). Both studies calculated a G+C content of 46.2% and identified seven ribosomal RNA (rRNA) operons and 72 transfer RNAs (tRNAs). Altogether, 4,286 (Veith et al., 2004) or 4,208 (Rey et al., 2004) protein-coding genes were annotated, including regions coding for exoenzymes such as α-amylases, pectate lyases and cellulases. Furthermore, *B. licheniformis* has the capacity to produce several extracellular proteases, e.g.

Subtilisin Carlsberg, to make environmental proteins accessible as carbon and nitrogen sources. In total, 296 proteins were predicted to carry an N-terminal signal peptide for the secretion to the extracellular space (Voigt et al., 2006).

The ability of the Gram-positive model organism *B. subtilis* to differentiate into distinct, co-existing cell types such as motile, sporulating or competent cells, cells that secrete extracellular enzymes or cells that produce biofilms and cannibalize their siblings, has been studied intensely (Dubnau and Losick, 2006; Shank and Kolter, 2011; Veening et al., 2008). Therefore, the formation of different subpopulations discussed in this thesis is based on studies targeting *B. subtilis* and supplemented with the incomplete knowledge of the process in *B. licheniformis*. To coordinate the activation and regulation of the developmental programs of cell differentiation, *B. subtilis* possesses Spo0A, DegU and ComA, which are the master regulators of a complex regulatory network (Davidson et al., 2012; Kuchina et al., 2011; López and Kolter, 2010; Schultz et al., 2009). The phosphorylated state of each master regulator activates the expression of genes required for a distinct developmental pathway (López and Kolter, 2010). The onset of differentiation is dependent on a certain threshold of the regulator, which – once it is reached – will promote an exponential activation of the respective regulon by positive autoregulation or mutually repressing repression (Lopez et al., 2009). Therefore, only a fraction of the cell population can form a specific subpopulation whereas other fractions differentiate into separate subpopulations and thereby build the multi-cellular community (Lopez et al., 2009).

Various approaches to enhance the performance of *B. licheniformis* in terms of the previously mentioned industrial protease production have addressed many different targets. (i) The improvement of *B. licheniformis* as an industrial production strain has mainly been performed by random mutagenesis and subsequent rationalized screening procedures (Gupta et al., 2002a; Hintz, 2003; Parekh et al., 2000). In recent years, attention has also been paid on directed genetic alterations targeting genetic accessibility (Hoffmann et al., 2010; Rachinger, 2010; Waschkau et al., 2008), secretion capability (Waldeck et al., 2007a), sporulation and biological containment (Borgmeier et al., 2012; Nahrstedt et al., 2005; Waldeck et al., 2007b). (ii) The optimization approaches did not only address the host organism, but also the commercial enzymes. Strategies for the improvement of subtilisin by random or site-directed mutagenesis, gene shuffling and phage display have facilitated the production of enzymes with altered pH optima and substrate specificities and with enhanced stability and resistance

to oxidative agents (Maurer, 2004; Schallmey et al., 2004). Additionally, engineering of signal peptides led to significantly increased yields of secreted protease (Degering et al., 2010). (iii) The bioprocess itself has also been the target of optimization efforts concerning the oxygen transfer rate (Çalık et al., 1999, 2000a, 2000b), pH value (Hornbaek et al., 2004a; Çalık et al., 2002), inoculum quality (Hornbaek et al., 2004b), initial glucose concentration (Çalık et al., 2003a), and the fermentation medium composition (Enshasy and Azaly, 2008; Çalık et al., 2003b). (iv) Furthermore, purification techniques have been improved or newly developed to allow efficient and resource-saving downstream processing (Gupta et al., 2002b).

All approaches targeting the internal optimization of the organism suffer from the drawback that they do not account for the complex interactions throughout the entire pathway network (Becker and Wittmann, 2012). As optimization on a global scale has shown tremendous improvements of productivity (Leprince et al., 2012; Park et al., 2008) such an approach could also allow increasing the performance of *B. licheniformis*. To achieve this goal, it is necessary to gain profound knowledge of the cellular processes during protease production, for example by studying the proteome and transcriptome of the cellular community at all stages of the fermentation process. Therefore, one major focus of this thesis was the analysis of the *B. licheniformis* transcriptome during an industrial fermentation process. To introduce this topic, an overview of the current knowledge in the fields of transcriptome analysis and post-transcriptional regulation will be given in the following chapter.

A.2 The bacterial transcriptome

The term “transcriptome” was initially introduced in 1996 and describes the complete set of transcripts in a cell at a specific developmental state or physiological condition (Piétu et al., 1999). It usually consists of *de novo* synthesized and post-transcriptionally modified messenger RNA (mRNA) and non-coding RNA species like rRNA, tRNA, and regulatory RNA (van Vliet, 2010).

A.2.1 Post-transcriptional regulation

Over the past decades, the regulation of bacterial transcription was thought to be quite well understood and generally considered to be accomplished by the interaction of different RNA polymerase σ factors and a variety of transcriptional repressors and activators with the DNA

strand. In contrast, the last decade has revealed plenty additional mechanisms (Sesto et al., 2013). On DNA level, methylation patterns, nucleoid compaction, RNA polymerase-associated regulatory proteins, and operon-internal promoters have been shown to have an impact on transcriptional regulation (Güell et al., 2011). Furthermore, it is now known that post-transcriptional regulation has a fundamental part in the organization of gene expression and can affect transcription as well as translation (Güell et al., 2009; Marguerat and Bähler, 2010). In general, post-transcriptional control relies on regulatory RNAs either located in untranslated regions (UTRs) of mRNA transcripts (A.2.2) or transcribed as independent, non-coding RNAs acting on their targets by base-pairing mechanisms (A.2.2.1 & A.2.2.2).

A.2.2 Regulatory RNAs localized in untranslated regions

One main mechanism of post-transcriptional and translational regulation is located in the 5'untranslated regions of mRNA transcripts. Several 5'UTRs harbor control systems that are

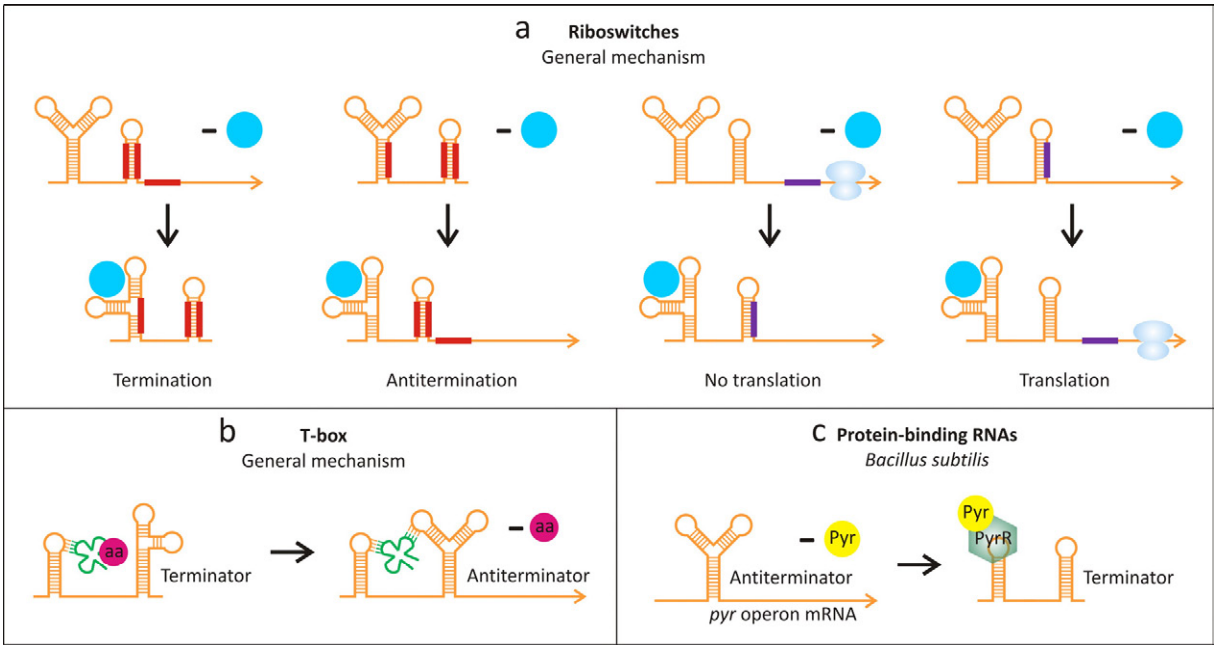


Figure 1 Mechanisms of 5'UTR-intrinsic regulatory elements - Part A
(a) Riboswitches. Upon metabolite (blue) binding to the sensor domain, transcriptional regulation is achieved by the formation of terminating or antiterminating stem loops by the *expression platform* domain (second stem loop). Translational regulation is achieved by different stem loop structures which either release or bind the ribosome binding site (purple; Breaker, 2012; Romby and Charpentier, 2010). **(b) T-boxes.** The formation of a terminator is forced upon the binding of a charged tRNA (green & magenta), whereas the binding of an uncharged tRNA (green) induces the formation of an antiterminating stem loop to enable the transcription of aminoacyl-tRNA ligase and related genes (Green et al., 2010; Gutiérrez-Preciado et al., 2009; Mäder et al., 2012). **(c) Protein-binding RNAs.** During conditions of high pyrimidine availability, the regulatory protein PyrR binds to the RNA and triggers the formation of an intrinsic terminator to halt further transcription of the operon (Winkler, 2012).
The components of the figure were modified from Waters and Storz (2009), Green et al. (2010), and Winkler (2012).

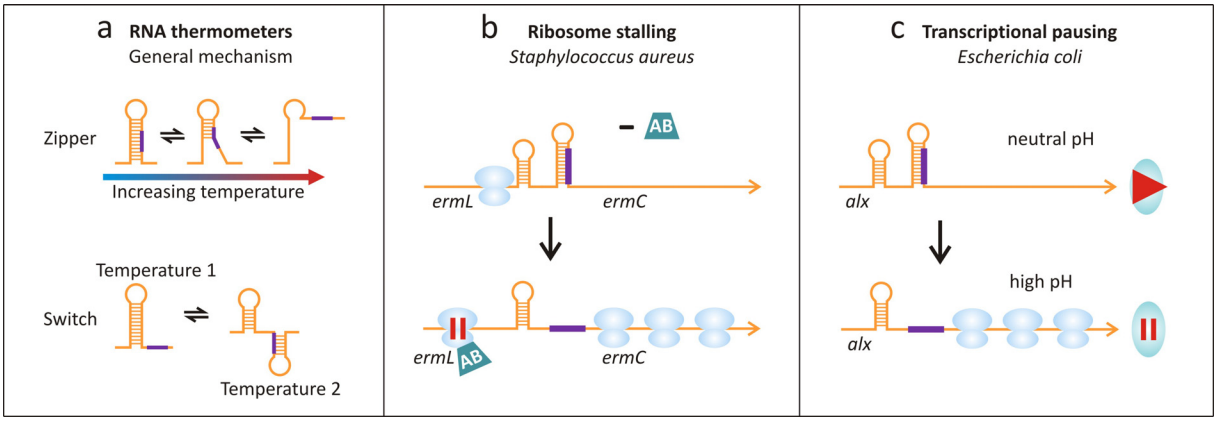


Figure 2 Mechanisms of 5'UTR-intrinsic regulatory elements - Part B
Ribosome binding sites are indicated in purple. **(a) RNA thermometers.** RNA-based thermoregulation of bacterial genes is accomplished by translation regulation by two different mechanisms: RNA zippers and RNA switches (Dethoff et al., 2012; Kortmann and Narberhaus, 2012). **(b) Ribosome stalling.** The ribosome induces the formation of a stem loop that sequesters the RBS located upstream of the erythromycin resistance gene *ermC* during translation of ErmL. When erythromycin is present, the ribosome stalls during translation of ErmL. Consequently, the stem loop which sequesters the RBS is not formed and erythromycin resistance can be developed (Cruz-Vera et al., 2011; Ramu et al., 2009). **(c) Transcriptional pausing.** Under conditions of neutral pH, the leader peptide of *alx* forms a structure, which sequesters the RBS and inhibits translation. At alkaline conditions, the RNA polymerase pauses during the transcription of the leader peptide and thereby promotes the development of a RBS-releasing conformation (Nechooshtan et al., 2009).
The components of the figure were modified from Kortmann and Narberhaus (2012), Cruz-Vera et al. (2011), and Nechooshtan et al. (2009).

based on *cis*-acting regulatory elements capable of adopting different RNA structures. Even though it is not always possible to make a clear distinction between all types of *cis*-acting regulators (Breaker, 2011), the elements can roughly be divided into the six classes of riboswitches, T-boxes, protein-binding RNAs, RNA thermometers, ribosome stalling and transcriptional pausing (Figures 1&2; Cruz-Vera et al., 2011; Dethoff et al., 2012; Green et al., 2010; Gutiérrez-Preciado et al., 2009; Kortmann and Narberhaus, 2012; Landick, 2006; Mäder et al., 2012; Nechooshtan et al., 2009; Ramu et al., 2009; Winkler, 2012). Since they all rely on similar 5'UTR-internal structural rearrangements with analogous effects on transcription or translation, only the well-studied group of riboswitches will be exemplarily introduced in more detail.

The term "riboswitch" is exclusively used for RNA structures that bind metabolites or inorganic ions without the involvement of any protein (Breaker, 2011; Serganov and Nudler, 2013). They consist of two functional domains: the sensor domain is a highly conserved structure that specifically recognizes defined ligands, whereas the *expression platform* domain enables regulation of the downstream coding sequence (Breaker, 2012; Romby and Charpentier, 2010). Riboswitches have four functionalities (Figure 1a): either promoting or terminating transcription or translation (Waters and Storz, 2009). Furthermore, some ribo-

switches with rare mechanisms were identified, including dual control of transcription and translation (Serganov and Nudler, 2013), self-cleaving ribozyme action (Ferré-D'Amaré, 2010), and tandem riboswitches containing at least two sensor domains (Welz and Breaker, 2007). It has also been shown that after transcription termination, a *cis*-regulatory element can act in *trans* as sRNA (A.2.2.1; Loh et al., 2009; Serganov and Nudler, 2013).

In contrast, the less understood 3'UTRs of bacterial mRNAs mainly seem to contain transcription terminators aiding the protection against RNase-mediated degradation of the transcript (Gripenland et al., 2010). As many long 3'UTRs extend beyond their intrinsic terminator structures, regulatory effects of these long transcripts are also assumed. They might, for example, trigger mRNA localization or the regulation of adjacent genes located on the opposite strand (A.2.2.2; Gripenland et al., 2010; Sorek and Cossart, 2010).

A.2.2.1 *Trans*-encoded small RNAs

The most extensively studied group of regulatory base-pairing RNAs, the *trans*-encoded base-pairing small RNAs (sRNAs), range in size from 50 to 400 nt and are mainly involved in adjusting the cell to environmental changes (Balasubramanian and Vanderpool, 2013; Desnoyers et al., 2013). They are transcribed from a chromosomal locus distal, hence in *trans*, to their target RNA (Brantl, 2009). It should be mentioned that most, but not all sRNAs are strictly "non-coding", as some regulatory RNAs also encoding small peptides with complementary functions have been identified (Gottesman and Storz, 2011; Vanderpool et al., 2011).

In general, *trans*-encoded sRNAs interact with their target RNAs by imperfect base-pairing, triggered by the binding of a seed region of at least six contiguous nucleotides to the target (Gottesman and Storz, 2011; Waters and Storz, 2009). Upon the first contact of the two RNA molecules, additional base pairs can form, often in conjunction with a rearrangement of the RNA structure (Storz et al., 2011). Until today, various regulatory mechanisms of *trans*-encoded sRNAs were elucidated, and more are expected to be discovered (Desnoyers et al., 2013). In Gram-positive bacteria, five mechanisms are known so far (Figure 3a-e): upon binding of the sRNA, inhibition of translation can be achieved by structural changes downstream of the ribosome binding site (RBS) or by direct blocking of the RBS coupled with mRNA degradation. Instead, translation activation is induced by mRNA stabilization, mRNA processing or revelation of the RBS (Brantl, 2012a, 2012b). In Gram-negative bacteria, additional mechanisms have been observed, including independent direct blocking of

the RBS or mRNA degradation, blocking of ribosome standby sites and translation enhancers, and differential degradation of polycistronic mRNAs (Brantl, 2012b; Desnoyers et al., 2013). As the sRNA research on Gram-positive bacteria is still in its infancy, it cannot be excluded that the latter mechanisms also play a role in post-transcriptional regulation in Gram-positive bacteria (Brantl, 2012b).

It is known that some mRNAs, which mainly encode transcription regulators, are targeted by multiple sRNAs (Storz et al., 2011). Furthermore, many sRNAs were shown to regulate not one, but a multitude of genes (Schmiedel et al., 2012; Storz et al., 2011). It is now evident that such sRNA regulons are integrated in global regulatory networks to transduce environmental stimuli in order to orchestrate the cellular response to specific environmental challenges (Beisel and Storz, 2010).

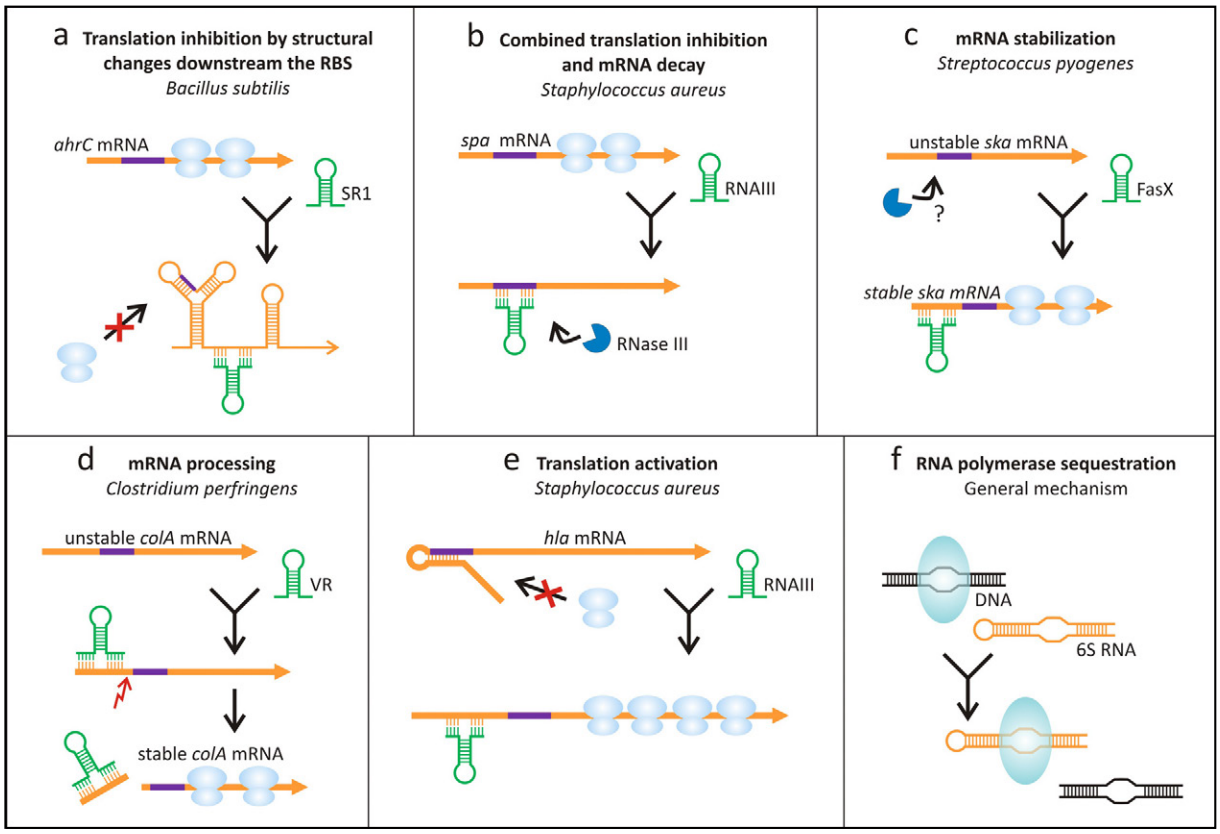


Figure 3 *Trans*-encoded regulatory sRNAs in Gram-positive bacteria
Ribosome binding sites are indicated in purple, ribosomes are light blue and RNA polymerases are turquoise. (**a**) **Translation inhibition by structural changes downstream of the RBS** (Heidrich et al., 2007). (**b**) **Combined translation inhibition and mRNA decay.** First example of the multiple targets of the Staphylococcus aureus sRNA RNAIII (Boisset et al., 2007). (**c**) **mRNA stabilization** (Ramirez-Peña et al., 2010). (**d**) **mRNA processing** (Obana et al., 2010). (**e**) **Translation activation.** Second example of RNAIII (Morfeldt et al., 1995). **(f) RNA polymerase sequestration.** Example of a protein-binding sRNA. The 6S RNA forms a double-stranded hairpin with a critical bubble that mimics the formation of DNA in an open complex promoter. Therefore, an increased level of 6S RNA titrates the σB factor of RNA polymerase and decreases the transcription from resembling promoters (Gottesman and Storz, 2011).
The components of the figure were modified from Waters and Storz (2009) and Brantl (2012a).

Next to the numerous regulatory RNAs that act by base-pairing mechanisms, the group of *trans*-encoded sRNAs also encompasses several protein-binding regulatory sRNAs, which can be divided into two different groups (Romby and Charpentier, 2010). The first group contains protein-binding RNAs (RNase P, tmRNA, Scr), which form ribonucleoprotein complexes with housekeeping functions (Romby and Charpentier, 2010; Waters and Storz, 2009). The second group encompasses CsrB, GlmY and 6S RNA, of which only the latter is found in *B. licheniformis* (Figure 3f). These sRNAs are mimicking the structure of another nucleic acid and thus sequester the respective target protein (Gottesman and Storz, 2011; Romby and Charpentier, 2010).

A.2.2.2 Cis-encoded antisense RNAs

In contrast to the group of *trans*-encoded sRNAs, *cis*-encoded antisense RNAs (asRNAs) are transcribed directly from the opposite strand of their target RNA (Georg and Hess, 2012). Therefore, they are completely complementary to the sense RNA and form complete duplexes between both molecules (Brantl, 2012a).

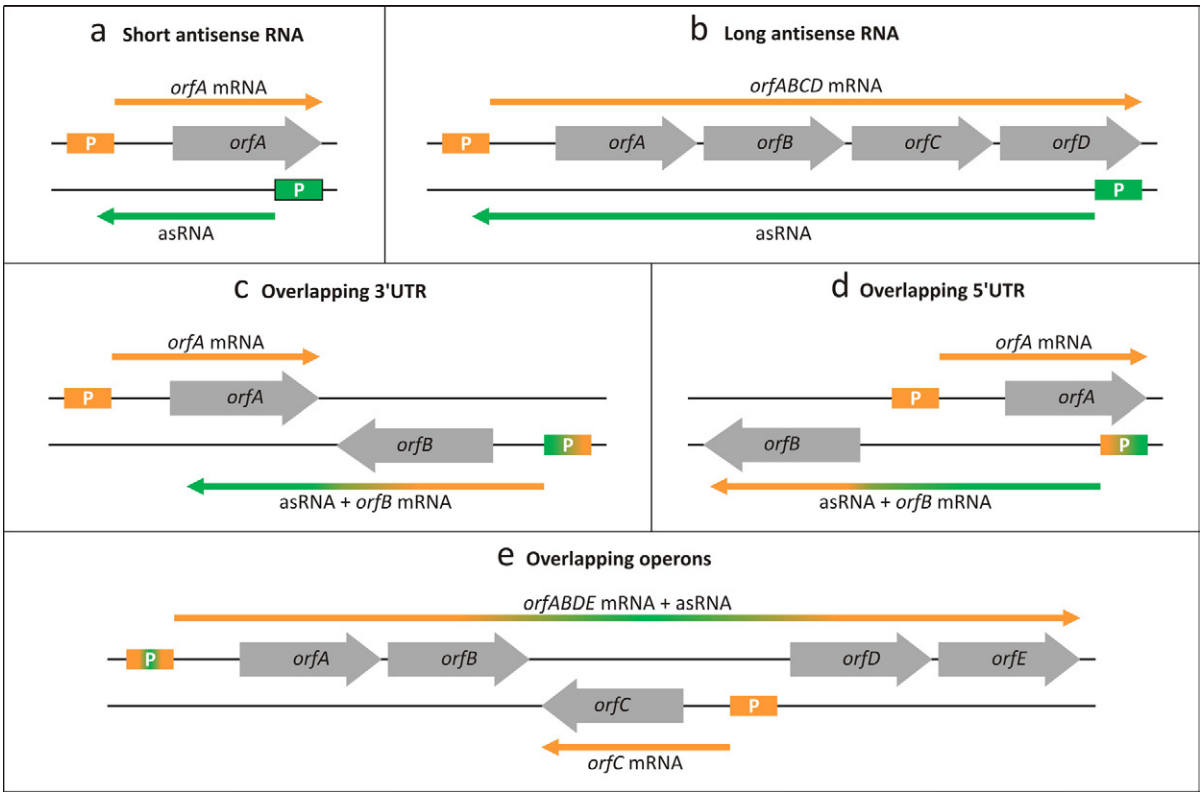


Figure 4 Categories of *cis*-encoded antisense RNAs
(a) Short antisense RNA. (b) Long antisense RNA. (c) Overlapping 3'UTR. (d) Overlapping 5'UTR. (e) Overlapping operons.
The components of the figure were modified from Sesto et al. (2013) and Lasa et al. (2012).

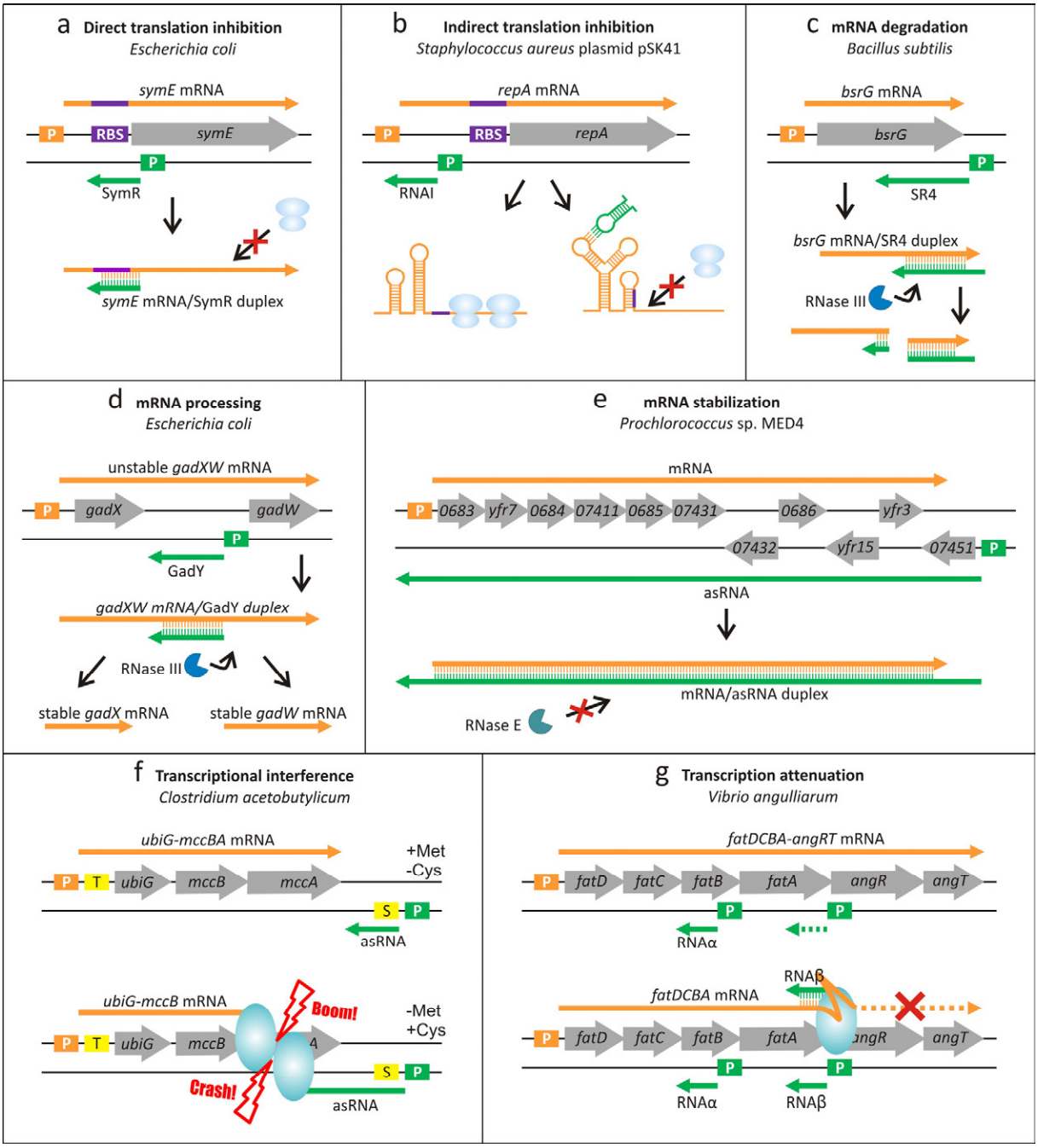


Figure 5 Mechanisms of *cis*-encoded antisense RNAs
Ribosome binding sites are indicated in purple, T- and S-boxes are yellow, ribosomes are light blue and RNA polymerases are turquoise. **(a) Direct translation inhibition** (Kawano et al., 2007)**. (b) Indirect translation inhibition** (Kwong et al., 2006)**. (c) mRNA degradation** (Jahn et al., 2012)**. (d) mRNA processing** (Opdyke et al., 2004, 2011)**. (e) mRNA stabilization** (Stazic et al., 2011)**. (f) Transcriptional interference.** In the presence of methionine cysteine a long asRNA is transcribed, which then induces the collision and subsequent dissociation of the RNA polymerase transcribing the mRNA (André et al., 2008; Georg and Hess, 2011). **(g) Transcription attenuation.** The binding of RNAβ to its complementary mRNA during transcription leads to a structural change and the formation of a stem loop, which forces transcription termination (Stork et al., 2007).
The components of the figure were modified from Sesto et al. (2013), Georg and Hess (2011), Lasa et al. (2012), and Brantl (2012a, 2012b).

In general, *cis*-encoded asRNAs, transcribed from autonomous promoters located on the opposite strand of the sense target, do not encode proteins and vary tremendously in size (Figure 4a&b; Lasa et al., 2012; Sesto et al., 2013). Short asRNAs of 100 to 300 nt have been observed as well as long asRNA, ranging in size up to 7,000 nt (Georg and Hess, 2011; Stazic et al., 2011). AsRNAs can originate from elongated 5' and 3'UTRs of mRNAs that are transcribed in reverse direction of the target mRNA (Figure 4c&d; Georg and Hess, 2011; Sesto et al., 2013). They can also be transcribed from intergenic operon regions, which overlap a single gene encoded on the opposite strand (Figure 4e; Lasa et al., 2012). Hence, *cis*-encoded asRNAs are located on non-coding as well as on coding RNA transcripts.

So far, six different regulatory mechanisms employed by *cis*-encoded asRNAs were determined (Georg and Hess, 2012; Sesto et al., 2013). Inhibition of translation can be achieved by direct or indirect blocking of the RBS (Figure 5a&b; Kiley Thomason and Storz, 2010), the translation process can additionally be affected by degradation (Figure 5c), processing (Figure 5d) or stabilization (Figure 5e) of the target mRNA (Brantl, 2007, 2012b; Georg and Hess, 2012). Regulatory effects that target transcription are either based on RNA polymerase interferences (Figure 5f) or transcription attenuation (Figure 5g; Georg and Hess, 2011; Kiley Thomason and Storz, 2010). In addition to these regulatory effects on protein-coding genes, some asRNAs overlap and thereby affect regulatory sRNAs (Sesto et al., 2013).

A.2.3 Analysis of the bacterial transcriptome

The study of the transcriptome, referred to as transcriptomics, is necessary for the interpretation of functional genomic elements and the revelation of the RNA content of cells in different developmental stages and under different conditions (Wang et al., 2009). In order to achieve these goals, the key aims of transcriptomics are: (i) cataloging of all species of RNA transcripts, (ii) determination of transcriptional structures of genes, in terms of transcription start sites (TSS), 5′ and 3′ends, or putative post-transcriptional modifications, and (iii) quantification of changes in the expression levels of each transcript (Wang et al., 2009).

Over the past two decades, several methods to infer and quantify the transcriptome were developed. First, Sanger sequencing of Expressed Sequence Tags (EST; Adams et al., 1991) and full-length cDNA (Strausberg, 1999) as well as Serial Analysis of Gene Expression (SAGE; Velculescu et al., 1995) and Massively Parallel Signature Sequencing (MPSS; Brenner et al., 2000) have granted insights into the transcribed regions of a genome. Other

ways to address quantification are the use of different hybridization-based array techniques. After the development of ORF-based microarrays in the mid 1990s (Brown and Botstein, 1999; Schena et al., 1995), nowadays whole genome tiling arrays (Mockler et al., 2005; Yamada et al., 2003) are the method of choice for high-throughput transcriptome analyses (Mäder et al., 2011; Nicolas et al., 2012). However, the onset of new sequencing technologies during the past years has allowed the emergence of further methods.

After the unveiling of the human genome (Lander et al., 2001) by classic Sanger chain termination sequencing (first-generation sequencing) had produced costs of over three billion US$ in the 1990s and early 2000s (Schadt et al., 2010), the future needs for advanced sequencing technologies became obvious (Groß, 2011; Schloss, 2008). As a result, second-generation sequencers like 454 GS20, Solexa GA and SOLiD were released from 2005 on (MacLean et al., 2009; Rothberg and Leamon, 2008). Albeit the underlying biochemistries of these sequencing machines differ, all share an underlying PCR amplification step and require extensive washing steps (Liu et al., 2012; Schadt et al., 2010). These features are drawbacks in terms of sequence reliability and processing time, therefore, sequencing approaches overcoming these drawbacks are considered as third-generation sequencing systems, but are not commonly employed yet (Mason and Elemento, 2012; Schadt et al., 2010).
Next to genome sequencing, metagenomics, and the analysis of methylation patterns and single nucleotide polymorphisms (MacLean et al., 2009), one broad scope of second-generation sequencing is the analysis of transcriptomes by RNA sequencing (RNA-Seq; van Vliet, 2010; Wang et al., 2009). Since its initial application on a bacterial transcriptome in 2009 (Passalacqua et al., 2009; Yoder-Himes et al., 2009), this method has enabled new insights into the regulation and composition of the bacterial transcriptome (Güell et al., 2011).

During preparation of RNA-Seq libraries for second-generation sequencing, as shown in Figure 6, several challenges have to be overcome. **(i) RNA isolation.** High quality RNA is required. Ideally, the RNA should be highly pure, not degraded, include all RNA species and keep its natural proportions (Mäder et al., 2011). Hence, isolation methods have to be adapted to the respective organism and the specific aims of the study. Special attention must be paid to ensure that the protocol allows the isolation of RNA transcripts <200 nt (Koo et al., 2011). **(ii) rRNA depletion.** Bacterial RNA preparations usually contain more than 80% rRNA (Filiatrault, 2011; van Vliet, 2010). To increase the amount of mRNA and other RNA species of interest in the sequencing library, many protocols rely on the depletion of rRNA (Perkins et

al., 2009; Vivancos et al., 2010). However, as sequencing capacities are expected to rise tremendously upon developments in third-generation sequencing, the depletion step may not be necessary anymore in future protocols (Croucher and Thomson, 2010). **(iii) Strand specificity.** It is now standard to keep the strand specificity of the RNA during library preparation to gain information valuable for accurate determination of gene expression and evaluation of all kinds of antisense transcription (Filiatrault, 2011). Next to the approach described in Figure 6, several further methods were developed (Croucher et al., 2009; Filiatrault et al., 2010; Marguerat and Bähler, 2010). Until today, none of those methods became prevalent, as every method has its own specific advantages as well as drawbacks and inherent biases.

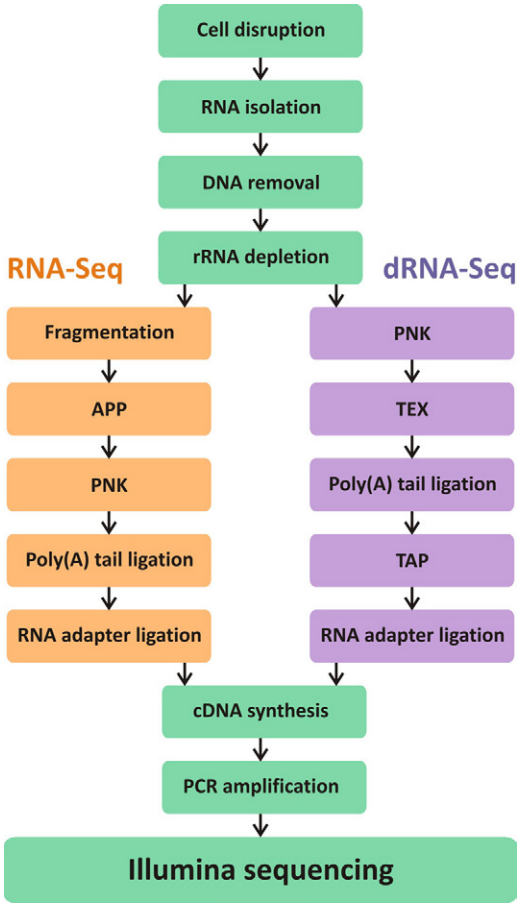


Figure 6 Library preparation for whole transcriptome and differential RNA-Seq
RNA-Seq. After establishing monophosphorylated 5'ends by antarctic phosphatase (APP) and polynucleotide kinase (PNK) treatment, followed by ligation of a 3'poly(A) tail and a 5'RNA adapter, whole transcriptome RNA-Seq allows the sequencing of all transcripts available after the removal of ribosomal RNAs. **dRNA-Seq.** In contrast, the dRNA-Seq procedure allows the enrichment of 5'triphosphorylated fragments originating from 5'ends of unprocessed and non-degraded transcripts (Sharma et al., 2010): treatment with PNK and 5'phosphate-dependent exonuclease (TEX) promotes the degradation of all non- and monophosphorylated transcripts. The subsequent pyrophosphatase (TAP) step generates 5'monophosphorylated fragments to enable ligation of the 5'RNA adapter. A detailed description of the process will be given in Chapter B.

RNA-Seq does not only bear the possibility to analyze whole transcriptomes (Dötsch et al., 2012; Høvik et al., 2012), but can also be utilized to specifically study small RNAs (Arnvig et al., 2011; Liu et al., 2009) or protein-RNA interactions (Gatewood et al., 2012; Sittka et al., 2008). Furthermore, techniques to directly sequence the 5'region of RNA molecules have been developed (Fouquier d'Hérouel et al., 2011; Sharma et al., 2010). They allow the identification of putative transcription start sites in order to aid the annotation of transcript boundaries and the prediction of promoter sites (Albrecht et al., 2010; Mitschke et al., 2011). The first method to accomplish this goal - differential RNA-Seq (dRNA-Seq) - was introduced in 2010 by Sharma et al. (2010) and is outlined in Figure 6.

Although tiling arrays are a mature technology, which still is more efficient than RNA-Seq, they have several drawbacks suggesting RNA-Seq as advantageous technique (Mäder et al., 2011). Firstly, RNA-Seq has a very low, if any, background signal because sequences can be unambiguously mapped to the genome and no cross-hybridization can occur (van Vliet, 2010; Wang et al., 2009). Secondly, since the measurement cannot be saturated, RNA-Seq has no upper limit of quantification. Hence, transcripts with a high dynamic range spanning several orders of magnitude can be detected (Vivancos et al., 2010; Wang et al., 2009). Both effects enhance the accuracy of RNA-Seq regarding quantification of gene expression (van Vliet, 2010; Wang et al., 2009). The accuracy is also elevated by the provided single-base resolution (Güell et al., 2011; Sorek and Cossart, 2010). Moreover, RNA-Seq is highly reproducible (Wang et al., 2009) and requires only small amounts of input RNA (Mutz et al., 2013), particularly when considering upcoming developments in third-generation sequencing.

A.3 Aim of the thesis

The main goals of the project, in which this thesis is embedded, are to gain profound knowledge of the regulatory processes in wild type and production strains of *B. licheniformis* and, based on this understanding, to identify targets, which can be modified to improve the efficiency of industrial fermentation processes.

In this framework, the aims of the present thesis were the analysis of the transcriptome and proteome of *B. licheniformis* DSM13 sampled from different time points of an industry-oriented fermentation for the production of detergent proteases. After whole transcriptome RNA-Seq and differential RNA-Seq were accomplished, the obtained data were assessed for the determination of RNA-based regulatory features, transcription start points and operon boundaries as well as for an improvement of the functional annotation of the *B. licheniformis* genome sequence. Furthermore, gene expression and protein abundance data allowed the analysis of pathways involved in carbon and nitrogen metabolism, stress response, protein secretion and cell differentiation under fermentation conditions. All gained data were evaluated with a special emphasis to putative targets for bioprocess optimization. Further goals of the thesis were the development of prediction algorithms to enable the comprehensive analysis of the generated transcriptome data, as well as the sequencing and annotation of the genome of *Geobacillus* sp. GHH01, another biotechnologically promising member of the genus *Bacillus*.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656.

Agerholm, J.S., Jensen, N.E., Giese, S.B., and Jensen, H.E. (1997). A preliminary study on the pathogenicity of *Bacillus lichenformis* bacteria in immunodepressed mice. APMIS *105*, 48–54.

Albrecht, M., Sharma, C., Reinhardt, R., Vogel, J., and Rudel, T. (2010). Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res *38*, 868–877.

André, G., Even, S., Putzer, H., Burguière, P., Croux, C., Danchin, A., Martin-Verstraete, I., and Soutourina, O. (2008). S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. Nucleic Acids Res *36*, 5955–5969.

Arnvig, K.B., Comas, I., Thomson, N.R., Houghton, J., Boshoff, H.I., Croucher, N.J., Rose, G., Perkins, T.T., Parkhill, J., Dougan, G., et al. (2011). Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. PLoS Pathog *7*, e1002342.

Ash, C., Farrow, J.A.E., Dorsch, M., Stackebrandt, E., and Collins, M.D. (1991). Comparative Analysis of *Bacillus antracis*, *Bacillus cereus*, and Related Species on the Basis of Reverse Transcriptase Sequencing of 16S rRNA. Int J Syst Bacteriol *41*, 343–346.

Balasubramanian, D., and Vanderpool, C.K. (2013). New developments in post-transcriptional regulation of operons by small RNAs. RNA Biol *10*, 337–341.

Becker, J., and Wittmann, C. (2012). Systems and synthetic metabolic engineering for amino acid production - the heartbeat of industrial strain development. Curr Opin Biotechnol *23*, 718–726.

Beisel, C.L., and Storz, G. (2010). Base pairing small RNAs and their roles in global regulatory networks. FEMS Microbiol Rev *34*, 866–882.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y., Jacquier, A., et al. (2007). *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Genes Dev *21*, 1353–1366.

Borgmeier, C., Bongaerts, J., and Meinhardt, F. (2012). Genetic analysis of the *Bacillus licheniformis degSU* operon and the impact of regulatory mutations on protease production. J Biotechnol *159*, 12–20.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr Opin Microbiol *10*, 102–109.

Brantl, S. (2009). Bacterial chromosome-encoded small regulatory RNAs. Future Microbiol *4*, 85–103.

Brantl, S. (2012a). Acting antisense: plasmid- and chromosome-encoded sRNAs from Gram-positive bacteria. Future Microbiol *7*, 853–871.

Brantl, S. (2012b). Small Regulatory RNAs (sRNAs): Key Players in Prokaryotic Metabolism, Stress Response, and Virulence. In Regulatory RNAs, B. Mallick, and Z. Ghosh, eds. (Berlin, Heidelberg: Springer), pp. 73–109.

Breaker, R.R. (2011). Prospects for riboswitch discovery and analysis. Mol Cell *43*, 867–879.

Breaker, R.R. (2012). Riboswitches and the RNA world. Cold Spring Harb Perspect Biol *4*, a003566.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol *18*, 630–634.

Brown, P.O., and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. Nat Genet *21*, 33–37.

Çalık, P., Çalık, G., Takaç, S., and Ozdamar, T.H. (1999). Metabolic flux analysis for serine alkaline protease fermentation by *Bacillus licheniformis* in a defined medium: effects of the oxygen transfer rate. Biotechnol Bioeng *64*, 151–167.

Çalık, P., Çalık, G., Takaç, S., and Özdamar, T. (2000a). Metabolic flux analyses for serine alkaline protease production. Enzyme Microb Tech *27*, 793–805.

Çalık, P., Çalık, G., and Ozdamar, T.H. (2000b). Oxygen-transfer strategy and its regulation effects in serine alkaline protease production by *Bacillus licheniformis*. Biotechnol Bioeng *69*, 301–311.

Çalık, P., Bilir, E., Çalık, G., and Özdamar, T.H. (2002). Influence of pH conditions on metabolic regulations in serine alkaline protease production by *Bacillus licheniformis*. Enzyme Microb Tech *31*, 685–697.

Çalık, P., Tomlin, G.C., Oliver, S.G., and Özdamar, T.H. (2003a). Overexpression of a serine alkaline protease gene in *Bacillus licheniformis* and its impact on the metabolic reaction network. Enzyme Microb Tech *32*, 706–720.

Çalık, P., Çelik, E., Telli, İ.E., Oktar, C., and Özdemir, E. (2003b). Protein-based complex medium design for recombinant serine alkaline protease production. Enzyme Microb Tech *33*, 975–986.

Croucher, N.J., and Thomson, N.R. (2010). Studying bacterial transcriptomes using RNA-seq. Curr Opin Microbiol *13*, 619–624.

Croucher, N.J., Fookes, M.C., Perkins, T.T., Turner, D.J., Marguerat, S.B., Keane, T., Quail, M.A., He, M., Assefa, S., Bähler, J., et al. (2009). A simple method for directional transcriptome sequencing using Illumina technology. Nucleic Acids Res *37*, e148.

Cruz-Vera, L.R., Sachs, M.S., Squires, C.L., and Yanofsky, C. (2011). Nascent polypeptide sequences that influence ribosome function. Curr Opin Microbiol *14*, 160–166.

Davidson, F., Seon-Yi, C., and Stanley-Wall, N. (2012). Selective Heterogeneity in Exoprotease Production by *Bacillus subtilis*. PLoS One *7*, e38574.

Degering, C., Eggert, T., Puls, M., Bongaerts, J., Evers, S., Maurer, K.-H., and Jaeger, K.-E. (2010). Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. Appl Environ Microbiol *76*, 6370–6376.

Desnoyers, G., Bouchard, M.-P., and Massé, E. (2013). New insights into small RNA-dependent translational regulation in prokaryotes. Trends Genet *29*, 92–98.

Dethoff, E.A., Chugh, J., Mustoe, A.M., and Al-Hashimi, H.M. (2012). Functional complexity and regulation through RNA dynamics. Nature *482*, 322–330.

Dubnau, D., and Losick, R. (2006). Bistability in bacteria. Mol Microbiol *61*, 564–572.

Dötsch, A., Eckweiler, D., Schniederjans, M., Zimmermann, A., Jensen, V., Scharfe, M., Geffers, R., and Häussler, S. (2012). The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. PLoS One *7*, e31092.

Enshasy, E., and Azaly, E. (2008). Optimization Of The Industrial Production Of Alkaline Protease By *Bacillus licheniformis* In Different Production Scales. Aus J Basic Appli Sci *2*, 583–593.

Euzéby, J.P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. Int J Syst Bacteriol *47*, 590–592.

Ferré-D'Amaré, A.R. (2010). The *glmS* ribozyme: use of a small molecule coenzyme by a gene-regulatory RNA. Q Rev Biophys *43*, 423–447.

Filiatrault, M.J. (2011). Progress in prokaryotic transcriptomics. Curr Opin Microbiol *14*, 579–586.

Filiatrault, M.J., Stodghill, P.V., Bronstein, P.A., Moll, S., Lindeberg, M., Grills, G., Schweitzer, P., Wang, W., Schroth, G.P., Luo, S., et al. (2010). Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. J Bacteriol *192*, 2359–2372.

Fouquier d'Hérouel, A., Wessner, F., Halpern, D., Ly-Vu, J., Kennedy, S.P., Serror, P., Aurell, E., and Repoila, F. (2011). A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen *Enterococcus faecalis*. Nucleic Acids Res *39*, e46.

Galanos, J., Perera, S., Smith, H., Neal, D.O., Sheorey, H., and Waters, M.J. (2003). Bacteremia Due to Three *Bacillus* Species in a Case of Munchausen ' s Syndrome. J Clin Microbiol *41*, 2247–2248.

Gatewood, M.L., Bralley, P., Weil, M.R., and Jones, G.H. (2012). The transcriptome of *Streptomyces coelicolor*: RNA-seq and RNA immunoprecipitation identify substrates for RNase III. J Bacteriol *149*, 2228.

Georg, J., and Hess, W.R. (2011). *cis*-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol R *75*, 286–300.

Georg, J., and Hess, W.R. (2012). Natural Antisense Transcripts in Bacteria. In Regulatory RNAs, B. Mallick, and Z. Ghosh, eds. (Berlin, Heidelberg: Springer), pp. 95–108.

Gottesman, S., and Storz, G. (2011). Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol *3,* a0003798.

Green, N.J., Grundy, F.J., and Henkin, T.M. (2010). The T box mechanism: tRNA as a regulatory molecule. FEBS Letters *584*, 318–324.

Gripenland, J., Netterling, S., Loh, E., Tiensuu, T., Toledo-Arana, A., and Johansson, J. (2010). RNAs: regulators of bacterial virulence. Nat Rev Microbiol *8*, 857–866.

Groß, M. (2011). Genomsequenzierer. Die dritte Generation. Chemie in Unserer Zeit *45*, 184–187.

Gupta, R., Beg, Q.K., and Lorenz, P. (2002a). Bacterial alkaline proteases: molecular approaches and industrial applications. Appl Microbiol Biotechnol *59*, 15–32.

Gupta, R., Beg, Q.K., Khan, S., and Chauhan, B. (2002b). An overview on fermentation, downstream processing and properties of microbial alkaline proteases. Appl Microbiol Biotechnol *60*, 381–395.

Gutiérrez-Preciado, A., Henkin, T.M., Grundy, F.J., Yanofsky, C., and Merino, E. (2009). Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. Microbiol Mol Biol R *73*, 36–61.

Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. Science *326*, 1268–1271.

Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. (2011). Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nat Rev Microbiol *9*, 658–669.

Haydushka, I.A., Markova, N., Kirina, V., and Atanassova, M. (2012). Recurrent Sepsis Due To *Bacillus Licheniformis*. J Glob Infect Dis *4*, 82–83.

Heidrich, N., Moll, I., and Brantl, S. (2007). In vitro analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA. Nucleic Acids Res *35*, 4331–4346.

Hintz, M. (2003). Untersuchungen zur Subtilisin-Produktion mit *Bacillus licheniformis* und Konstruktion eines alternativen Selektionssystems. *PhD thesis*. Institut für molekulate Enzymtechnologie, Heinrich-Heine-Universität Düsseldorf, Germany.

Hoffmann, K., Wollherr, A., Larsen, M., Rachinger, M., Liesegang, H., Ehrenreich, A., and Meinhardt, F. (2010). Facilitation of direct conditional knockout of essential genes in *Bacillus*

licheniformis DSM13 by comparative genetic analysis and manipulation of genetic competence. Appl Environ Microbiol *76*, 5046–5057.

Hornbaek, T., Jakobsen, M., Dynesen, J., and Nielsen, A.K. (2004a). Global transcription profiles and intracellular pH regulation measured in *Bacillus licheniformis* upon external pH upshifts. Arch Microbiol *182*, 467–474.

Hornbaek, T., Nielsen, A.K., Dynesen, J., and Jakobsen, M. (2004b). The effect of inoculum age and solid versus liquid propagation on inoculum quality of an industrial *Bacillus licheniformis* strain. FEMS Microbiol Lett *236*, 145–151.

Høvik, H., Yu, W.-H., Olsen, I., and Chen, T. (2012). Comprehensive transcriptome analysis of the periodontopathogenic bacterium *Porphyromonas gingivalis* W83. J Bacteriol *194*, 100–114.

Jahn, N., Preis, H., Wiedemann, C., and Brantl, S. (2012). BsrG/SR4 from *Bacillus subtilis*--the first temperature-dependent type I toxin-antitoxin system. Mol Microbiol *83*, 579–598.

Kawano, M., Aravind, L., and Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol Microbiol *64*, 738–754.

Kiley Thomason, M., and Storz, G. (2010). Bacterial antisense RNAs: how many are there, and what are they doing? Annu Rev Genet *44*, 167–188.

Kirk, O. (2002). Industrial enzyme applications. Curr Opin Biotechnol *13*, 345–351.

Koo, J.T., Alleyne, T.M., Schiano, C.A., Jafari, N., and Lathem, W.W. (2011). Global discovery of small RNAs in *Yersinia pseudotuberculosis* identifies *Yersinia*-specific small, noncoding RNAs required for virulence. Proc Natl Acad Sci USA *108*, E709–17.

Kortmann, J., and Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers and switches. Nat Rev Microbiol *10*, 255–265.

Kuchina, A., Espinar, L., Çağatay, T., Balbin, A.O., Zhang, F., Alvarado, A., Garcia-Ojalvo, J., and Süel, G.M. (2011). Temporal competition between differentiation programs determines cell fate choice. Mol Syst Biol *7*, 557.

Kwong, S.M., Skurray, R.A., and Firth, N. (2006). Replication control of staphylococcal multiresistance plasmid pSK41: an antisense RNA mediates dual-level regulation of Rep expression. J Bacteriol *188*, 4404–4412.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans *34*, 1062–1066.

Lasa, I., Toledo-Arana, A., and Gingeras, T.R. (2012). An effort to make sense of antisense transcription in bacteria. RNA Biol *9*, 1039–1044.

Leprince, A., van Passel, M.W.J., and Martins dos Santos, V.A.P. (2012). Streamlining genomes: toward the generation of simplified and stabilized microbial systems. Curr Opin Biotechnol *23*, 651–658.

Liu, J., Livny, J., Lawrence, M., Kimball, M., Waldor, M., and Camilli, A. (2009). Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. Nucleic Acids Res *37*, e46.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J Biomed Biotechnol *2012*, 251364.

Logan, N.A., and de Vos, P. (2008). Genus I. Bacillus Cohn 1872, 174^{AL}. In Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes, P. de Vos, G.M. Garrity, D. Jones, N.R.

Krieg, W. Ludwig, F.A. Rainey, K.-H. Schleifer, and W.B. Whitman, eds. (Dordrecht Heidelberg London New York:Springer), pp. 21-128.

Loh, Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. Cell *139*, 770–779.

Lopez, D., Vlamakis, H., and Kolter, R. (2009). Generation of multiple cell types in *Bacillus subtilis*. FEMS Microbiol Rev *33*, 152–163.

López, D., and Kolter, R. (2010). Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. FEMS Microbiol Rev *34*, 134–149.

MacLean, D., Jones, J., and Studholme, D. (2009). Application of "next-generation" sequencing technologies to microbial genetics. Nat Rev Microbiol *7*, 287–296.

Madslien, E.H., Olsen, J.S., Granum, P.E., and Blatny, J.M. (2012). Genotyping of *B. licheniformis* based on a novel multi-locus sequence typing (MLST) scheme. BMC Microbiol *12*, 230.

Marguerat, S., and Bähler, J. (2010). RNA-seq: from technology to biology. Cell Mol Life Sci *67*, 569–579.

Mason, C.E., and Elemento, O. (2012). Faster sequencers, larger datasets, new challenges. Genome Biol *13*, 314.

Maurer, K.-H. (2004). Detergent proteases. Curr Opin Biotechnol *15*, 330–334.

Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voß, B., Steglich, C., Wilde, A., Vogel, J., et al. (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. Proc Natl Acad Sci USA *108*, 1–6.

Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. Genomics *85*, 1–15.

Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol *24*, 22–30.

Mäder, U., Nicolas, P., Richard, H., Bessières, P., and Aymerich, S. (2011). Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. Curr Opin Biotechnol *22*, 32–41.

Mäder, U., Schmeisky, A.G., Flórez, L.A., and Stülke, J. (2012). SubtiWiki--a comprehensive community resource for the model organism *Bacillus subtilis*. Nucleic Acids Res *40*, D1278–87.

Morfeldt, E., Taylor, D., von Gabain, A., and Arvidson, S. (1995). Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded. EMBO J *14*, 4569–4577.

Nahrstedt, H., Waldeck, J., Gröne, M., Eichstädt, R., Feesche, J., and Meinhardt, F. (2005). Strain development in *Bacillus licheniformis*: Construction of biologically contained mutants deficient in sporulation and DNA repair. J Biotechnol *119*, 245–254.

Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V.V., Osipov, G.A., Belyaev, S.S., et al. (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen . nov ., sp . nov . and *Geobacillus uzenensis* sp . nov . from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatu, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G.thermoglucosidasius* and *G. thermodenitrificans*. Int J Syst Evol Micr *51*, 433–446.

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., and Altuvia, S. (2009). A pH-responsive riboregulator. Gene Dev *23*, 2650–2662.

Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., et al. (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. Science *335*, 1103–1106.

Obana, N., Shirahama, Y., Abe, K., and Nakamura, K. (2010). Stabilization of *Clostridium perfringens* collagenase mRNA by VR-RNA-dependent cleavage in 5' leader sequence. Mol Microbiol *77*, 1416–1428.

Opdyke, J.A., Kang, J., and Storz, G. (2004). GadY , a Small-RNA Regulator of Acid Response Genes in *Escherichia coli*. J Bacteriol *186*, 6698-705.

Opdyke, J.A., Fozo, E.M., Hemm, M.R., and Storz, G. (2011). RNase III participates in GadY-dependent cleavage of the *gadX-gadW* mRNA. J Mol Biol *406*, 29–43.

Parekh, S., Vinci, V.A., and Strobel, R.J. (2000). Improvement of microbial strains and fermentation processes. Appl Microbiol Biotechnol *54*, 287–301.

Park, J., Lee, S., Kim, T., and Kim, H. (2008). Application of systems biology for bioprocess development. Trends Biotechnol *26*, 404–412.

Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., and Bergman, N.H. (2009). Structure and complexity of a bacterial transcriptome. J Bacteriol *191*, 3203–3211.

Perkins, T., Kingsley, R., Fookes, M., Gardner, P., James, K., Yu, L., Assefa, S., He, M., Croucher, N., Pickard, D., et al. (2009). A Strand-Specific RNA-Seq Analysis of the Transcriptome of the Typhoid Bacillus *Salmonella* Typhi. PLoS Genet *5*, e1000569.

Piétu, G., Mariage-Samson, R., Fayein, N.-A., Matingou, C., Eveno, E., Houlgatte, R., Decraene, C., Vandenbrouck, Y., Tahi, F., Devignes, M.-D., et al. (1999). The Genexpress IMAGE Knowledge Base of the Human Brain Transcriptome: A Prototype Integrated Resource for Functional and Computational Genomics. Genome Res *9*, 195–209.

Rachinger, M. (2010). Stammdesign in *B. licheniformis*. *PhD thesis*. Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Germany.

Ramirez-Peña, E., Treviño, J., Liu, Z., Perez, N., and Sumby, P. (2010). The group A *Streptococcus* small regulatory RNA FasX enhances streptokinase activity by increasing the stability of the *ska* mRNA transcript. Mol Microbiol *78*, 1332–1347.

Ramu, H., Mankin, A., and Vazquez-Laslop, N. (2009). Programmed drug-dependent ribosome stalling. Mol Membr Biol *71*, 811–824.

Rey, M., Ramaiya, P., Nelson, N., and Brody-Karpin, S. (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. Genome Biol *5*, R77.

Romby, P., and Charpentier, E. (2010). An overview of RNAs with regulatory functions in gram-positive bacteria. Cell Mol Life Sci *67*, 217–237.

Rooney, A.P., Price, N.P.J., Ehrhardt, C., Swezey, J.L., and Bannan, J.D. (2009). Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. Int J Syst Evol Micr *59*, 2429–2436.

Rothberg, J., and Leamon, J. (2008). The development and impact of 454 sequencing. Nat Biotechnol *26*, 1117–1124.

Sayem, S.M.A., Manzo, E., Ciavatta, L., Tramice, A., Cordone, A., Zanfardino, A., De Felice, M., and Varcamonti, M. (2011). Anti-biofilm activity of an exopolysaccharide from a sponge-associated strain of *Bacillus licheniformis*. Microb Cell Fact *10*, 74.

Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. Hum Mol Genet *19*, R227–40.

Schallmey, M., Singh, A., and Ward, O. (2004). Developments in the use of *Bacillus* species for industrial production. Can J Microbiol *50*, 1–17.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467–470.

Schloss, J.A. (2008). How to get genomes at one ten-thousandth the cost. Nat Biotechnol *26*, 1113–1115.

Schmiedel, J.M., Axmann, I.M., and Legewie, S. (2012). Multi-target regulation by small RNAs synchronizes gene expression thresholds and may enhance ultrasensitive behavior. PLoS One *7*, e42296.

Schultz, D., Wolynes, P.G., Ben Jacob, E., and Onuchic, J.N. (2009). Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis*. Proc Natl Acad Sci USA *106*, 21027–21034.

Serganov, A., and Nudler, E. (2013). A decade of riboswitches. Cell *152*, 17–24.

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat Rev Microbiol *11*, 75–82.

Shank, E.A., and Kolter, R. (2011). Extracellular signaling and multicellularity in *Bacillus subtilis*. Curr Opin Microbiol *14*, 741–747.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature *464*, 250–255.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C., Rolle, K., Binnewies, T., Hinton, J., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet *4*, e1000163.

Slepecky, R.A., and Hemphill, H.E. (2006). The Genus Bacillus - Nonmedical. In The Prokaryotes: Vol. 4: Bacteria: Firmicutes, Cyanobacteria: A Handbook on the Biology of Bacteria, S. Falkow, E. Rosenberg, K.-H. Schleifer, E. Stackebrandt, and M. Dworkin, eds. (Dordrecht Heidelberg London New York: Springer), pp. 530-630.

Sorek, R., and Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet *11*, 9–16.

Stazic, D., Lindell, D., and Steglich, C. (2011). Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. Nucleic Acids Res *39*, 4890–4899.

Stork, M., Di Lorenzo, M., Welch, T.J., and Crosa, J.H. (2007). Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. J Bacteriol *189*, 3479–3488.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell *43*, 880–891.

Strausberg, R.L. (1999). The Mammalian Gene Collection. Science *286*, 455–457.

Sugar, A.M., and McCloskey, R.V. (1977). *Bacillus licheniformis* Sepsis. JAMA *238*, 1180–1181.

United States Food and Drug Administration (1999-2013). GRAS Notice Inventory No. 22, 24, 72, 79, 265, 277, and 472. [http://www.accessdata.fda.gov/scripts/fcn/fcnNavigation.cfm?filter=licheniformis&sortColumn=&rpt=grasListing]

United States Evironmental Protection Agency (1997). Bacillus Licheniformis Final Risk Assessment. [http://www.epa.gov/oppt/biotech/pubs/fra/fra005.htm]

Vanderpool, C.K., Balasubramanian, D., and Lloyd, C.R. (2011). Dual-function RNA regulators in bacteria. Biochimie *93*, 1943–1949.

Van Vliet, A.H.M. (2010). Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol Lett *302*, 1–7.

Veening, J.-W., Smits, W.K., and Kuipers, O.P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol *62*, 193–210.

Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K.-H., Ehrenreich, P., Bäumer, S., Henne, A., Liesegang, H., Merkl, R., et al. (2004). The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. J Mol Microbiol Biotechnol *7*, 204–211.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484–487.

Vivancos, A.P., Güell, M., Dohm, J.C., Serrano, L., and Himmelbauer, H. (2010). Strand-specific deep sequencing of the transcriptome. Genome Res *20*, 989–999.

Voigt, B., Schweder, T., Sibbald, M., Albrecht, D., Ehrenreich, A., Bernhardt, J., Feesche, J., Maurer, K.-H., Gottschalk, G., van Dijl, J., et al. (2006). The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions. Proteomics *6*, 268–281.

Waldeck, J., Meyer-Rammes, H., Wieland, S., Feesche, J., Maurer, K.-H., and Meinhardt, F. (2007a). Targeted deletion of genes encoding extracellular enzymes in *Bacillus licheniformis* and the impact on the secretion capability. J Biotechnol *130*, 124–132.

Waldeck, J., Meyer-Rammes, H., Nahrstedt, H., Eichstädt, R., Wieland, S., and Meinhardt, F. (2007b). Targeted deletion of the *uvrBA* operon and biological containment in the industrially important *Bacillus licheniformis*. Appl Microbiol Biotechnol *73*, 1340–1347.

Wang, L.-T., Lee, F.-L., Tai, C.-J., and Kasai, H. (2007). Comparison of *gyrB* gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the *Bacillus subtilis* group. Int J Syst Evol Micr *57*, 1846–1850.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57–63.

Waschkau, B., Waldeck, J., Wieland, S., Eichstädt, R., and Meinhardt, F. (2008). Generation of readily transformable *Bacillus licheniformis* mutants. Appl Microbiol Biotechnol *78*, 181–188.

Waters, L., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell *136*, 615–628.

Welz, R., and Breaker, R.R. (2007). Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. RNA *13*, 573–582.

Winkler, W.C. (2012). RNA-mediated Regulation in *Bacillus subtilis*. In *Bacillus*: Cellular and Molecular Biology (Second Edition), P. Graumann, ed. (Norfolk, England: Caister Academic Press), pp. 167-214.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., et al. (2003). Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science *302*, 842–846.

Yoder-Himes, D., Chain, P., Zhu, Y., Rubin, E., Wurtzel, O., Tiedje, J., and Sorek, R. (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. Proc Natl Acad Sci USA *106*, 3976–3981.

CHAPTER B

RNA-SEQ OF *BACILLUS LICHENIFORMIS*: ACTIVE REGULATORY RNA FEATURES EXPRESSED WITHIN A PRODUCTIVE FERMENTATION

Sandra Wiegand, Sascha Dietrich, Robert Hertel, Johannes Bongaerts, Stefan Evers, Sonja Volland, Rolf Daniel and Heiko Liesegang

BMC Genomics (2013), 14:667

Authors' contributions to this work

Performed the experiments: SW, RH
Analyzed data: SW, HL
Developed the prediction algorithms: SW, SD
Programmed the analysis tools: SD
Provided fermentation facilities: JB, SE
Submitted data: SW, SV
Wrote the paper: SW, HL, RD
Provided research facilities: RD
Conceived and designed the experiments: HL

Wiegand et al. BMC Genomics 2013, 14:667
http://www.biomedcentral.com/1471-2164/14/667

BMC Genomics

RESEARCH ARTICLE Open Access

RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation

Sandra Wiegand[1], Sascha Dietrich[1], Robert Hertel[1], Johannes Bongaerts[2], Stefan Evers[2], Sonja Volland[1], Rolf Daniel[1] and Heiko Liesegang[1*]

Abstract

Background: The production of enzymes by an industrial strain requires a complex adaption of the bacterial metabolism to the conditions within the fermenter. Regulatory events within the process result in a dynamic change of the transcriptional activity of the genome. This complex network of genes is orchestrated by proteins as well as regulatory RNA elements. Here we present an RNA-Seq based study considering selected phases of an industry-oriented fermentation of *Bacillus licheniformis*.

Results: A detailed analysis of 20 strand-specific RNA-Seq datasets revealed a multitude of transcriptionally active genomic regions. 3314 RNA features encoded by such active loci have been identified and sorted into ten functional classes. The identified sequences include the expected RNA features like housekeeping sRNAs, metabolic riboswitches and RNA switches well known from studies on *Bacillus subtilis* as well as a multitude of completely new candidates for regulatory RNAs. An unexpectedly high number of 855 RNA features are encoded antisense to annotated protein and RNA genes, in addition to 461 independently transcribed small RNAs. These antisense transcripts contain molecules with a remarkable size range variation from 38 to 6348 base pairs in length. The genome of the type strain *B. licheniformis* DSM13 was completely reannotated using data obtained from RNA-Seq analyses and from public databases.

Conclusion: The hereby generated data-sets represent a solid amount of knowledge on the dynamic transcriptional activities during the investigated fermentation stages. The identified regulatory elements enable research on the understanding and the optimization of crucial metabolic activities during a productive fermentation of *Bacillus licheniformis* strains.

Keywords: dRNA-Seq, RNA-based regulation, UTR, ncRNA, sRNA, Antisense RNA, Subtilisin, Transcription start site, Operon prediction, Reannotation

Background

Bacillus licheniformis is a spore-forming soil bacterium closely related to the Gram-positive model organism *Bacillus subtilis.* The species' saprophytic life style, based on the secretion of biopolymer-degrading enzymes, predestinates strains of *B. licheniformis* as ideal candidates for the large-scale industrial production of exoenzymes, such as amylases and peptide antibiotics [1]. Especially its high capacity of secreting overexpressed alkaline serine proteases has made *B. licheniformis* one of the most important bacterial workhorses in industrial enzyme production [2]. Due to their high stability and relatively low substrate specificity, alkaline serine proteases like subtilisins are crucial additives to household detergents and the greatest share on the worldwide enzyme market [2,3]. Attempts to optimize the productivity have addressed the fermentation process [4,5], protein-engineering [3,6,7], and cellular influences on protein quality and quantity [2,8]. Since the 4.2 Mb circular genome of the type strain

* Correspondence: hlieseg@gwdg.de
[1]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077, Göttingen, Germany
Full list of author information is available at the end of the article

BioMed Central
© 2013 Wiegand et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

B. licheniformis DSM13 was published in 2004 [1,9], several genome-based studies targeting strain improvement have been performed successfully [10,11]. However, genome-based studies are limited to information directly accessible from the DNA sequence and cannot benefit from knowledge of the active transcriptome. Considering that the regulatory network represented by protein- and RNA-based regulators determines the performance of an industrial-oriented fermentation process [12] RNA-Seq data might contribute to further optimization approaches.

RNA-based regulatory elements are involved in the regulation of metabolism, growth processes, the adaptation to stress and varying culture conditions [13] and can be divided into two main categories. The first category comprises non-coding RNAs (ncRNAs). *Trans*-encoded ncRNAs, often referred to as small RNAs (sRNAs), are encoded independently from protein genes and are able to modulate the mRNA of genes located at distant chromosomal loci or to interact with target proteins [14]. Upon formation of secondary structures, *trans*-encoded ncRNAs interact with their target RNAs by imperfect base pairing, which is triggered by the binding of a seed region of at least six contiguous nucleotides [15,16]. This mechanism allows the sRNA to address multiple targets, thus orchestrating different members of one regulon [14,17]. It has been shown that sRNAs affect mRNA degradation and translation and modulate protein activity [14,16]. A second class of regulatory ncRNAs is encoded in *cis,* which means that these ncRNAs are transcribed from the antisense strand of protein-coding genes [18]. Hence, they are complementary in full-length and can therefore form RNA duplexes with the mRNA of the targeted genes [19]. Most described examples of these *cis*-encoded antisense RNAs (asRNA) range from 100 to 300 nt in size, but some asRNAs are also shown to be substantially longer [18,20]. Antisense RNAs have been proven to either positively or negatively affect transcription, translation and mRNA stability [16]. In addition, a *cis*-encoded asRNA might work as a trans-encoded sRNA for another target [19]. Antisense transcription has been detected in multiple organisms [21] and, with the growing number of explored species, it is assumed that antisense transcripts can be found for ~10 to 20% of the bacterial genes [22]. A second class of RNA-based regulators encompasses *cis*-regulatory elements, mainly present at the 5′ untranslated region (5′UTR) of mRNA transcripts, e.g. riboswitches, T-boxes or thermosensors [23]. Whereas both 5′ as well as 3'untranslated regions can bear signals for the initiation and termination of translation [24,25], respectively, 5′UTRs additionally have the ability to fine-tune translation by *cis*-regulatory elements. They can be prone to RNA-binding proteins or antisense RNAs, carry attenuation systems [14,23] and play a role in mRNA stability [26]. In contrast, 3'UTRs are not as well understood and have escaped the attention of most transcriptomic studies [27]. It is known that long UTRs can be localized antisense to adjacent genes on the opposite strand; in fact some of these overlapping UTRs have been demonstrated to act as negative regulators for genes encoded on the opposite strand [20].

The development of next-generation sequencing techniques including RNA sequencing (RNA-Seq) enabled the genome-wide identification of RNA-based regulatory elements in an unprecedented depth. The high dynamic range of RNA-Seq allows the identification of transcripts which are expressed at vastly different levels. Also, this method does not exhibit background noise and is therefore appropriate for the identification of lowly abundant transcripts [28]. RNA-Seq analyses targeting ncRNA in particular, have been published for e.g. *Mycobacterium tuberculosis* [29], *Streptomyces coelicolor* [30] and *Sinorhizobium meliloti* [31].

The major goal of the project in which this study is embedded is the improvement of production strains and thus ultimately the enhancement of enzyme production. This study is targeted on the identification of active regulatory RNA elements within the different phases of a productive fermentation process. Therefore samples from crucial stages of an industrial-oriented *B. licheniformis* subtilisin fermentation process have been examined by strand-specific RNA-Seq and differential RNA-Seq (dRNA-Seq) [32]. A comprehensive analysis of the data revealed a multitude of RNA features which correlate to the physiology and the growth phases during the process. The combination of genomic data and RNA features provides an excellent basis to understand the regulatory events within an industrial fermentation process.

Results and discussion

B. licheniformis MW3Δspo, a germination deficient mutant of *B. licheniformis* DSM13, transformed with an expression plasmid encoding an alkaline serine protease, was grown in fed-batch mode in 6 L cultures. The fermentations were carried out in complex amino acid broth under conditions resembling the parameters used in industrial fermentation processes (Figure 1). To enhance the reliability of the analysis, the experiments were carried out in triplicate (L, R and M). Samples were taken at five selected time points of the fermentation process, which were chosen to follow the initial cell growth (sampling points I, II and III) and to determine the decisive changes within the early (IV) and the late stage (V) of the protease-producing states (Figure 1). Total RNA from each sample was prepared for strand-specific whole transcriptome sequencing [33]. RNA from samples L-I to L-V was additionally prepared for

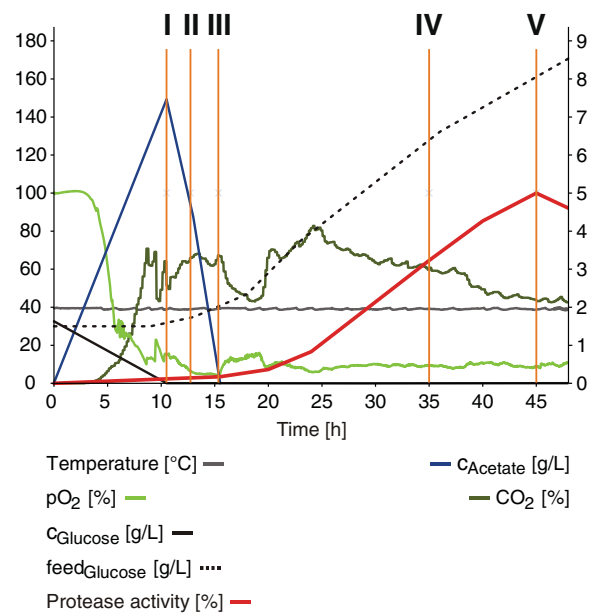


Figure 1 Protease production and process parameters. Process parameters are shown for fermentation L (the parameters for the replicate fermentations R and M are corresponding, data not shown). Temperature T [°C], oxygen partial pressure pO_2 [%], glucose concentration $c_{Glucose}$ [g/L], supplied glucose $feed_{Glucose}$ [g/L] and normalized protease activity [%] are displayed on left y-axis, whereas acetate concentration $c_{Acetate}$ [g/L] and carbon dioxide content CO_2 [%] are scaled on the right y-axis. Process time t [h] is given on the x-axis. The sampling points I to V are indicated by orange lines.

differential RNA-Seq for determination of transcription start sites (TSS), as described by Sharma et al. [32].

Whole transcriptome sequencing

Strand-specific deep sequencing of the whole transcriptome of 15 *B. licheniformis* samples yielded more than 500 million reads with a specific length of 50 nucleotides. The number of reads for each library ranged from 2.4×10^7 to 4.3×10^7.

After the application of a strict quality processing (see Methods), 77.3 to 93.9% of these reads have been found to map to the chromosome and the expression plasmid used in this study (for details see Additional file 1: Figure S1 and Additional file 2: Table S1). Due to repeat regions, 1.45% of the *B. licheniformis* genome is not precisely mappable when considering the applied read length of 50 nucleotides. Thus, all reads mapping completely to such repeat regions have been excluded from further analysis. This pertains mainly to those 68.5 to 88.8% of reads which map to tRNA and rRNA genes. The majority of these rRNA matching reads can be assigned to 5S rRNA genes, which is in accordance with the fact that the applied depletion targets especially 16S and 23S rRNAs. Also, all reads mapping to the plasmid were removed from the dataset, as this analysis is focused on the transcriptional activity of the chromosome. Finally, 4.4 to 12.0% of the initial reads were taken for further analyses. These reads enabled the identification of transcriptional units and the determination of their boundaries to assign the transcriptional activity of coding as well as non-coding regions of the chromosome (see Methods).

To facilitate the comparison of different transcription levels between samples, we introduce the **n**ucleotide activity **p**er **k**ilobase of exon model per **m**illion mapped reads (NPKM) value as single nucleotide-resolution measure of transcriptional activity (see Methods). NPKMs for each RNA feature and for every gene were calculated and are available at Additional file 2: Table S2 and Table S3.

Transcription start site determination and operon prediction

Differential RNA-Seq (dRNA-Seq) has been designed by Sharma et al. [32] to allow selective enrichment of native 5′ ends of transcripts for the determination of transcription start sites (TSS). The method is based on the observation that 5′ triphosphorylated RNA fragments are originating from native 5′ ends. In contrast, 5′ monophosphorylated RNAs are products of RNA decay or processing and do not contain information of transcription initiation. The dRNA-Seq approach includes a treatment with 5′ phosphate-dependent exonuclease (TEX), which results in the depletion of all monophosphorylated transcripts. It has been shown that TSS identification based on dRNA-Seq data is superior to an estimation of transcript boundaries based on whole transcriptome RNA-Seq reads [32].

The differential sequencing of samples L-I to L-V resulted in 22,047,373 reads (Additional file 2: Table S4). A total of 2522 putative TSS was predicted (see Methods), 1500 of which were detected in at least two samples (Additional file 2: Table S5). A comparison of the latter with the transcript boundaries obtained by whole transcriptome sequencing (Additional file 2: Table S6) shows that 412 identified TSS confirm the RNA-Seq data, whereas the other findings introduce TSS not detectable by conventional RNA-Seq. To allow the assignment of the identified TSS to their putative origin, an allocation to four different classes was accomplished (Figure 2A) [34]. Naturally, the affiliation of TSS according to this schema is ambiguous as some TSS sort to multiple classes, e. g. some TSS are located in a promoter region and within the upstream gene as well. The distribution of the identified TSS to each class is shown in Figure 2B. 1092 TSS were detected in promoter regions, 72 genes are bearing more than one putative TSS in this region. The dRNA-Seq data enabled conclusions for TSS determination in cases in which read-through transcription of the upstream gene caused by leaky termination prohibits the identification of downstream TSS by conventional RNA-Seq data. Of the

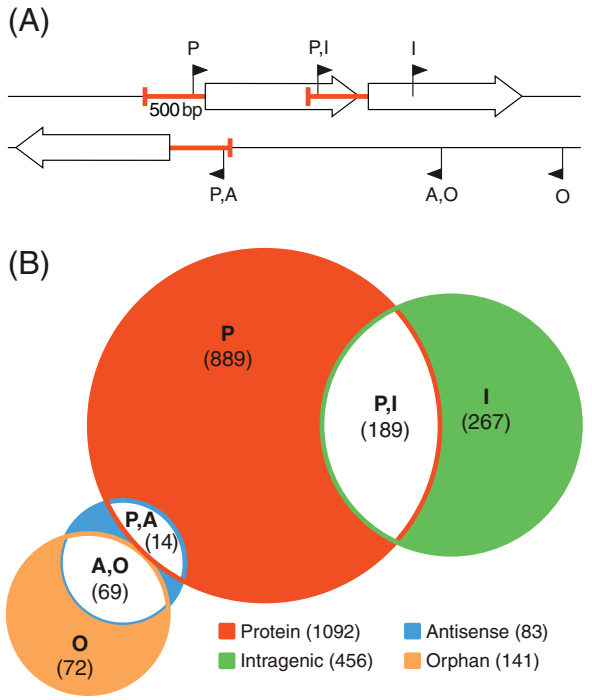


Figure 2 Classification and distribution of TSS. (A) Classification scheme of transcription start sites adapted from Dötsch et al. [34]. White arrows indicate genes. P: Protein-coding gene-dependent TSS located within a 500 bp range upstream of annotated start codons. I: Intragenic TSS situated within an annotated gene on the same strand. A: TSS localized antisense to an annotated gene. O: Orphan TSS not located in a promoter region or a gene on the same strand. **(B)** Distribution of transcription start sites identified in this study. Numbers in brackets give the amount of instances for each class. Numbers in the legend give the total amount of every class.

identified 456 intragenic TSS, 267 are not located in the 500 bp promoter region of the downstream gene, reflecting a high number of putative internal promoters. Orphan TSS may indicate potential start sites of yet unknown genes or non-coding RNAs, this is supported by the finding that 76 of the 141 detected orphan TSS could be allocated to identified ncRNAs.

Operon prediction based on RNA-Seq and dRNA-Seq data resulted in 2510 putative operons structuring the genome of *B. licheniformis* (Additional file 2: Table S7). While most operons are monocistronic (66.8%) or bicistronic (18.3%), seven operons seem to encompass more than ten genes (Additional file 1: Figure S2). This small number of long operons is not in accordance with the operon prediction made by Kristoffersen et al. [35] for *B. cereus*. The difference is due to the varying operon concept employed here. Especially the consideration of internal TSS in combination with distinct shifts of expression resulted in an increase of shortened operons in this study.

Reannotation

The first genome annotation of *B. licheniformis* DSM13 has been published in 2004 [1]. It has been shown previously that mapping of RNA-Seq data to genomes allows the correction of open reading frames and supports the identification of not-annotated protein genes [36]. Therefore, we performed a complete reannotation of the genome in order to integrate the RNA-Seq data provided by this study as well as the progress in gene prediction and annotation of the recent years. Distinct transcription start sites determined by dRNA-Seq and RNA-Seq-based whole transcriptome data have been used to identify putative mis-annotated genes (Figure 3A). These findings were validated by length comparisons to genes deposited in public databases and confirmation of ribosomal binding sites and −10 and −35 promoter regions. This approach enabled the correction of reading frames of 23 protein genes, 25 pseudo genes, 21 rRNA genes and two tRNA genes (Additional file 2: Table S8). Moreover, 60 previously not-annotated protein genes were identified based on transcriptional activity and protein conservation (Figure 3B, Additional file 2: Table S9). 52 genes (Additional file 2: Table S10) were removed from the annotation as these previously predicted ORFs could not be verified by detailed genome analysis and comparisons to public databases. In total, the reannotation approach resulted in a dataset containing 4297 ORFs. Comparisons to the annotation of *B. licheniformis* DSM13 by Rey et al. [9] showed that 16 of the newly annotated genes have not been described for this organism before. 18 of the removed genes were annotated in both genomes. More than 2000 gene annotations have been improved. These improvements mainly comprise former hypothetical proteins now assigned to a function and proteins with altered gene symbols. In addition to gene-associated improvements, seven genomic regions were identified as prophage regions based on GC content deviations, significant similarities to known prophage genes and the presence of insertion repeats. The transcriptional activity of the prophage regions was rather low, which is consistent with the observation that many prophages are induced during SOS response, which should not occur within a fermentation process [37].

5′ and 3′untranslated regions

In this study, 1433 5′untranslated regions (Figure 4) with a mean length of 117 nt (Figure 5A, Additional file 2: Table S6) could be identified. Thirty of these 5′ UTRs are shorter than 11 nt, implicating that leaderless transcription, commonly found in many bacteria [38], is not an abundant mechanism in *B. licheniformis.* Correspondingly, low occurrence of leaderless transcription has also been suggested for other members of the phylum *Firmicutes* [39]. The most strongly

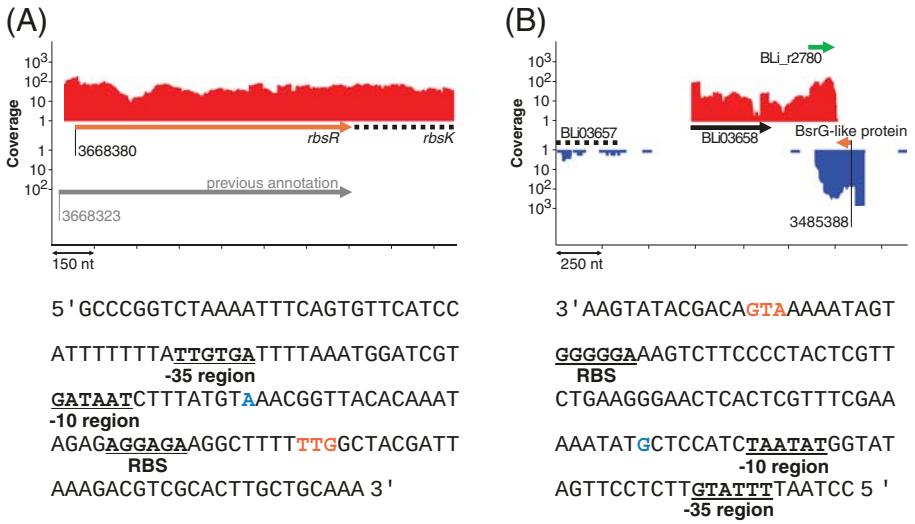


Figure 3 Correction and insertion of annotated genes. (A) Correction of start codons. *(Upper panel)* Transcriptional activity of *pooled RNA-Seq data*. The grey arrow displays the coordinates of the ribose operon repressor RbsR (BLi03840) according to Veith et al. [1]. Based on the transcriptional data, the start codon has been reassigned 57 bp downstream of the former position (orange arrow). *(Lower panel)* The new start codon is marked in orange and the transcription start site in blue. The location of patterns of a ribosomal binding site and −10 and −35 regions of the *rbsR*- regulating σ^A upstream of the gene provide additional confirmation to the new annotation. **(B)** Insertion of new genes. *(Upper panel)* Transcriptional activities (sample L-I) of BLi03658 (black arrow) and *indep* RNA BLi_r2780 (green arrow). The previously not detected [1,9] protein gene BLi05038 (orange arrow) was annotated as BsrG-like peptide (see also chapter *Comparative transcriptomics*). *(Lower panel)* The start codon of the new gene is marked in orange, and the transcription start site in green. The location of patterns of a ribosomal binding site and a σ^A -10 and −35 promoter region provide additional confirmation of the new annotation.

transcribed 5'UTRs ≥150 nt and the 5'UTRs discussed in the following passage are listed in Table 1.

At sampling points I, II, and III, the gene of the sporulation inhibitor KapD (BLi03329) reveals a 5'UTR of 113 nt (Figure 6A), whereas an alternative, dRNA-Seq-supported 5'UTR (BLi_r2510) with a length of 2226 nt is present at the later stages of the fermentation process. The TSS of both 5'UTRs seem to be preceded by a σ^A recognition site. At sampling point IV, the transcriptional activity of the gene is higher than the activity of the 5'UTR region. In *B. subtilis*, growth phase-dependent differentiation into subpopulations of distinct cell types with different gene expression patterns is well described [40,41]. The divergent expression levels of *kapD* and the long 5'UTR in *B. licheniformis* might therefore result from different usage of promoter sites dependent on the respective cell type. However, the observed effect could also derive from slow decay rates of the short form of the *kapD* mRNA transcribed earlier. 52 further 5'UTRs exhibited antisense activity towards upstream genes ($A_{5'UTR}$; Figure 4), as shown for the untranslated region BLi_r1609 upstream of the glutamate synthase operon *gltAB* (BLi02161/62; Figure 6B). The observed 5'UTR is completely antisense to the gene of the corresponding transcriptional activator GltC (BLi02163). The dRNA-Seq data suggest the presence of only one TSS. This finding might be an example for a regulatory linkage between adjacent genes localized on different strands. This concept has recently been termed the excludon by Sesto et al. [20], who demonstrated that long 5'UTRs can act negatively on the transcription of the opposite gene. Following this idea in the case of the glutamate synthase operon, the preceding 5'UTR might establish a negative feedback regulation of the transcriptional activator GltC. A corresponding elongated UTR of the *gltAB* operon has not been found in *B. subtilis* [42,43], which indicates different regulations of glutamate homeostasis in the two species.

Next to regulatory effects based on antisense orientation, 5'UTRs can bear intrinsic, so-called *cis*-regulatory elements [44]. At the time of this study, 62 *cis*-regulatory elements have been predicted for *B. licheniformis* DSM13 by covariance models [45,46]. All elements have been shown to be transcriptionally active during the fermentation process (Additional file 2: Table S11), although some are not located in 5'UTRs but in intergenic read-through regions. Three new T-boxes, located upstream of the serine acetyltransferase gene *cysE* and the tRNA ligase subunit genes *glyQ* and *pheS*, could be identified by comparison to the Rfam database. In *B. subtilis*, 92 *cis*-regulatory elements have been described [43], comprising RNA switches as well as protein-binding RNAs. For 76 of these instances, transcription could be shown at orthologous loci in *B. licheniformis* (Additional file 2: Table S12).

1365 3'UTRs (Additional file 2: Table S6) with an average length of 276 nt have been identified according to Figure 4, the most strongly transcribed 3'UTRs ≥150 nt

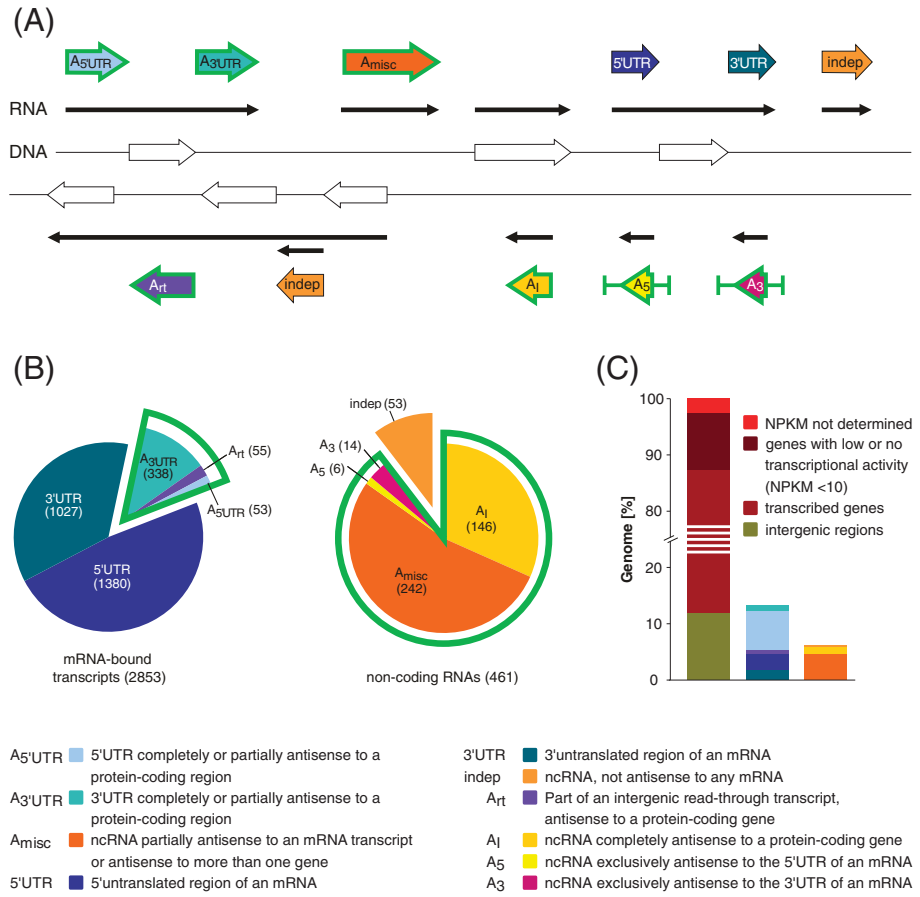


Figure 4 Classification and distribution of RNA features. (A) Classification scheme of ten RNA feature classes. White arrows indicate CDS and black arrows represent RNA transcripts. All antisense transcripts are framed green. RNA features which are part of an mRNA are denoted *5'UTRs* or *3'UTRs*. Antisense transcripts that are mRNA-bound were classified as $A_{5'UTR}$, $A_{3'UTR}$ and A_{rt}. Non-coding antisense transcripts were classified as A_5, A_3 and A_I and comprise antisense transcripts opposite to 5' and 3'UTRs or to protein-coding regions of the mRNA. A_{misc} designates antisense ncRNAs that target more than one gene or are only partially antisense. Independently transcribed ncRNAs without any antisense localization are designated *indep*. **(B)** Quantitative affiliation of identified RNA features. **(C)** *(Left)* Proportional distribution of intergenic regions or annotated genes with different transcriptional activities within the complete *B. licheniformis* DSM13 genome. *(Middle + Right)* Percentage of the total genome covered by the defined RNA classes.

and the 3'UTR discussed in this chapter are listed in Table 2. Of the identified 3'UTRs, 42% exceed 100 nt and 16% even exceed 500 nt in length (Figure 5B). In total, 338 3'UTRs are localized antisense to adjacent genes ($A_{3'UTR}$; Figure 4). A detailed manual inspection revealed that all 3'untranslated regions longer than 1000 nt seem to be protruding after incomplete termination [18,32]. Altogether, 684 3'UTRs with internal termination sites could be determined, whereas 511 3'UTRs end at predicted termination sites. These findings suggest that the effect of fading-out at the end of operons due to imperfect termination might be a common effect in *B. licheniformis*. An example is the 3965 nt 3'UTR (BLi_r2654) downstream of the cell envelope stress response operon *liaIHGFSR* (BLi03492-97; Figure 6C). The mRNA transcript of this operon protrudes beyond a termination signal, which is located directly behind the stop codon of *liaR*. This protruding mRNA sequence is antisense to the next four genes which comprise the germination receptor operon *gerAAABAC* (BLi03488-90) and a hypothetical protein (BLi03491). A second terminator structure can be found 370 nt upstream of the end of the transcript.

Non-coding RNA features

Non-coding RNAs were identified in non-coding regions of the chromosome, for example in intergenic regions or localized in antisense direction to protein genes (see Methods). The boundaries of the identified transcripts were determined by upshifts or downshifts of transcriptional activity. All identified RNA features were checked for similarities to complete protein genes as well as protein domains to ensure that they indeed represent non-coding RNAs.

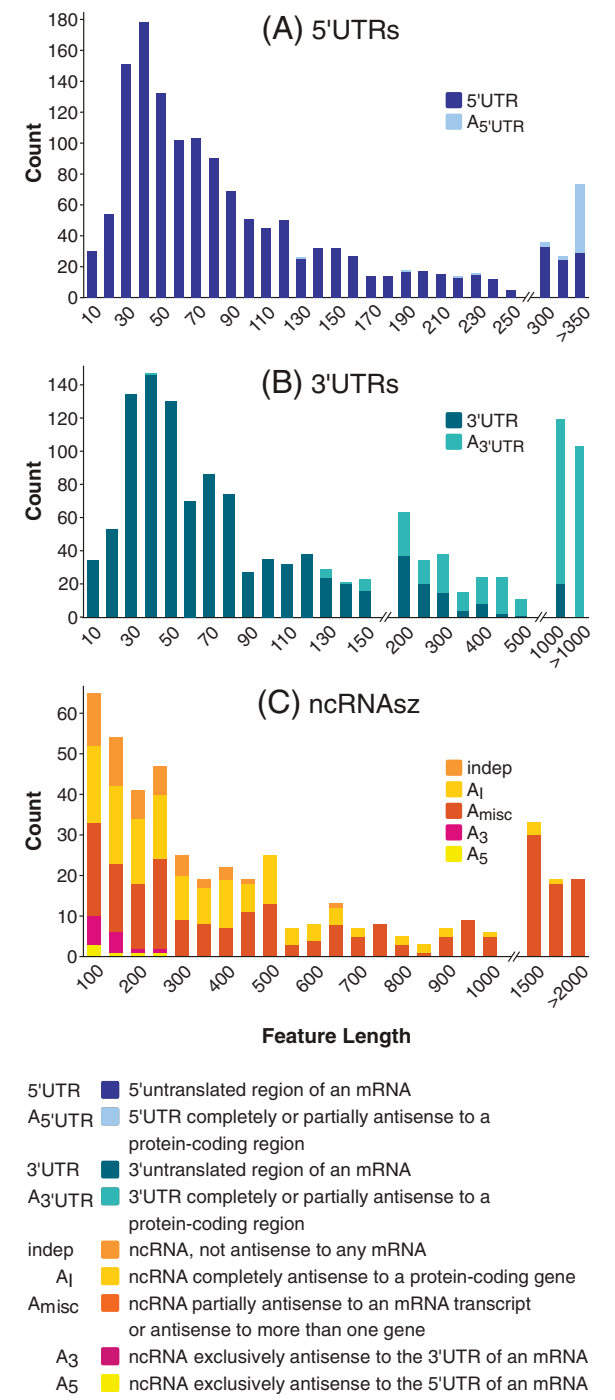


Figure 5 Length distribution of RNA features. Size range of **(A)** 1433 identified 5′ untranslated regions, **(B)** 1365 identified 3′ untranslated regions and **(C)** 461 identified non-coding RNAs. Please note that the classification scheme corresponds to Figure 4.

To extract the basic types of ncRNA expression profiles during the examined fermentation process, cluster analysis based on the *k*-means algorithm [47] was applied to those 273 ncRNAs with highly reliable replicates. In total, 15 clusters of divergent expression profiles were generated (Figure 7, Additional file 2: Table S13). Cluster 1 contains 36% of the applied ncRNAs and 50% of all ncRNAs >1000 nt. It displays a strong up-shift of transcriptional activity at sampling point IV followed by a decrease at sampling point V. The high portion of transcripts in this cluster prompts the conclusion that RNA-based regulation is especially important during the later stages of the fermentation process. Other ncRNAs exhibiting up-shifts of transcriptional activity are displayed in clusters 2 to 4, whereas clusters 5 to 8 include transcripts with activity down-shifts. The further clusters comprise ncRNAs with expression shifts during the early fermentation process, as well as an activity up-shift at sampling point V in clusters 10 to 12.

All assigned non-coding RNAs were categorized according to the scheme displayed in Figure 4 and subdivided into the classes A_5, A_3, A_I, A_{misc} and *indep*. Selected ncRNAs are listed in Table 3, whereas an overview of all identified features is given in Additional file 2: Table S6. Several ncRNAs have been selected for validation by Northern blotting (Additional file 1: Figure S3). The analyzed ncRNAs were chosen as they are exemplarily for their respective class. The occurrence of eight ncRNAs could be verified, especially ncRNAs <500 nt are in good accordance with the results gained by RNA-Seq. The results for transcripts >2000 nt are indicative for RNA degradation or processing and leaky transcription termination. However, three ncRNAs could not be validated, which is most probably due to their low expression levels.

Indep ncRNAs

As depicted in Figure 4, *indep* transcripts are defined as non-coding RNAs not localized antisense to any mRNA. Instead they can be found in intergenic regions or any other position of the chromosome. In total, 53 *indep* RNAs with sizes between 51 and 602 nt have been identified, of which 40 have TSS verified by dRNA-Seq. Within this group five housekeeping sRNAs could be annotated: the tmRNA SsrA (BLi_r2758), the 6S RNAs BsrA (BLi_r2163) and BsrB (BLi_r1454), the RNA component of RNase P RnpB (BLi_r1808) and the signal recognition particle Scr (BLi_r0016) [13]. 87% of the *indep* transcripts exhibited NPKM values ≥100 in at least three samples, reflecting a strong transcriptional activity of the encoding genomics regions. For example the sRNA Scr, an essential part of the protein secretion system [48], reaches a maximal NPKM value of almost 400,000. This is in perfect accordance with the fact that the cells are derived from a fermentation process optimized for protein secretion. Interestingly, 39 *indep* ncRNAs seem to be transcribed constitutively under the examined conditions (Figure 8A), whereas only thirteen *indep* RNAs show differential expression (likelihood value ≥0,99) [49]), as illustrated exemplarily for BLi_r1424 (Figure 8B).

Table 1 Selected 5'untranslated regions (5'UTRs)

RNA feature	Start	Stop	Length	Downstream gene	Antisense genes	*cis*-regulatory element	NPKM value*
BLi_r0085	210580	210926	347	*thrZ*		T-box	1966
BLi_r0356	542296	542520	225	BLi00536		*ydaO-yuaA* leader	1368
BLi_r0498	712264	712587	324	*yybP*		*yybP-ykoY* leader	1144
BLi_r0691	942026	942290	306	*thiC*		TPP	5030
BLi_r0744	998913	999033	121	*glpD*			5693
BLi_r0943	1207762	1207542	180	*yitJ*		SAM	2782
BLi_r0982	1243825	1243498	328	*trpS*		T-box	865
BLi_r0983	1244045	1244232	188	*oppA*			1465
BLi_r1011	1271178	1271391	255	*tenA*		TPP	9963
BLi_r1028	1291984	1292289	306	*metI*		SAM	3605
BLi_r1168	1487545	1487279	226	*mtnK*		SAM	2207
BLi_r1196	1510226	1511237	1012	BLi05023	BLi01539, BLi01540		651
BLi_r1485	1973018	1973209	192	BLi02027			1035
BLi_r1609	2118173	2117083	1091	*gltA*	*gltC*		34
BLi_r1634	2146236	2145933	304	*expZ*			1225
BLi_r1709	2204969	2205111	143	*dhaS*			2816
BLi_r1801	2295779	2295571	168	*xpt*		Purine	1540
BLi_r1835	2356768	2356478	291	*hbs*			3478
BLi_r1850	2382265	2381914	352	*ribU*		FMN	2367
BLi_r1871	2409288	2409010	238	*ribD*		FMN	2512
BLi_r2045	2616129	2615830	300	*glyQ*		T-box	787
BLi_r2142	2742406	2742106	301	*yrzI*			860
BLi_r2241	2878256	2877902	313	*lysC*		Lysine	751
BLi_r2286	2949768	2949565	204	*citZ*			1563
BLi_r2389	3060789	3060456	292	*leuS*		T-box	998
BLi_r2510	3188213	3185988	2226	*kapD*	*yuxJ, pbpD*		13
BLi_r2628	3302655	3302393	221	*metN2*		SAM	2661
BLi_r3184	4014316	4014539	224	*yxjG*		SAM	5296
BLi_r3195	4037045	4036819	185	BLi04205		TPP	9303
BLi_r3196	4037110	4037236	127	BLi04206			2693

*(pooled RNA-Seq data).

Antisense ncRNAs

In contrast to the class of indep ncRNAs, the antisense ncRNAs (asRNAs) A_1, A_3, A_5, and A_{misc} comprise non-coding transcripts localized antisense to annotated protein-coding genes. They either target the protein-coding region of a gene (A_1) or the 5' and 3' untranslated regions (A_5 and A_3). Furthermore, ncRNAs that target more than one gene or that are only partially antisense are classified (A_{misc}). In this study, 242 A_{misc} RNAs (Figure 8C/D) could be identified. Approximately 150 of them (and also all A_5 ncRNAs) are located opposite to ribosome binding sites and could therefore function as inhibitors of translation, a very common mechanism of *cis*-encoded asRNAs [50]. The length distribution of the non-coding RNA features is shown in Figure 5C, and illustrates that 42% of the A_{misc} RNAs are less than 400 nt in length. Twenty-seven of these short A_{misc} RNAs reach maximal NPKM values ≥100, suggesting putative sRNA mechanisms. However, some A_{misc} RNAs are much longer, i.e. BLi_r2246 is 6348 nt in length and spans six genes. The occurrence of antisense transcripts of such length is not unexpected, as asRNAs with very diverse sizes, reaching more than 7000 nt, have been described for several species [18]. Furthermore, 146 A_1 transcripts (Figure 8E) could be assigned, ranging in size from 54 to 1572 nt. Over 95% of the A_1 transcripts exhibit maximal NPKM values ≤100, 68% even ≤20, due to the low coverage only 20 TSS could be verified by dRNA-Seq for these asRNAs. In total, 408 non-coding asRNAs were determined, comprising 89% of all identified non-coding RNA

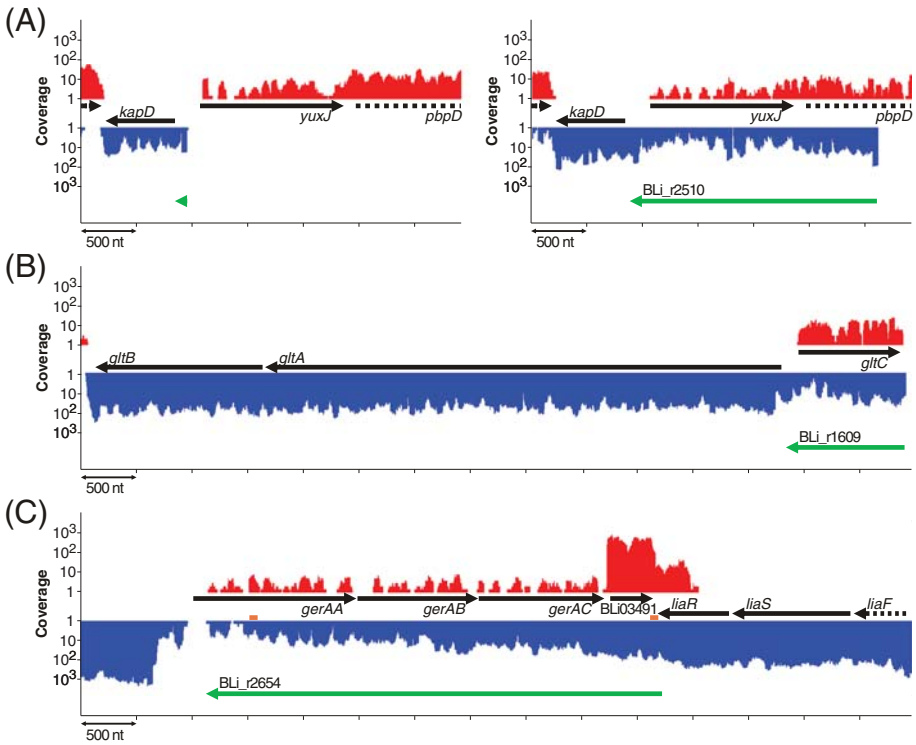


Figure 6 Untranslated regions (UTRs). Transcriptional activities of UTR regions. Black arrows indicate genes and green arrows the identified UTRs. **(A)** 5'UTR of *kapD* at sampling point II (left) and sampling point IV (right). **(B)** 5'UTR BLi_t1609 at sampling point IV. **(C)** 3'UTR BLi_r2654 (*pooled RNA-Seq data*). Predicted terminator sequences are marked orange.

transcripts and targeting 15% of all genes. The number of identified antisense ncRNAs is in accordance to previous studies which assume that antisense transcription concerns ~10 to 20% of the bacterial genes [22].

Antisense transcripts with putative impact on productivity

The general aim of our group is the identification of productivity related features. Thus, a special focus has been set on the identification of antisense transcripts with a putative impact on protease production as targets for strain improvement.

The alkaline serine protease Subtilisin Carlsberg (*apr*, BLi01109) represents the major secreted feeding protease of *B. licheniformis*. Thus, it competes for energetic and secretory resources with the production protease. We identified a *cis*-encoded 144 nt A_I asRNA (BLi_r0872) which is located at the 3'end of the *apr* mRNA (Figure 9). A highly active TSS determined from the dRNA-Seq data and a terminator structure downstream of the adjacent gene *yhfN* confirm the characterization of the transcript as independently transcribed asRNA. BLi_r0872 is highly expressed at all fermentation stages, whereas the transcriptional activity of the Subtilisin Carlsberg gene increases at the productive stages of the process. The presence of the *cis*-encoded asRNA opposite to the 3'end of the target mRNA resembles the *B. subtilis* RatA/*txpA* toxin/antitoxin system or the *Escherichia coli* GadY/*gadX* system in which an antisense RNA promotes either mRNA degradation or stability [19]. To elucidate the impact of the detected asRNA, further analyses will be necessary, especially as a corresponding transcript is absent in the transcriptome of *B. subtilis* [43].

Further antisense transcripts against genes involved in cell differentiation, cell stress response, and thiamine and folate biosynthesis could be observed and are presented in Additional file 1: Figure S4.

It is exciting to think about a regulatory impact of the mentioned ncRNAs, but there are also some noteworthy limitations to putative effects. (i) The completed RNA-Seq experiments cannot discern if the sense and the antisense transcript are transcribed in the same type of differentiated cells, which especially challenges stoichiometric estimations of asRNAs and their mRNA targets. Whether they can influence each other or fulfill different purposes in different cell types has to be a topic of single cell targeted investigations. (ii) It has been reported that functional sRNAs are produced in excess amounts over the targeted mRNA [16,51]. Therefore, a regulatory mechanism of poorly transcribed antisense RNA cannot be assumed *bona fide*, but has to be evaluated carefully. Nonetheless, our data implicate that there might be a biological function assignable to the RNA features, especially

Table 2 Selected 3'untranslated regions (3'UTRs)

RNA feature	Start	Stop	Length	Upstream gene	Antisense genes	NPKM value*
BLi_r0075	198675	198433	243	*citM*		127
BLi_r0671	919040	918595	446	*ygzB*	*perR1*	50
BLi_r0688	938151	938591	441	BLi00936		496
BLi_r0817	1054694	1054523	172	*msmX*	BLi01051	75
BLi_r0859	1099760	1099445	316	*ynzH*	*yhfE*	2310
BLi_r0949	1209837	1210136	300	BLi01196		188
BLi_r1013	1278149	1277370	780	*cotZ*	*fabI, cotO*	116
BLi_r1145	1465905	1468238	2334	*ykoM*	*ykoU, ykoV*	264
BLi_r1333	1655028	1654895	134	*ylaL*		195
BLi_r1357	1680492	1679988	505	*ylbP*	*gerR*	62
BLi_r1521	2008172	2008031	142	BLi02067		224
BLi_r1720	2216049	2215342	708	*odhB*	*yocS*	41
BLi_r1750	2244661	2244519	143	*yodL*	*yoyE*	16
BLi_r1797	2291386	2291184	203	*ypbQ*		268
BLi_r1927	2463046	2465026	1981	BLi02544	BLi02545,*ymaC*	213
BLi_r1985	2537707	2537586	122	*mntR*		151
BLi_r1995	2550986	2550862	125	*tasA*		482
BLi_r2041	2606579	2606461	119	*cccA*		431
BLi_r2067	2662102	2661914	189	BLi02768	*yrhD*	87
BLi_r2141	2741955	2741415	541	*yrzI*		348
BLi_r2152	2754987	2754859	129	*yrzB*		140
BLi_r2178	2787395	2787201	195	*yrbF*		1520
BLi_r2292	2952407	2952263	145	*pyk*		520
BLi_r2582	3255915	3256420	506	*yutI*	*yuxL*	324
BLi_r2620	3292613	3292481	133	*sufB*		405
BLi_r2654	3332662	3328698	3965	*liaR*	*gerAA, gerAB, gerAC,* BLi03491	11
BLi_r2700	3398622	3398373	250	*copA*		166
BLi_r2729	3427607	3428159	553	BLi05033		345
BLi_r2752	3464447	3464713	267	BLi03635		286
BLi_r2855	3591778	3591637	142	*cccB*		80

*(pooled RNA-Seq data).

when they are conserved within related species as *B. subtilis*. (iii) At last, it has to be experimentally excluded, especially for low abundant instances, that the found ncRNAs originate from spurious transcriptional events, for instance driven by alternative sigma factors [43].

Comparative transcriptomics

In total, we determined 461 candidate non-coding RNA transcripts, including antisense, as well as *indep* ncRNAs (see *Non-coding RNA features*). For *Synechocystis* sp. PCC6803, *Sinorhizobium meliloti* and the archaea *Sulfolobus solfataricus* P2 and *Methanosarcina mazei* Gö1 between 50 and 107 non-coding RNAs per Mb were identified [31,52-54], matching our result of 109 ncRNAs/Mb. For *B. subtilis*, the close relative of *B. licheniformis*, Nicolas et al. [43] have found 472 non-coding RNA features in a tiling array-based, condition-dependent transcriptome study. The majority (68%) of these features are intergenic transcripts determined by promoter analysis, whereas only 32% are derived from independently transcribed (antisense) RNAs. In contrast, the majority of ncRNAs identified in *B. licheniformis* are antisense RNAs (89%), transcribed independently from protein-coding genes. The identification of more antisense transcripts in *B. licheniformis* might be accounted to the reduced background noise in RNA-Seq in comparison to tiling arrays, which allows a better detection of low abundant transcripts [28]. 167 of the *B. licheniformis* ncRNAs are located in regions with high sequence similarity to *B. subtilis* [55] and 126 ncRNAs are encoded at the frontiers

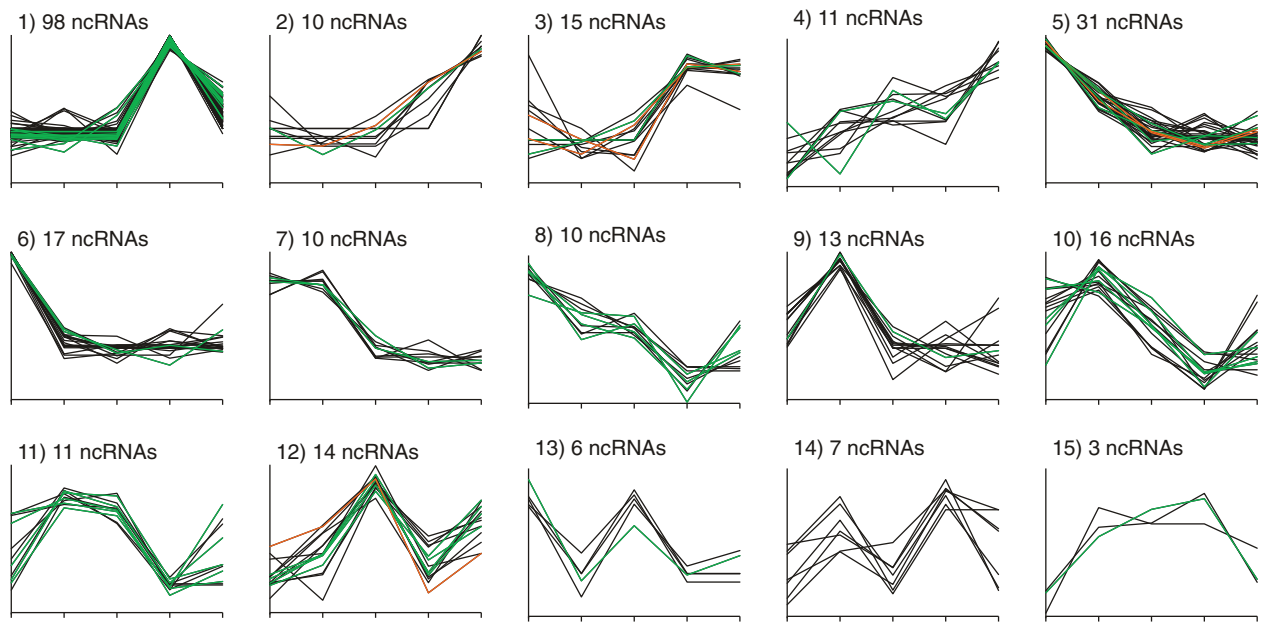


Figure 7 Cluster analysis of ncRNA expression profiles. Expression profiles of ncRNAs after *k*-means clustering (Additional file 2: Table S13). The x-axis shows sampling points I to V from left to right and the y-axis gives the expression strength in z-score transformed mean NPKM values of each replicate. Clusters are numbered and captioned with the count of included ncRNAs. Transcripts with a maximal NPKM value >100,000 are marked in orange and transcripts with a maximal NPKM value >1000 are marked in green.

of conserved and not conserved regions of the two genomes. Based on sequence similarity, only 43 (Additional file 2: Table S14) out of the, in total, 293 ncRNAs located in these regions seem to occur in the *B. subtilis* transcriptome [43], emphasizing the differences of the two closely related species. Comparisons to two earlier *B. subtilis* transcriptome studies show similar low levels of accordance [56,57]. However, as mentioned above, it is also possible that the identified antisense ncRNAs partly derive from spurious transcription events [43], and hence do not introduce a species-specific effect.

For *B. subtilis*, 22 sRNAs have been validated experimentally [43,58]. Comparison to Rfam and/or comparison of genomic locations facilitated the detection of eleven of these sRNAs in the transcriptome of *B. licheniformis* (Additional file 2: Table S15). These include, in addition to the mentioned five housekeeping sRNAs [13], two regulatory RNAs with well-known function in *B. subtilis*: SR1 and RnaA [59,60]. The other RNAs found in *B. licheniformis* are BsrI, CsfG, SurC and RsaE [61-64]. The *B. subtilis* sRNAs which could not be confirmed in *B. licheniformis* originate from loci with no conserved gene pattern in this organism and thus may contribute to the differences between the two species. Jahn et al. [65] described the toxin-antitoxin system BsrG/SR4 located in the SPβ prophage region of *B. subtilis*. Although *B. licheniformis* does not harbor a homolog of the SPβ prophage, two distinct transcripts were found to encode peptides similar to the BsrG toxin (Additional file 1: Figure S5). Additionally, the transcriptional activity of the corresponding loci revealed pairs of overlapping transcripts from both strands (Figure 3 and Additional file 1: Figure S5) as shown for the BsrG/SR4 type toxin-antitoxin system. Therefore both newly identified ORFs were annotated as BsrG-like peptides (BLi05015 and BLi05038). Furthermore, the antisense transcripts (*indep* RNAs BLi_r1034 and BLi_r2780) resemble the SR4 antitoxin, especially in stem loops SL3, SL4 and TSL [65] directly antisense to the BsrG-encoding mRNA.

Conclusions

The presented study generated substantial data on the transcriptional activity of *B. licheniformis* within five relevant growth stages of an industrial-oriented fermentation process. A detailed analysis of the transcriptome data enabled us to accomplish a high quality functional genome reannotation of *B. licheniformis* DSM13 (Figure 10, Ring 4). The integration of the reannotation and the transcriptionally active regions (Figure 10, Ring 1&2) resulted in the identification and quantification of hundreds of RNA based regulatory elements as well as protein encoding genes.

In total, 3314 RNA features have been sorted into ten functional classes (Figure 4). 1433 5'UTRs and 1365 3'UTRs (Figure 10, Ring 8) as well as 461 ncRNAs (Figure 10, Ring 7) and 55 antisense intergenic read-through (A_{rt}) transcripts have been identified. A striking observation was the identification of 855 RNA features, which mapped antisense to annotated genomic features

Table 3 Selected non-coding RNAs (ncRNAs)

RNA feature	Start	Stop	Length	Class	Rfam	Upstream gene	Downstream gene	Antisense genes	NPKM value*
BLi_r0016	30440	30837	398	*indep*	Scr	*tadA*	*dnaX*		223117
BLi_r0026	54282	49490	4793	A_{misc}	RnaA			*metS, yabD, yabE, rnmV*	18
BLi_r0086	212998	213153	156	*indep*		*thrZ*	BLi00235		242952
BLi_r0253	413569	412086	1484	A_{misc}				BLi00412, BLi00413	14
BLi_r0415	617488	617027	462	A_{misc}				*thiL*	6
BLi_r0451	653178	652888	291	A_{misc}				BLi00649	15139
BLi_r0844	1082413	1082091	323	*indep*		*yhaA1*	*hit*		8039
BLi_r0872	1108968	1109111	144	A_{I}				*apr*	25689
BLi_r1000	1262387	1262504	118	*indep*	RsaE	*pepF*	*yjbL*		5791
BLi_r1034	1300088	1300311	224	*indep*		*pbpE1*	BLi01297		1902
BLi_r1306	1639742	1639946	205	A_{misc}	SR1			*speA*	502
BLi_r1347	1673741	1673635	107	A_{misc}	CsfG			*ylbG, ylbH*	5366
BLi_r1424	1898847	1898597	251	*indep*		*yqeD*	BLi01936		2023
BLi_r1454	1929530	1929709	180	*indep*	BsrB				26604
BLi_r1474	1960434	1960112	323	*indep*		BLi02008			61592
BLi_r1596	2101575	2102388	814	A_{I}				*cysP2*	12
BLi_r1645	2156680	2156597	84	*indep*		*yobS*	*yndG*		13545
BLi_r1808	2302203	2301803	401	*indep*	RnpB	*gpsB*	*ypsC*		55930
BLi_r1834	2355791	2356137	347	A_{misc}				*folE*	1
BLi_r2049	2634028	2634231	204	A_{misc}	SurC			*dnaK*	29
BLi_r2163	2770133	2769931	203	*indep*	BsrA	*aspS*	*yrvM*		389442
BLi_r2390	3060847	3062662	1816	A_{misc}				*ytvB, yttB*	11
BLi_r2624	3299320	3299025	296	*indep*	BsrI	*yurZ*	BLi03452		514
BLi_r2645	3325524	3323636	1889	$A_{3'UTR}$		*yirB*		*cssR, cssS*	112
BLi_r2758	3469610	3469009	602	*indep*	SsrA	*smpB*	BLi03638		78287
BLi_r2780	3485163	3485306	144	*indep*		BLi03658	BLi03670		3747
BLi_r2828	3552518	3552467	52	*indep*		*trxB*	*yvcI*		15924
BLi_r2863	3611192	3612078	887	A_{misc}				*degU, degS*	5
BLi_r2925	3692599	3692663	65	A_{I}				BLi03865	14110
BLi_r3203	4050815	4050900	86	A_{misc}				*bglP*	35711

*(pooled RNA-Seq data).

(Figure 10, Ring 6). Notably antisense RNA features have been found in each of the functional classes and include transcripts of a size range from 38 to 6348 base pairs in length. We have identified both: constitutively as well as growth phase dependently expressed RNA features.

Our data represent a solid amount of knowledge on regulatory elements which orchestrate the cellular activities of *B. licheniformis* during the succession of growth phases within a productive fermentation. To generate an overview of the functional diversity of the identified RNA features, all instances have been screened against the Rfam database. This approach resulted in hits to experimentally well characterized RNA features known from *B. subtilis* and other relatives, as well as in a multitude of so far unknown RNA features without any Rfam hit. The knowledge on genes and regulatory RNA features which are transcriptionally active during an industrial-oriented fermentation enables an excellent access to a rational strain design approach for the optimization of *B. licheniformis* as industrial workhorse. Especially the regulatory features which represent differences to the model organism *B. subtilis* give new insights to the still open question what makes strains of the species *B. licheniformis* superior to *B. subtilis* strains in terms of protease production capacity in industrial applications [2]. In the future it may be promising to correlate the transcriptional activity of the RNA features to the corresponding protein expression patterns.

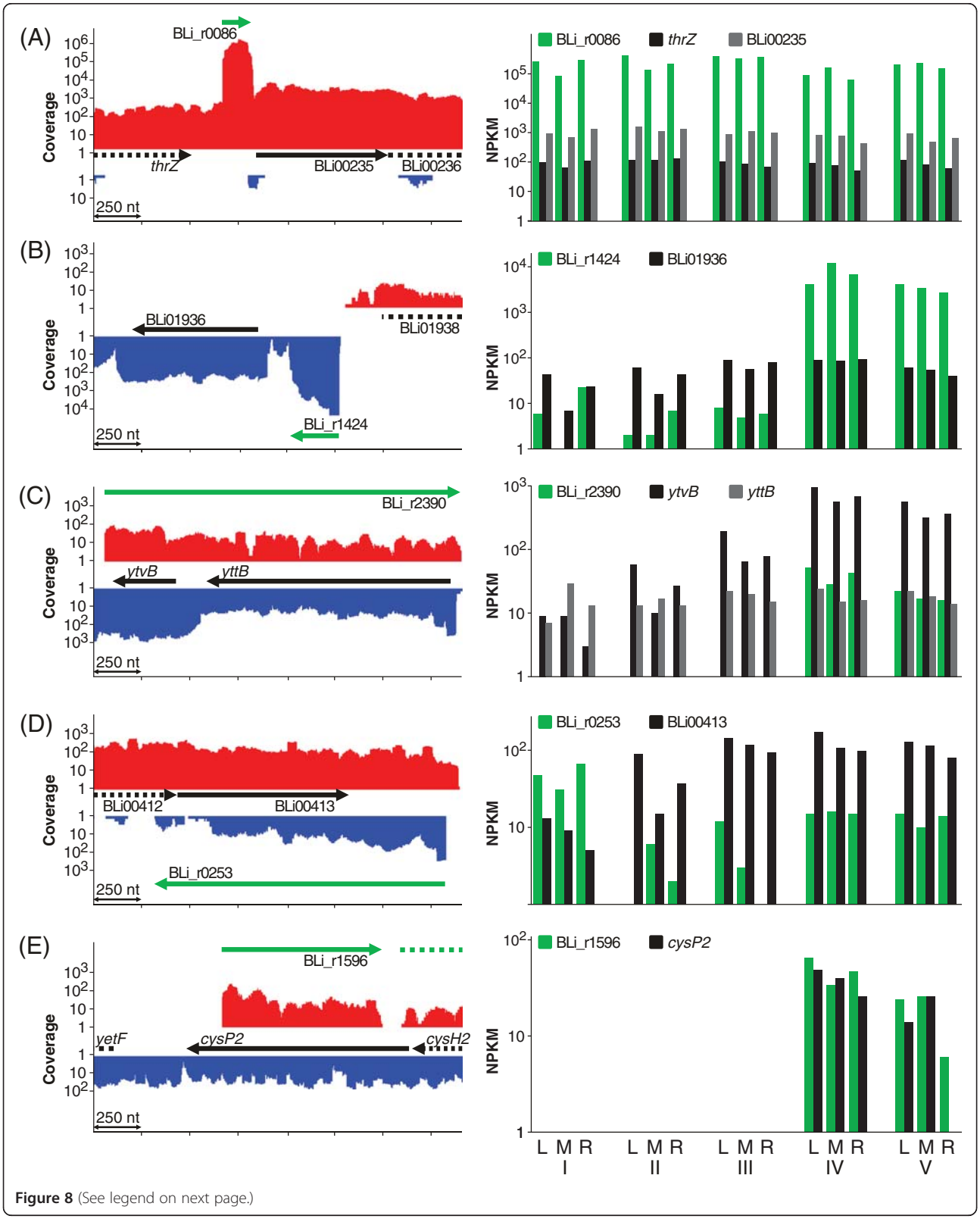


Figure 8 (See legend on next page.)

(See figure on previous page.)
Figure 8 Non-coding RNAs (ncRNAs). *(Left)* Sum of transcriptional activities from all 15 replicates (*pooled RNA-Seq data*). Black arrows indicate genes and green arrows the identified ncRNAs. *(Right)* Log-transformed NPKM values of ncRNAs and adjacent genes for single samples. **(A)** *Indep* RNA BLi_r0086 is transcribed constitutively with a length of 156 nt and located between the genes of threonyl-tRNA synthetase (*thrZ*, BLi00234) and a hypothetical protein (BLi00235). Both adjacent genes are also transcribed constitutively, but are less abundant by four and three orders of magnitude, respectively. **(B)** The differentially expressed *indep* transcript BLi_r1424 is located between the gene of a hypothetical protein (BLi01936) and a pseudogene (*yobN*, BLi01938) with a length of 251 nt. The TSS could be confirmed by dRNA-Seq. In the three early conditions the BLi_r1424 transcription level is low, but NPKM values of more than 12,000 were recorded during the productive stages of the fermentation process. A direct transcriptional connection to the adjacent BLi01936 is not visible from the shown NPKM values. **(C)** BLi_r2390 antisense to *ytvB* (BLi03176) and *yttB* (BLi03177) is an example for long antisense ncRNAs. The A_{misc} RNA occurs only in the later stages of the fermentation process, parallel to a distinct increase in transcriptional activity of *ytvB*, but does not exceed it regarding the NPKM value. **(D)** One example suggesting a regulatory function of A_{misc} RNAs is BLi_r0253, oriented antisense to BLi00413. In the earliest stage the asRNA shows stronger transcription than the corresponding gene, but in all later stages the asRNA is only weakly transcribed. This might indicate a silencing effect in the exponential stage. **(E)** A_i RNA BLi_r1596 localized antisense to the gene of the sulfate permease CysP2 (BLi02153). The transcription of both, the ncRNA and the protein-coding gene, starts during the late stages of the fermentation process.

Methods

Bacterial strain and fermentation conditions

Bacillus licheniformis MW3Δspo (kindly provided by F. Meinhardt and St. Wemhoff, University of Münster) was used for the fermentation experiments. *B. licheniformis* MW3Δspo is a derivate of the *B. licheniformis* wild type strain DSM13, bearing three deletions: Δ*hsdR* and Δ*hsdR2* coding for restriction endonucleases [10] and Δ*yqfD* [66] to prevent the production of viable spores and thus the long-term contamination of the used fermenters [67].

Fermentation was carried out for 46 h in aerated 16 L fermenters with a culture volume of 6 L at 39°C. Medium contained 12% w/v of a complex nitrogen source, 57 mM KH_2PO_4, 21 mM $(NH_4)_2SO_4$, 0.53 mM $Mn(II)SO_4$, 0.17 mM $Fe(II)SO_4$, 2.0 mM $CaCl_2$ * 2 H_2O, 5.7 mM $MgSO_4$, 0.4% v/v PPG200, 0.03 mM tetracycline and 3% w/v glucose. The pH value was regulated to a set point of 7.9 with sodium hydroxide solution. Glucose-feed was started after exceeding the point of biphasic growth.

RNA isolation and preparation

5 mL of the harvested cells were mixed with 5 mL of RNAprotect Bacteria Reagent (Qiagen) directly upon sampling. After 10 min incubation at room temperature the samples were centrifuged at 4500× *g*, the supernatant was removed, the sample was snap-frozen in liquid nitrogen and finally stored at −80°C. The cells were separated from the remainders of the fermentation broth by washing repeatedly with Buffer RLT (Qiagen). Subsequent RNA isolation was carried out with a modified protocol of the RNeasy Midi Kit (Purification of RNA including small RNAs using the RNeasy Midi Kit RY39 Apr-09, Qiagen) to retain short RNAs. The cells were disintegrated with the ball mill Mikro-Dismembrator U (B. Braun Biotech) in 400 μL Buffer RLT and afterwards resuspended in 1.4 mL Buffer RLT and 2.7 mL pure ethanol. The initial washing step of the column was done using 4 mL Buffer RWT (Qiagen). The DNA was digested successively with two different DNases (TURBO™ DNase, Ambion and DNase I recombinant, Roche), with a purification step after the first treatment. Purification was performed with a protocol adapted for small RNA purification of the RNeasy MinElute Cleanup Kit (Qiagen). Instead of 250 μL, 675 μL pure ethanol were added to the RNA before binding to the column to shift the binding capacity of the column. A control PCR with 35 cycles was conducted to confirm complete DNA removal. Depletion of rRNA was obtained using the MICROBExpress™ Bacterial mRNA Enrichment Kit (Ambion) according to manufacturer's instructions. The following purification step was also carried out with the described adaption to the RNeasy MinElute Cleanup Kit.

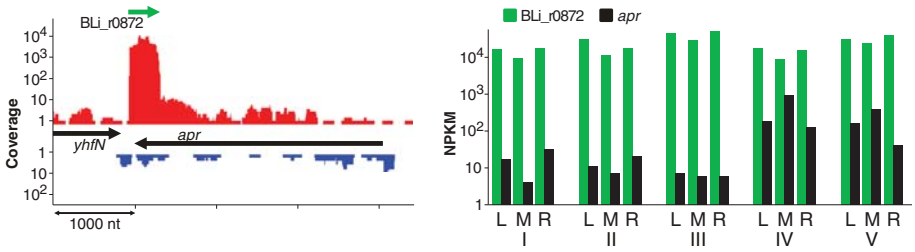


Figure 9 Antisense RNA against Subtilisin Carlsberg. *(Left)* Transcriptional activities (sample L-III) of *apr*, the gene which encodes Subtilisin Carlsberg, and the A_i RNA BLi_r0872 (green), which is antisense to the 3'UTR of *apr*. *(Right)* Log-transformed NPKM values of BLi_r0872 and *apr*.

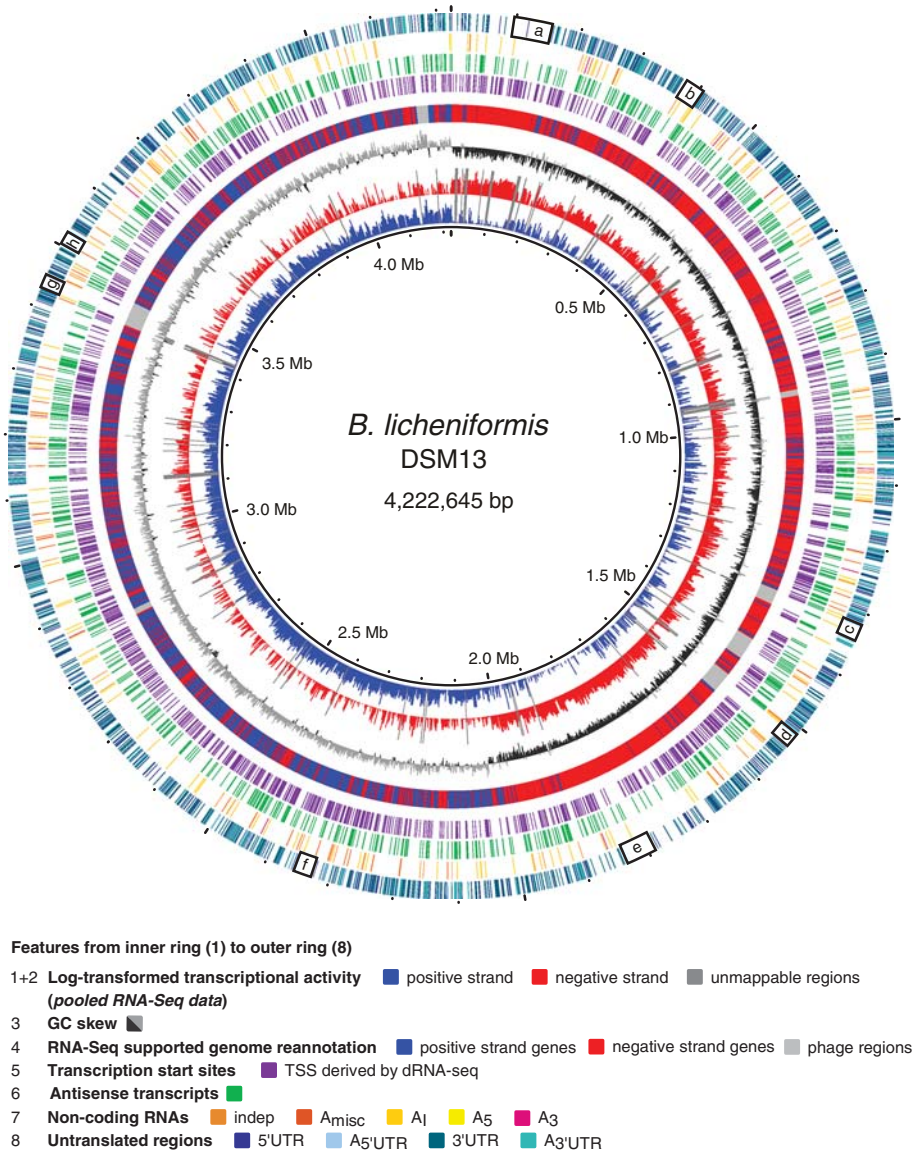


Figure 10 Circular plot of transcriptional activity and identified RNA features. Combined depiction of reannotated genes and transcriptional activity of *B. licheniformis*. Unmappable regions, GC skew, transcription start sites, non-coding RNAs, untranslated regions and antisense transcripts are also shown. 5'UTRs and 3'UTRs are evenly distributed over the whole chromosome of *B. licheniformis*, except for regions a – h: these regions contain long operon structures (a: ribosomal superoperon, b: *lch* operon, e: *fla/che* operon, f: *trp* operon, h: *eps* operon) or prophage regions with low transcriptional activity (c, d and g). The classification scheme corresponds to Figure 4. (*indep*) ncRNA, not antisense to any mRNA; (A_{misc}) ncRNA partially antisense to an mRNA transcript or antisense to more than one gene; (A_I) ncRNA completely antisense to a protein-coding gene; (A_5) ncRNA exclusively antisense to the 5'UTR of an mRNA; (A_3) ncRNA exclusively antisense to the 5'UTR of an mRNA; (5'UTR) 5'untranslated region of an mRNA; ($A_{5'UTR}$) 5'UTRs completely or partially antisense to a protein-coding region; (3'UTR) 3'untranslated region of an mRNA transcript; ($A_{3'UTR}$) 3'UTRs completely or partially antisense to a protein-coding region.

Library construction and sequencing

cDNA libraries were prepared by vertis Biotechnologie AG, Germany (www.vertis-biotech.com). For whole transcriptome libraries [28,33] the RNA samples were fragmented by ultrasound and dephosphorylated with Antarctic phosphatase. After polynucleotide kinase treatment the RNA was poly(A)tailed and an RNA adapter was ligated to the 5'phosphate. cDNA synthesis was accomplished by the use of poly(T) adapters and M-MLV reverse transcriptase. The subsequent PCR was carried out with cycle numbers between nine and twelve. The construction of the libraries for the dRNA-Seq was performed as described by Sharma et al. [32], supplemented by an additional treatment with polynucleotide kinase after the fragmentation step to allow removal of fragments previously not phosphorylated. The samples were incubated with Terminator™ 5'-Phosphate-

Dependent Exonuclease (Epicentre) and a poly(A) tail was ligated to the 3'end of the transcripts. Hereafter an incubation step with tobacco acid pyrophosphatase and the ligation of an RNA adapter to the 5'end was conducted. Reverse transcription was processed as described above, the cycle numbers of the following PCR were 14 or 15. The RNA-Seq libraries as well as the libraries for dRNA-Seq were size fractioned in the range of 200 to 400 nt on agarose gels and then sequenced on an Illumina HiSeq 2000 machine with a read length of 50 nt.

In silico sequence read processing

Initially, all sequence reads mapping to *B. licheniformis* rRNA and tRNA genes according to BLAST analysis were removed. The remaining reads were processed in a multi-step procedure to ensure the reliability of the read mappings used for the analysis of the transcriptional activity of the genome and to estimate the quality of the RNA-Seq data. All reads which mapped over the full read length of 50 bases with 98% or sequence identity were used for further analyses. Additionally, a distinct bit score was required to ensure an unambiguous assignment to one locus. All discarded reads were screened with relaxed similarity quality criteria vs. the *B. licheniformis* genome. 75% of these reads generated hits and were therefore assigned as bad quality *B. licheniformis* reads. The remaining reads (approximately 3% of the total generated sequence) cannot be mapped on the genome. A detailed sequence analysis of these unmappable reads revealed that they mainly contain poly(A) tails or concatenated adapters and therefore represent methodic artifacts. All datasets were depleted for plasmid-mapping reads and have been deposited in NCBI Sequence Read Archive database under accession number SRP018744 (Additional file 2: Table S16).

To obtain the maximal number of features, a dataset containing all reads of the 15 samples was prepared and is referred to as *pooled RNA-Seq data*. To reduce putative background noise, all reads with coverage of one and no intersecting or adjacent reads were omitted prior to combination of the datasets. This is done to reduce transcriptional activity that was not replicated within a dataset in order to avoid incorrect extension of predicted features (e.g. transcriptional activity from leaky termination masking or extending 5'UTR extensions due to overlap). The generation of five datasets describing each sampling point was processed accordingly.

Expression strength values

The analytical methods used to process the 15 generated RNA-Seq datasets require the use of single nucleotide activities instead of read mappings. This makes RPKMs [68] inapplicable as a measure of transcriptional activity. Instead, we defined the **n**ucleotide activity **p**er **k**ilobase of exon model per **m**illion mapped reads (NPKM) value. An NPKM is defined as:

$$NPKM\ (n,m) = 10^9 \frac{\sum_{i=n}^{m} f(i)}{\sum_{i=1}^{m} g(i)(m-n)}$$

Where n and m are the start and stop of the region of interest, $f(i)$ is the base activity of base i on a specific strand and $g(i)$ is the sum of the activities of base i of positive and negative strands.

NPKM values are a derivate of RPKMs [68], adapted to per base nucleotide activities. They are designed to be functionally equivalent to RPKMs, albeit they are more accurate due to the single base-resolution. We are aware that RPKMs and therefore NPKMs do not account for sequencing-based bias [69]. Although sequencing-based bias produces some local errors, the overall comparability of active genomic regions is still possible.

Untranslated regions

5' and 3' UTRs were considered as regions of continuous, non-interrupted transcriptional activity upstream or downstream of annotated genomic features, respectively. The boundary of an identified 5'UTRs was set at the point of the rising of the continuous transcript from zero transcriptional activity. The boundary of a 3' UTRs was accordingly set at the point of the downshift of the continuous transcript to zero transcriptional activity.

The analysis of 5' and 3'untranslated regions was aimed to find the longest UTR, as the longest transcript should cover all possible alternative UTRs and contain all transcribed regulatory elements. Therefore, the computational analysis was based on the *pooled RNA-Seq data*. Few 5' and 3'UTRs were manually extended on account of adjacent transcripts which are only separated from the UTR by a very short downshift and potentially are part of the UTR. To exclude that the resulting UTRs correspond to previously not annotated protein genes, searches versus the InterPro and the UniProtKB/Swiss-Prot databases were performed [70,71].

5' and 3'UTRs which are antisense to an adjacent gene on the opposite strand were classified as $A_{5'UTR}$ and $A_{3'UTR}$. The respective UTRs were computationally examined and assigned to be antisense when their overlap to an opposite gene exceeded 100 nt in length.

Intergenic read-through transcripts localized antisense to an opposite gene were determined manually and classified as A_{rt}.

Non-coding RNA features

The RNA-Seq data were scanned for transcriptionally active regions that were clearly separated from the transcripts corresponding to any annotated gene or its untranslated regions. This primary computational search

identified transcripts which were either located on the opposite strand of a protein-coding gene, in intergenic regions or any other region of the chromosome. The boundaries of the identified transcripts were set to those nucleotides with the first and last occurrence of transcriptional activity higher than zero of the corresponding transcriptional unit. NPKM values for the resulting loci were generated from each of the 15 datasets. Subsequently, all results from the computational search were evaluated as depicted in Additional file 1: Figure S6A to approve the reliability of the identified ncRNAs. Searches vs. the InterPro and the UniProtKB/Swiss-Prot databases were performed to exclude the possibility that the resulting non-coding RNA features correspond to non-annotated protein genes [70,71]. Subsequently, the non-coding transcripts were subdivided into the ncRNA classes described in Figure 4 (A_3, A_5, A_I, A_{misc}, and *indep*).

The class *indep* comprises all identified ncRNAs that are not located antisense to any protein-coding gene or its respective untranslated regions. Several transcripts of this category were added manually as this class comprises RNA transcripts which could not clearly be distinguished from surrounding mRNAs by complete downshifts of transcriptional activity, but were detected by their remarkably higher abundance.

The categories A_3, A_5, A_I and A_{misc} comprise ncRNAs which are localized antisense to protein-coding genes or their respective untranslated regions. The class A_I contains all ncRNAs with an antisense localization solely towards a protein-coding gene. The class A_5 contains all ncRNAs with an antisense localization solely towards the 5'UTR of an opposite mRNA. The class A_3 contains all ncRNAs with an antisense localization solely towards the 3'UTR of an opposite mRNA. The class A_{misc} contains all ncRNAs with an antisense localization towards more than one protein-coding gene and all ncRNAs which are only partially antisense to an mRNA transcript.

Analysis of dRNA-Seq reads

Transcriptional start sites were determined by the identification of significant increases of the log-scaled expression strength of the dRNA-Seq data from succeeding bases greater than ln 4. The reference value of ln 4 was empirically determined based on the observation that ln 4 represents the smallest expression strength increase for TSS present across all samples of one sampling point. In a second step, all TSS in promoter regions of rRNA or tRNA genes and all TSS being apart less than 20 bp were excluded. TSS matching the boundaries of RNA-Seq predicted 5'UTRs or ncRNAs were determined accordingly to the flow chart depicted in Additional file 1: Figure S6B.

Transcriptome Viewer

Additionally, the gained RNA-Seq data were used to generate logarithmic scaled, color coded graphs representing strand-specific transcription.

Operon prediction

Operon predictions based on whole transcriptome sequencing, dRNA-Seq transcription start sites, and operon and transcription terminator site determination with DOOR [72], OperonDB [73], and TransTermHP [74]. Operon predictions were curated manually as described by Sharma et al. [32], regarding especially level shifts in transcriptional activity.

Reannotation

Functional reannotation was carried out using the ERGO software tool (Integrated Genomics, Chicago, USA) [75] and the IMG/ER (Integrated Microbial Genomes/Expert Review) system [45]. Subsequent manual curation was based on the results of a bidirectional BLAST analysis comprising *B. subtilis*, *B. pumilus* and related, manually annotated organisms, the comparisons to UniProtKB/Swiss-Prot and UniProtKB/TrEMBL databases [71] and the analysis of functional domains with InterProScan [70]. The annotation of new genes and the correction of reading frames was based on transcriptional activity and was performed upon analysis of GC frame plots, ribosome-binding sites and −10 and −35 promoter regions using Artemis v12 [76] and comparisons to UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, and InterProScan [70,71]. The removal of gene annotations relied on the combined evaluation of GC frame plots, ribosome-binding sites and −10 and −35 promoter regions using Artemis v12 [76] and comparisons to UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, and InterProScan [70,71]. The absence of transcriptional activity was not used to support the removal of gene annotations. Prophage regions have been annotated by an initial bioinformatic search using Prophagefinder [77] followed by manual evaluation of the candidate regions. Based on the existence of GC content deviations, genes in these regions with significant similarities to known prophages and the identification of insertion repeats, genomic regions were assigned as prophages. The annotation followed the principles of prophage annotation outlined by Casjens [78]. The reannotated data set has been used to update the *B. licheniformis* DSM13 genome data initially submitted by Veith et al. [1] and is now available at NCBI under accession number AE017333.1.

Clustering of ncRNAs

Cluster analysis to elucidate the fundamental types of ncRNA expression profiles was performed based on the respective NPKM values (Additional file 2: Table S2). To

ensure that the data of each replicate are sufficiently reliable, t-tests were performed with MeV [79]. For at least three out of the five samples, the respective ncRNA had to have a P value <0.15 to be taken into further analysis, as described by Koburger et al. [80]. Furthermore, all ncRNAs taken into analysis had to have a minimal NPKM value >10. Means of the replicates of each sampling point were built and z-score transformation was performed. The number of clusters was determined by *Figure of merit* (FOM) analysis, which basically is an estimate of the predictive power of a clustering algorithm [81]. Clusters were generated by employing *k*-means clustering [47] with Euclidian distances in the MeV software [79] and subsequent manual curation.

Utilized software and databases

ACT and Mauve

The comparison of RNA features from *B. licheniformis* with the reference genome *B. subtilis* was based on sequence similarity analyzed with ACT v11, the Artemis comparison tool [82]. Quantification of ncRNAs located in conserved or not-conserved loci, was done employing the progressive Mauve alignment tool [83].

baySeq

Determination of constitutive or differential expression of the RNA features was employed with baySeq [49], which uses an empirical Bayes approach assuming a negative binomial distribution and is capable of dealing with multi-group experimental designs. Input data were generated by counting the reads referring to every gene.

DOOR and OperonDB

Predictions for operons were thankfully downloaded from the DOOR Database of prOkaryotic OpeRons [72] and OperonDB [73].

Gem mappability

The determination of the genome mappability was calculated for a read length of 50 nt with the Gem mappability program [84].

MeV

Cluster analysis was performed using the Multiexperiment Viewer v4.8 [79].

Rfam

Annotation of *cis*-regulatory elements and small RNAs was carried out by Infernal searches [85] of RNA features versus the Rfam database [46].

TransTermHP

Transcription terminators pre-computed with TransTermHP v2.07 were gratefully downloaded from transterm.cbcb.umd.edu [74]. 3'UTRs were checked for terminators as described by Martin et al. [86]. Terminators were considered as internal if they were located at least 50 nt upstream of the end of the transcript.

Northern blot analysis

B. licheniformis DSM13 was cultivated at 37°C and 160 rpm in a 5 L Erlenmeyer flask on defined minimal medium [87]. Cells were harvested at OD_{600} 1 and 4.5 and after having reached the stationary phase for at least 2 h. *Escherichia coli* DH5α was cultivated in Luria broth at 37°C and 180 rpm to an OD_{600} of 2. RNA was isolated as described in *RNA isolation and preparation*. Digoxigenin-labeled RNA probes were prepared by *in vitro* transcription with T7 RNA polymerase (DIG Northern Starter Kit, Roche). Templates for *in vitro* transcription were generated by PCR using primer pairs (Additional file 2: Table S17) containing a primer flanked with the T7 promoter sequence. Gel electrophoresis of the RNA was carried out using a 1% agarose formaldehyde MOPS gel [88] with 100 V applied for 2,5 h. RNA was transferred to the membrane (Nylon Membranes, positively charged, Roche) via vacuum blotting with the Amersham VacuGene XL Vacuum Blotting System (GE Healthcare) using the recommended protocol. The RNA probe hybridization procedure was performed following the manufacturer's instructions (DIG Northern Starter Kit, Roche). Detection was accomplished with ChemoCam Imager (Intes). RiboRuler High Range RNA Ladder (Thermo Scientific) ranging from 200 to 6000 nt was used as RNA marker.

Additional files

Additional file 1: Figure S1. Distribution of whole transcriptome sequencing reads. **Figure S2:** Comparative operon prediction. **Figure S3:** Northern blot confirmation of non-coding RNAs. **Figure S4:** Antisense RNAs with putative impact on productivity. **Figure S5:** BsrG/SR4-like loci in *B. licheniformis*. **Figure S6:** Work flow charts.

Additional file 2: Table S1. Whole transcriptome sequencing reads. **Table S2:** NPKM values of RNA features. **Table S3:** NPKM values of all genes. **Table S4:** Differential RNA-Seq reads. **Table S5:** Transcription start sites. **Table S6:** Identified RNA features. **Table S7:** Predicted operons. **Table S8:** Corrected genes. **Table S9:** New genes. **Table S10:** Removed genes. **Table S11:** Predicted *cis*-regulatory elements. **Table S12:** Comparison of *cis*-regulatory elements known from *B. subtilis*. **Table S13:** Cluster analysis of ncRNA expression profiles. **Table S14:** Comparison of ncRNAs to *B. subtilis*. **Table S15:** Comparison of small RNAs from *B. subtilis* to identified ncRNAs. **Table S16:** Sequence Read Archive accession. **Table S17:** Primer pairs for Northern blots.

Abbreviations

°C: Degrees Celsius; μL: Microliter; A: Adenine; asRNA: Antisense RNA; bp: Base pairs; C: Cytosine; cDNA: Complementary DNA; dRNA-Seq: Differential RNA sequencing; g: Gram; *g*: Gravitational constant; G: Guanine; h: Hours; L: Liter; ln: Natural logarithm; Mb: Megabase pairs; min: Minute; mL: Milliliter; mM: Millimolar; mRNA: Messenger RNA; ncRNA: Non-coding RNA; NPKM: Nucleotide activity per kilobase of exon model per million mapped reads; nt: Nucleotides; OD: Optical density; ORF: Open reading frame; PCR: Polymerase chain reaction; RNA-Seq: RNA sequencing; rpm: Revolutions per minute; rRNA: Ribosomal RNA; T: Thymine;

TAP: Tobacco Acid Pyrophosphatase; TEX: Terminator™ 5′-Phosphate-Dependent Exonuclease; tmRNA: Transfer-messenger RNA; tRNA: Transfer RNA; TSS: Transcription start sites; UTR: Untranslated region; V: Volt; v/v: Volume per volume; w/v: Weight per volume.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SW performed the experiments, analyzed data and wrote paper, SD developed the analysis tools, RH performed northern blots, JB and SE provided industrial fermentation facilities and performed the fermentation, SV prepared submission of genome and transcriptome data, RD wrote paper and provided research facilities, HL wrote paper, designed research and analyzed data. All authors read and approved the final version of the manuscript.

Acknowledgements
This study was funded by the Bundesministerium für Bildung und Forschung (FKZ-0315387).
The authors would like to thank the Henkel Company for kind access to their fermentation facility. We are grateful for expert technical assistance by Ayhan Aydemir and Maik Schlieper.

Author details
[1]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077, Göttingen, Germany. [2]Henkel AG & Co. KGaA, Henkelstraße 67, D-40191, Düsseldorf, Germany.

Received: 25 February 2013 Accepted: 25 September 2013
Published: 1 October 2013

References
1. Veith B, Herzberg C, Steckel S, Feesche J, Maurer K-H, Ehrenreich P, Bäumer S, Henne A, Liesegang H, Merkl R, Ehrenreich A, Gottschalk G: **The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential.** *J Mol Microbiol Biotechnol* 2004, **7:**204–211.
2. Schallmey M, Singh A, Ward O: **Developments in the use of *Bacillus* species for industrial production.** *Can J Microbiol* 2004, **50:**1–17.
3. Maurer K-H: **Detergent proteases.** *Curr Opin Biotechnol* 2004, **15:**330–334.
4. Çalık P, Takac S, Çalık G, Özdamar T: **Serine alkaline protease overproduction capacity of *Bacillus licheniformis*.** *Enzyme Microb Tech* 2000, **26:**45–60.
5. Çalık P, Çalık G, Takaç S, Ozdamar TH: **Metabolic flux analysis for serine alkaline protease fermentation by *Bacillus licheniformis* in a defined medium: effects of the oxygen transfer rate.** *Biotechnol Bioeng* 1999, **64:**151–167.
6. Degering C, Eggert T, Puls M, Bongaerts J, Evers S, Maurer K-H, Jaeger K-E: **Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides.** *Appl Environ Microbiol* 2010, **76:**6370–6376.
7. Gupta R, Beg QK, Lorenz P: **Bacterial alkaline proteases: molecular approaches and industrial applications.** *Appl Microbiol Biotechnol* 2002, **59:**15–32.
8. Nielsen AK, Breüner A, Krzystanek M, Andersen JT, Poulsen TA, Olsen PB, Mijakovic I, Rasmussen MD: **Global transcriptional analysis of *Bacillus licheniformis* reveals an overlap between heat shock and iron limitation stimulon.** *J Mol Microbiol Biotechnol* 2010, **18:**162–173.
9. Rey M, Ramaiya P, Nelson N, Brody-Karpin S: **Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species.** *Genome Biol* 2004, **5:**R77.
10. Waschkau B, Waldeck J, Wieland S, Eichstädt R, Meinhardt F: **Generation of readily transformable *Bacillus licheniformis* mutants.** *Appl Microbiol Biotechnol* 2008, **78:**181–188.
11. Nahrstedt H, Waldeck J, Gröne M, Eichstädt R, Feesche J, Meinhardt F: **Strain development in *Bacillus licheniformis*: Construction of biologically contained mutants deficient in sporulation and DNA repair.** *J Biotechnol* 2005, **119:**245–254.
12. Madan Babu M, Teichmann SA, Aravind L: **Evolutionary dynamics of prokaryotic transcriptional regulatory networks.** *J Mol Biol* 2006, **358:**614–633.
13. Romby P, Charpentier E: **An overview of RNAs with regulatory functions in gram-positive bacteria.** *Cell Mol Life Sci* 2010, **67:**217–237.
14. Storz G, Vogel J, Wassarman KM: **Regulation by small RNAs in bacteria: expanding frontiers.** *Mol Cell* 2011, **43:**880–891.
15. Gottesman S, Storz G: **Bacterial small RNA regulators: versatile roles and rapidly evolving variations.** *Cold Spring Harb Perspect Biol* 2011, **3:**a003798.
16. Waters L, Storz G: **Regulatory RNAs in bacteria.** *Cell* 2009, **136:**615–628.
17. Schmiedel JM, Axmann IM, Legewie S: **Multi-target regulation by small RNAs synchronizes gene expression thresholds and may enhance ultrasensitive behavior.** *PLoS One* 2012, **7:**e42296.
18. Georg J, Hess WR: ***cis*-antisense RNA, another level of gene regulation in bacteria.** *Microbiol Mol Biol R* 2011, **75:**286–300.
19. Brantl S: **Regulatory mechanisms employed by *cis*-encoded antisense RNAs.** *Curr Opin Microbiol* 2007, **10:**102–109.
20. Sesto N, Wurtzel O, Archambaud C, Sorek R, Cossart P: **The excludon: a new concept in bacterial antisense RNA-mediated gene regulation.** *Nat Rev Microbiol* 2013, **11:**75–82.
21. Sorek R, Cossart P: **Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.** *Nat Rev Genet* 2010, **11:**9–16.
22. Güell M, Yus E, Lluch-Senar M, Serrano L: **Bacterial transcriptomics: what is beyond the RNA horiz-ome?** *Nat Rev Microbiol* 2011, **9:**658–669.
23. Breaker RR: **Prospects for riboswitch discovery and analysis.** *Mol Cell* 2011, **43:**867–879.
24. Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, Burns SC, Penalva LOF: **Before It Gets Started: Regulating Translation at the 5′ UTR.** *Comp Funct Genom* 2012, **2012:**475731.
25. Zoll J, Heus H, van Kuppeveld F, Melchers W: **The structure-function relationship of the enterovirus 3′-UTR.** *Virus Res* 2009, **139:**209–216.
26. Kaberdin VR, Bläsi U: **Translation initiation and the fate of bacterial mRNAs.** *FEMS Microbiol Rev* 2006, **30:**967–979.
27. Evguenieva-Hackenberg E, Klug G: **New aspects of RNA processing in prokaryotes.** *Curr Opin Microbiol* 2011, **14:**587–592.
28. Vivancos AP, Güell M, Dohm JC, Serrano L, Himmelbauer H: **Strand-specific deep sequencing of the transcriptome.** *Genome Res* 2010, **20:**989–999.
29. Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, Rose G, Perkins TT, Parkhill J, Dougan G, Young DB: **Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*.** *PLoS Pathog* 2011, **7:**e1002342.
30. Vockenhuber M-P, Sharma CM, Statt MG, Schmidt D, Xu Z, Dietrich S, Liesegang H, Mathews DH, Suess B: **Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*.** *RNA Biol* 2011, **8:**468–477.
31. Schlüter J-P, Reinkensmeier J, Daschkey S, Evguenieva-Hackenberg E, Janssen S, Jänicke S, Becker JD, Giegerich R, Becker A: **A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*.** *BMC Genomics* 2010, **11:**245.
32. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J: **The primary transcriptome of the major human pathogen *Helicobacter pylori*.** *Nature* 2010, **464:**250–255.
33. Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH: **Structure and complexity of a bacterial transcriptome.** *J Bacteriol* 2009, **191:**3203–3211.
34. Dötsch A, Eckweiler D, Schniederjans M, Zimmermann A, Jensen V, Scharfe M, Geffers R, Häussler S: **The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing.** *PLoS One* 2012, **7:**e31092.
35. Kristoffersen SM, Haase C, Weil MR, Passalacqua KD, Niazi F, Hutchison SK, Desany B, Kolsto A-B, Tourasse NJ, Read TD, Okstad OA: **Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium.** *Genome Biol* 2012, **13:**R30.
36. Wurtzel O, Sesto N, Mellin JR, Karunker I, Edelheit S, Bécavin C, Archambaud C, Cossart P, Sorek R: **Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species.** *Mol Syst Biol* 2012, **8:**1–14.
37. Lemire S, Figueroa-Bossi N, Bossi L: **Bacteriophage crosstalk: coordination of prophage induction by trans-acting antirepressors.** *PLoS Genet* 2011, **7:**e1002149.
38. Zheng X, Hu G-Q, She Z-S, Zhu H: **Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes.** *BMC Genomics* 2011, **12:**361.

39. Nakagawa S, Niimura Y, Miura K, Gojobori T: **Dynamic evolution of translation initiation mechanisms in prokaryotes.** *Proc Natl Acad Sci U S A* 2010, **107:**6382–6387.
40. Veening J-W, Igoshin OA, Eijlander RT, Nijland R, Hamoen LW, Kuipers OP: **Transient heterogeneity in extracellular protease production by *Bacillus subtilis*.** *Mol Syst Biol* 2008, **4:**184.
41. Shank EA, Kolter R: **Extracellular signaling and multicellularity in *Bacillus subtilis*.** *Curr Opin Microbiol* 2011, **14:**741–747.
42. Gunka K, Commichau FM: **Control of glutamate homeostasis in *Bacillus subtilis*: a complex interplay between ammonium assimilation, glutamate biosynthesis and degradation.** *Mol Microbiol* 2012, **85:**213–224.
43. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Hartig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Le Chat L, Lecointe F, Lewis P, Liebermeister W, *et al*: **Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*.** *Science* 2012, **335:**1103–1106.
44. Brantl S: **Acting antisense: plasmid- and chromosome-encoded sRNAs from Gram-positive bacteria.** *Future Microbiol* 2012, **7:**853–871.
45. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC: **IMG ER: a system for microbial genome annotation expert review and curation.** *Bioinformatics* 2009, **25:**2271–2278.
46. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, Bateman A: **Rfam: Wikipedia, clans and the "decimal" release.** *Nucleic Acids Res* 2011, **39:**D141–D145.
47. Sherlock G: **Analysis of large-scale gene expression data.** *Curr Opin Immunol* 2000, **12:**201–205.
48. Pool MR: **Signal recognition particles in chloroplasts, bacteria, yeast and mammals (Review).** *Mol Membr Biol* 2005, **22:**3–15.
49. Hardcastle TJ, Kelly KA: **baySeq: empirical Bayesian methods for identifying differential expression in sequence count data.** *BMC Bioinformatics* 2010, **11:**422.
50. Kiley Thomason M, Storz G: **Bacterial antisense RNAs: how many are there, and what are they doing?** *Annu Rev Genet* 2010, **44:**167–188.
51. Levine E, Zhang Z, Kuhlman T, Hwa T: **Quantitative characteristics of gene regulation by small RNA.** *PLoS Biol* 2007, **5:**e229.
52. Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, Bantscheff J, Voß B, Steglich C, Wilde A, Vogel J, Hess WR: **An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803.** *Proc Natl Acad Sci U S A* 2011, **108:**1–6.
53. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons B, Sorek R: **A single-base resolution map of an archaeal transcriptome.** *Genome Res* 2010, **20:**133–141.
54. Jäger D, Sharma C, Thomsen J, Ehlers C, Vogel J, Schmitz R: **Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability.** *Proc Natl Acad Sci U S A* 2009, **106:**21878–21882.
55. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Médigue C, Danchin A: **From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later.** *Microbiology* 2009, **155:**1758–1775.
56. Rasmussen S, Nielsen H, Jarmer H: **The Transcriptionally Active Regions in the Genome of *Bacillus subtilis*.** *Mol Microbiol* 2009, **73:**1043–1057.
57. Irnov I, Sharma C, Vogel J, Winkler W: **Identification of regulatory RNAs in *Bacillus subtilis*.** *Nucleic Acids Res* 2010, **38:**6637–6651.
58. Mäder U, Schmeisky AG, Flórez LA, Stülke J: **SubtiWiki–a comprehensive community resource for the model organism *Bacillus subtilis*.** *Nucleic Acids Res* 2012, **40:**D1278–D1287.
59. Eiamphungporn W, Helmann JD: **Extracytoplasmic function sigma factors regulate expression of the *Bacillus subtilis yabE* gene via a *cis*-acting antisense RNA.** *J Bacteriol* 2009, **191:**1101–1105.
60. Gimpel M, Heidrich N, Mäder U, Krügel H, Brantl S: **A dual-function sRNA from *B. subtilis*: SR1 acts as a peptide encoding mRNA on the *gapA* operon.** *Mol Microbiol* 2010, **76:**990–1009.
61. Geissmann T, Chevalier C, Cros M-J, Boisset S, Fechter P, Noirot C, Schrenzel J, François P, Vandenesch F, Gaspin C, Romby P: **A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation.** *Nucleic Acids Res* 2009, **37:**7239–7257.
62. Marchais A, Duperrier S, Durand S, Gautheret D, Stragier P: **CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers.** *RNA Biol* 2011, **8:**358–364.
63. Saito S, Kakeshita H, Nakamura K: **Novel small RNA-encoding genes in the intergenic regions of *Bacillus subtilis*.** *Gene* 2009, **428:**2–8.
64. Silvaggi J, Perkins J, Losick R: **Genes for Small, Noncoding RNAs under Sporulation Control in *Bacillus subtilis*.** *J Bacteriol* 2006, **188:**532–541.
65. Jahn N, Preis H, Wiedemann C, Brantl S: **BsrG/SR4 from *Bacillus subtilis*–the first temperature-dependent type I toxin-antitoxin system.** *Mol Microbiol* 2012, **83:**579–598.
66. Feucht A, Evans L, Errington J: **Identification of sporulation genes by genome-wide analysis of the σE regulon of *Bacillus subtilis*.** *J Mol Biol* 2003, **149:**3023–3034.
67. Wemhoff S: *Deletionsmutagenese des yqfCD/phoH-Operons in Bacillus licheniformis - Untersuchungen zur Auswirkung auf die Sporulation*, Master thesis. Westfälische Wilhelms-Universität Münster, Institut für Molekulare Mikrobiologie und Biotechnologie; 2008.
68. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5:**621–628.
69. Risso D, Schwartz K, Sherlock G, Dudoit S: **GC-content normalization for RNA-Seq data.** *BMC Bioinformatics* 2011, **12:**480.
70. Zdobnov EM, Apweiler R: **InterProScan–an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17:**847–848.
71. The UniProt Consortium: **The Universal Protein Resource (UniProt) 2009.** *Nucleic Acids Res* 2009, **37:**D169–D174.
72. Mao F, Dam P, Chou J, Olman V, Xu Y: **DOOR: a database for prokaryotic operons.** *Nucleic Acids Res* 2009, **37:**D459–D463.
73. Pertea M, Ayanbule K, Smedinghoff M, Salzberg SL: **OperonDB: a comprehensive database of predicted operons in microbial genomes.** *Nucleic Acids Res* 2009, **37:**D479–D482.
74. Kingsford CL, Ayanbule K, Salzberg SL: **Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake.** *Genome Biol* 2007, **8:**R22.
75. Overbeek R: **The ERGOTM genome analysis and discovery system.** *Nucleic Acids Res* 2003, **31:**164–171.
76. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16:**944–945.
77. Bose M, Barber RD: **Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences.** *In Silico Biol* 2006, **6:**223–227.
78. Casjens S: **Prophages and bacterial genomics: what have we learned so far?** *Mol Microbiol* 2003, **49:**277–300.
79. Saeed A, Bhagabati N, Braisted J, Liang W, Sharov V, Howe E, Li J, Thiagarajan M, White J, Quackenbush J: **TM4 microarray software suite.** *Methods Enzymol* 2006, **411:**134.
80. Koburger T, Weibezahn J, Bernhardt J, Homuth G, Hecker M: **Genome-wide mRNA profiling in glucose starved *Bacillus subtilis* cells.** *Mol Genet Genomics* 2005, **274:**1–12.
81. Hahne H, Mäder U, Otto A, Bonn F, Steil L, Bremer E, Hecker M, Becher D: **A comprehensive proteomics and transcriptomics analysis of *Bacillus subtilis* salt stress adaptation.** *J Bacteriol* 2010, **192:**870–882.
82. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool.** *Bioinformatics* 2005, **21:**3422–3423.
83. Darling AE, Mau B, Perna NT: **ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.** *PLoS One* 2010, **5:**e11147.
84. Marco-Sola S, Sammeth M, Guigó R, Ribeca P: **The Gem mapper: fast, accurate and versatile alignment by filtration.** *Nat Methods* 2012, **9:**1186–1188.
85. Nawrocki EP, Kolbe DL, Eddy SR: **Infernal 1.0: inference of RNA alignments.** *Bioinformatics* 2009, **25:**1335–1337.
86. Martin J, Zhu W, Passalacqua K, Bergman N, Borodovsky M: ***Bacillus anthracis* genome organization in light of whole transcriptome sequencing.** *BMC Bioinformatics* 2010, **11:**S10.
87. Schwarzer M: *Physiologische Untersuchungen zur Regulation des Aminosäure-Stoffwechsels von Bacillus licheniformis DSM13*, PhD thesis. Georg-August-Universität Göttingen, Institut für Mikrobiologie und Genetik; 2010.
88. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning - A Laboratory Manual.* Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1989.

doi:10.1186/1471-2164-14-667
Cite this article as: Wiegand *et al.*: RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation. *BMC Genomics* 2013 **14**:667.

Additional information

Additional file 1

Figure S1 Distribution of whole transcriptome sequencing reads 45
Figure S2 Comparative operon prediction 45
Figure S3 Northern blot confirmation of non-coding RNAs 46
Figure S4 Antisense RNAs with putative impact on productivity 54
Figure S5 BsrG/SR4-like loci in *B. licheniformis* 56
Figure S6 Work flow charts 57

Additional file 2

Table S1 Whole transcriptome sequencing reads 58
Table S2 NPKM values of RNA features 59
Table S3 NPKM values of all genes 59
Table S4 Differential RNA-Seq reads 59
Table S5 Transcription start sites 60
Table S6 Identified RNA features 60
Table S7 Predicted operons 60
Table S8 Corrected genes 60
Table S9 New genes 62
Table S10 Removed genes 63
Table S11 Predicted *cis*-regulatory elements 65
Table S12 Comparison of *cis*-regulatory elements known from *B. subtilis* 65
Table S13 Cluster analysis of ncRNA expression profiles 65
Table S14 Comparison of ncRNAs to *B. subtilis* 65
Table S15 Comparison of small RNAs from *B. subtilis* to identified ncRNAs 67
Table S16 Sequence Read Archive accession 69
Table S17 Primer pairs for Northern blots 70

References 71

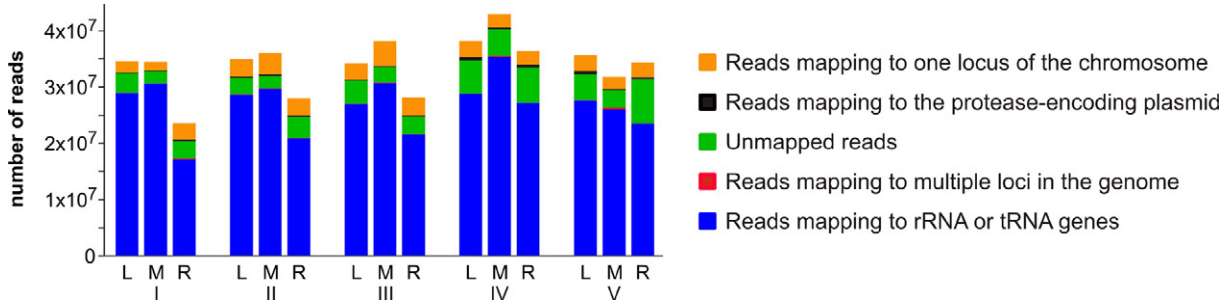


Figure S1 Distribution of whole transcriptome sequencing reads
Mapping results of the RNA-Seq reads to the *B. licheniformis* genome based on sequence similarity. Results are shown for replicates of the sampling points I to V. The category *unmapped reads* comprises reads derived from experimental artifacts like poly(A) tails or concatenated RNA adapters, as well as reads with more than one sequencing error per 50 bp. Please note that not a single read which has been tested from this fraction can be assigned to other organisms than *B. licheniformis*.

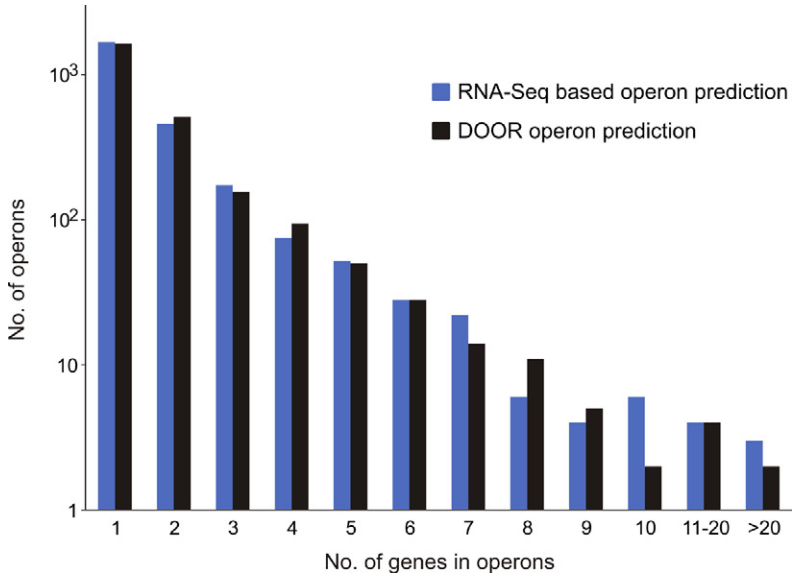
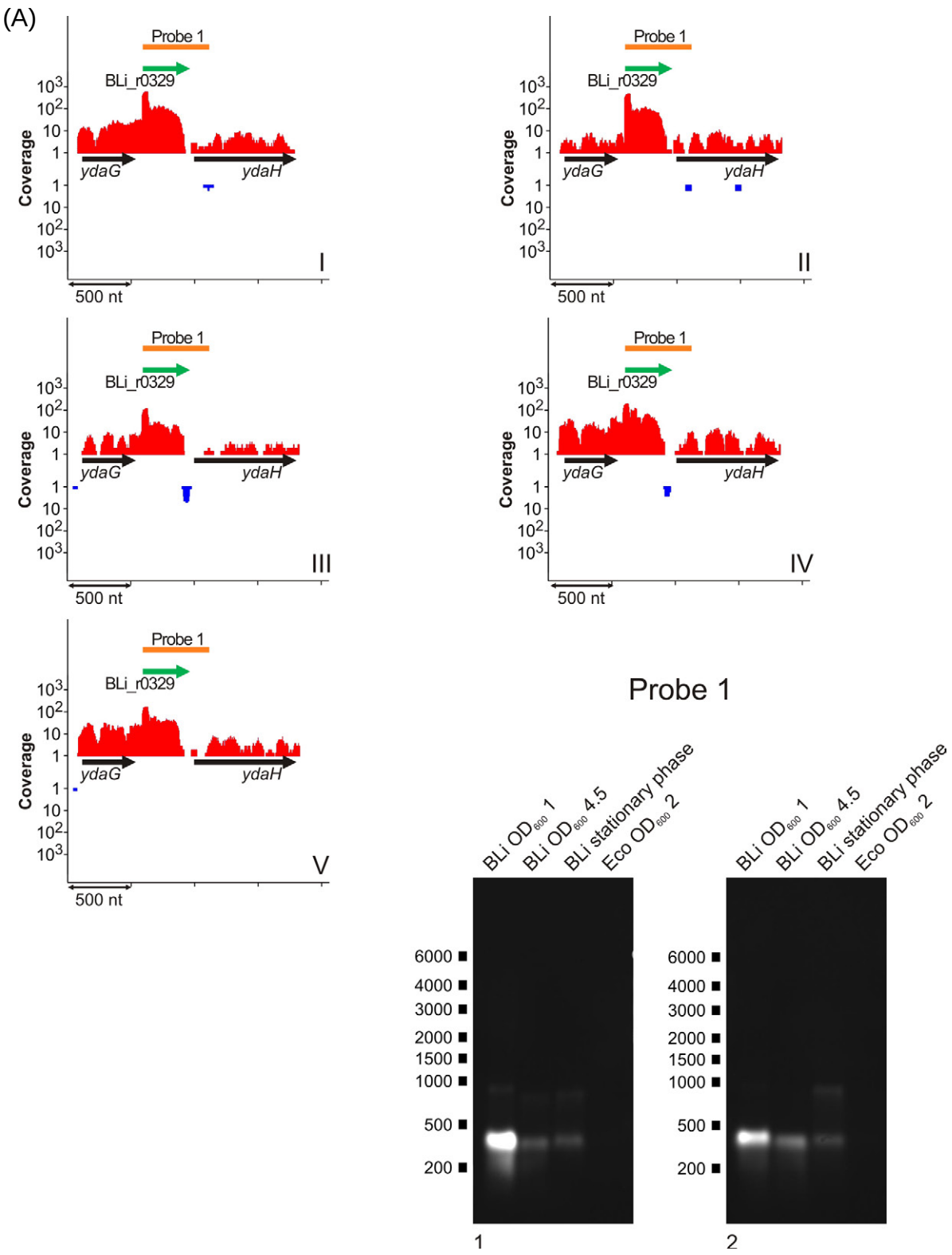
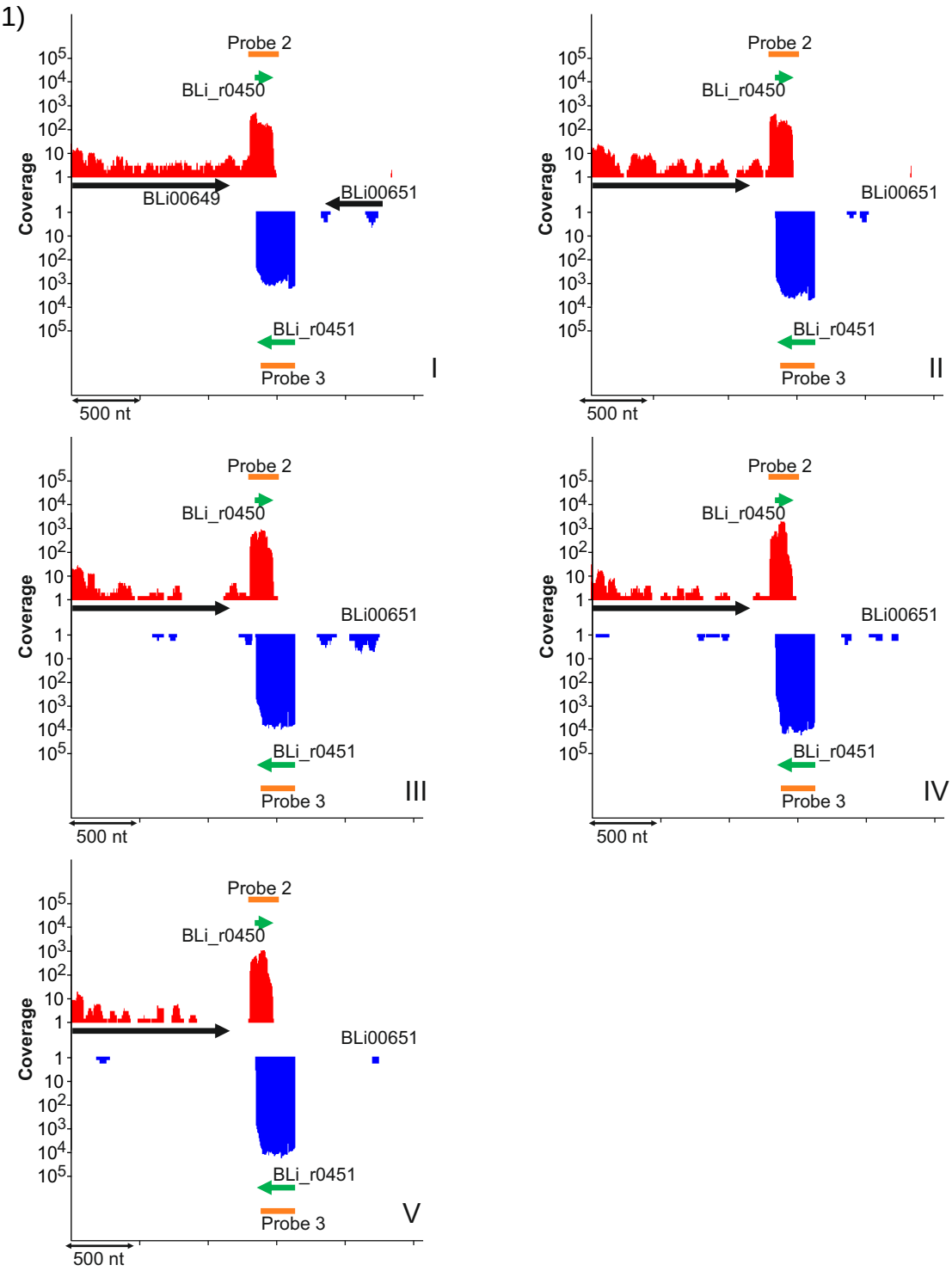


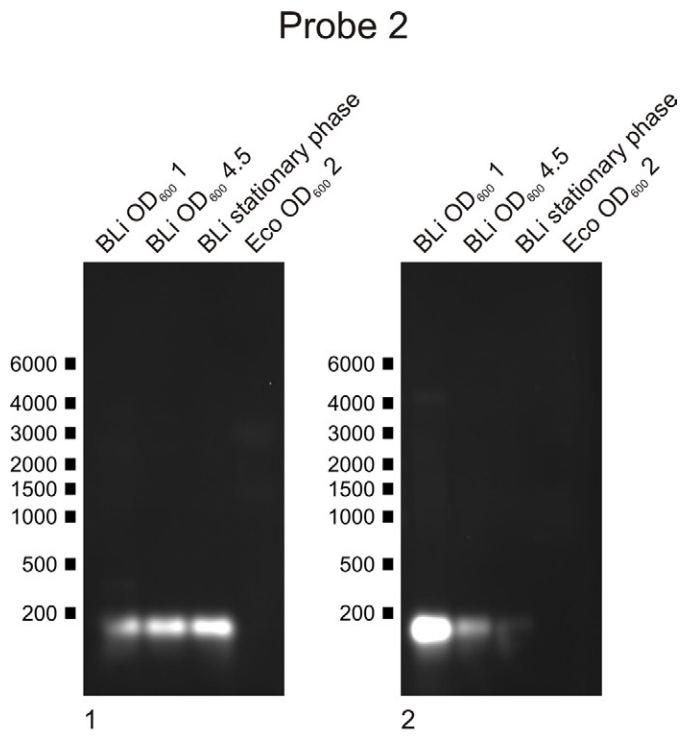
Figure S2 Comparative operon prediction
Number of genes in predicted operons are shown for the *in silico* prediction available at DOOR (black; Mao et al., 2009) and the manually curated prediction based on RNA-Seq provided in this study (blue). The deviations especially in the longer operons are due to the here employed expression profile-accounting method of operon prediction. Please note that results are given on log-transformed scale.

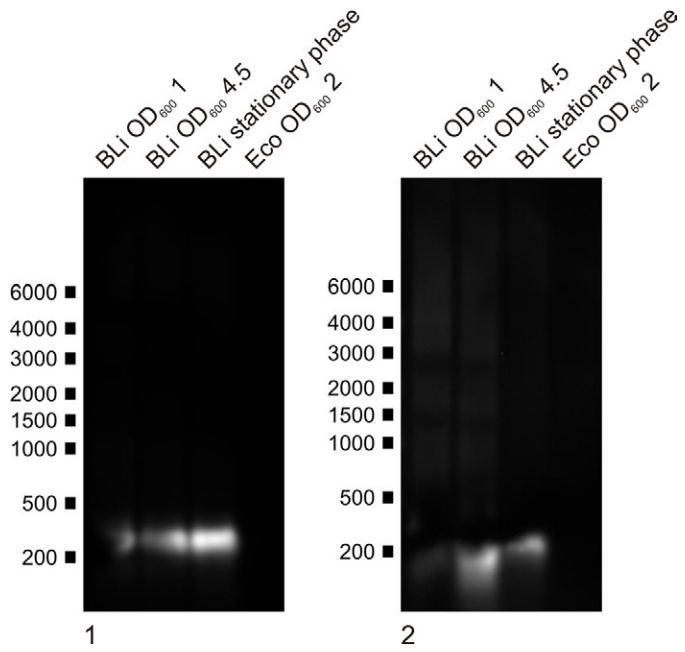

(A)
Probe 1
BLi_r0329
ydaG
ydaH
Coverage
500 nt
I
II
III
IV
V
Probe 1
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2

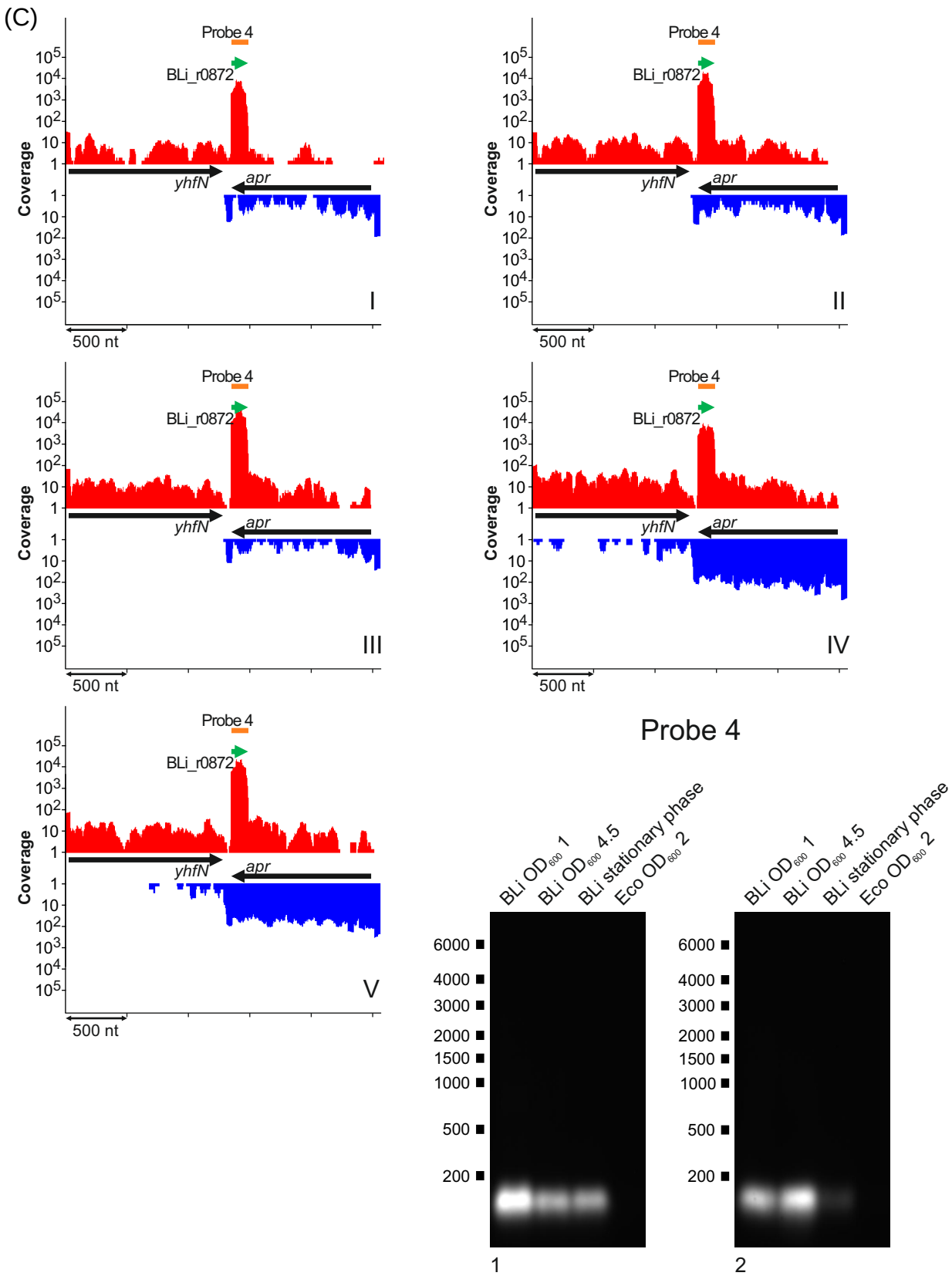
(B.1)

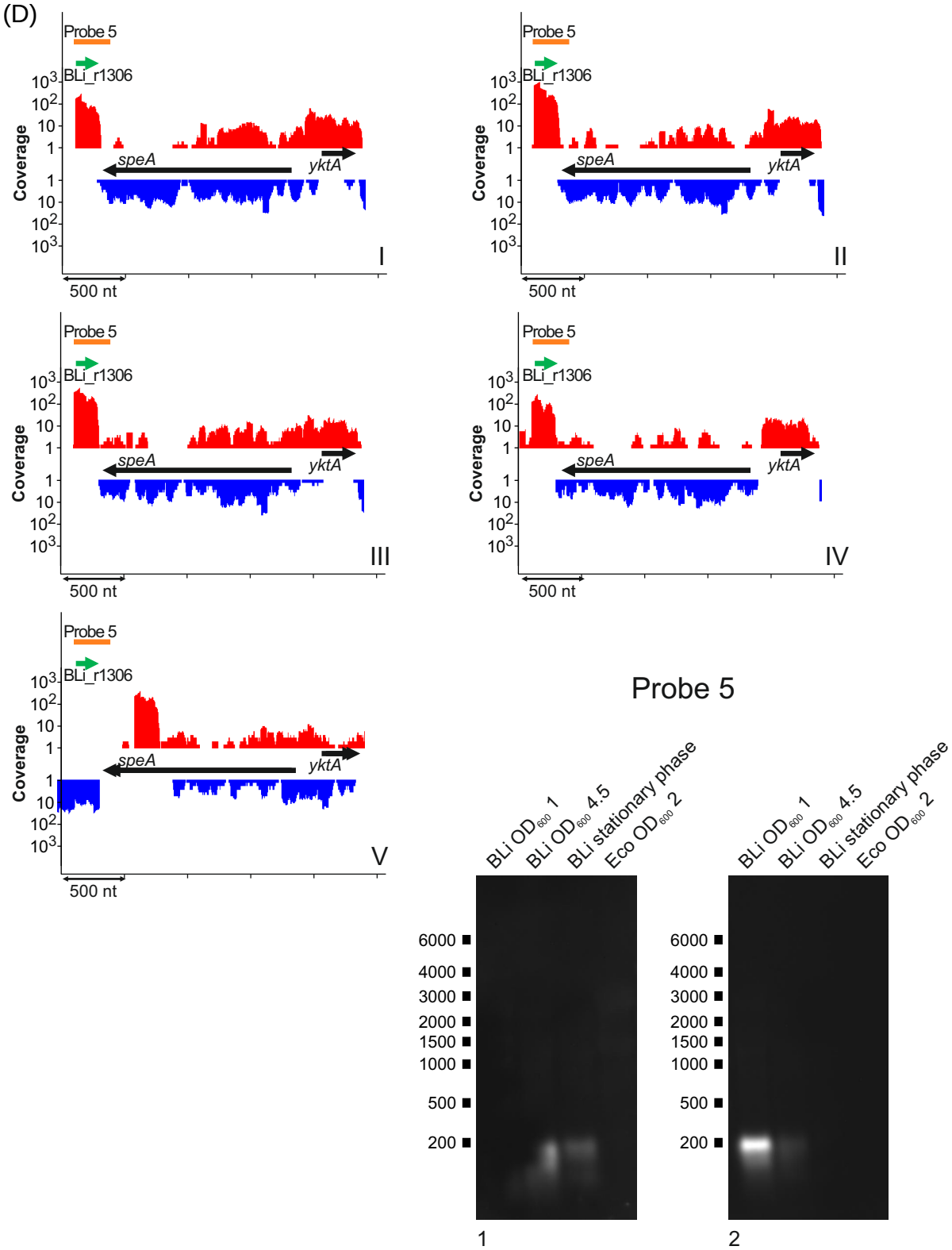

Probe 2
BLi_r0450
BLi00649
BLi00651
BLi_r0451
Probe 3
Coverage
500 nt
I
II
III
IV
V

(B.2)

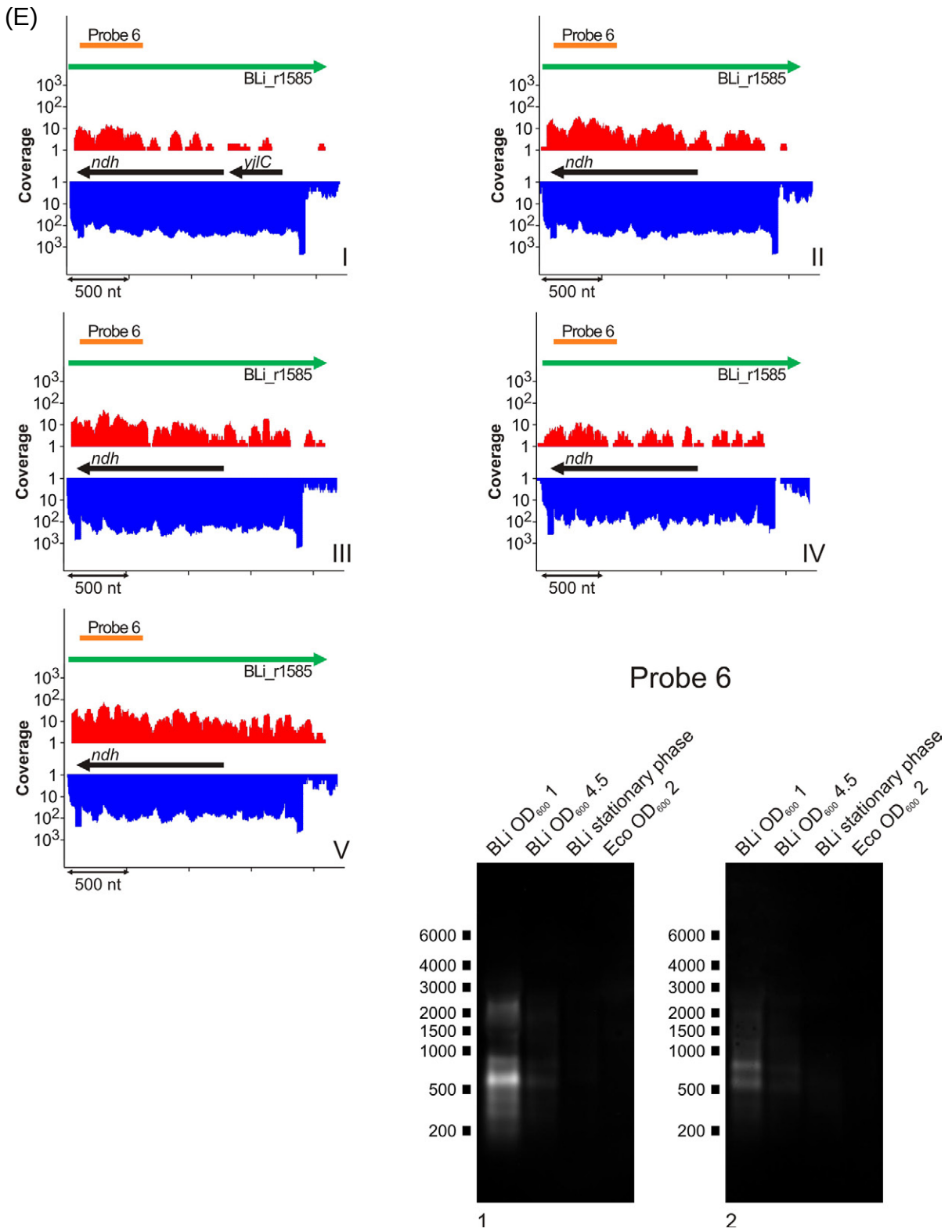

Probe 2
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2


Probe 3
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2

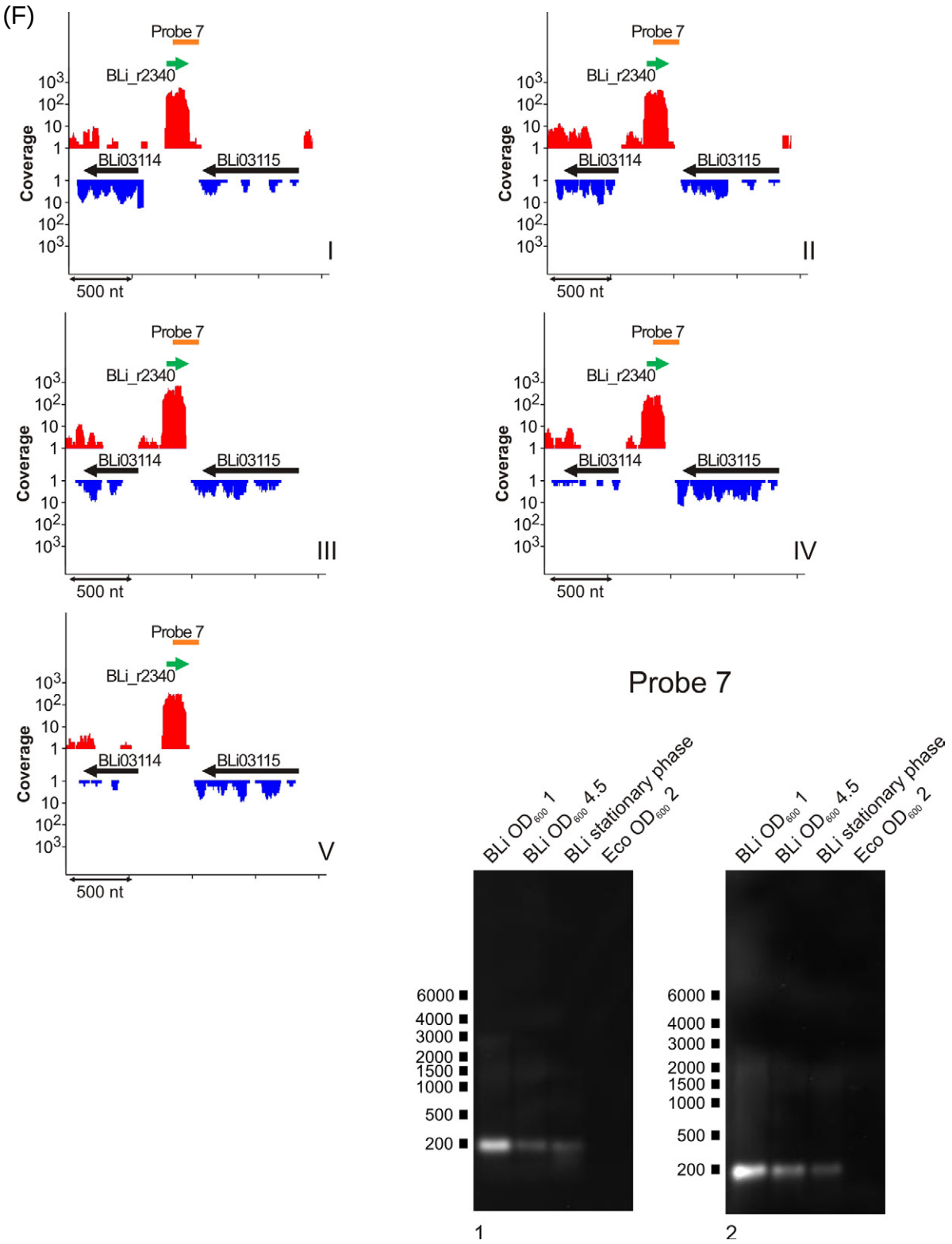

(C)
Probe 4
BLi_r0872
yhfN
apr
Coverage
500 nt
I
II
III
IV
V
Probe 4
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2


(D)
Probe 5
BLi_r1306
Coverage
speA
yktA
500 nt
I
II
III
IV
V
Probe 5
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2

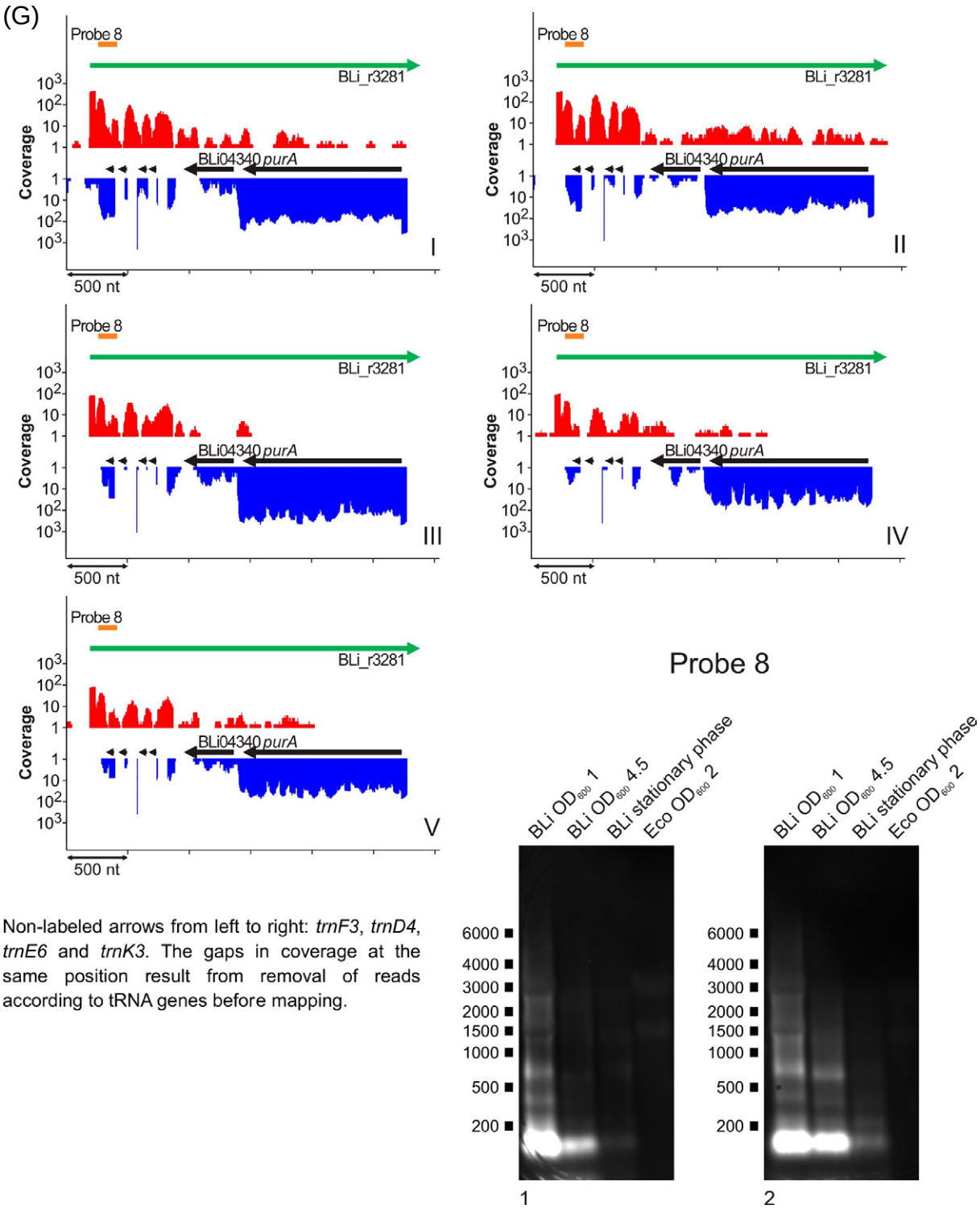
(E)


Probe 6
BLi_r1585
Coverage
ndh
yjlC
500 nt
I
II
III
IV
V
Probe 6
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2

(F)

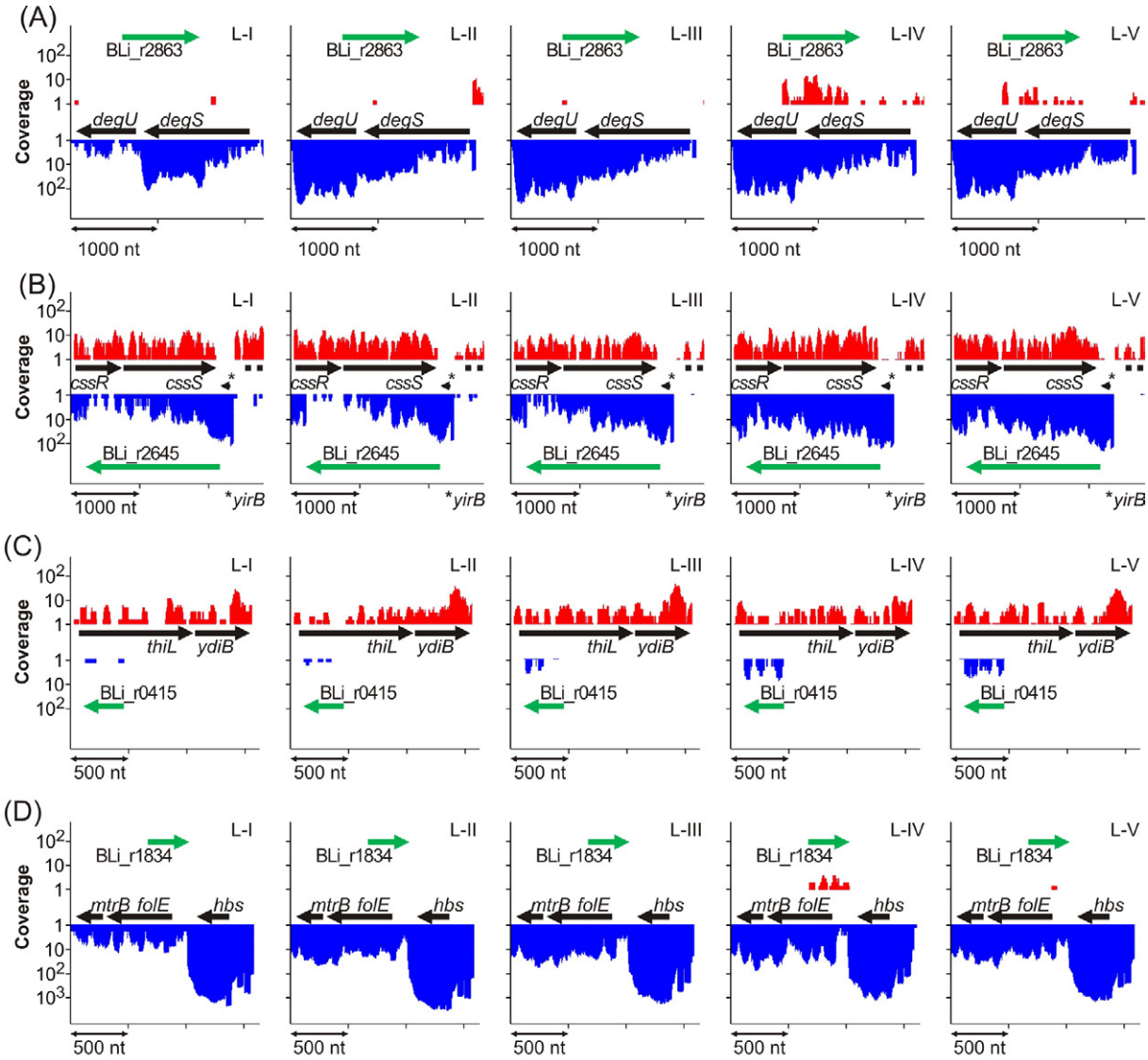

Probe 7
BLi_r2340
BLi03114
BLi03115
Coverage
500 nt
I
II
III
IV
V
Probe 7
BLi OD600 1
BLi OD600 4.5
BLi stationary phase
Eco OD600 2
6000
4000
3000
2000
1500
1000
500
200
1
2

(G)



Non-labeled arrows from left to right: *trnF3*, *trnD4*, *trnE6* and *trnK3*. The gaps in coverage at the same position result from removal of reads according to tRNA genes before mapping.

Figure S3 Northern blot confirmation of non-coding RNAs
For each blot the corresponding genomic loci and the transcriptional activity is shown for all sampling points. The black arrows show genes, the green arrow the respective ncRNA. Probes (Table S17) are marked orange. Northern blotting was carried out as described in *Methods*. The size range is given in base pairs and the number under each picture signifies the different replicates. **(A)** *indep* RNA BLi_r0329 (expected size: 373 bp). **(B)** *indep* RNAs BLi_r0450 and BLi_r0451 (expected size: 127 bp and 291 bp, respectively). **(C)** A_I RNA BLi_r0872 (expected size: 144 bp). This ncRNA is of special interest since it is encoded antisense to the alkaline protease Subtilisin Carlsberg. **(D)** A_{misc} RNA BLi_r1306 (expected size: 205 bp). Please note that within the same transcript an ORF was annotated "SR1-like protein", according to the finding in *B. subtilis,* where SR1 acts as functional sRNA and encodes a protein. **(E)** A_{misc} RNA BLi_r1585 (expected size: 2072 bp). The observed bands indicate processing or degradation of the RNA transcript. **(F)** *indep* RNA BLi_r2340 (expected size: 188 bp) **(G)** A_{misc} RNA BLi_r3281 (expected size: 2688 bp). The observed bands may indicate a fading-out of the transcription after leaky termination.



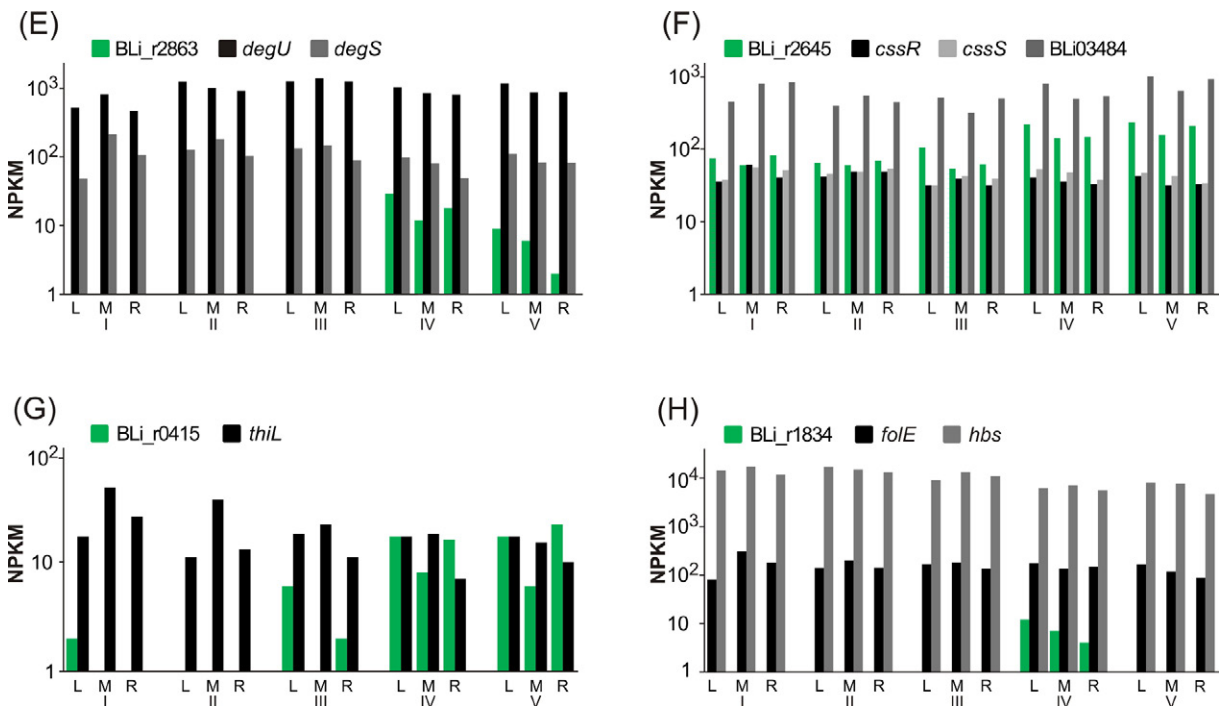


Figure S4 Antisense RNAs with putative impact on productivity
Antisense transcriptional activities at different fermentation stages. Black arrows indicate genes and green arrows identify antisense RNAs. **(A)** A_{misc} RNA BLi_r2863, **(B)** $A_{3'UTR}$ BLi_r2645, **(C)** A_{misc} RNA BLi_r0415 and **(D)** A_{misc} BLi_r1834 are shown for L-I to L-V. NPKM values are displayed for all replicates of the sampling points I to V for: **(E)** BLi_r2863 and the *degSU* operon, a two-component regulatory system which is involved in the manifestation of multicellular communities (Murray et al., 2009) and regulates biofilm formation, genetic competence, swarming motility, polyglutamic acid production as well as exoprotease secretion (Davidson et al., 2012; Hoffmann et al., 2010). As the transcript is present in the productive stages of the fermentation process, an influence on protease secretion via an impact on the *degSU* mRNA should be carefully considered. In *B. subtilis*, an ncRNA antisense to *degSU* is also annotated (Nicolas et al., 2012), but varies in length and start position due to a disparate genomic context. **(F)** BLi_r2645 and the *cssRS* operon, a two-component regulatory system in control of the serine proteases HtrA and HtrB (Darmon et al., 2002), which bear proteolytical as well as chaperone activity in response to secretion stress (Marciniak et al., 2012). The overlapping 3'UTR antisense to the *cssRS* operon is generated by the proteolytical anti-adaptor protein YirB. **(G)** BLi_r0415 and *thiL*, which encodes a thiamine-monophosphate kinase; **(H)** BLi_r1834 and *folE*, which encodes a GTP cyclohydrolase I involved in folate biosynthesis (plus *hbs*, the adjacent DNA-binding protein HBsu). Please note that results are given on log-transformed scale.

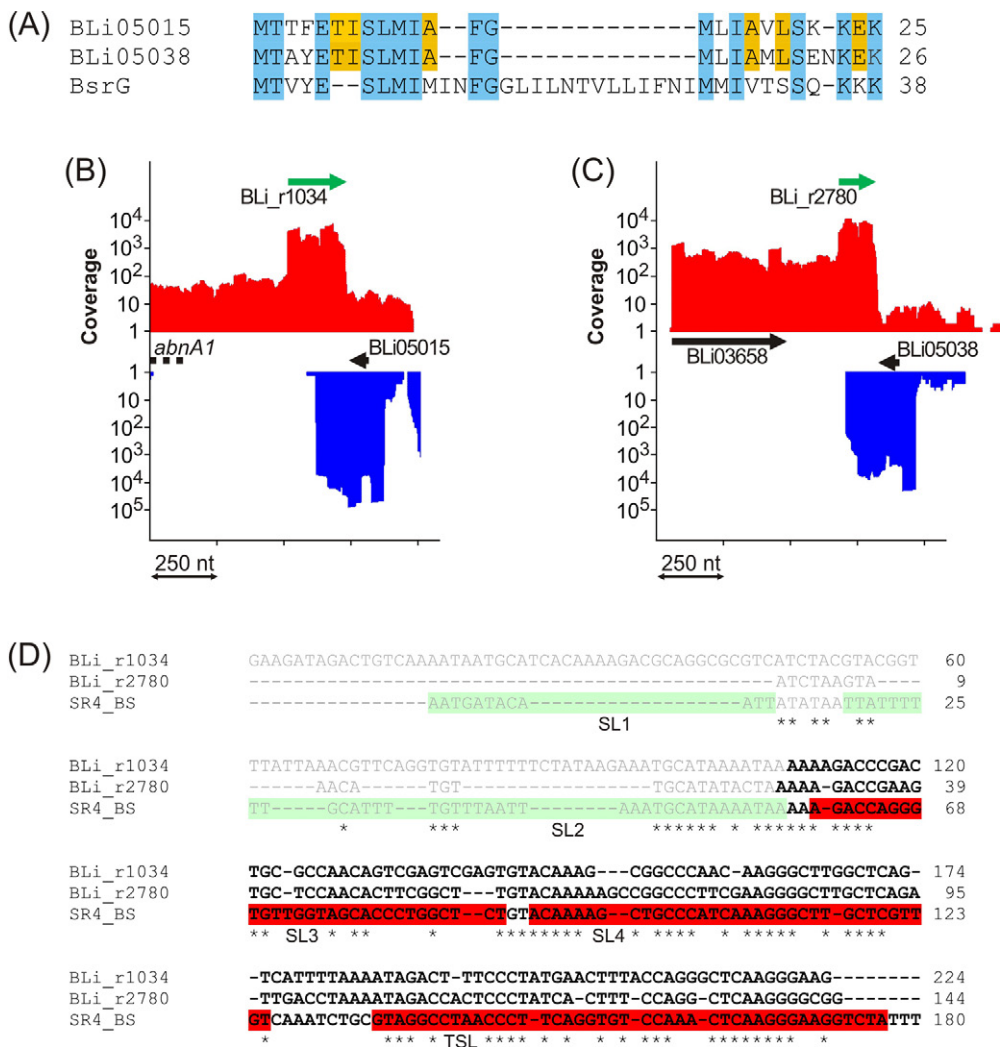


Figure S5 BsrG/SR4-like loci in *B. licheniformis*

(A) ClustalW2 alignment of BsrG from *B. subtilis* (Jahn et al., 2012) and the peptides deduced from the newly annotated genes BLi05015 and BLi05038. Blue boxes show amino acids conserved in all three peptides and orange boxes amino acids only conserved between the two *B. licheniformis* peptides. **(B)** BLi05015 (1300390-1300316) indicated by the black arrow and BLi_r1034 (green arrow). **(C)** Transcriptional activities of BLi05038 (3485388-3485308) indicated by the black arrow and BLi_r2780 (green arrow). **(D)** ClustalW2 alignment of SR4 from *B. subtilis* (Jahn et al., 2012) and *indep* RNAs BLi_t1034 and BLi_r2780. Nucleotides overlapping the opposite mRNA are pictured black and bold, non-overlapping nucleotides are marked gray. The green and red boxes indicate the stem loops SL1 to SL4 and TSL identified for SR4 (Jahn et al., 2012) in the non-overlapping and the overlapping region, respectively.

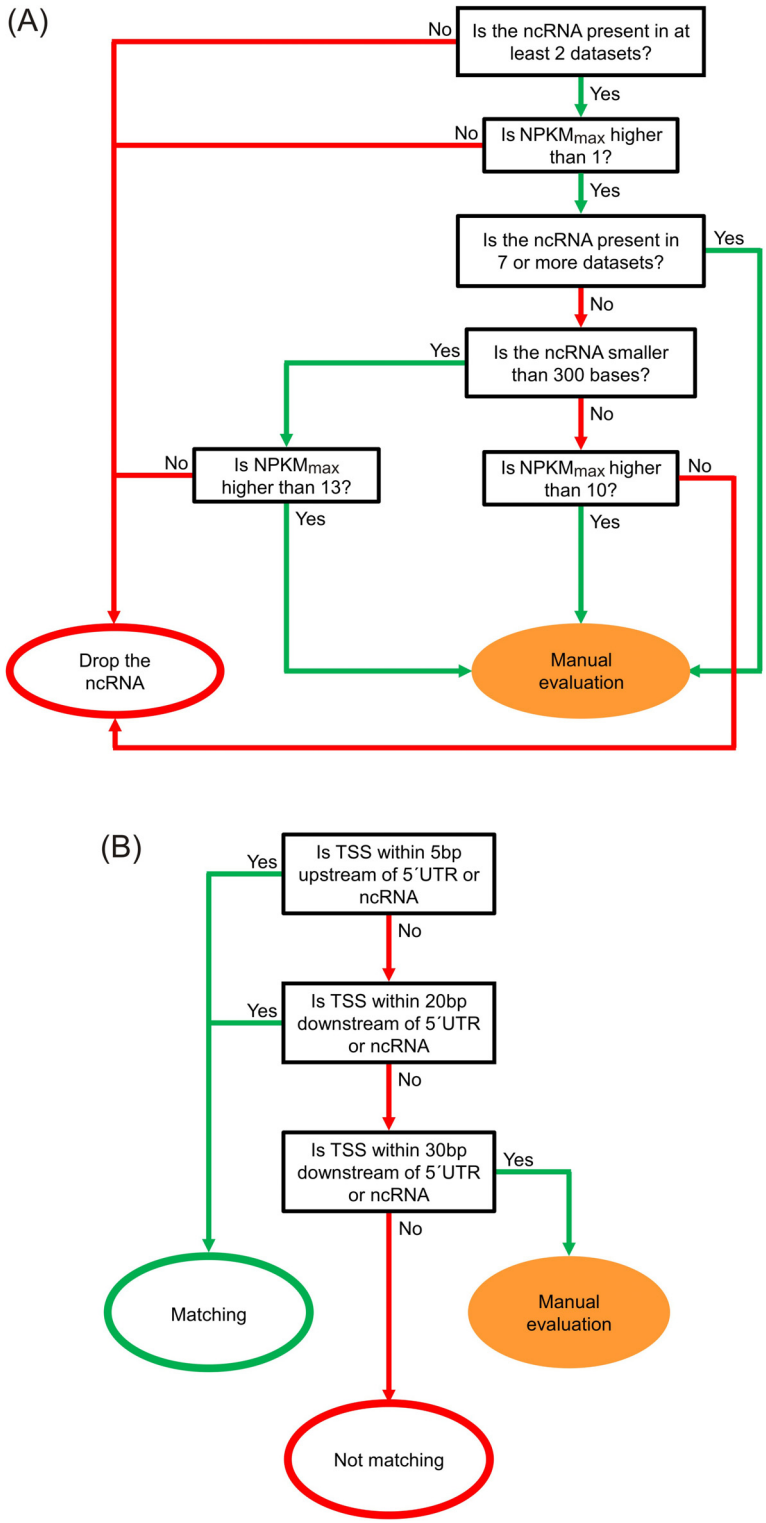


Figure S6 Work flow charts
(A) Work flow of non-coding RNA evaluation. **(B)** Work flow of transcription start site evaluation.

Table S1 Whole transcriptome sequencing reads
The total of all reads could be assigned to: reads mapping to rRNA or tRNA genes; reads mapping to multiple loci in the genome; unmapped reads; reads mapping to the protease-encoding plasmid and reads mapping to one locus of the chromosome.

Sample	∑ reads	reads against tRNA or rRNA genes	reads against tRNA or rRNA genes [%]	multihit reads	multihit reads [%]	un-mapped reads	un-mapped reads [%]	plasmid-mapping reads	plasmid-mapping reads [%]	chromosome-mapping reads	chromosome-mapping reads [%]
L-I	34.553.369	28.969.888	83,8	32.013	0,09	3.424.018	9,91	174.110	0,50	1.953.340	5,7
R-I	34.491.633	30.642.606	88,8	13.956	0,04	2.148.822	6,23	161.458	0,47	1.524.791	4,4
M-I	23.583.582	17.140.657	72,7	138.359	0,59	3.157.703	13,39	309.415	1,31	2.837.448	12,0
L-II	34.935.270	28.636.703	82,0	60.336	0,17	2.963.179	8,48	189.401	0,54	3.085.651	8,8
R-II	36.024.461	29.680.210	82,4	98.178	0,27	2.203.136	6,12	289.793	0,80	3.753.144	10,4
M-II	28.018.016	20.903.311	74,6	81.313	0,29	3.689.827	13,17	235.161	0,84	3.108.404	11,1
L-III	34.152.665	27.024.138	79,1	32.914	0,10	4.084.170	11,96	154.977	0,45	2.856.466	8,4
R-III	38.187.230	30.719.713	80,4	81.366	0,21	2.746.909	7,19	183.528	0,48	4.455.714	11,7
M-III	28.174.975	21.615.689	76,7	49.622	0,18	3.099.499	11,00	149.313	0,53	3.260.852	11,6
L-IV	38.200.610	28.824.001	75,5	22.027	0,06	5.878.876	15,39	579.663	1,52	2.896.043	7,6
R-IV	42.959.023	35.377.857	82,4	150.077	0,35	4.734.060	11,02	327.074	0,76	2.369.955	5,5
M-IV	36.414.040	27.221.034	74,8	23.275	0,06	6.205.296	17,04	523.290	1,44	2.441.145	6,7
L-V	35.656.360	27.619.551	77,5	30.045	0,08	4.611.200	12,93	578.175	1,62	2.817.389	7,9
R-V	31.798.624	26.061.998	82,0	280.073	0,88	3.075.680	9,67	224.030	0,70	2.156.843	6,8
M-V	34.376.322	23.540.363	68,5	28.510	0,08	7.807.537	22,71	362.828	1,06	2.637.084	7,7

Table S2 NPKM values of RNA features

Please refer to “Chapter B_Additional information” on digital medium.

Table S3 NPKM values of all genes

Please refer to “Chapter B_Additional information” on digital medium.

Table S4 Differential RNA-Seq reads

The total of all reads could be assigned to: reads mapping to multiple loci in the genome; unmapped reads; reads mapping to the protease-encoding plasmid and reads mapping to one locus of the chromosome. The majority of the multihit reads can be retraced to originate from rRNA and tRNA genes.

Sample	∑ reads	multihit reads	multihit reads [%]	unmapped reads	unmapped reads [%]	plasmid-mapping reads	plasmid-mapping reads [%]	chromosome-mapping reads	chromosome-mapping reads [%]
L-I	4.612.696	3.915.364	84,9	457.689	9,9	11.987	0,26	227.656	4,9
L-II	5.037.992	3.965.058	78,7	630.907	12,5	14.930	0,30	427.097	8,5
L-III	4.107.133	2.633.377	64,1	1.163.040	28,3	9.947	0,24	300.769	7,3
L-IV	3.879.347	2.231.724	57,5	1.293.062	33,3	43.331	1,12	311.230	8,0
L-V	4.410.205	2.623.351	59,5	1.376.881	31,2	41.902	0,95	368.071	8,3

Table S5 Transcription start sites

Please refer to "Chapter B_Additional information" on digital medium.

Table S6 Identified RNA features

Please refer to "Chapter B_Additional information" on digital medium.

Table S7 Predicted operons

Please refer to "Chapter B_Additional information" on digital medium.

Table S8 Corrected genes

Locus tag	New start	Old start	Stop	Strand	Type	Gene symbol	Function
BLi00008	9713	9710	11250	+	rRNA	*rrsA*	16s_rRNA
BLi00011	11594	11592	14522	+	rRNA	*rrlA*	23s_rRNA
BLi00012	14656	14644	14770	+	rRNA	*rrfA*	5s_rRNA
BLi00022	26352	26355	26444	+	tRNA	*trnS1*	tRNA-Ser-TCA
BLi00033	34411	34408	35948	+	rRNA	*rrsB*	16s_rRNA
BLi00036	36292	36290	39220	+	rRNA	*rrlB*	23s_rRNA
BLi00037	39355	39343	39469	+	rRNA	*rrfB*	5s_rRNA
BLi00098	95153	95150	96691	+	rRNA	*rrsC*	16s_rRNA
BLi00099	96873	96871	99801	+	rRNA	*rrlC*	23s_rRNA
BLi00100	99936	99924	100050	+	rRNA	*rrfC*	5s_rRNA
BLi00176	158109	158106	159647	+	rRNA	*rrsD*	16s_rRNA
BLi00177	159829	159827	162757	+	rRNA	*rrlD*	23s_rRNA
BLi00178	162892	162880	163006	+	rRNA	*rrfD*	5s_rRNA
BLi00209	187962	187899	188210	+	CDS		HTH-type transcriptional regulator
BLi00232	210427	210227	208350	-	pseudo	*yyaL*	hypothetical protein
BLi00252	228113	227626	227318	-	pseudo	*ybfI*	putative HTH-type transcriptional regulator
BLi00291	266093	266181	266786	+	pseudo	*ycbL*	two-component response regulator
BLi00392	368757	367582	367079	-	pseudo	*tlpC*	methyl-accepting chemotaxis protein
BLi00421	421011	421104	419599	-	CDS	*rocR*	arginine utilization regulatory protein
BLi00575	575330	575351	575091	-	CDS		putative phage protein
BLi00606	611805	611802	613342	+	rRNA	*rrsE*	16s_rRNA
BLi00607	613524	613522	616452	+	rRNA	*rrlE*	23s_rRNA
BLi00608	616587	616575	616701	+	rRNA	*rrfE*	5s_rRNA
BLi00848	867352	867574	866801	-	CDS		hypothetical protein
BLi00870	889249	889631	889780	+	pseudo	*yfiT*	putative metal-dependent hydrolase
BLi00903	920785	920782	922323	+	rRNA	*rrsF*	16s_rRNA
BLi00904	922505	922503	925433	+	rRNA	*rrlF*	23s_rRNA
BLi00905	925606	925617	925720	+	rRNA	*rrfF*	5s_rRNA
BLi01034	1035023	1035056	1034319	-	CDS	*yhdW1*	putative glycerophosphoryl diester phosphodiesterase

Table S8 Continued

Locus tag	New start	Old start	Stop	Strand	Type	Gene symbol	Function
BLi01269	1279238	1279289	1278750	-	CDS	*cotY*	spore coat protein
BLi01304	1307931	1307450	1307058	-	pseudo	*yoaU*	putative HTH-type transcriptional regulator
BLi01427	1422023	1421963	1422430	+	CDS		hypothetical protein
BLi01506	1480354	1480300	1480515	+	CDS		hypothetical protein
BLi01754	1713319	1713361	1713990	+	CDS		hypothetical protein
BLi01765	1722163	1722085	1722624	+	CDS	*lspA*	lipoprotein signal peptidase
BLi01938	1899053	1899497	1900495	+	pseudo	*yobN*	putative L-amino-acid oxidase
BLi01942	1903533	1904017	1904790	+	pseudo	*tetB*	tetracycline/Na+ resistance protein
BLi02004	1958074	1957906	1958553	+	CDS	*yoaO*	hypothetical protein
BLi02061	2000669	2000744	1999578	-	CDS	*araR1*	arabinose metabolism transcriptional repressor
BLi02350	2299348	2299435	2297543	-	CDS	*ypvA*	putative ATP-dependent helicase
BLi02482	2415316	2415118	2416212	+	CDS	*ypuA*	hypothetical protein
BLi02508	2435454	2436059	2436928	+	pseudo	*pucI*	putative allantoin permease
BLi02548	2466810	2466849	2465857	-	CDS	*yqjP*	putative beta-lactamase
BLi02561	2482119	2481190	2480804	-	pseudo	*gmuC*	glucomannan-specific EIIC component
BLi02576	2496082	2495412	2495314	-	pseudo	*artP*	arginine ABC transporter binding protein
BLi02698	2599042	2598575	2598318	-	pseudo	*ywqL*	putative endonuclease
BLi02777	2669421	2669256	2668447	-	pseudo	*ansA2*	L-asparaginase
BLi02839	2725814	2725883	2724579	-	CDS	*ytbD*	major facilitator superfamily protein
BLi02893	2784448	2784295	2786001	+	CDS	*spoVB*	stage V sporulation protein
BLi03090	2976780	2976858	2976097	-	CDS	*ytfI*	DUF2953 family protein
BLi03240	3116592	3116589	3116500	-	tRNA	*trnS5*	tRNA-Ser-TCA
BLi03253	3118373	3118385	3118259	-	rRNA	*rrfG*	5s_rRNA
BLi03254	3121435	3121437	3118507	-	rRNA	*rrlG*	23s_rRNA
BLi03255	3123152	3123155	3121614	-	rRNA	*rrsG*	16s_rRNA
BLi03276	3142332	3143342	3143512	+	pseudo	*oxdD*	oxalate decarboxylase
BLi03570	3416061	3415480	3414884	-	pseudo		amidase
BLi03655	3483214	3482574	3481447	-	pseudo	*yvbJ*	hypothetical protein
BLi03659	3486205	3485812	3485573	-	pseudo		hypothetical protein
BLi03675	3502474	3502543	3501758	-	CDS	*lutA*	lactate utilization protein
BLi03794	3612592	3612643	3611435	-	CDS	*degS*	two-component sensor histidine kinase
BLi03840	3668380	3668323	3669360	+	CDS	*rbsR*	ribose operon repressor
BLi04044	3870416	3870668	3871042	+	pseudo	*ywqN*	putative oxidoreductase
BLi04053	3879902	3879579	3878986	-	pseudo		putative HTH-type transcriptional regulator
BLi04062	3886860	3886533	3885913	-	pseudo		glyoxylate reductase
BLi04122	3942251	3942338	3942060	-	CDS	*lanR*	putative regulatory protein
BLi04144	3970756	3971145	3971444	+	pseudo	*yxdJ2*	two-component response regulator
BLi04152	3978373	3977693	3977478	-	pseudo		hypothetical protein
BLi04196	4027880	4027259	4025799	-	pseudo	*katE*	catalase
BLi04297	4145251	4147104	4147199	+	pseudo		hypothetical protein
BLi04300	4148379	4148596	4149111	+	pseudo	*yvfS*	putative ABC transporter permease
BLi04345	4192151	4192211	4191933	-	CDS		hypothetical protein

Table S9 New genes

Locus tag	Start	Stop	Strand	Gene symbol	Function
BLi05000	23278	22148	-	*glxK*	glycerate kinase
BLi05001	39637	39828	+	*gin*	forespore-specific protein
BLi05002	217298	217065	-		hypothetical protein
BLi05003	247064	245634	-	*glnt*	sodium:glutamine symporter
BLi05004	629808	629461	-		hypothetical protein
BLi05005	774580	775293	+		hypothetical protein
BLi05006	897768	897598	-		hypothetical protein
BLi05007	928712	928849	+		hypothetical protein
BLi05008	928965	929090	+		hypothetical protein
BLi05009	929505	929807	+		hypothetical protein
BLi05010	934163	934459	+		putative RlfA like protein
BLi05011	1047444	1047184	-		putative heterocycle-containing bacteriocin
BLi05012	1101815	1101714	-		hypothetical protein
BLi05013	1252084	1251422	-	*yjbE*	membrane protein
BLi05014	1284826	1285083	+		hypothetical protein
BLi05015	1300390	1300313	-		BsrG-like peptide
BLi05016	1426080	1425856	-		phage putative repressor
BLi05017	1426652	1426888	+		phage putative repressor
BLi05018	1430073	1430177	+		phage related membrane protein
BLi05019	1434830	1434919	+		putative phage protein
BLi05020	1436519	1436725	+		putative phage protein
BLi05021	1444952	1445098	+		putative phage protein
BLi05022	1450290	1450733	+		putative phage protein
BLi05023	1511238	1511426	+		putative phage protein
BLi05024	1511517	1511735	+		putative phage protein
BLi05025	1512367	1512537	+		putative phage protein
BLi05026	1515821	1516159	+		putative phage protein
BLi05027	2055295	2055023	-		hypothetical protein
BLi05028	2195153	2194911	-		spore coat protein-like protein
BLi05029	2447588	2447373	-		hypothetical protein
BLi05030	2511511	2511215	-		hypothetical protein
BLi05031	2894531	2894830	+		hypothetical protein
BLi05032	3393454	3393789	+		hypothetical protein
BLi05033	3427079	3427606	+		putative phage protein
BLi05034	3454164	3453790	-		hypothetical protein
BLi05035	3454940	3454716	-	*cotD3*	spore coat protein D
BLi05037	3467283	3467921	+		putative phage protein
BLi05038	3485388	3485308	-		BsrG-like peptide
BLi05039	3701264	3701154	-	*usd*	putative spoIIID leader peptide

Table S9 Continued

Locus tag	Start	Stop	Strand	Gene symbol	Function
BLi05040	3798142	3798408	+	*ywjC*	general stress protein
BLi05041	3842580	3842290	-		hypothetical protein
BLi05042	3949514	3949296	-	*lanA1*	lichenicidin prepeptide
BLi05043	4157493	4157263	-		putative phage protein
BLi05044	4161050	4160448	-		putative phage protein
BLi05045	4169347	4168853	-		putative phage protein
BLi05046	763186	763308	+	*phrK*	response regulator aspartate phosphatase (RapK) regulator
BLi05047	1142428	1142544	+	*phrG*	response regulator aspartate phosphatase (RapG) regulator
BLi05048	1280933	1281289	+	*yjcA*	sporulation protein
BLi05049	1281576	1281785	+		hypothetical protein
BLi05050	1465091	1465270	+	*ykoL*	stress response protein
BLi05051	2244490	2244624	+	*yoyE*	hypothetical protein
BLi05052	2310333	2310437	+	*sspM*	small acid-soluble spore protein
BLi05053	2760326	2760195	-	*yrzQ*	hypothetical protein
BLi05054	3097490	3097651	+	*ytzL*	hypothetical protein
BLi05055	4040502	4040365	-		hypothetical protein
BLi05056	1639788	1639907	+		Sr1-like peptide
BLi05057	2246052	2246186	+	*yoyF*	hypothetical protein
BLi05058	2742105	2741956	-	*yrzI*	hypothetical proteinI
BLi05059	3089961	3090104	+		hypothetical protein
BLi05060	3863368	3863484	+		hypothetical protein

Table S10 Removed genes

Locus tag	Start	Stop	Strand	Gene symbol
BLi00233	210427	210266	-	*yyaO*
BLi00253	228113	227673	-	*ybfI2*
BLi00290	266093	266191	+	
BLi00393	368757	367579	-	*tlpC*
BLi00577	577425	577535	+	
BLi00653	654973	655200	+	
BLi00752	763203	763397	+	
BLi00869	889249	889626	+	*yfiT*
BLi00961	959291	959205	-	
BLi01073	1076310	1076396	+	
BLi01084	1082262	1082564	+	
BLi01273	1281616	1280960	-	
BLi01278	1285095	1284778	-	

Table S10 Continued

Locus tag	Start	Stop	Strand	Gene symbol
BLi01305	1307931	1307431	-	*yoaU2*
BLI01492	1465399	1465154	-	
BLi01753	1713116	1713325	+	
BLi01764	1722009	1722134	+	
BLi01937	1899053	1899475	+	
BLi01941	1903533	1904096	+	
BLi02507	2435454	2436377	+	*ywoE*
BLi02517	2442450	2442752	+	
BLi2518	2442701	2442468	-	*yqkE*
BLi02562	2482119	2481226	-	
BLi02577	2496082	2495543	-	*yqiX*
BLi02632	2547446	2547240	-	
BLi02699	2599042	2598599	-	*ywqL*
BLi02772	2665310	2665447	+	
BLi02773	2665444	2665578	+	
BLi02778	2669421	2669257	-	*yccC2*
BLi02800	2685947	2686267	+	
BLi03252	3118011	3117814	-	
BLi03275	3142332	3143408	+	*yoaN1*
BLi03523	3365908	3366282	+	
BLi03564	3409984	3410073	+	
BLi03571	3416061	3415396	-	
BLi03584	3422949	3423176	+	
BLi03620	3453653	3453579	-	
BLi03656	3483214	3482549	-	*yvbJ2*
BLi03660	3486205	3485876	-	*yoaZ2*
BLi03727	3552312	3552584	+	
BLi04043	3870416	3870661	+	
BLi04054	3879902	3879576	-	*ywbI2*
BLi04063	3886860	3886555	-	
BLi04112	3934943	3934791	-	
BLi04143	3970756	3971130	+	
BLi04153	3978373	3977690	-	
BLi04186	4014225	4014064	-	
BLi04197	4027880	4027263	-	*katE2*
BLi04295	4145251	4145511	+	
BLi04296	4145904	4146893	+	
BLi04299	4148379	4148927	+	*yvfS1*
BLi04309	4156999	4157157	+	

Table S11 Predicted *cis*-regulatory elements

Please refer to "Chapter B_Additional information" on digital medium.

Table S12 Comparison of *cis*-regulatory elements known from *B. subtilis*

Please refer to "Chapter B_Additional information" on digital medium.

Table S13 Cluster analysis of ncRNA expression profiles

Please refer to "Chapter B_Additional information" on digital medium.

Table S14 Comparison of ncRNAs to *B. subtilis*
All ncRNAs identified in this study were compared to ncRNAs identified for *B. subtilis* by Nicolas et al. (2012). Last and second last columns show the names for matching transcripts and the allocated classes in *B. subtilis*.

RNA feature	Start	Stop	Strand	Class	*B. subtilis* transcript tag	*B. subtilis* class
BLi_r0001	945	5	-	A_{misc}	S4, S2	indep, Indep-nt
BLi_r0016	30440	30837	+	indep	S17	5'
BLi_r0026	54282	49490	-	A_{misc}	S25	Indep-NT
BLi_r0035	76356	74559	-	A_{misc}	S37	indep
BLi_r0329	518777	519149	+	indep	S140	indep
BLi_r0373	571429	571162	-	A_{misc}	S166	Indep-NT
BLi_r0598	837456	837589	+	A_{misc}	S274	inter
BLi_r0706	959393	959171	-	A_{misc}	S313	indep
BLi_r0736	991982	992128	+	indep	S1029	indep
BLi_r0844	1082413	1082091	-	indep	S357	indep
BLi_r1000	1262387	1262504	+	indep	S414	inter
BLi_r1003	1263961	1264435	+	A_{misc}	S416	Indep-NT
BLi_r1125	1438764	1438856	+	indep	S433	inter
BLi_r1235	1558438	1556635	-	A_{misc}	S503	Indep-NT
BLi_r1259	1588610	1588725	+	indep	S512	indep
BLi_r1347	1673741	1673635	-	A_{misc}	S547	indep
BLi_r1370	1710228	1709186	-	A_{misc}	S562	indep
BLi_r1500	1984654	1984873	+	A_{misc}	S665	indep
BLi_r1622	2128868	2128723	-	A_{misc}	S708	Indep-NT
BLi_r1777	2271835	2271887	+	A_{misc}	S821	indep
BLi_r1808	2302203	2301803	-	indep	RnpB	
BLi_r1826	2320181	2320265	+	A_{misc}	S849	indep
BLi_r1838	2359724	2359631	-	indep	S863	indep
BLi_r1955	2492524	2492751	+	A_{misc}	S907	indep
BLi_r2042	2607228	2608674	+	A_{misc}	S951	Indep-NT
BLi_r2043	2608736	2609200	+	A_I	S951	Indep-NT
BLi_r2163	2770133	2769931	-	indep	BSU_misc_RNA_41	
BLi_r2295	2957279	2960402	+	A_{misc}	S1105	Indep-NT

Table S14 Continued

RNA feature	Start	Stop	Strand	Class	*B. subtilis* transcript tag	*B. subtilis* class
BLi_r2390	3060847	3062662	+	A_{misc}	S1157	Indep-NT
BLi_r2442	3114363	3114648	+	A_I	S1180	Indep-NT
BLi_r2497	3179098	3178992	-	A_3	S1476	inter
BLi_r2500	3179557	3179966	+	A_{misc}	S1201	inter
BLi_r2546	3229393	3229473	+	A_I	S1225	Indep-NT
BLi_r2548	3232552	3232651	+	A_3	S1227	Indep-NT
BLi_r2624	3299320	3299025	-	indep	BsrI	
BLi_r2758	3469610	3469009	-	indep	SsrA	
BLi_r2771	3479954	3480425	+	A_I	S1299	inter
BLi_r2863	3611192	3612078	+	A_{misc}	S1354	Indep-NT
BLi_r2868	3615456	3617050	+	A_{misc}	S1359	Indep-NT
BLi_r2883	3640577	3640465	-	indep	S1372	inter
BLi_r3204	4051076	4050982	-	indep	S1514	inter
BLi_r3205	4051704	4054845	+	A_{misc}	S1520	Indep-NT
BLi_r3313	4220164	4220560	+	A_I	S1579	indep

Table S15 Comparison of small RNAs from *B. subtilis* to identified ncRNAs
B. subtilis small RNAs were taken from Nicolas et al. (2012) and SubtiWiki (Mäder et al., 2012). Detection was derived by comparison to Rfam (Gardner et al., 2011) or locus conservation.

B. subtilis sRNA	*B. licheniformis* RNA feature	*B. licheniformis* Class	Start	Stop	Strand	Detected by	Locus in *B. subtilis*	Locus in *B. licheniformis*
6S RNA								
BsrA	BLi_r2163	indep	2770133	2769931	-	Rfam+Locus	*yrvM<<*<<aspS*	*yrvM<<*<<aspS*
BsrB	BLi_r1454	indep	1929530	1929709	+	Rfam	*yocI<<*<<yocJ*	*cwlC<>*><*putative transmembrane protein
tmRNA								
SsrA	BLi_r2758	indep	3469610	3469009	-	Rfam+Locus	*yvaG<<*<<smpB*	phage integrase *><*<<smpB*
small cytoplasmatic RNA (protein secretion)								
Scr	BLi_r0016	indep	30440	30837	+	Rfam+Locus	*tadA>>*>>dnaX*	*tadA>>*>>dnaX*
RNA component of RNase P								
RnpB	BLi_r1808	indep	2302203	2301803	-	Rfam+Locus	*ypsC<<*<<gpsB*	*ypsC<<*<<gpsB*
regulatory RNAs with known function								
FsrA	Locus not or only partially conserved						*ykuI>>*>_>ykuJ*	
RatA	Locus not or only partially conserved						*yqdB><*<_>bsrH*	
SR1	BLi_r1306[#]	A_{misc}	1639742	1639946	+	Locus	*slp<_>*><speA*	*slp<>*><*SR1-like protein
SR4	please refer to Results and Discussion						*yolA<>*><yokL*	
rnaA	BLi_r0026	A_{misc}	54282	49490	-	Locus	*yabE><*<>rnmV*	*yabE><*<>rnmV*
small RNAs with unknown functions								
BsrC	No transcript [##]						*ydaG><*<>ydaH*	
BsrD	Locus not or only partially conserved						*yddM>?*?<yddN*	
BsrE	Locus not or only partially conserved						*yoyA<<*<<yobJ*	

Table S15 Continued

***B. subtilis* sRNA**	***B. licheniformis* RNA feature**	***B. licheniformis* Class**	**Start**	**Stop**	**Strand**	**Detected by**	**Locus in *B. subtilis***	**Locus in *B. licheniformis***
BsrF	Locus not or only partially conserved						*yobO*>_>*><*csaA*	
BsrH	Locus not or only partially conserved						*ratA*<_>*><*yqbM*	
BsrI	BLi_r2624	indep	3299320	3299025	-	Locus	*sufC*<<*<<*yurZ*	cation efflux facilitator<<*<<*yurZ*
CsfG	BLi_r1347	A_{misc}	1673741	1673635	-	Locus	*ylbG*><*<>*ylbH*	*ylbG*><*<>*ylbH*
SurA	Locus not or only partially conserved						*yndK*><*<>*yndL*	
SurC	BLi_r2049	A_{misc}	2634028	2634231	+	Locus	*dnaJ*<>*><*dnaK*	*dnaJ*<>*><*dnaK*
RsaE	BLi_r1000	indep	1262387	1262504	+	Rfam+Locus	*yizD*<>*><*yjbH*	*yizD*<>*><*yjbH*
RnaB	Locus not or only partially conserved						*desR*><*<<*yocH*	
RnaC	Locus not or only partially conserved						*yrhK*>>*><*cypB*	

[#]: sRNA SR1 and SR1-like protein BLi05056 ((Gimpel et al., 2010; Sr1: small regulatory RNA controlling ahrC expression, peptide controlling the stability of the cggR-gapA mRNA)

[##]: A positive Rfam result (Gardner et al., 2011) has been found, but an active transcript has been found exclusively on the opposite strand. Nevertheless, this results match the transcriptome pattern of *B. subtilis* in this chromosomal region (Nicolas et al., 2012).

Table S16 Sequence Read Archive accession

Type	Name	Accession number
STUDY	*Bacillus licheniformis* DSM13 Transcriptome	SRP018744
SAMPLE	M-I	SRS396769
EXPERIMENT	whole transcriptome RNA-Seq M-I	SRX242882
SAMPLE	M-II	SRS396772
EXPERIMENT	whole transcriptome RNA-Seq M-II	SRX242885
SAMPLE	M-III	SRS396773
EXPERIMENT	whole transcriptome RNA-Seq M-III	SRX242886
SAMPLE	M-IV	SRS396774
EXPERIMENT	whole transcriptome RNA-Seq M-IV	SRX242887
SAMPLE	M-V	SRS396775
EXPERIMENT	whole transcriptome RNA-Seq M-V	SRX242888
SAMPLE	R-I	SRS396777
EXPERIMENT	whole transcriptome RNA-Seq R-I	SRX242890
SAMPLE	R-II	SRS396778
EXPERIMENT	whole transcriptome RNA-Seq R-II	SRX242891
SAMPLE	R-III	SRS396779
EXPERIMENT	whole transcriptome RNA-Seq R-III	SRX242892
SAMPLE	R-IV	SRS396781
EXPERIMENT	whole transcriptome RNA-Seq R-IV	SRX242893
SAMPLE	R-V	SRS396782
EXPERIMENT	whole transcriptome RNA-Seq R-V	SRX242894
SAMPLE	L-I	SRS396783
EXPERIMENT	whole transcriptome RNA-Seq L-I	SRX242895
EXPERIMENT	differential RNA-Seq L-I	SRX242900
SAMPLE	L-II	SRS396784
EXPERIMENT	whole transcriptome RNA-Seq L-II	SRX242896
EXPERIMENT	differential RNA-Seq L-II	SRX242901
SAMPLE	L-III	SRS396785
EXPERIMENT	whole transcriptome RNA-Seq L-III	SRX242897
EXPERIMENT	differential RNA-Seq L-III	SRX242902
SAMPLE	L-IV	SRS396786
EXPERIMENT	whole transcriptome RNA-Seq L-IV	SRX242898
EXPERIMENT	differential RNA-Seq L-IV	SRX242903
SAMPLE	L-V	SRS396787
EXPERIMENT	whole transcriptome RNA-Seq L-V	SRX242899
EXPERIMENT	differential RNA-Seq L-V	SRX242904

Table S17 Primer pairs for Northern blots

Probe	Figure	Target	Primer (5'-3')	Position	Strand
1	S1A	BLi_r0329	AACAGGAACACGCAAAAGAGCAA	518774	+
			CTAATACGACTCACTATAGGGAGAGCCGAAGCG GTCTCAATTATC	519297	-
2	S1B	BLi_r0450	GGAAATCATGATATATTCAAGGTGTATG	652827	+
			CTAATACGACTCACTATAGGGAGACGTAGTGGCC ATTCTGTCATC	653057	-
3	S1B	BLi_r0451	CTAATACGACTCACTATAGGGAGAGGAAGCAAC AGAGAGGCTCC	652922	+
			GTCGGCTCATCCTCCATCTC	653148	-
4	S1C	BLi_r0872	CACTAGCTTTTTCTATATGCCATTTG	1108930	+
			CTAATACGACTCACTATAGGGAGACAGCTTCACA AGTCCGCAAC	1109069	-
5	S1D	BLi_r1306	GATTTAAATGTGTTATACAATTTACCGTTGAC	1639680	+
			CTAATACGACTCACTATAGGGAGAAAAGAACAG CAGGCAATCCTGTAAA	1639975	-
6	S1E	BLi_r1585	GAGCGCTTCATTTCCATTAATAAGCC	2082882	+
			CTAATACGACTCACTATAGGGAGACCTGTCACAA ATGTAGAAGGAAACG	2083388	-
7	S1F	BLi_r2340	CATTCATTTCGCCTCCGTAGCTC	2997767	+
			CTAATACGACTCACTATAGGGAGACTTCTTCGAG ACCAGAGATGTAATC	2997975	-
8	S1G	BLi_r3281	CAAAGCCGTCTATATAGTAACGTC	4187044	+
			CTAATACGACTCACTATAGGGAGATTCCGTCCCG AGCCACTC	4187205	-

References

Darmon, E., Noone, D., Masson, A., Bron, S., Kuipers, O.P., Devine, K.M., and van Dijl, J.M. (2002). A Novel Class of Heat and Secretion Stress-Responsive Genes Is Controlled by the Autoregulated CssRS Two-Component System of *Bacillus subtilis*. J Bacteriol *184*, 5661.

Davidson, F., Seon-Yi, C., and Stanley-Wall, N. (2012). Selective Heterogeneity in Exoprotease Production by *Bacillus subtilis*. PLoS One *7*, e38574.

Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., et al. (2011). Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res *39*, D141–5.

Hoffmann, K., Wollherr, A., Larsen, M., Rachinger, M., Liesegang, H., Ehrenreich, A., and Meinhardt, F. (2010). Facilitation of direct conditional knockout of essential genes in *Bacillus licheniformis* DSM13 by comparative genetic analysis and manipulation of genetic competence. Appl Environ Microbiol *76*, 5046–5057.

Jahn, N., Preis, H., Wiedemann, C., and Brantl, S. (2012). BsrG/SR4 from *Bacillus subtilis*--the first temperature-dependent type I toxin-antitoxin system. Mol Microbiol *83*, 579–598.

Mäder, U., Schmeisky, A.G., Flórez, L.A., and Stülke, J. (2012). SubtiWiki--a comprehensive community resource for the model organism *Bacillus subtilis*. Nucleic Acids Res *40*, D1278–87.

Mao, F., Dam, P., Chou, J., Olman, V., and Xu, Y. (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Res *37*, D459–63.

Marciniak, B.C., Trip, H., van-der Veek, P.J., and Kuipers, O.P. (2012). Comparative transcriptional analysis of *Bacillus subtilis* cells overproducing either secreted proteins, lipoproteins or membrane proteins. Microb Cell Fact *11*, 66.

Murray, E.J., Kiley, T.B., and Stanley-Wall, N.R. (2009). A pivotal role for the response regulator DegU in controlling multicellular behaviour. Microbiology+ *155*, 1–8.

Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., et al. (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. Science *335*, 1103–1106.

CHAPTER C

FERMENTATION STAGE-DEPENDENT ADAPTATIONS OF *BACILLUS LICHENIFORMIS* DURING ENZYME PRODUCTION

Sandra Wiegand, Birgit Voigt, Dirk Albrecht, Johannes Bongaerts, Stefan Evers, Michael Hecker, Rolf Daniel and Heiko Liesegang

Microbial Cell Factories (2013), 12:120

Authors' contributions to this work

Performed the experiments: SW
Analyzed data: SW, BV, HL
Supervised proteomics experiments: BV
Performed mass spectrometry analyses: DA
Provided fermentation facilities: JB, SE
Provided research facilities: MH, RD
Wrote paper: SW, HL, RD
Designed research: HL

Wiegand *et al. Microbial Cell Factories* 2013, **12**:120
http://www.microbialcellfactories.com/content/12/1/120

MICROBIAL CELL FACTORIES

RESEARCH Open Access

Fermentation stage-dependent adaptations of *Bacillus licheniformis* during enzyme production

Sandra Wiegand[1], Birgit Voigt[2], Dirk Albrecht[2], Johannes Bongaerts[3], Stefan Evers[3], Michael Hecker[2], Rolf Daniel[1] and Heiko Liesegang[1*]

Abstract

Background: Industrial fermentations can generally be described as dynamic biotransformation processes in which microorganisms convert energy rich substrates into a desired product. The knowledge of active physiological pathways, reflected by corresponding gene activities, allows the identification of beneficial or disadvantageous performances of the microbial host. Whole transcriptome RNA-Seq is a powerful tool to accomplish in-depth quantification of these gene activities, since the low background noise and the absence of an upper limit of quantification allow the detection of transcripts with high dynamic ranges. Such data enable the identification of potential bottlenecks and futile energetic cycles, which in turn can lead to targets for rational approaches to productivity improvement. Here we present an overview of the dynamics of gene activity during an industrial-oriented fermentation process with *Bacillus licheniformis*, an important industrial enzyme producer. Thereby, valuable insights which help to understand the complex interactions during such processes are provided.

Results: Whole transcriptome RNA-Seq has been performed to study the gene expression at five selected growth stages of an industrial-oriented protease production process employing a germination deficient derivative of *B. licheniformis* DSM13. Since a significant amount of genes in *Bacillus* strains are regulated posttranscriptionally, the generated data have been confirmed by 2D gel-based proteomics. Regulatory events affecting the coordinated activity of hundreds of genes have been analyzed. The data enabled the identification of genes involved in the adaptations to changing environmental conditions during the fermentation process. A special focus of the analyses was on genes contributing to central carbon metabolism, amino acid transport and metabolism, starvation and stress responses and protein secretion. Genes contributing to lantibiotics production and Tat-dependent protein secretion have been pointed out as potential optimization targets.

Conclusions: The presented data give unprecedented insights into the complex adaptations of bacterial production strains to the changing physiological demands during an industrial-oriented fermentation. These are, to our knowledge, the first publicly available data that document quantifiable transcriptional responses of the commonly employed production strain *B. licheniformis* to changing conditions over the course of a typical fermentation process in such extensive depth.

Keywords: Differential gene expression, Transcriptomics, Proteomics, RNA-Seq, Subtilisin Carlsberg, Industrial production, Stress response, Sporulation, Lichenicidin

* Correspondence: hlieseg@gwdg.de
[1]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany
Full list of author information is available at the end of the article

BioMed Central

© 2013 Wiegand et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

For several decades, strains of the *Bacillus subtilis* group [1] have been exploited for industrial purposes. The scope of applications includes the production of amylases, proteases and antibiotics by strains of *B. subtilis*, *B. amyloliquefaciens*, *B. pumilus* or *B. licheniformis* [2]. High capacities of product secretion, high growth rates, and the GRAS (generally regarded as safe) status of many strains have contributed to the employment of these species as biotechnological workhorses [2]. In general, the production process can be considered as an energy consuming biotransformation in which a nutrient rich substrate is converted into the desired product by a member of the genus *Bacillus*.

The productive process examined in this study is based on the production platform *B. licheniformis*, which has been proven to perform well for the production of alkaline proteases and in particular subtilisins, which are used in all types of laundry detergents [3]. Therefore, research efforts have been focused on the *B. licheniformis* subtilisin fermentation process and the resulting yield of active enzyme. A major aspect has been monitoring and improvement of bioprocess parameters such as oxygen transfer rate [4-6], pH value [7,8], inoculum quality [9] and initial glucose concentration [10], whereas other studies addressed the optimization of the fermentation medium [11,12]. Strategies for the molecular biological improvement of subtilisin [13] and its secretion [14] have been described. Attention has also been paid to strain optimization by generation of deletion mutants targeting transfer of genetic material [15,16], secretion capability [17], sporulation and biological containment [18-20]. Investigation of *B. licheniformis* under different stress conditions by proteomics and microarray-based transcriptomics have been applied to identify marker genes [21-24], to enable the detection of stressors during a productive fermentation process. However, rational strain or bioprocess optimization requires potential targets and therefore the knowledge of genomic activities during the crucial stages of a fermentation process under industry-oriented conditions is essential.

An RNA-Seq-based study targeting the identification of *B. licheniformis* DSM13 RNA-based regulatory elements such as non-coding and antisense RNAs under production-oriented growth conditions has recently been published by our group [25]. The application of RNA-Seq allows the quantification of transcripts with a hitherto unmatched dynamic range spanning several orders of magnitude [26], therefore enabling in depth analysis of differential expression between physiological conditions or developmental states. Further advantages of RNA-Seq are the low background noise, the provided single base resolution and the high reproducibility [26,27]. Therefore, RNA-Seq, especially when coupled with other "omics" techniques like 2D gel-based proteomics, provides the opportunity for global investigation of microbial gene expression. However, although recent advantages in RNA-Seq technology have greatly enhanced the efficiency and availability of this approach, no such data on industrial fermentations of *B. licheniformis* have been made publicly available to this day.

To identify gene activities of *B. licheniformis* directly related to the productivity of a subtilisin fermentation process, we present a high-resolution quantitative and dynamic exploration of the transcriptional responses of *B. licheniformis* confirmed by proteome data. Special attention was given to production stage-related adaptions of *B. licheniformis*. The RNA abundances and the cytoplasmic proteome composition of all samples were determined by RNA-Seq experiments and by 2D gel electrophoresis [25], respectively. As measure of gene expression, the normalized amount of sequenced nucleotides per gene is expressed in single-base resolution by the NPKM (**n**ucleotide activity **p**er **k**ilobase of exon model per **m**illion mapped reads) value [25], which is closely related to the more common RPKM value [28]. These data provide a first analytical framework to gain better understanding of the dynamics during such fermentations, and to enable the identification of potential physiological and genetic bottlenecks. Furthermore, the data are intended as a reference for subsequent comparisons with transcriptome data from other fermentation procedures employing related *Bacillus* strains, in order to guide rational approaches for the optimization of production processes.

Results and discussion

In this study, transcriptome and proteome data of selected samples from an industry-oriented fermentation have been analyzed with focus on physiological changes during the process. The samples were taken in triplicate at five time points (sampling points I-V) during growth within a subtilisin fermentation process of *B. licheniformis* MW3Δspo (Figure 1; Additional file 1: Figure S1). This strain is a germination deficient mutant of *B. licheniformis* DSM13, transformed with an expression plasmid encoding a subtilisin protease. Sampling point I represents the growth in presence of glucose, whereas sampling points II and III correspond to the subsequent phase of glucose starvation. Sampling points IV and V represent the productive stages of the process in which the alkaline protease is synthesized and secreted.

Previously, we curated the annotation of 4172 protein-coding genes and determined the respective transcript abundances in all samples [25,29], resulting in NPKM values from 0 for lacking transcripts to 85.267 for the most abundant transcripts (Figure 2). Analysis of the obtained data with baySeq [30] and ANOVA revealed that 980 and 1016 genes, respectively, are differentially expressed at the different sampling points. In total, 1395

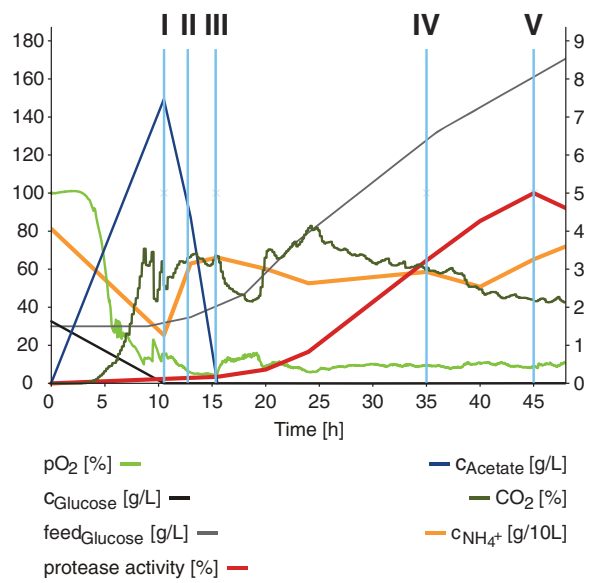


Figure 1 Protease production and process parameters. Process parameters are shown for fermentation L (please refer to Additional file 1: Figure S1 for replicate fermentations R and M). Oxygen partial pressure pO_2 [%], glucose concentration $c_{Glucose}$ [g/L], supplied glucose $feed_{Glucose}$ [g/L] and normalized protease activity [%] are displayed on the left y-axis, whereas acetate concentration $c_{Acetate}$ [g/L], carbon dioxide content CO_2 [%], and ammonium concentration c_{NH4+} [g/10 L] are scaled on the right y-axis. The process time t [h] is given on the x-axis. The sampling points I to V are indicated by light blue lines. The figure was modified from Wiegand et al. [25] where a detailed analysis of RNA-Seq data of the same experimental setup aiming at the identification of regulatory RNAs was performed.

genes were determined as differentially expressed by at least one method and utilized for further analysis. Generic GO slim enrichment analysis [31,32] revealed that genes assigned to regulation, protein modification and metabolism, DNA metabolism, cell cycle and translation are underrepresented within this dataset of differentially expressed genes. Overrepresented genes were assigned to protein transport, response to external stimuli, carbohydrate metabolic processes and cell differentiation.

The transcriptome data allowed the assignment of 3567 genes to 23 clusters by *k*-means cluster analysis based on the determined differentially expressed genes (Figure 3 and Figure 4; Additional file 2: Table S1) [36]. Each cluster was also examined for over- and underrepresented groups of genes by GO term-based enrichment analysis (Additional file 2: Table S2) [32,37]. Clusters A-H and N-Q comprise genes which are more abundant at the early stages of the process than at the later sampling points. In this group, overrepresented genes are mainly involved in gene expression and translation, biosynthetic processes, transport and metabolism of amino acids, or central carbon metabolism including glycolysis and TCA cycle. Another pattern can be found for clusters I-M and R + S which contain genes displaying higher transcript abundances in the productive stage of the fermentation (sampling points IV and V). In clusters with the highest measured transcript abundances at stage IV (K-M) genes were predominantly involved in sporulation processes. Transcripts more abundant in stage V than in stage IV are depicted in clusters I and J and, among others, encompass genes for phosphate ABC transporter PstABC and nitrate reductase NarGHIJ.

Detailed analyses of transcript and protein abundances (Additional file 2: Table S3) concerning important factors of bacterial growth and productivity (amino acid transport and metabolism, central carbon metabolism, starvation and stress responses, and protein secretion) will be presented in the following passages.

Amino acid transport and metabolism

The examined fermentation process was performed in the presence of a complex nitrogen source initially supplemented with glucose. To elucidate how *B. licheniformis* utilizes the supplied peptide substrate, the transcript (Figure 5 and Figure 6) and protein (Additional file 1: Figure S2 and Figure S3) abundances of the major amino acid metabolism-related genes were examined in the context of their metabolic network.

The genome of *B. licheniformis* encodes six unambiguous operons encoding peptide ABC transporters (*app1*, *app2*, *dpp*, *opp*, BLi00892-96, BLi02527-31) [25,29], four of them showing transcript abundance under the examined conditions (Figure 5). The *app1* and the *opp* operon each encode oligopeptide ABC transporter systems. They are transcribed during all stages of the fermentation process, but show top transcript abundances at the earlier sampling points, particularly at sampling point II (NPKM values >500). In contrast, the *dpp* operon encoding a dipeptide ABC transporter displays increased RNA abundance over time with maximal levels at sampling point IV. Furthermore, transcripts of the dipeptide/oligopeptide ABC transporter operon BLi00892-96 are only abundant at the later fermentation stages. Regarding the RNA abundances of the *opp* and *dpp* operons, similar patterns of activation and repression during cell growth and sporulation have been observed in *B. subtilis* [38-40]. In contrast, the *app1* operon, which is orthologous to the *app* operon of *B. subtilis*, does not resemble the sporulation-dependent regulation in the model organism [39,40], indicating a different regulation of this operon in *B. licheniformis.*

The RNA abundances of the ABC transporter operons seem to reveal a fermentation stage-dependent pattern, promoting the idea that oligopeptides are imported primarily, whereas dipeptides are presumably consumed after oligopeptides are exhausted. This transcriptional pattern may be influenced by the fact that dipeptides should become more available over time due to the activity of extracellular protease secreted within the fermentation process. The RNA abundances of further

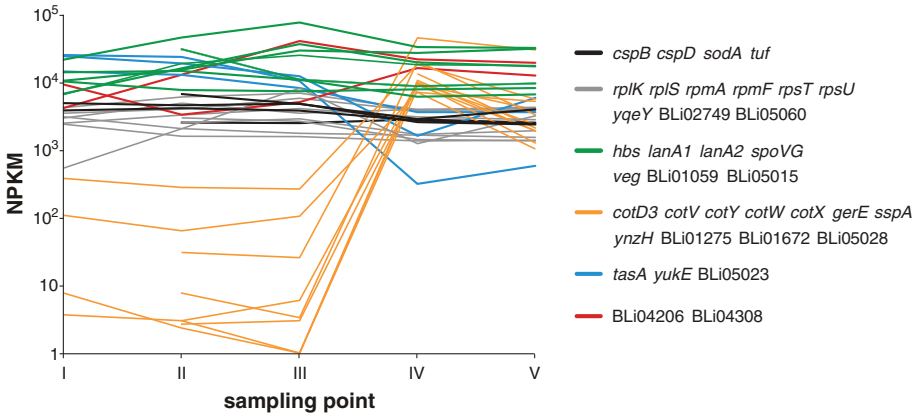


Figure 2 Most abundant gene transcripts. Mean NPKM values of the most abundant gene transcripts are plotted against sampling points. Colored lines indicate classes of similarly abundant transcripts. Grey and black lines represent genes with high RNA abundances in all 15 samples (NPKM values 1000–5000); black lines additionally indicate genes referred to in the text. Green color marks those genes, whose transcripts are most abundant (NPKM values >5000) throughout all sampling points, whereas genes with highly abundant transcripts (NPKM values >5000) only at specific sampling points are shown in orange, red and blue. The most abundant transcript was assigned to *lanA2* in sample R-III. This gene encodes a component of the two-peptide lantibiotic lichenicidin [33] and is transcribed with NPKM values >5000 at all sampling points. Accordingly high transcript abundance can also be observed for *lanA1*, which codes for the second prepeptide of this lantibiotic. Further genes with similar abundances are coding for the BsrG-like peptide (BLi05015) [25,34], the sporulation protein SpoVG, the DNA-binding protein Hbsu, BLi01059 and Veg [35]. The gene encoding the oxygen detoxification protein SodA, the cold shock responsive genes *cspB* and *cspD*, the gene for the elongation factor Tu, and genes coding for components of the translation machinery are transcribed with NPKM values >1000 at all sampling points. Transcripts which are highly abundant exclusively in the later, productive stages of the process are associated to spore formation, whereas transcripts which are highly abundant during the early stages of the process encode ribosomal proteins, proteins of the TCA cycle and ATP synthase subunits. For an illustration of highly abundant proteins please refer to Additional file 1: Figure S8.

amino acid transporters are shown in Additional file 1: Figure S4.

Sampling point I

At this fermentation stage of carbon excess (Figure 1), transcripts of the operon of the glutamate synthase (GOGAT) *gltAB* and the gene of the glutamine synthetase (GS) *glnA* are highly abundant (Figure 6). This prompts the conclusion of a strong glutamate production, which is fed by 2-oxoglutarate provided by the catabolism of glucose (see *Central carbon metabolism*). The high transcript abundance of genes involved in the glutamate-dependent anabolism of proline, aspartate, alanine and aromatic and branched-chain amino acids (Figure 5 and Figure 6) indicates that the produced glutamate is utilized for the synthesis of other amino acids [41,42], despite the given complex amino acid broth.

Further active genes have been assigned to aspartate degradation for pyrimidine biosynthesis (Figure 3), arginine and S-adenosyl methionine (SAM) metabolism for putrescine synthesis, and cysteine degradation releasing sulfur-containing compounds (Figure 5 and Figure 6).

Sampling point II

Upon glucose exhaustion (Figure 1), the transcriptome indicates drastic changes in the fluxes of the amino acid metabolism. Transcripts of genes for glutamate-releasing catabolic processes are highly abundant, as it can also be observed for transcripts of genes promoting the degradation of proline, arginine and branched-chain amino acids (Figure 5 and Figure 6). In reverse, the transcripts of the glutamate-consuming pathways abundant at sampling point I have declined. Glutamate now seems to be metabolized to 2-oxoglutarate by the glutamate dehydrogenase (GDH) GudB and channeled into the TCA cycle. Complementarily, the transcripts of GS and GOGAT are also less abundant [41]. Furthermore, the observed transcript abundances indicate that threonine is metabolized to glycine which is then degraded by the glycine cleavage complex, in order to gain reducing equivalents and C1 compounds while serine is degraded to pyruvate by L-serine dehydratase SdaAAAB to provide further energy sources.

Sampling point III

During the later glucose exhaustion stage (Figure 1) most genes involved in amino acid metabolism show reduced transcript abundances. Elevated abundances are nearly exclusively found in pathways involved in serine degradation, such as the above mentioned conversion of serine to pyruvate, the metabolization to glycine for subsequent degradation by the glycine cleavage complex, and the conversion to cysteine which is then further metabolized to pyruvate via the intermediate alanine (Figure 5 and Figure 6).

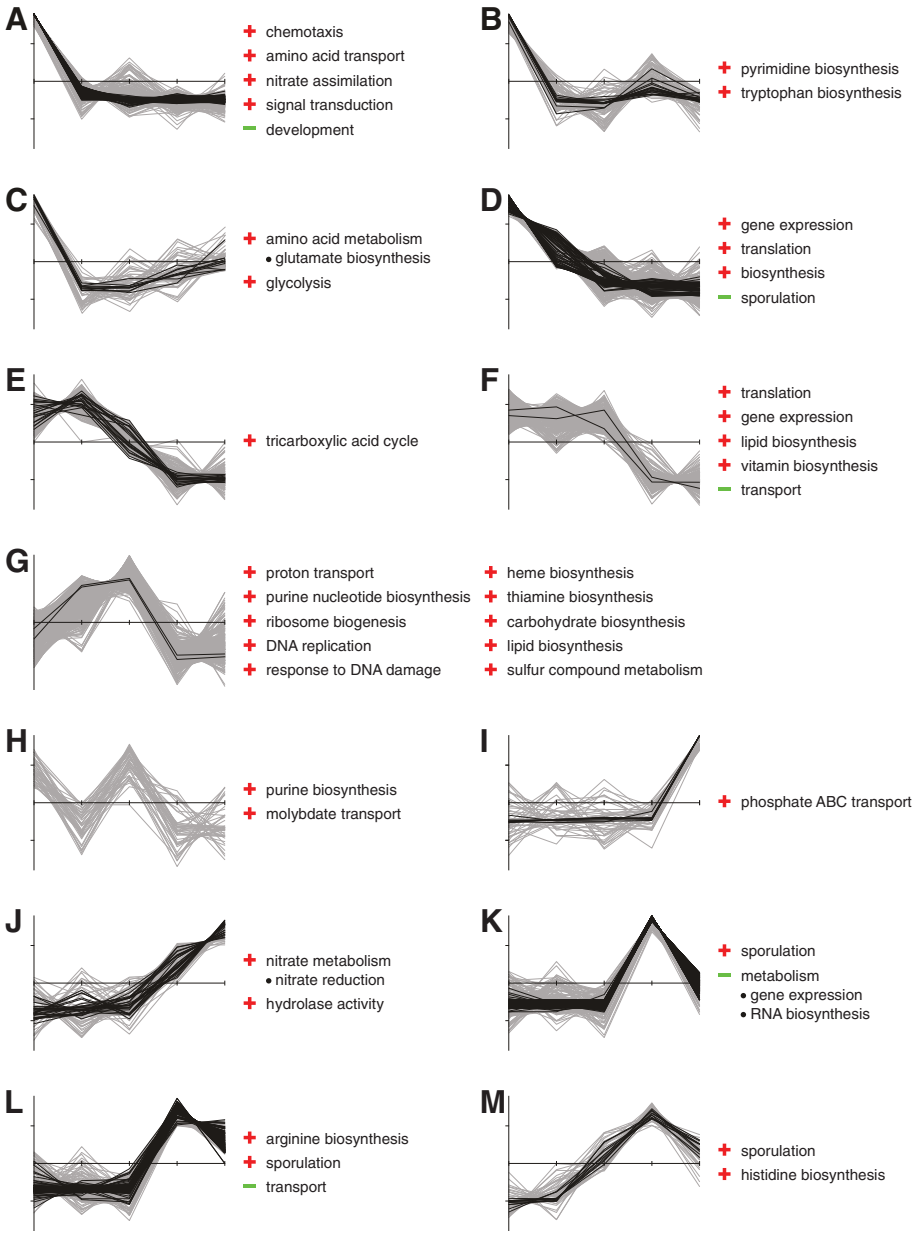


Figure 3 *k*-means clustering of expression profiles with assigned GO terms. Thirteen gene clusters determined by *k*-means clustering are depicted with corresponding, significantly enriched GO terms [32] (see also Figure 4). Over- and underrepresented GO terms are indicated by red and green symbols, respectively. Black lines mark genes with a baySeq [30] likelihood value >0.99, which indicates differential expression; all other genes are colored in grey. All values are depicted by Z-score transformed NPKM values (y-axis) versus sampling points I to V (x-axis).

The transcript abundances of *purA* and *purB* – and other genes associated with purine biosynthesis (Figure 3) - indicate the degradation of aspartate in order to provide building blocks for this pathway.

Sampling points IV & V

In the productive stages of the fermentation process (Figure 1), the determined RNA abundances show that the amino acid metabolism has progressed to the glutamate-consuming synthesis of proline and the nitrogen-rich amino acids arginine and histidine (Figure 6). The reason for this reaction may lie in the previous high induction of genes mediating the degradation of proline and arginine during the earlier stage of glucose exhaustion. The glutamate required for these anabolic reactions is delivered by the glutamate dehydrogenase GdhA [43], and the

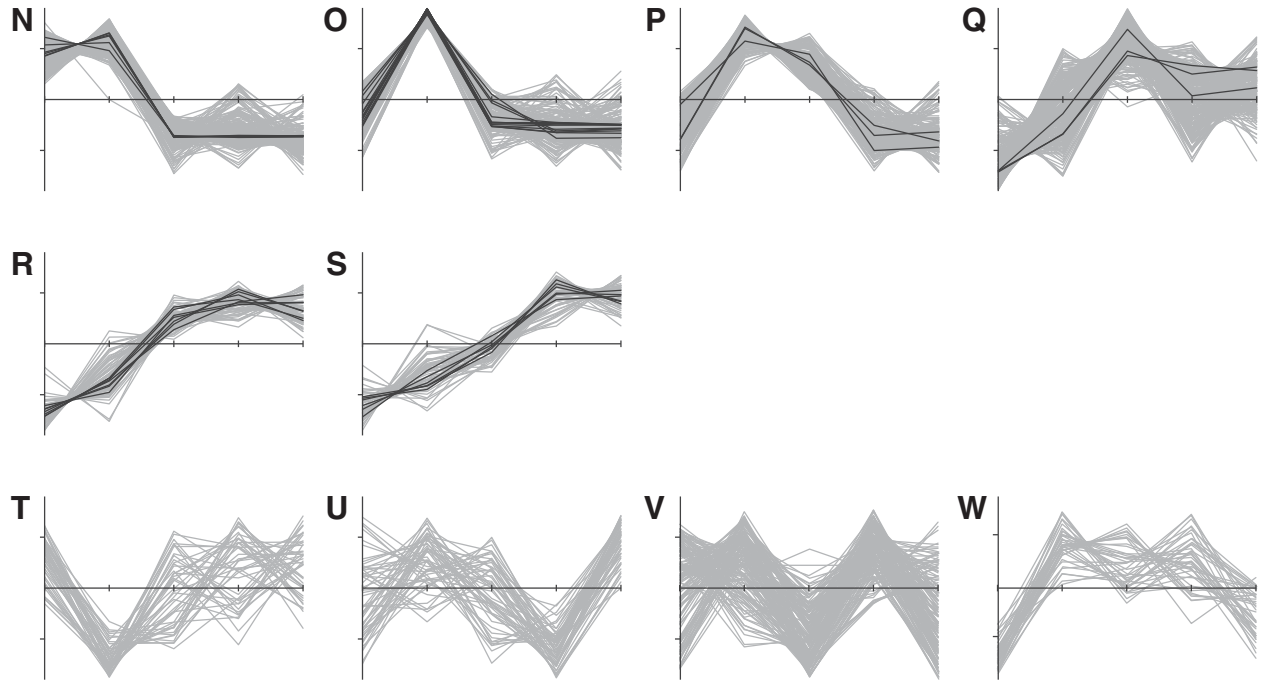


Figure 4 *k*-means clustering of expression profiles. Ten gene clusters determined by *k*-means clustering with no significantly enriched GO terms (see also Figure 3). Black lines mark genes with a baySeq [30] likelihood value >0.99, which indicates differential expression; all other genes are colored in grey. All values are depicted by Z-score transformed NPKM values (y-axis) versus sampling points I to V (x-axis).

GOGAT/GS system which becomes slightly re-induced upon amino acid consumption and the applied glucose feed [41].

Further pathways whose transcripts are abundant at these fermentation stages include the synthesis of threonine via the anabolism of homoserine, and the conversion of valine to alanine (Figure 5 and Figure 6). Also, the transcripts of genes for the degradation of lysine are highly abundant; as members of the σ^E regulon, they are activated by the initiation of sporulation [44]. In addition to the high transcript abundance of sporulation-related genes shown in Figure 3, this is evidence for active sporulation within the fermenter population.

Of course, the conditions in the fermenter do not cause a response to nitrogen limitation as described by Voigt et al. [21]. However, the shut-down of branched-chain amino acid degradation during the phase of glucose exhaustion at sampling point III might be accounted to a limitation effect, as in *B. subtilis* the orthologous transcriptional regulator for activation of this pathway is induced by the presence of such amino acids [45]. Additionally, as transcripts of several amino acid synthesis pathways of are abundant during the later fermentation stages, these amino acids are seemingly not available in excess.

Central carbon metabolism

The production process was initially supplemented with glucose. Upon depletion of this sugar and its derivates, a pulsed glucose feed was established in order to enhance the available energy. Thus, enzymes relevant for sugar catabolism (Figure 7; Additional file 1: Figure S5) and sugar transport (Additional file 1: Figure S6) are required to maintain an optimal energy supply throughout the fermentation. The transcriptional changes of those enzymes were analyzed at the different sampling points and will be described in the following passages.

Sampling point I

Shortly before the total depletion of the initially supplied glucose (Figure 1), the genes for glycolytic enzymes are highly transcribed (NPKM values from 359 to 2473). High transcript abundance has also been recorded for the genes of the oxidative pentose phosphate pathway and the TCA, but not for the embedded glyoxylate bypass [29] (Figure 7). Furthermore, the *alsSD* operon for acetoin synthesis and the phosphate acetyltransferase gene *pta* for acetate production were maximally transcribed, indicating the channeling of carbon to the production of overflow metabolites (Figure 1) [46]. However, varying transcript abundances of the acetate kinase gene (*ackA*), also involved in acetate synthesis, were observed. This is due to slightly asynchronous samples for this fermentation stage (Figure 1; Additional file 1: Figure S1) and restricts the determination of reproducible NPKM values for this gene and sampling point (Figure 7).

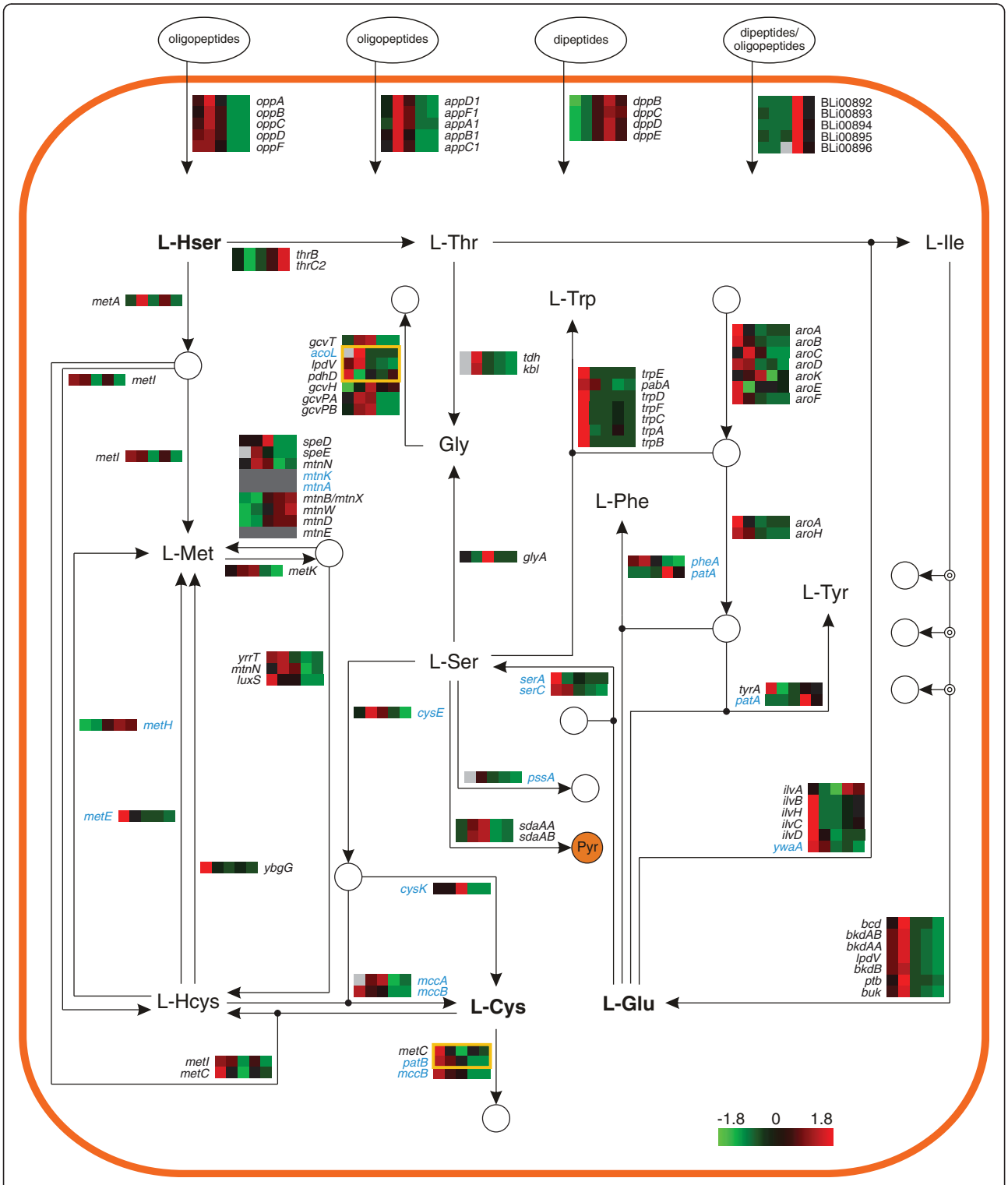


Figure 5 Transcriptome of the amino acid metabolism - Part I. Heat map representation of Z-score transformed NPKM values of genes involved in amino acid transport and metabolism (see also Figure 6). Genes with an assigned antisense RNA [25] are marked in blue, genes with NPKM values <10 at all sampling points are indicated by dark grey boxes and statistically not significant values are indicated by light grey boxes. Amino acids written in bold can also be found in Figure 6. Yellow frames indicate reactions with multiple assigned enzymes of which only one is strictly necessary. Pyr: Pyruvate. For the corresponding proteome data please refer to Additional file 1: Figure S2.

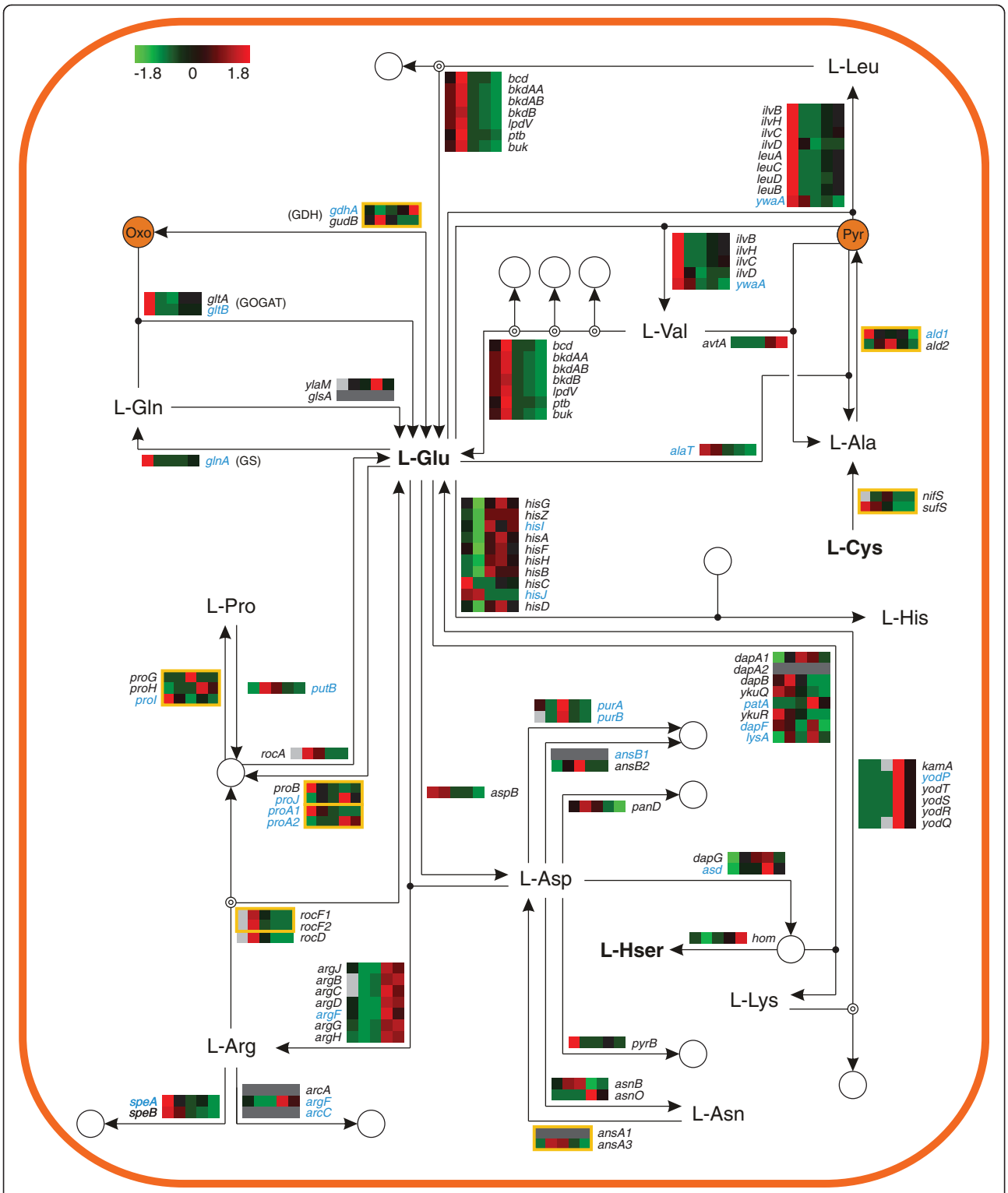


Figure 6 Transcriptome of the amino acid metabolism - Part II. Heat map representation of Z-score transformed NPKM values of genes involved in amino acid transport and metabolism (see also Figure 5). Genes with an assigned antisense RNA [25] are marked in blue, genes with NPKM values <10 at all sampling points are indicated by dark grey boxes and statistically not significant values are indicated by light grey boxes. Amino acids written in bold can also be found in Figure 5. Yellow frames indicate reactions with multiple assigned enzymes of which only one is strictly necessary. Pyr: Pyruvate, Oxo: 2-Oxoglutarate. For the corresponding proteome data please refer to Additional file 1: Figure S3.

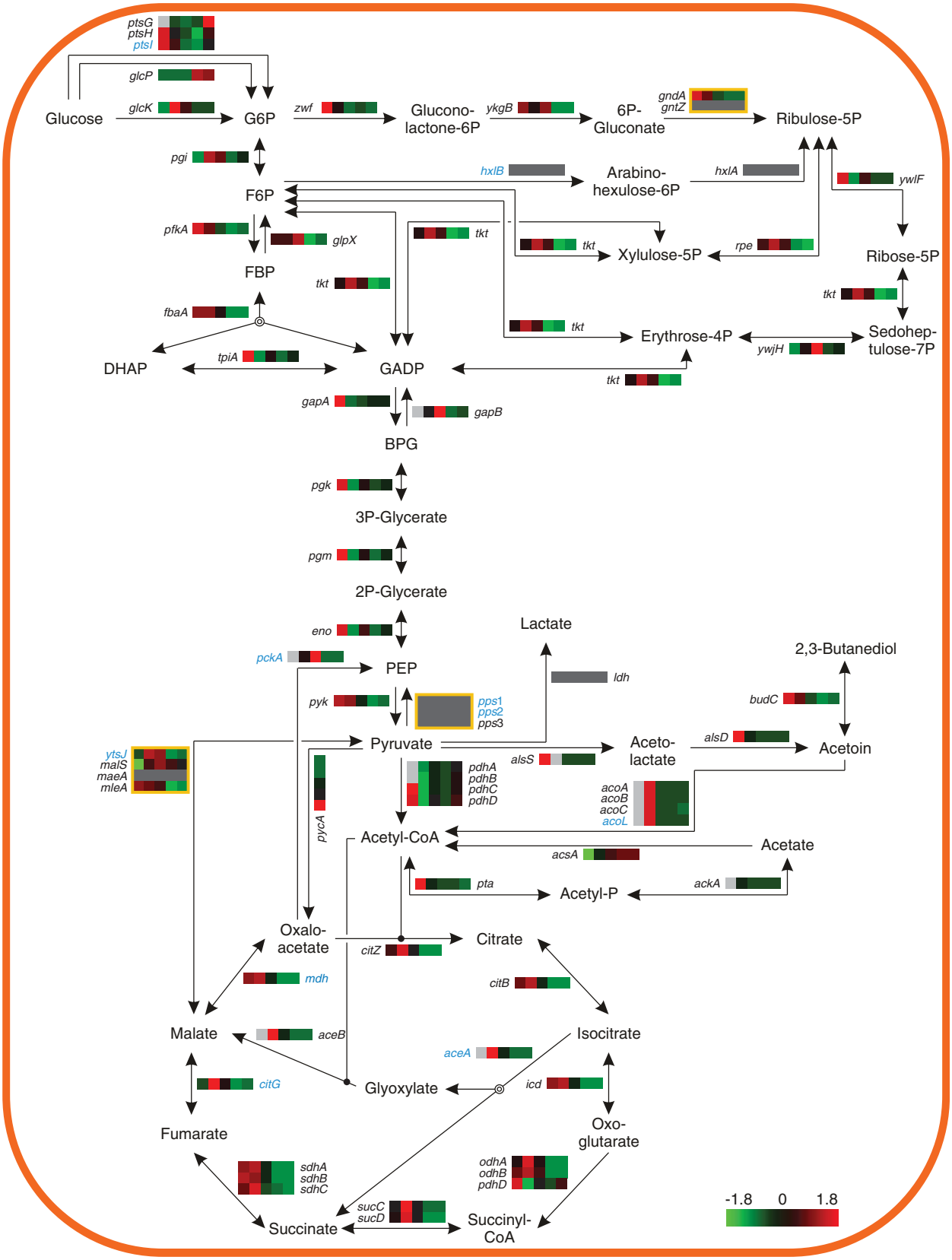


Figure 7 (See legend on next page.)

(See figure on previous page.)
Figure 7 Transcriptome of the central carbon metabolism. Heat map representation of Z-score transformed NPKM values of genes involved in central carbon metabolism. Genes with an assigned antisense RNA [25] are marked in blue, genes with NPKM values <25 at all sampling points are indicated by dark grey boxes and statistically not significant values are indicated by light grey boxes. Yellow frames indicate reactions with multiple assigned enzymes of which only one is strictly necessary. For the corresponding proteome data please refer to Additional file 1: Figure S5.

Sampling point II

After exhaustion of the carbohydrate source (Figure 1), the transcript abundances of the genes of acetate and acetoin synthesis decline (NPKM values <72) (Figure 7). This regulatory effect can be explained by the fact that the expression of the acetate synthesis genes *pta* and *ackA* is influenced by CcpA triggered carbon catabolite activation, as shown for *B. subtilis* [47-49], which ceases with glucose depletion. Furthermore, the acetoin synthesis operon *alsSD* is activated by the transcriptional regulator AlsR in the presence of acetate [50,51]. Therefore, it exhibits reduced transcription when acetate concentration decreases due to the dissimilation of the products formed during overflow metabolism (Figure 1). The dissimilatory reaction is caused by the termination of carbon catabolite repression, allowing an increasing transcript abundance of *acsA* (Figure 7), which encodes an acetyl-CoA synthetase for the conversion of acetate to acetyl-CoA [52]. Transcript abundance of the *acuABC* operon, which has been shown to lead to in- and reactivation of AcsA in *B. subtilis* [53,54], is also increased upon cessation of carbon catabolite repression (Additional file 1: Figure S7). However, an influence of this operon on the acetate or acetoin metabolism of *B. licheniformis* has not been revealed [55]. Furthermore, the transcript abundance of the *acoABCL* operon has strongly increased (NPKM values >2500) (Figure 7). The expression of the corresponding transcriptional activator gene, *acoR*, depends on induction by acetoin [55,56]. Therefore, a high concentration of acetoin is indicated by the high transcript abundance of this operon. In contrast, the gene of the acetoin reductase/2,3-butanediol dehydrogenase *budC* is only weakly transcribed (NPKM value <60) throughout the production process. Thus, it appears that no significant 2,3-butanediol production occurred under the given conditions.

Negative regulatory effects could be observed for the genes of the *gapA* and the *pdh* operon (Figure 7), which are repressed as reaction to glucose starvation [21]. In contrast, the genes coding for the isocitrate lyase AceA and the malate synthase AceB, reach their top level of transcript abundances at this sampling point. Both genes belong to the glyoxylate bypass, allowing *B. licheniformis* not only to gain energy by C2 compound oxidation, but also to grow on acetate and acetoin as sole carbon sources by bypassing the oxidative, CO_2 evolving steps of the TCA cycle [55,57]. Additionally, the high transcript abundance of the other genes of the TCA cycle enables the utilization of 2-oxoglutarate provided from amino acid catabolism (see *Amino acid transport and metabolism*).

In general, the registered changes in metabolism during this process stage are in good accordance with results presented by a previous study on glucose starvation in *B. licheniformis* [21]. However, this is the first time that expression of these production-relevant genes [3] is shown during growth of *B. licheniformis* in rich medium.

Sampling point III

At this stage of the fermentation process, the C2 compounds were completely depleted and the cells entered a short phase of reduced metabolic activity (Figure 1). The genes coding for glycolytic enzymes also involved in gluconeogenesis, which have shown decreased transcript abundances at sampling point II, are slightly increased (Figure 7). This is confirmed by the amount of the corresponding proteins (Additional file 1: Figure S5). Furthermore, transcripts of exclusively gluconeogenic genes *gapB* and *glpX* show their maximal abundances at this sampling point (Figure 7). Phosphoenolpyruvate (PEP), the building block for gluconeogenesis, seems to be converted from oxaloacetate, as the gene for the phosphoenolpyruvate carboxykinase PckA is maximally transcribed. Contrarily, the genes for the phosphoenolpyruvate synthases Pps1, Pps2 and Pps3 [25], promoting PEP synthesis from pyruvate, show only low levels of transcript abundance (NPKM values <25). Additionally, the genes of the non-oxidative pentose phosphate pathway show their highest transcript abundances during the phase of glucose starvation (sampling point II and III). This regulatory effect is in accordance with previous observations in *B. licheniformis* [21], and is remarkable as no glucose-dependent regulation of this pathway has been found in *B. subtilis* [58].

Taken together, the observations indicate that the C2 compounds catabolized via the glyoxylate bypass are utilized for the generation of glucose and other sugars.

Sampling points IV & V

The last two samples were taken during the subtilisin production stage of the fermentation process. At these sampling points, glucose was added to the fermenter in pulsed feeding steps and channeled to energy metabolism via glycolysis and TCA cycle (Figure 7). Although the RNA abundances of both pathways are reduced compared to the previous sampling points, they are still abundant

(NPKM values >100). Similar results were also obtained for the transcript abundances of genes involved in gluconeogenesis and the non-oxidative pentose phosphate pathway. These findings, together with the above described glutamate-consumption by anabolic amino acid pathways during the late stages of the fermentation process, indicate that the pulsed supply with glucose during this fermentation stage is not only sufficient for provision of reducing equivalents, but also for facilitation of 2-oxoglutarate formation needed for glutamate synthesis.

Starvation and stress responses

Bacterial cells react to declining nutrient concentrations or changing environmental conditions by exhibiting well orchestrated starvation or stress responses. To elucidate whether *B. licheniformis* suffers of any of these situations during the fermentation process, we compared the obtained transcriptomic and proteomic data to described starvation and stress responses of *B. licheniformis* and *B. subtilis* [23,24,59-62].

In *B. licheniformis*, oxidative stress induced by hydrogen peroxide results in increased RNA abundances of the PerR, Spx, Fur, and SOS regulon [23]. In our study, we found the PerR as well as the Spx regulon (Figure 8) temporarily induced during the early stages of the fermentation process (sampling points I to III). The transcript as well as the gene product of the superoxide dismutase-encoding *sodA* were highly abundant at all sampling points (see also Figure 2 and Additional file 1: Figure S8), leading to an accumulation of the SodA protein over time. A similar pattern of transcript abundance and protein accumulation could also be observed for the putative thiol peroxidase Tpx, but not for the vegetative catalase KatA. In contrast to these results, the Fur regulon and the SOS regulon did not show distinct RNA abundances. In *B. subtilis*, the SOS regulon has been described to be more responsive to hydrogen peroxide than to superoxide exposure [59]. Therefore, and also considering the high transcript abundance of *sodA*, we infer a cellular response to a potentially toxic superoxide load at the early stages of the fermentation process.

Strikingly, we found the *lan* gene cluster [33,63], which encodes the genes for lantibiotic production and immunity, to show a high transcript abundance during the early stages of the fermentation process (Additional file 1: Figure S9). The highly abundant transcripts of the lichenicidin prepeptide genes *lanA2* and *lanA1* depicted in Figure 2 are members of this cluster. Although the transcript abundances decline at the later fermentation stages, they remain on a high level, indicating a substantial lichenicidin production during the fermentation process. In consistence with the elevated lichenicidin challenge, cell envelope stress responsive operons like *liaRSFGHI* [24]

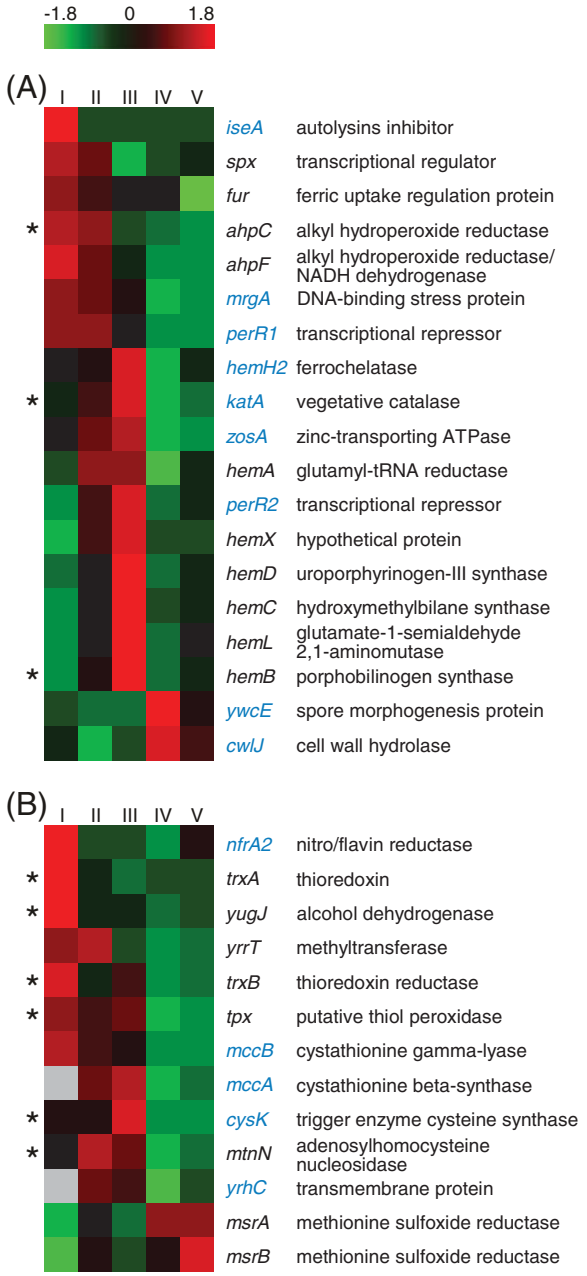


Figure 8 Peroxide stress response. Heat map representation of Z-score transformed NPKM values. The depicted **(A)** PerR and **(B)** Spx regulons have been identified as *B. licheniformis* peroxide stress markers by Schroeter et al. [23]. Genes with an assigned antisense RNA [25] are marked in blue, asterisks indicate a detected protein spot for the respective gene (Additional file 2: Table S3) and statistically not significant values are marked by grey boxes.

display transcript abundances during the early stages of the process (Additional file 1: Figure S10). In the following stages, the level of transcripts declines to NPKM values still >100; corresponding to the high levels of lantibiotic-coding mRNA over the complete fermentation process. It remains elusive why the cells channel energy to the

production of antimicrobial compounds and the corresponding immunity response while grown in pure culture.

The performed cluster analysis indicated emerging transcript abundances for sporulation-dependent genes at sampling point IV (Figure 3; Additional file 1: Figure S11). In general, sporulation is observed as a complex, energy-consuming response to nutrient limitation, which is activated by the master regulator of sporulation Spo0A [64]. Obviously, sporulation is rather unproductive in terms of industrial fermentation and thus undesired during such processes. Unfortunately, it has been shown that deletion of Spo0A does not only result in a sporulation-deficient phenotype, but also in increased cell lysis [65]. In *B. subtilis*, it has been shown that activation of Spo0A is influenced by a cascade of different regulatory systems, including the potassium leakage-sensing kinase KinC [66,67]. Strikingly, one known effect of two-peptide lantibiotics like the aforementioned *lan* cluster is indeed the induction of potassium leakage [68,69]. Thus, our data indicate a lichenicidin-mediated sporulation induction within the fermentation. The deletion of the *lan* system may present a promising approach to strain optimization, as this should lead to reduced levels of phosphorylated Spo0A with less probability of exceeding the sporulation-inducing threshold.

A detailed inspection of iron starvation and heat shock response did not reveal any distinct activation patterns (Additional file 1: Figure S12 and Figure S13), whereas a notable phosphate starvation response could be identified at the latest stage of one fermentation (Additional file 1: Figure S14).

Protein secretion

In the Gram-positive model organism *B. subtilis,* secretion of subtilisin is directed via the secretory (Sec) pathway [14]. Subtilisin is synthesized as preproenzyme [70] containing a Sec-dependent signal peptide [14] and a propeptide, which serves as intramolecular chaperone [71]. The nascent protein chain is recognized by the signal recognition particle (SRP) and transferred to the membrane [72] where it is forwarded to the Sec translocase and transported across the membrane [73]. After cleavage of the signal peptide, the subsequent protein folding into prosubtilisin is aided by the propeptide [74] and the extracytoplasmic chaperone PrsA [75]. Following the autoprocessed cleavage of the propeptide, it is degraded *in trans* and the active enzyme is released into the extracellular space [71]. The heat map depicted in Figure 9A shows the RNA abundances of the required orthologous genes in *B. licheniformis*. Interestingly, the transcript abundances of these genes decline in the late stages of the process in which the main amount of subtilisin is secreted. This observation is also supported on protein level by the abundance of PrsA, which declines at the later sampling points (Additional file 2: Table S3). The only exception to this pattern is the highly abundant SRP component *scr* (4.5S RNA) which shows increased RNA abundance at the later sampling points.

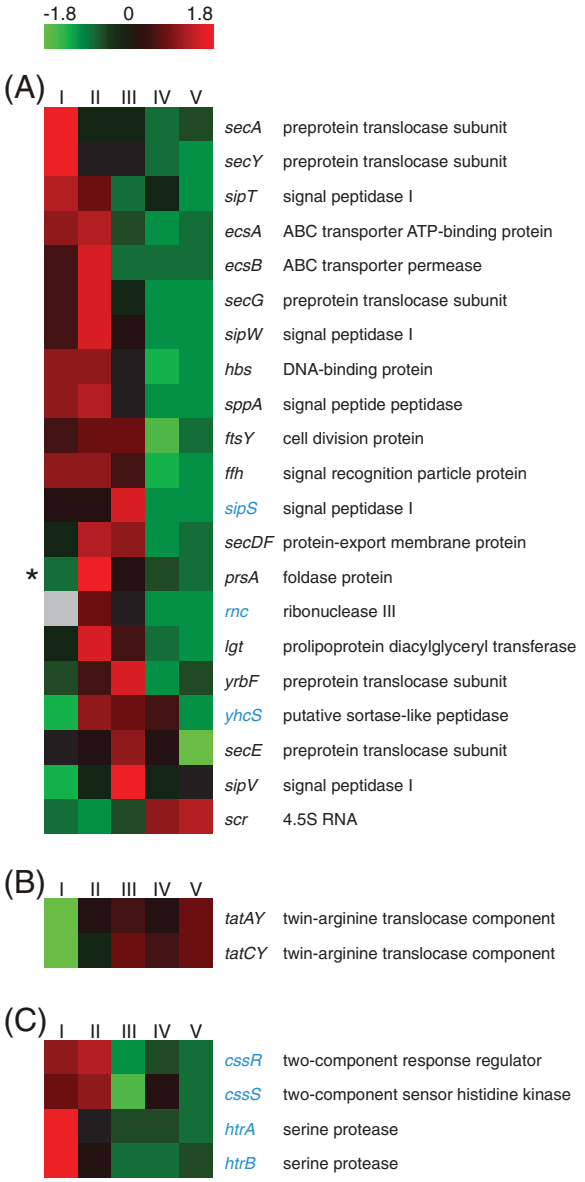


Figure 9 Secretion. Heat map representation of Z-score transformed NPKM values. The depicted genes belong to **(A)** the Sec pathway, **(B)** the Tat pathway and **(C)** the secretion stress CssR regulon [73,76-79]. Genes with an assigned antisense RNA [25] are marked in blue, asterisks indicate a detected protein spot for the respective gene (Additional file 2: Table S3) and statistically not significant values are marked by grey boxes. The genes *tatAD* and *tatCD* are not shown due to lacking transcript abundances (NPKM values <10).

In contrast to genes of the Sec pathway, the transcript abundances of the genes of the twin-arginine translocation (Tat) system TatAyCy double from sampling point I

to II (Figure 9B) and increase further to maximal abundance (NPKM values 1221 and 1155) at the latest sampling point. In contrast to TatAyCy, transcripts of the second Tat system TatAdCd do not show any abundance during the fermentation process. Five proteins were predicted to contain a corresponding Tat signal peptide. However, the pattern of transcript abundances of these proteins does not indicate that they are the main secretion targets for the strongly transcribed TatAyCy system. It has recently been shown that the extracellular *B. subtilis* lipase BSU02700, which is Sec-dependently secreted under standard conditions, is translocated by the *B. subtilis* Tat pathway in a hyper-secreting strain [80]. Hence, this phenomenon has been assumed to be an overflow mechanism [80]. Considering the RNA abundances shown for the secretion machinery in this study (Figure 9), it is tempting to speculate that this overflow mechanism may also play a role in the secretion of subtilisin in *B. licheniformis*. The fact that no typical Tat signal peptide is attached to the subtilisin proenzyme seems to argue against this hypothesis. However, it was shown that the conservation of the RR motif of the signal peptide [81] is not essential for Tat-dependent secretion [82,83]; an RK motif, as present in the Subtilisin Carlsberg prepeptide, can likewise facilitate Tat-dependent secretion (Additional file 1: Figure S15) [82,83]. The blurred boundaries of Tat- and Sec-dependent secretion are additionally pointed out by the facts that proteins with a Tat signal peptide can, *vice versa*, be secreted by the Sec pathway and that the Tat system is accessible for originally Sec-dependent proteins fused to Tat signal peptides [84,85]. This was demonstrated by the detection of TatAyCy-dependent export of active subtilisin [85] and gives reason to assume that Tat-dependent subtilisin secretion is not an obstacle for proper folding of the enzyme.

In *B. subtilis*, a CssRS-dependent response to protein secretion stress triggered by both, homologous and heterologous proteins, has been described [86-88]. The two-component regulatory system CssRS reacts to secretion stress by activating the transcription of *htrA* and *htrB* [87], which encode membrane-anchored serine proteases that trigger refolding or degradation of misfolded or aggregated proteins [89]. The here determined RNA abundances of the genes of the CssR regulon are given in Figure 9C. High transcription rates can be observed for the genes of the serine proteases HtrA and HtrB at the first sampling point, but these rates decline at later stages of the process. This reaction could be due to various reasons: (i) even highly synthesized pre-subtilisin does not aggregate in the cells, (ii) the cells are highly tolerant to large amounts of (aggregated) pre-subtilisin or (iii) the preprotein is efficiently exported in the late production stages, maybe even supported by the Tat pathway.

Conclusion & outlook

The presented data give unprecedented insights into the complex adaptations of the bacterial production strain *B. licheniformis* DSM13 to the changing physiological demands during an industrial-oriented fermentation using the example of a detergent protease production process. We thereby provide reference data for a better understanding and possible optimization of industrial fermentation processes.

These insights enabled us to pinpoint physiological adaptations within the bioprocess, many of which could be confirmed by proteome analysis. Cluster analysis clearly revealed strong growth phase dependencies of many genes as well as some phase-independent genes. RNA as well as protein abundances of the central carbon metabolism and the amino acid metabolism are in accordance with the initial glucose-driven metabolism. Main changes in the corresponding pathways regard the overflow metabolism and the subsequent catabolism of the thereby produced C2 compounds as well as the alternating synthesis and degradation of glutamate and its derivates. Changes in RNA abundances reflect a transition of sustenance from more complex molecules like peptides and oligomers to amino acids. This emphasizes the importance of the secreted protease and its activity on the substrate as a functional component of the productive fermentation.

By comparing our data to previous transcriptome studies focusing on stress conditions, we were able to reliably identify potential stress factors within the process. A detailed inspection of the associated transcripts revealed oxidative stress and increasing phosphate limitation as important factors. Notably, the transcripts of the lichenicidin biosynthesis-related genes *lanA1* and *lanA2* are highly abundant throughout all sampling points. The high abundances of the complete *lan* gene cluster and the majority of sporulation-involved genes indicate a substantial production of antimicrobial compounds and a responsive, KinC-enhanced induction of sporulation. This energy-consuming behavior is clearly not supportive in terms of productivity.

An interesting finding concerning protein secretion pathways is the increase in abundance of the Tat pathway components *tatAY* and *tatCY* from sampling point I to V, which is in contrast to the abundance patterns of genes of the Sec secretory system. Thus, the cells seem to increase the overall secretion capacity during the production process by including non-typical secretory pathways.

The presented findings enabled the identification of important physiological and genetic switches of *B. licheniformis* which limit the overall productivity. The data indicate several opportunities to improve the strains performance in the production of subtilisin. The observed adaptions to the changing substrate supply during the successive metabolization of media components suggest

that an optimization of the non-optimal amino acid composition or phosphate supply may lead to better reproducibility, increased efficiency, cycle time reduction, and finally a diminished employment of resources. Optimization of the deployed strain should also be achieved by the introduction of genetic modifications. For instance, the observed strong expression of the *lan* gene cluster which encodes an undesirable cell wall stress inducing byproduct marks it as promising target for a gene deletion. Another approach might be the modulation of the subtilisin signal peptide to channel subtilisin to the putative Tat-dependent secretion pathway.

Methods

Bacterial strains and fermentation

The samples for the proteome analysis were derived from fermentation experiments carried out for *Bacillus licheniformis* MW3Δspo described earlier. For detailed description of fermentation conditions, sampling points and sequencing of the transcriptome please refer to Wiegand et al. [25].

Preparation of cytosolic protein extracts

50 mL of harvested cells were supplemented with 0.5 mL of protease inhibitor (3758.1, Carl Roth, Germany) directly upon sampling. Centrifugation was carried at 4500× *g* and 4°C for 10 min. The supernatant was removed and the cells were stored at −80°C.

For preparation of the cytosolic protein extracts the insoluble components of the fermentation medium were removed from the bacterial pellet by washing at least three times in ice cold 100 mM Tris/HCl, pH 7.5 buffer at 10000× *g* and 4°C for 10 min. After the last washing step the pellet was resuspended in 600 μL TE buffer (10 mM Tris/HCl, pH 7.5, 1 mM EDTA) containing 1.4 mM phenylmethylsulfonyl fluoride (PMSF). After addition of 250 μL glass beads (0.25-0.5 mm) the cells were disrupted by using RiboLyser cell mill (30 s at 6800 rpm, 5 min on ice, 30 s at 6800 rpm; Hybaid, UK). Glass beads and cell debris were removed by two centrifugation steps at 13000× *g* and 4°C for 30 min. To remove ions, which could disturb the isoelectric focusing, the protein extracts were purified with Amicon Ultra 3 K Centrifugal Filters (Millipore, Germany). The protein concentration of the protein extracts was determined using Roti®NanoQuant (Carl Roth, Germany).

Two dimensional gel electrophoresis, imaging and quantification

Isoelectric focusing (IEF) of the cytosolic protein extracts was performed according to [90]. IPG *Blue*Strips 4–7 (SERVA, Germany) were loaded with 150 μg protein extract, which was adjusted to 340 μL with 2 M thiourea/8 M urea buffer and 34 μL CHAPS solution (20 mM DTT, 1% w/v CHAPS detergent) and rehydrated over night. IEF was carried out on a Multiphor II unit (Amersham Pharmacia Biotech) employing the following step gradient: 150 V for 150 Vh,300 V for 300 Vh, 600 V for 600 Vh, 1500 V for 1500 Vh, and a final phase of 3000 V for 57.5 kVh at 20°C. Before separation in the second dimension the IPG strips were incubated in 3.5 mL equilibration buffer A (2.4 M urea, 12% v/v glycerol, 4% v/v 0,5 M Tris pH 6.8, 55.5 mM SDS, 9 mM DTT) and equilibration buffer B (2.4 M urea, 12% v/v glycerol, 4% v/v 0,5 M Tris pH 6.8, 55 mM SDS, 9 mM DTT, 97 mM iodacetamide, 0.15 mM bromphenol blue), each for 15 min on an orbital shaker. Electrophoresis of the proteins was carried out on 12.16% acrylamide/0.34% bisacrylamide gels at 40 W for 1 h and 16 W for additional 16.5 h at 12°C. Gels were stained with Flamingo™ Fluorescent Gel Stain (Bio-Rad Laboratories, USA) following the manufacturer's instructions. The gels were imaged with a Typhoon Imager 9400 (GE Health Care, UK). Spot detection was performed semi-manually with the Delta2D software version 4.1 (Decodon, Germany). Spot quantification was also done with the Delta2D software as described by Wolf et al. [91]. Quantities for proteins represented by more than one distinct spot are given for each spot separately.

Identification of proteins from 2D gel spots

Selected protein spots were excised from 2-D gels using a spot cutter (Bio-Rad, USA). Digestion with trypsin and spotting on the MALDI-target was achieved using the Ettan Spot Handling Workstation (GE Health Care, UK) employing the manufacturers' protocol. Mass spectrometry was carried out with a Proteome Analyzer 4800 (Applied Biosystems, USA) according to Wolf et al. [91]. The spectra were recorded in a mass range from 900 to 3700 Da with a focus mass of 1600 Da. An internal calibration was performed automatically when the autolytic fragments of trypsin with the mono-isotopic $(M + H)^+$ m/z at 1045.556 or 2211.104 reached a signal to noise ratio (S/N) of at least 20.

Peak lists were created by using the script of the GPS Explorer TM Software Version 3.6 (Applied Biosystems, USA) with the following settings: mass range from 900 to 3700 Da, peak density of 15 peaks per 200 Da, minimal area of 100 and maximal 60 peaks per spot. The peak list was created for an S/N ratio of 15. MALDI-TOF-TOF measurements were carried out for the five strongest peaks of the TOF-spectrum. Using a random search pattern, 25 sub-spectra with 125 shots per sub-spectrum were accumulated for one main spectrum. The mono-isotopic arginine $(M + H)^+$ m/z at 175.119 or lysine $(M + H)^+$ m/z at 147.107 was used for internal calibration (one-point-calibration) when it reached a signal to noise ratio (S/N) of at least 5. Peak lists were created with the following settings: mass range from 60 to

precursor – 20 Da, peak density of 15 peaks per 200 Da, minimal area of 100 and maximal 65 peaks per precursor. The peak list was created for an S/N ratio of 10. For data base search the Mascot search engine version 2.4.0 (Matrix Science Ltd, UK) with a specific *Bacillus licheniformis* (http://www.uniprot.org/uniprot/?query=Bacillus+licheniformis&sort=score) database was used.

Statistical data analysis

NPKM (nucleotide activity per kilobase of exon model per million mapped reads) values [25] were computed for every protein-coding gene in all samples as a measure of RNA abundance. Based on the NPKM values analysis of differential expression was performed with baySeq [30] and one-way ANOVA [92]. Genes were assumed to be differentially expressed with a resulting baySeq likelihood value >0.9 or an ANOVA-based p-value <0.01 (False Discovery Rate (FDR) 2%).

k-means clustering

k-means clusters were generated to identify fermentation stage-dependent trends in gene expression. (i) To ensure that the data of each replicate are sufficiently reliable, t-tests [93] were performed using TM4 MeV v4.8 software [36]. At least three out of the five samples had to have a p-value <0.15 to be taken into further analysis. (ii) For setting up the clusters only transcripts with baySeq likelihood values >0.99 were applied to the next step. (iii) Means of the replicates of each sampling point were calculated and z-score transformation was performed to gain a mean expression value of 0 and a standard deviation of 1 [94]. Clusters A to K and M to S were generated with TM4 MeV v4.8 [36] employing k-means clustering with Euclidian distances after estimating the cluster number by Figure of merit (FOM) analysis [94]. Clusters were cured manually. (iv) Finally, expression profiles of all other genes (p-value <0.15) were added to the determined clusters. Therefore Gene Distance Matrices were computed with TM4 MeV v4.8 [36] in Euclidian distance for all remaining transcripts and the respective cluster means as point of reference. Transcripts were assigned to the previously determined clusters dependent on their scaled distance value. In general the scaled distance value had to be <0.3, exceptions are clusters G and Q with a scaled distance value <0.6. 312 of the remaining 332 expression profiles could be assigned to the newly defined clusters L and T - W.

GO

Gene Ontology terms [37] have been assigned to the genome of *B. licheniformis* using Blast2GO [32]. Enrichment analysis for every cluster was performed with the implemented Gossip [95] package running a two-tailed Fisher's Exact Test (FDR 0.05). Go terms over- or underrepresented were sorted to their respective child or parent using OBO-Edit. To enable a broad overview of enriched groups Generic GO slims [31] also were assigned and analyzed with Blast2GO.

Heat maps

Color codes presented in the heat maps are based on z-score transformed mean NPKM values for the transcriptome and mean spot quantities for the proteome. Figures 5, 6 and 7 were designed utilizing the CellDesigner™ v4.2 software [96] and employing the databases of SubtiWiki [76], BioCyc [97], KEGG [98] and IMG [99].

Tat signal prediction

Proteins with RR and KR motifs of Tat signal peptides where predicted employing the TatP v1.0 software [100].

Additional files

Additional file 1: Figures S1–S15. Figure S1: Protease production and process parameters. **Figure S2**: Proteome of the amino acid metabolism – Part I. **Figure S3**: Proteome of the amino acid metabolism – Part II. **Figure S4**: Amino acid transport. **Figure S5**: Proteome of the central carbon metabolism. **Figure S6**: Carbohydrate transport. **Figure S7**: Acetoin utilization operon *acuABC*. **Figure S8**: Most abundant proteins. **Figure S9**: Lichenicidin gene cluster. **Figure S10**: Cell envelope stress response. **Figure S11**: Sporulation. **Figure S12**: Iron starvation. **Figure S13**: Heat shock response. **Figure S14**: Phosphate stress response. **Figure S15**: Putative Tat signal peptide of Subtilisin Carlsberg.

Additional file 2: Tables S1–S3. Table S1: k-means clustering of expression profiles. **Table S2**: GO term enrichment analysis of k-means clusters. **Table S3**: Proteome data.

Abbreviations

(M + H)+: Protonated molecular ions; °C: Degrees Celsius; μL: Microliter; 2D: Two dimensional; BPG: 1,3-Biphosphoglycerate; C2: Organic molecule harboring two carbon atoms; CHAPS: 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate; CoA: Coenzyme A; Da: Dalton; DHAP: Dihydroxyacetone phosphate; DTT: Dithiothreitol; F6P: Fructose 6-phosphate; FBP: Fructose 1,6-bisphosphate; FDR: False Discovery Rate; g: Gravitational constant; G6P: Glucose 6-phosphate; GADP: Glyceraldehyde 3-phosphate; GDH: Glutamate dehydrogenase; GO: Genome ontology; GOGAT: Glutamate synthase; GS: Glutamine synthetase; H: Hours; H_2O_2: Hydrogen peroxide; HCl: Hydrogen chloride; IEF: Isoelectric focusing; IPG: Immobilized pH gradient; kVh: Kilovolt hour; M: Molar; m/z: Mass-to-charge ratio; MALDI: Matrix-assisted laser desorption/ionization; min: Minute; mL: Milliliter; mM: Millimolar; mm: Millimeter; NPKM: Nucleotide activity per kilobase of exon model per million mapped reads; Oxo: Oxoglutarate; P: Phosphate; PEP: Phosphoenolpyruvate; Pyr: Pyruvate; RK motif: Twin-arginine motif; RNA-Seq: RNA sequencing; Rrpm: Revolutions per minute; RR motif: Arginine-lysine motif; s: Second; S/N: Signal-to-noise ratio; SDS: Sodium dodecyl sulfate; SRP: Signal recognition particle; Tat: twin-arginine translocation; TAXI: TRAP-associated extracytoplasmic immunity protein; TCA: Tricarboxylic acid; TOF: Time-of-flight mass spectrometer; TRAP: Tripartite ATP-independent periplasmic dicarboxylate transporter; V: Volt; v/v: Volume per Volume; Vh: Volt hour; W: Watt; w/v: Weight per volume.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SW performed the experiments, analyzed data and wrote paper, BV supervised proteomics experiments and analyzed data, DA performed mass

spectrometry analyses, JB and SE provided industrial fermentation facilities and organized fermentation and sampling, MH granted access to facilities for proteome analysis, RD wrote paper and provided research facilities, HL wrote paper, designed research and analyzed data. All authors read and approved the final version of the manuscript.

Acknowledgements
This study was funded by the Bundesministerium für Bildung und Forschung (FKZ-0315387).
The authors would like to thank Sascha Dietrich and John Vollmers for fruitful discussions during manuscript preparation. We are grateful to Hendrik Hellmuth for evaluation of the manuscript, to Anja Poehlein for thorough proof-reading and to Ayhan Aydemir and Antje Fengler for expert technical assistance.

Author details
[1]Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany. [2]Division of Microbial Physiology and Molecular Biology, Institut für Mikrobiologie, Norddeutsches Zentrum für Mikrobielle Genomforschung, Ernst-Moritz-Arndt-Universität Greifswald, F.-L.-Jahnstr. 15, D-17487 Greifswald, Germany. [3]Henkel AG & Co. KGaA, Henkelstr. 67, D-40191 Düsseldorf, Germany.

Received: 26 August 2013 Accepted: 1 December 2013
Published: 6 December 2013

References

1. Rooney AP, Price NPJ, Ehrhardt C, Swezey JL, Bannan JD: **Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov.** *Int J Syst Evol Micr* 2009, **59:**2429–2436.
2. Schallmey M, Singh A, Ward O: **Developments in the use of *Bacillus* species for industrial production.** *Can J Microbiol* 2004, **50:**1–17.
3. Çalık P, Kalender N, Özdamar TH: **Overexpression of serine alkaline protease encoding gene in *Bacillus* species: performance analyses.** *Enzyme Microb Tech* 2003, **33:**967–974.
4. Çalık P, Çalık G, Takaç S, Özdamar T: **Metabolic flux analyses for serine alkaline protease production.** *Enzyme Microb Tech* 2000, **27:**793–805.
5. Çalık P, Çalık G, Ozdamar TH: **Oxygen-transfer strategy and its regulation effects in serine alkaline protease production by *Bacillus licheniformis*.** *Biotechnol Bioeng* 2000, **69:**301–311.
6. Çalık P, Çalık G, Takaç S, Ozdamar TH: **Metabolic flux analysis for serine alkaline protease fermentation by *Bacillus licheniformis* in a defined medium: effects of the oxygen transfer rate.** *Biotechnol Bioeng* 1999, **64:**151–167.
7. Çalık P, Bilir E, Çalık G, Özdamar TH: **Influence of pH conditions on metabolic regulations in serine alkaline protease production by *Bacillus licheniformis*.** *Enzyme Microb Technol* 2002, **31:**685–697.
8. Hornbaek T, Jakobsen M, Dynesen J, Nielsen AK: **Global transcription profiles and intracellular pH regulation measured in *Bacillus licheniformis* upon external pH upshifts.** *Arch Microbiol* 2004, **182:**467–474.
9. Hornbaek T, Nielsen AK, Dynesen J, Jakobsen M: **The effect of inoculum age and solid versus liquid propagation on inoculum quality of an industrial *Bacillus licheniformis* strain.** *FEMS Microbiol Lett* 2004, **236:**145–151.
10. Çalık P, Tomlin GC, Oliver SG, Özdamar TH: **Overexpression of a serine alkaline protease gene in *Bacillus licheniformis* and its impact on the metabolic reaction network.** *Enzyme Microb Tech* 2003, **32:**706–720.
11. Çalık P, Çelik E, Telli İE, Oktar C, Özdemir E: **Protein-based complex medium design for recombinant serine alkaline protease production.** *Enzyme Microb Tech* 2003, **33:**975–986.
12. Enshasy E, Azaly E: **Optimization of the industrial production of alkaline protease by *Bacillus licheniformis* in different production scales.** *Aus J Basic Appli Sci* 2008, **2:**583–593.
13. Gupta R, Beg QK, Lorenz P: **Bacterial alkaline proteases: molecular approaches and industrial applications.** *Appl Microbiol Biotechnol* 2002, **59:**15–32.
14. Degering C, Eggert T, Puls M, Bongaerts J, Evers S, Maurer K-H, Jaeger K-E: **Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides.** *Appl Environ Microbiol* 2010, **76:**6370–6376.
15. Waschkau B, Waldeck J, Wieland S, Eichstädt R, Meinhardt F: **Generation of readily transformable *Bacillus licheniformis* mutants.** *Appl Microbiol Biotechnol* 2008, **78:**181–188.
16. Hoffmann K, Wollherr A, Larsen M, Rachinger M, Liesegang H, Ehrenreich A, Meinhardt F: **Facilitation of direct conditional knockout of essential genes in *Bacillus licheniformis* DSM13 by comparative genetic analysis and manipulation of genetic competence.** *Appl Environ Microbiol* 2010, **76:**5046–5057.
17. Waldeck J, Meyer-Rammes H, Wieland S, Feesche J, Maurer K-H, Meinhardt F: **Targeted deletion of genes encoding extracellular enzymes in *Bacillus licheniformis* and the impact on the secretion capability.** *J Biotechnol* 2007, **130:**124–132.
18. Nahrstedt H, Waldeck J, Gröne M, Eichstädt R, Feesche J, Meinhardt F: **Strain development in *Bacillus licheniformis*: Construction of biologically contained mutants deficient in sporulation and DNA repair.** *J Biotechnol* 2005, **119:**245–254.
19. Waldeck J, Meyer-Rammes H, Nahrstedt H, Eichstädt R, Wieland S, Meinhardt F: **Targeted deletion of the *uvrBA* operon and biological containment in the industrially important *Bacillus licheniformis*.** *Appl Microbiol Biotechnol* 2007, **73:**1340–1347.
20. Borgmeier C, Bongaerts J, Meinhardt F: **Genetic analysis of the *Bacillus licheniformis degSU* operon and the impact of regulatory mutations on protease production.** *J Biotechnol* 2012, **159:**12–20.
21. Voigt B, Hoi LT, Jürgen B, Albrecht D, Ehrenreich A, Veith B, Evers S, Maurer K-H, Hecker M, Schweder T: **The glucose and nitrogen starvation response of *Bacillus licheniformis*.** *Proteomics* 2007, **7:**413–423.
22. Hoi LT, Voigt B, Jürgen B, Ehrenreich A, Gottschalk G, Evers S, Feesche J, Maurer K-H, Hecker M, Schweder T: **The phosphate-starvation response of *Bacillus licheniformis*.** *Proteomics* 2006, **6:**3582–3601.
23. Schroeter R, Voigt B, Jürgen B, Methling K, Pöther D-C, Schäfer H, Albrecht D, Mostertz J, Mäder U, Evers S, Maurer K-H, Lalk M, Mascher T, Hecker M, Schweder T: **The peroxide stress response of *Bacillus licheniformis*.** *Proteomics* 2011, **11:**2851–2866.
24. Wecke T, Veith B, Ehrenreich A, Mascher T: **Cell envelope stress response in *Bacillus licheniformis*: integrating comparative genomics, transcriptional profiling, and regulon mining to decipher a complex regulatory network.** *J Bacteriol* 2006, **188:**7500–7511.
25. Wiegand S, Dietrich S, Hertel R, Bongaerts J, Evers S, Volland S, Daniel R, Liesegang H: **RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation.** *BMC Genomics* 2013, **14:**667.
26. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10:**57–63.
27. Sorek R, Cossart P: **Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.** *Nat Rev Genet* 2010, **11:**9–16.
28. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5:**621–628.
29. Veith B, Herzberg C, Steckel S, Feesche J, Maurer K-H, Ehrenreich P, Bäumer S, Henne A, Liesegang H, Merkl R, Ehrenreich A, Gottschalk G: **The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential.** *J Mol Microbiol Biotechnol* 2004, **7:**204–211.
30. Hardcastle TJ, Kelly KA: **baySeq: empirical Bayesian methods for identifying differential expression in sequence count data.** *BMC Bioinformatics* 2010, **11:**422.
31. The Gene Ontology Consortium: **The Gene Ontology (GO) database and informatics resource.** *Nucleic Acids Res* 2004, **32:**D258–D261.
32. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21:**3674–3676.
33. Dischinger J, Josten M, Szekat C, Sahl H-G, Bierbaum G: **Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13.** *PLoS One* 2009, **4:**e6788.
34. Jahn N, Preis H, Wiedemann C, Brantl S: **BsrG/SR4 from *Bacillus subtilis*–the first temperature-dependent type I toxin-antitoxin system.** *Mol Microbiol* 2012, **83:**579–598.
35. Fukushima T: **Transcriptional, functional and cytochemical analyses of the *veg* gene in *Bacillus subtilis*.** *J Biochem* 2003, **133:**475–483.

36. Saeed A, Bhagabati N, Braisted J, Liang W, Sharov V, Howe E, Li J, Thiagarajan M, White J, Quackenbush J: **TM4 microarray software suite.** *Methods Enzymol* 2006, **411:**134.
37. The Gene Ontology Consortium: **Gene Ontology: tool for the unification of biology.** *Nat Genet* 2000, **25:**25–29.
38. Slack FJ, Mueller JP, Strauch MA, Mathiopoulos C, Sonenshein AL: **Transcriptional regulation of a *Bacillus subtilis* dipeptide transport operon.** *Mol Microbiol* 1991, **5:**1915–1925.
39. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Hartig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Le Chat L, Lecointe F, Lewis P, Liebermeister W, *et al*: **Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*.** *Science* 2012, **335:**1103–1106.
40. Koide A, Perego M, Hoch JA: **ScoC regulates peptide transport and sporulation initiation in *Bacillus subtilis*.** *J Bacteriol* 1999, **181:**1–5.
41. Sonenshein A: **Control of key metabolic intersections in *Bacillus subtilis*.** *Nat Rev Microbiol* 2007, **5:**917–927.
42. Gunka K, Commichau FM: **Control of glutamate homeostasis in *Bacillus subtilis*: a complex interplay between ammonium assimilation, glutamate biosynthesis and degradation.** *Mol Microbiol* 2012, **85:**213–224.
43. Bernlohr RW, Schreier HJ, Donohue TJ: **Enzymes of Glutamate and Glutamine Biosynthesis in *Bacillus licheniformis*.** *Curr Top Cell Regul* 1984, **24:**145–152.
44. Feucht A, Evans L, Errington J: **Identification of sporulation genes by genome-wide analysis of the σE regulon of *Bacillus subtilis*.** *J Mol Biol* 2003, **149:**3023–3034.
45. Debarbouille M, Gardan R, Arnaud M, Rapoport G: **Role of BkdR, a transcriptional activator of the SigL-dependent isoleucine and valine degradation pathway in *Bacillus subtilis*.** *J Bacteriol* 1999, **181:**2059.
46. Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S: **Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms.** *Microb Cell Fact* 2012, **11:**122.
47. Grundy FJ, Waters DA, Allen SHG, Henkin TM: **Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA.** *J Bacteriol* 1993, **175:**7348–7355.
48. Presecan-Siedel E, Galinier A, Longin R, Deutscher J, Danchin A, Glaser P, Martin-Verstraete I: **Catabolite regulation of the *pta* gene as part of carbon flow pathways in *Bacillus subtilis*.** *J Bacteriol* 1999, **181:**6889–6897.
49. Deutscher J, Galinier A, Martin-Verstraete I: **Carbohydrate Uptake and Metabolism.** In *Bacillus subtilis and its closest relatives—from genes to cells*. Edited by Sonenshein A, Hoch J, Losick R. Washington, D.C: ASM Press; 2001:129–150.
50. Renna MC, Najimudin N, Winik LR, Zahler SA: **Regulation of the *Bacillus subtilis alsS*, *alsD*, and *alsR* Genes involved In Post-Exponential-Phase Production of Acetoin.** *J Bacteriol* 1993, **175:**3863–3875.
51. Frädrich C, March A, Fiege K, Hartmann A, Jahn D, Härtig E: **The transcription factor AlsR binds and regulates the promoter of the *alsSD* operon responsible for acetoin formation in *Bacillus subtilis*.** *J Bacteriol* 2012, **194:**1100.
52. Grundy FJ, Turinsky AJ, Henkin TM: **Catabolite regulation of *Bacillus subtilis* acetate and acetoin utilization genes by CcpA.** *J Bacteriol* 1994, **176:**4527–4533.
53. Gardner JG, Escalante-Semerena JC: **In *Bacillus subtilis*, the sirtuin protein deacetylase, encoded by the *srtN* gene (formerly *yhdZ*), and functions encoded by the *acuABC* genes control the activity of acetyl coenzyme A synthetase.** *J Bacteriol* 2009, **191:**1749–1755.
54. Gardner JG, Escalante-Semerena JC: **Biochemical and mutational analyses of AcuA, the acetyltransferase enzyme that controls the activity of the acetyl coenzyme a synthetase (AcsA) in *Bacillus subtilis*.** *J Bacteriol* 2008, **190:**5132–5136.
55. Thanh TN, Jürgen B, Bauch M, Liebeke M, Lalk M, Ehrenreich A, Evers S, Maurer K-H, Antelmann H, Ernst F, Homuth G, Hecker M, Schweder T: **Regulation of acetoin and 2,3-butanediol utilization in *Bacillus licheniformis*.** *Appl Microbiol Biotechnol* 2010, **87:**2227–2235.
56. Ali NO, Bignon J, Rapoport G, Debarbouille M: **Regulation of the acetoin catabolic pathway is controlled by Sigma L in *Bacillus subtilis*.** *J Bacteriol* 2001, **183:**2497–2504.
57. Wolfe AJ: **The acetate switch.** *Microbiol Mol Biol R* 2005, **69:**12–50.
58. Blencke H-M, Homuth G, Ludwig H, Mäder U, Hecker M, Stülke J: **Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways.** *Metab Eng* 2003, **5:**133–149.
59. Mostertz J, Scharf C, Hecker M, Homuth G: **Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress.** *Microbiology+* 2004, **150:**497–512.
60. Helmann JD, Wu MFW, Kobel PA, Gamo F-J, Wilson M, Morshedi MM, Navre M, Paddon C: **Global transcriptional response of *Bacillus subtilis* to heat shock.** *J Bacteriol* 2001, **183:**7318.
61. Helmann JD, Wu MFW, Kobel PA, Gamo F-J, Wilson M, Morshedi MM, Navre M, Paddon C: **The global transcriptional response of *Bacillus subtilis* to peroxide stress is coordinated by three transcription factors.** *J Bacteriol* 2003, **185:**243.
62. Nielsen AK, Breüner A, Krzystanek M, Andersen JT, Poulsen TA, Olsen PB, Mijakovic I, Rasmussen MD: **Global transcriptional analysis of *Bacillus licheniformis* reveals an overlap between heat shock and iron limitation stimulon.** *J Mol Microbiol Biotechnol* 2010, **18:**162–173.
63. Begley M, Cotter PD, Hill C, Ross RP: **Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins.** *Appl Environ Microbiol* 2009, **75:**5451–5460.
64. Higgins D, Dworkin J: **Recent progress in *Bacillus subtilis* sporulation.** *FEMS Microbiol Rev* 2012, **36:**131–148.
65. Kodama T, Endo K, Ara K, Ozaki K, Kakeshita H, Yamane K, Sekiguchi J: **Effect of *Bacillus subtilis spo0A* Mutation on Cell Wall Lytic Enzymes and Extracellular Proteases , and Prevention of Cell Lysis.** *J Biosci Bioeng* 2007, **103:**13–21.
66. LeDeaux JR, Yu N, Grossman AD: **Different roles for KinA, KinB, and KinC in the initiation of sporulation in *Bacillus subtilis*.** *J Bacteriol* 1995, **177:**861–863.
67. López D, Fischbach MA, Chu F, Losick R, Kolter R: **Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*..** *Proc Natl Acad Sci USA* 2009, **106:**280–285.
68. Wiedemann I, Böttiger T, Bonelli RR, Wiese A, Hagge SO, Gutsmann T, Seydel U, Deegan L, Hill C, Ross P, Sahl H-G: **The mode of action of the lantibiotic lacticin 3147–a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II.** *Mol Microbiol* 2006, **61:**285–296.
69. Oman TJ, van der Donk WA: **Insights into the mode of action of the two-peptide lantibiotic haloduracin.** *ACS Chem Biol* 2009, **4:**865–874.
70. Wells JA, Ferrari E, Henner DJ, Estell DA, Chen EY: **Cloning, sequencing, and secretion of *Bacillus* amyloliquefaciens subtilisin in *Bacillus subtilis*.** *Nucleic Acids Res* 1983, **11:**7911–7925.
71. Forster BM, Marquis H: **Protein transport across the cell wall of monoderm Gram-positive bacteria.** *Mol Microbiol* 2012, **84:**405–413.
72. Harwood CR, Cranenburgh R: ***Bacillus* protein secretion: an unfolding story.** *Trends Microbiol* 2008, **16:**73–79.
73. Van Dijl JM, Bolhuis A, Tjalsma H, Jongbloed JDH, de Jong A, Bron S: **Protein Transport Pathways in *Bacillus subtilis*: a Genome-Based Road Map**. In *Bacillus subtilis and its closest relatives—from genes to cells*. Edited by Sonenshein A, Hoch J, Losick R. Washington, D.C: ASM Press; 2001:337–355.
74. Power SD, Adams RM, Wells JA: **Secretion and autoproteolytic maturation of subtilisin.** *Proc Natl Acad Sci USA* 1986, **83:**3096–3100.
75. Jacobs M, Andersen JB, Kontinen V, Sarvas M: ***Bacillus subtilis* PrsA is required in vivo as an extracytoplasmic chaperone for secretion of active enzymes synthesized either with or without pro-sequences.** *Mol Microbiol* 1993, **8:**957–966.
76. Mäder U, Schmeisky AG, Flórez LA, Stülke J: **SubtiWiki–a comprehensive community resource for the model organism *Bacillus subtilis*.** *Nucleic Acids Res* 2012, **40:**D1278–D1287.
77. Driessen AJM, Nouwen N: **Protein translocation across the bacterial cytoplasmic membrane.** *Annu Rev Biochem* 2008, **77:**643–667.
78. Fu LL, Xu ZR, Li WF, Shuai JB, Lu P, Hu CX: **Protein secretion pathways in *Bacillus subtilis:* implication for optimization of heterologous protein secretion.** *Biotechnol Adv* 2007, **25:**1–12.
79. Traag BA, Pugliese A, Eisen JA, Losick R: **Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*.** *J Bacteriol* 2013, **195:**253–260.
80. Kouwen TRHM, van der Ploeg R, Antelmann H, Hecker M, Homuth G, Mäder U, van Dijl JM: **Overflow of a hyper-produced secretory protein from the *Bacillus* Sec pathway into the Tat pathway for protein secretion as revealed by proteogenomics.** *Proteomics* 2009, **9:**1018–1032.
81. Palmer T, Berks BC: **The twin-arginine translocation (Tat) protein export pathway.** *Nat Rev Microbiol* 2012, **10:**483–496.
82. DeLisa MP, Samuelson P, Palmer T, Georgiou G: **Genetic analysis of the twin arginine translocator secretion pathway in bacteria.** *J Biol Chem* 2002, **277:**29825–29831.

83. Ize B, Gerard F, Zhang M, Chanal A, Voulhoux R, Palmer T, Filloux A, Wu L-F: **In vivo dissection of the tat translocation pathway in *Escherichia coli*.** *J Mol Biol* 2002, **317**:327–335.
84. Van der Ploeg R, Monteferrante CG, Piersma S, Barnett JP, Kouwen TRHM, Robinson C, van Dijl JM: **High-salinity growth conditions promote Tat-independent secretion of Tat substrates in *Bacillus subtilis*.** *Appl Environ Microbiol* 2012, **78**:7733–7744.
85. Kolkman MAB, van der Ploeg R, Bertels M, van Dijk M, van der Laan J, van Dijl JM, Ferrari E: **The twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: secretion of active subtilisin via the *B. subtilis* Tat pathway.** *Appl Environ Microbiol* 2008, **74**:7507–7513.
86. Hyyryläinen H-L, Bolhuis A, Darmon E, Muukkonen L, Koski P, Vitikainen M, Sarvas M, Bron S, van Dijl JM, Kontinen VP: **A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress.** *Mol Microbiol* 2001, **41**:1159–1172.
87. Darmon E, Noone D, Masson A, Bron S, Kuipers OP, Devine KM, van Dijl JM: **A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of *Bacillus subtilis*.** *J Bacteriol* 2002, **184**:5661.
88. Westers H, Westers L, Darmon E, van Dijl JM, Quax WJ, Zanen G: **The CssRS two-component regulatory system controls a general secretion stress response in *Bacillus subtilis*.** *FEBS J* 2006, **273**:3816–3827.
89. Noone D, Botella E, Butler C, Hansen A, Jende I, Devine KM: **Signal perception by the secretion stress-responsive CssRS two-component system in *Bacillus subtilis*.** *J Bacteriol* 2012, **194**:1800–1814.
90. Büttner K, Bernhardt J, Scharf C, Schmid R, Mäder U, Eymann C, Antelmann H, Völker A, Völker U, Hecker M: **A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*.** *Electrophoresis* 2001, **22**:2908–2935.
91. Wolf C, Hochgräfe F, Kusch H, Albrecht D, Hecker M, Engelmann S: **Proteomic analysis of antioxidant strategies of *Staphylococcus aureus*: diverse responses to different oxidants.** *Proteomics* 2008, **8**:3139–3153.
92. Raman B, McKeown CK, Rodriguez M, Brown SD, Mielenz JR: **Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation.** *BMC Microbiol* 2011, **11**:134.
93. Koburger T, Weibezahn J, Bernhardt J, Homuth G, Hecker M: **Genome-wide mRNA profiling in glucose starved *Bacillus subtilis* cells.** *Mol Genet Genomics* 2005, **274**:1–12.
94. Hahne H, Mäder U, Otto A, Bonn F, Steil L, Bremer E, Hecker M, Becher D: **A comprehensive proteomics and transcriptomics analysis of *Bacillus subtilis* salt stress adaptation.** *J Bacteriol* 2010, **192**:870–882.
95. Blüthgen N, Kiełbasa SM, Herzel H: **Inferring combinatorial regulation of transcription in silico.** *Nucleic Acids Res* 2005, **33**:272–279.
96. Funahashia A, Morohashia M, Kitanoa H, Tanimura N: **CellDesigner: a process diagram editor for gene-regulatory and biochemical networks.** *Biosilico* 2003, **1**:159–162.
97. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahrén D, Tsoka S, Darzentas N, Kunin V, López-Bigas N: **Expansion of the BioCyc collection of pathway/genome databases to 160 genomes.** *Nucleic Acids Res* 2005, **33**:6083–6089.
98. Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S, Kanehisa M: **KEGG Atlas mapping for global analysis of metabolic pathways.** *Nucleic Acids Res* 2008, **36**:W423–W426.
99. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC: **IMG: the integrated microbial genomes database and comparative analysis system.** *Nucleic Acids Res* 2012, **40**:D115–D122.
100. Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S: **Prediction of twin-arginine signal peptides.** *BMC Bioinformatics* 2005, **6**:167.

doi:10.1186/1475-2859-12-120
Cite this article as: Wiegand *et al.*: Fermentation stage-dependent adaptations of *Bacillus licheniformis* during enzyme production. *Microbial Cell Factories* 2013 **12**:120.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Additional information

Additional file 1

Figure S1 Protease production and process parameters 93
Figure S2 Proteome of the amino acid metabolism – Part I 94
Figure S3 Proteome of the amino acid metabolism – Part II 95
Figure S4 Amino acid transport 96
Figure S5 Proteome of the central carbon metabolism 98
Figure S6 Carbohydrate transport 99
Figure S7 Acetoin utilization operon *acuABC* 101
Figure S8 Most abundant proteins 101
Figure S9 Lichenicidin gene cluster 102
Figure S10 Cell envelope stress response 103
Figure S11 Sporulation 105
Figure S12 Iron starvation 113
Figure S13 Heat shock response 114
Figure S14 Phosphate starvation response 115
Figure S15 Putative Tat signal peptide of Subtilisin Carlsberg 116

Additional file 2

Table S1 *k-means* clustering of expression profiles 116
Table S2 GO term enrichment analysis of *k-means* clusters 116
Table S3 Proteome data 116

References 117

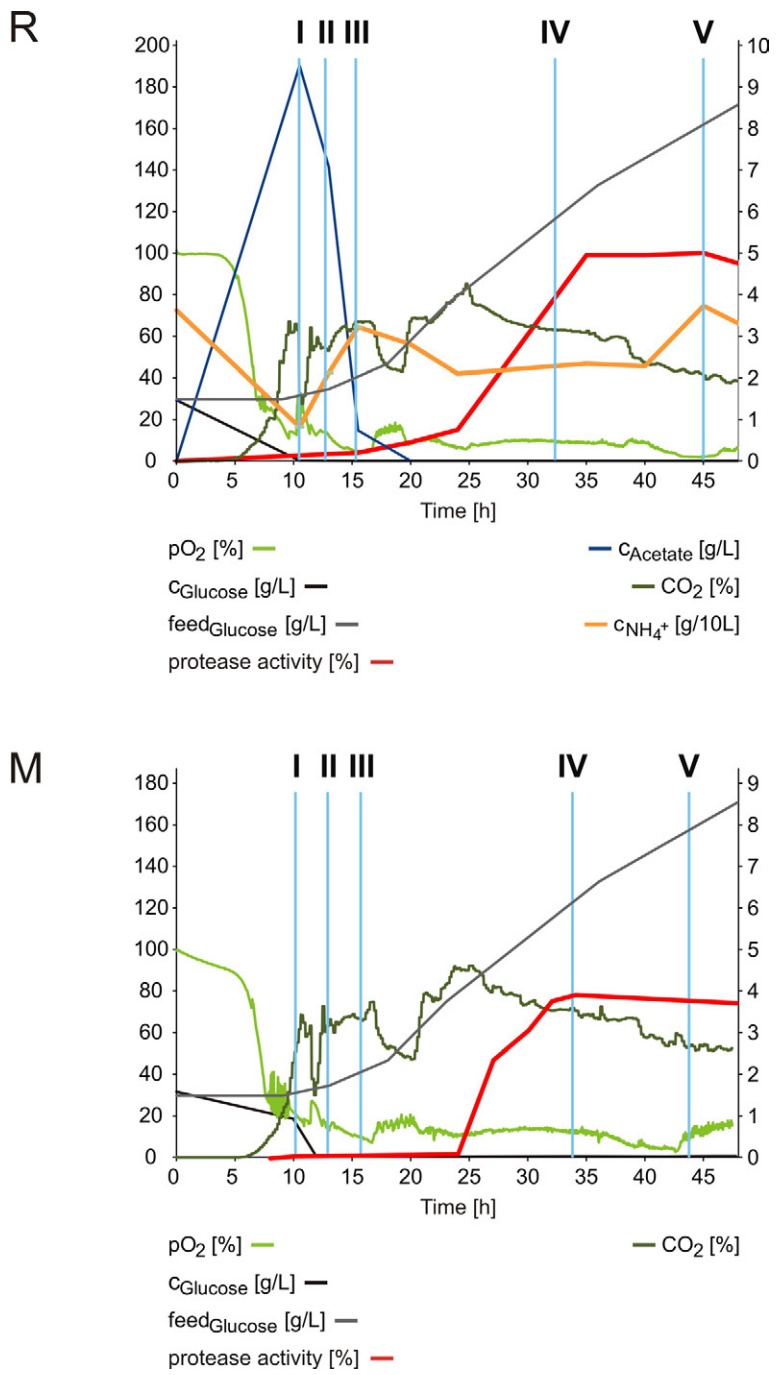


Figure S1 Protease production and process parameters
Process parameters are shown for fermentations R and M (please refer to Figure 1 for replicate L). Oxygen partial pressure pO_2 [%], glucose concentration $c_{Glucose}$ [g/L], supplied glucose $feed_{Glucose}$ [g/L] and normalized protease activity [%] are displayed on the left y-axis, whereas acetate concentration $c_{Acetate}$ [g/L] (only for fermentation R), carbon dioxide content CO_2 [%], and ammonium concentration c_{NH4+} [g/10L] (only for fermentation R) are scaled on the right y-axis. Process time t [h] is given on the x-axis. The sampling points I to V are indicated by light blue lines.

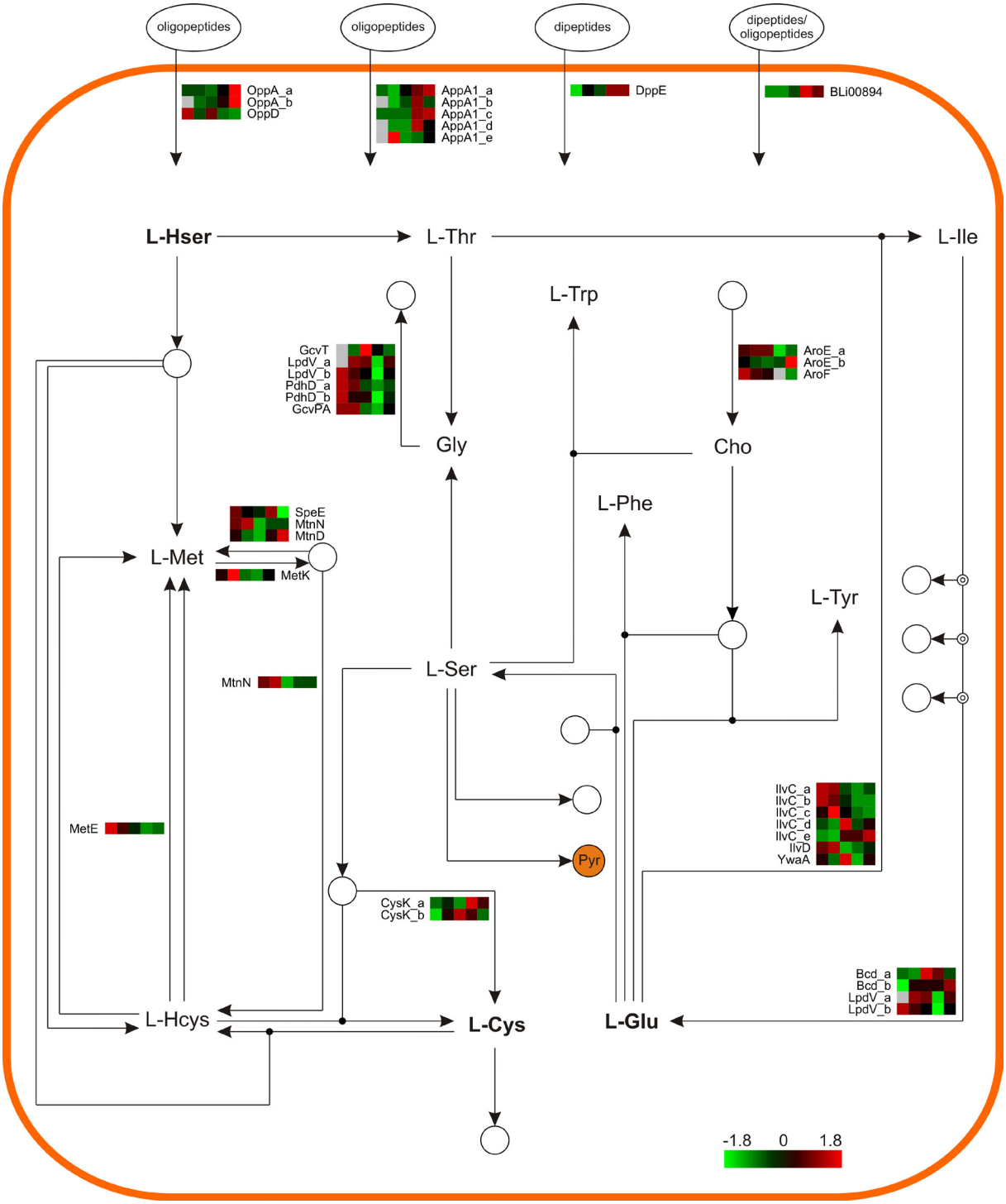


Figure S2 Proteome of the amino acid metabolism – Part I
Heat map representation of Z-score transformed protein spot volumes of proteins involved in amino acid transport and metabolism. In cases where a specific protein is assigned to more than one spot, the particular spots are indicated by an underscore, followed by an ordering letter. Statistically not significant values are indicated by light grey boxes. Pyr: Pyruvate. For the corresponding transcriptome data please refer to Figure 5.

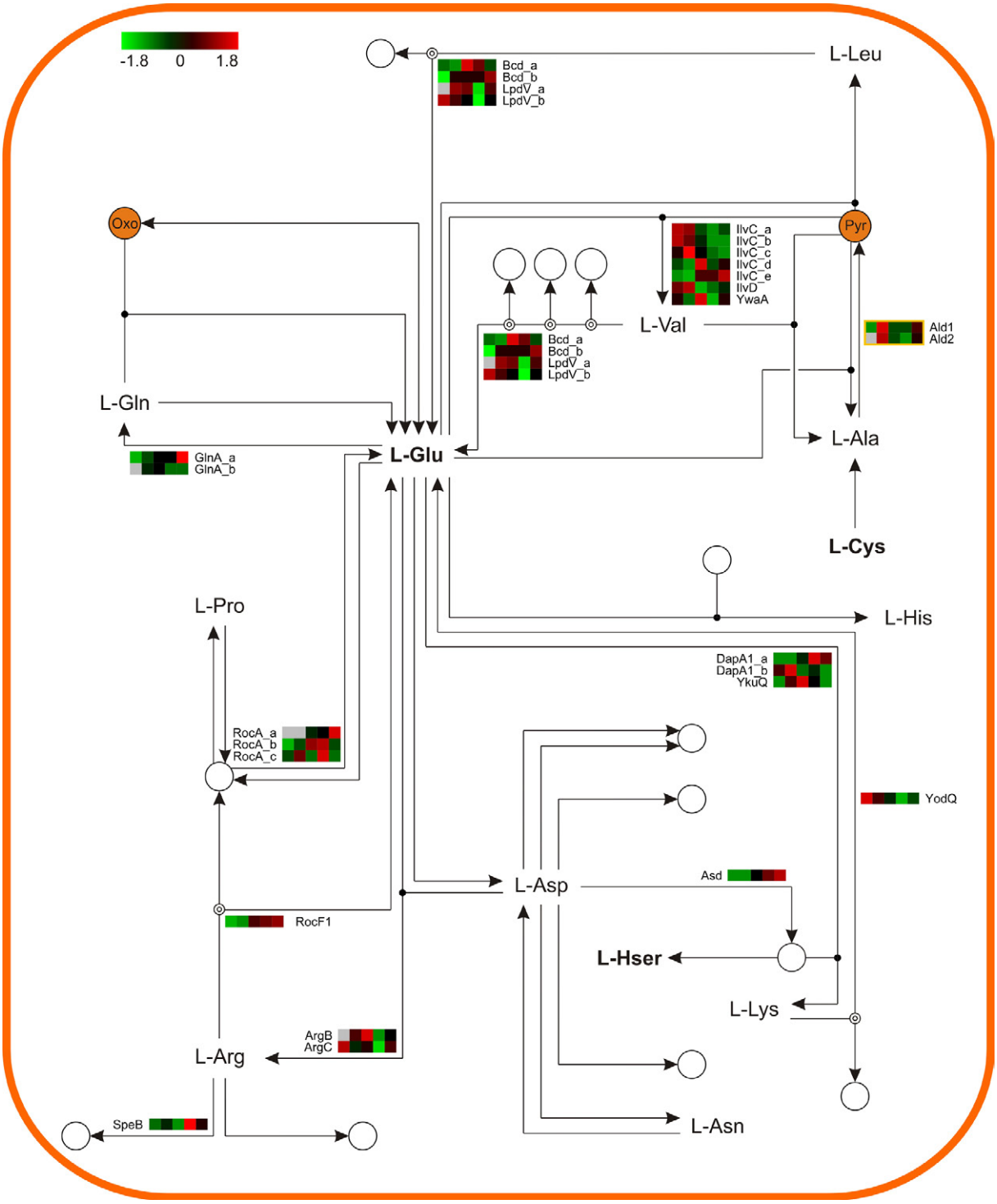
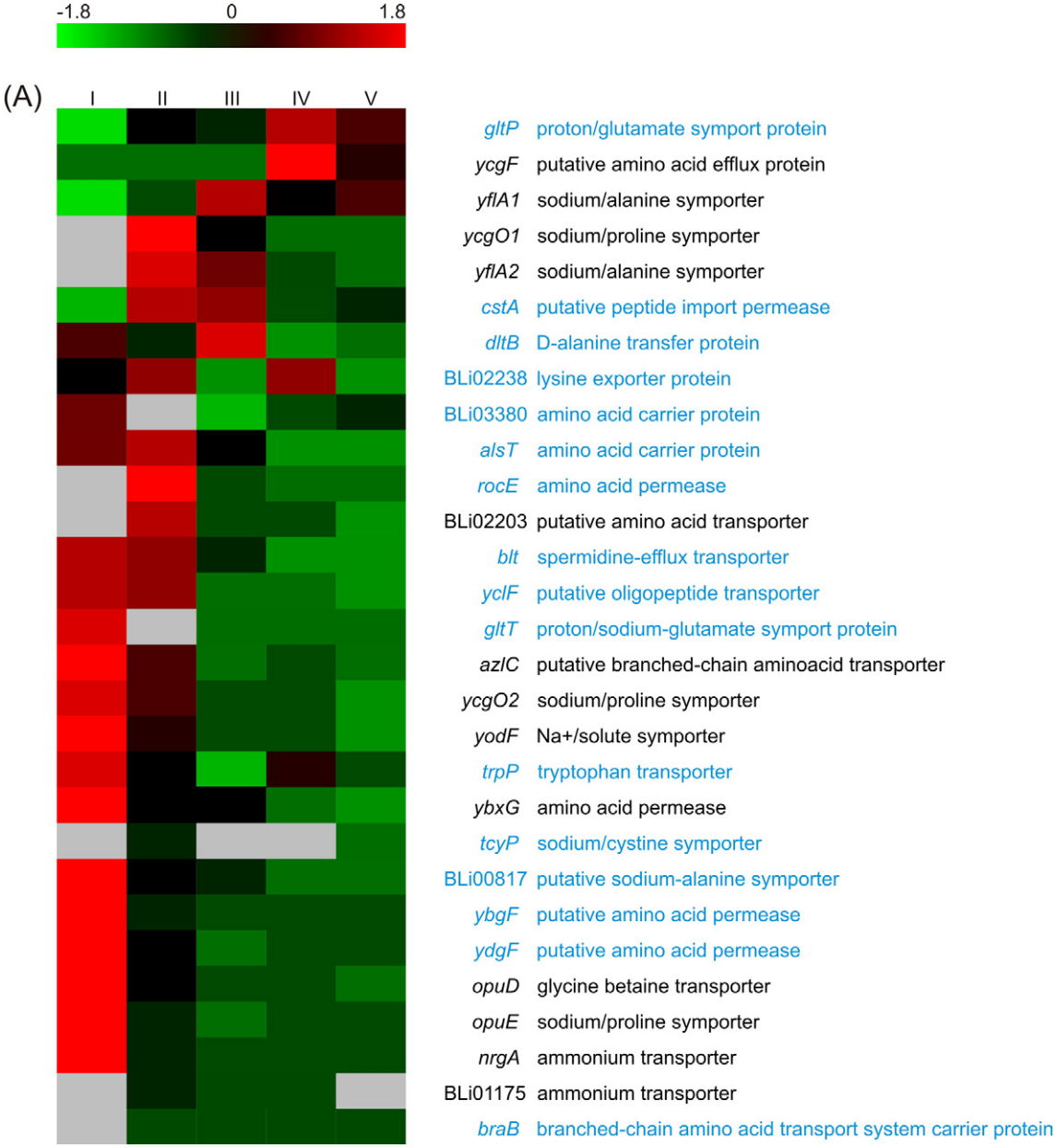


Figure S3 Proteome of the amino acid metabolism – Part II
Heat map representation of Z-score transformed protein spot volumes of proteins involved in amino acid metabolism. In cases where a specific protein is assigned to more than one spot, the particular spots are indicated by an underscore, followed by an ordering letter. Statistically not significant values are indicated by light grey boxes. Yellow frames indicate reactions with multiple assigned enzymes of which only one is strictly necessary. Pyr: Pyruvate, Oxo: 2-Oxoglutarate. For the corresponding transcriptome data please refer to Figure 6.


-1.8
0
1.8
(A)
I
II
III
IV
V
gltP proton/glutamate symport protein
ycgF putative amino acid efflux protein
yflA1 sodium/alanine symporter
ycgO1 sodium/proline symporter
yflA2 sodium/alanine symporter
cstA putative peptide import permease
dltB D-alanine transfer protein
BLi02238 lysine exporter protein
BLi03380 amino acid carrier protein
alsT amino acid carrier protein
rocE amino acid permease
BLi02203 putative amino acid transporter
blt spermidine-efflux transporter
yclF putative oligopeptide transporter
gltT proton/sodium-glutamate symport protein
azlC putative branched-chain aminoacid transporter
ycgO2 sodium/proline symporter
yodF Na+/solute symporter
trpP tryptophan transporter
ybxG amino acid permease
tcyP sodium/cystine symporter
BLi00817 putative sodium-alanine symporter
ybgF putative amino acid permease
ydgF putative amino acid permease
opuD glycine betaine transporter
opuE sodium/proline symporter
nrgA ammonium transporter
BLi01175 ammonium transporter
braB branched-chain amino acid transport system carrier protein

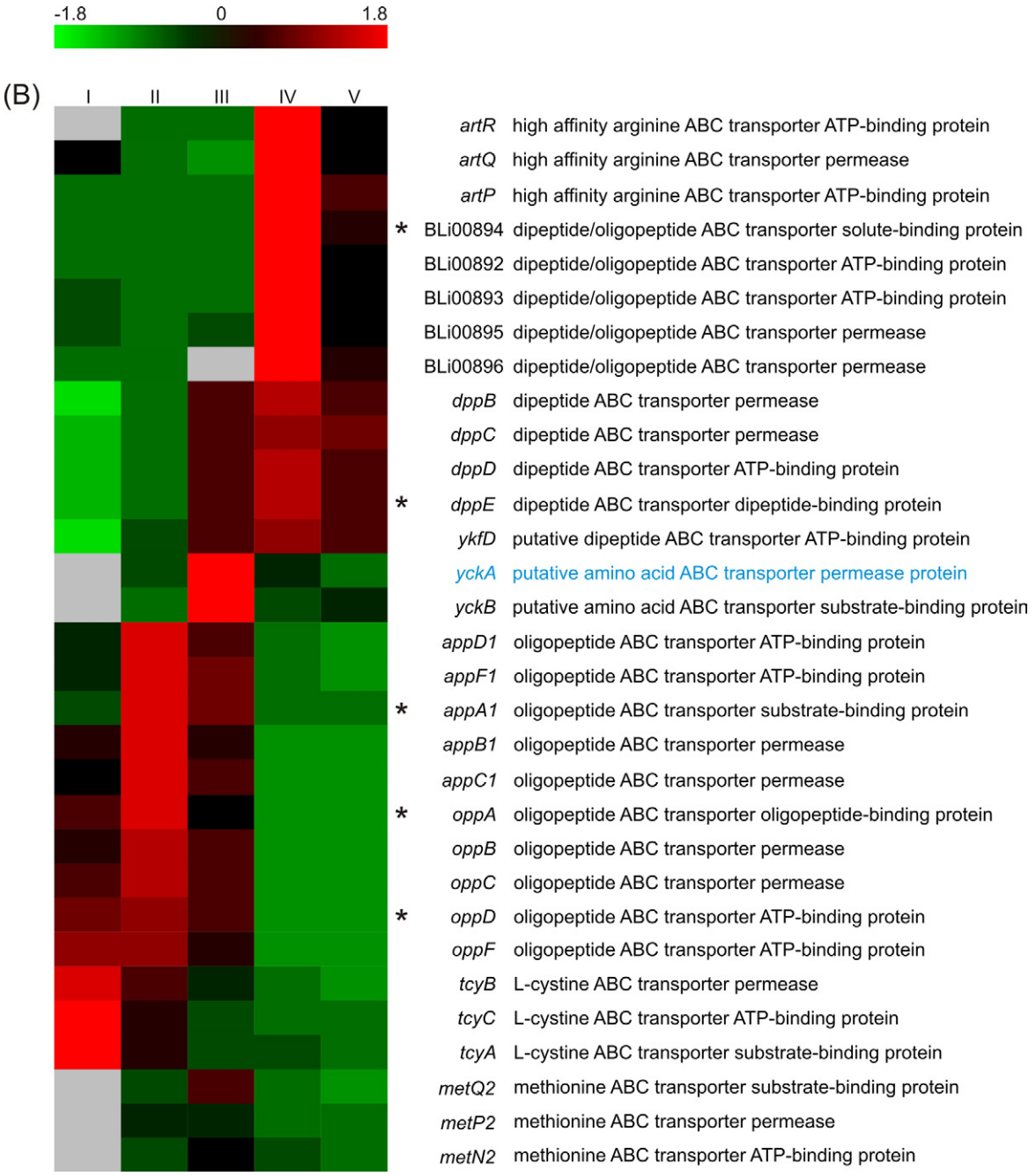


Figure S4 Amino acid transport

Heat map representation of Z-score transformed NPKM values. The depicted genes have been annotated as **(A)** amino acid or nitrogen transporters and **(B)** amino acid ABC transporter components. Please note that the figure does not give a complete list of genes involved in amino acid transport. Genes with an assigned antisense RNA (Chapter B) are marked in blue, asterisks indicate a detected protein spot for the respective gene (Table S3) and statistically not significant values are marked by grey boxes.

Transporter genes with high transcript abundances during the early stages of the fermentation process are for example encoding a tryptophan transporter (*trp*), cystine (*tcyABC*) and methionine (*metNPQ*) ABC transporters and diverse proteins for uptake of alanine or unspecific amino acids. Additionally, the transcript of the ammonium transporter NrgA, especially required for ammonium transport at low ammonium concentrations in *B. subtilis* (Detsch and Stülke, 2003), is highly abundant at sampling point I. The only operon besides the dipeptide ABC transporters mentioned in the main text with distinct transcript abundance at the later sampling points encodes a high-affinity arginine ABC transporter (*artPQR*).

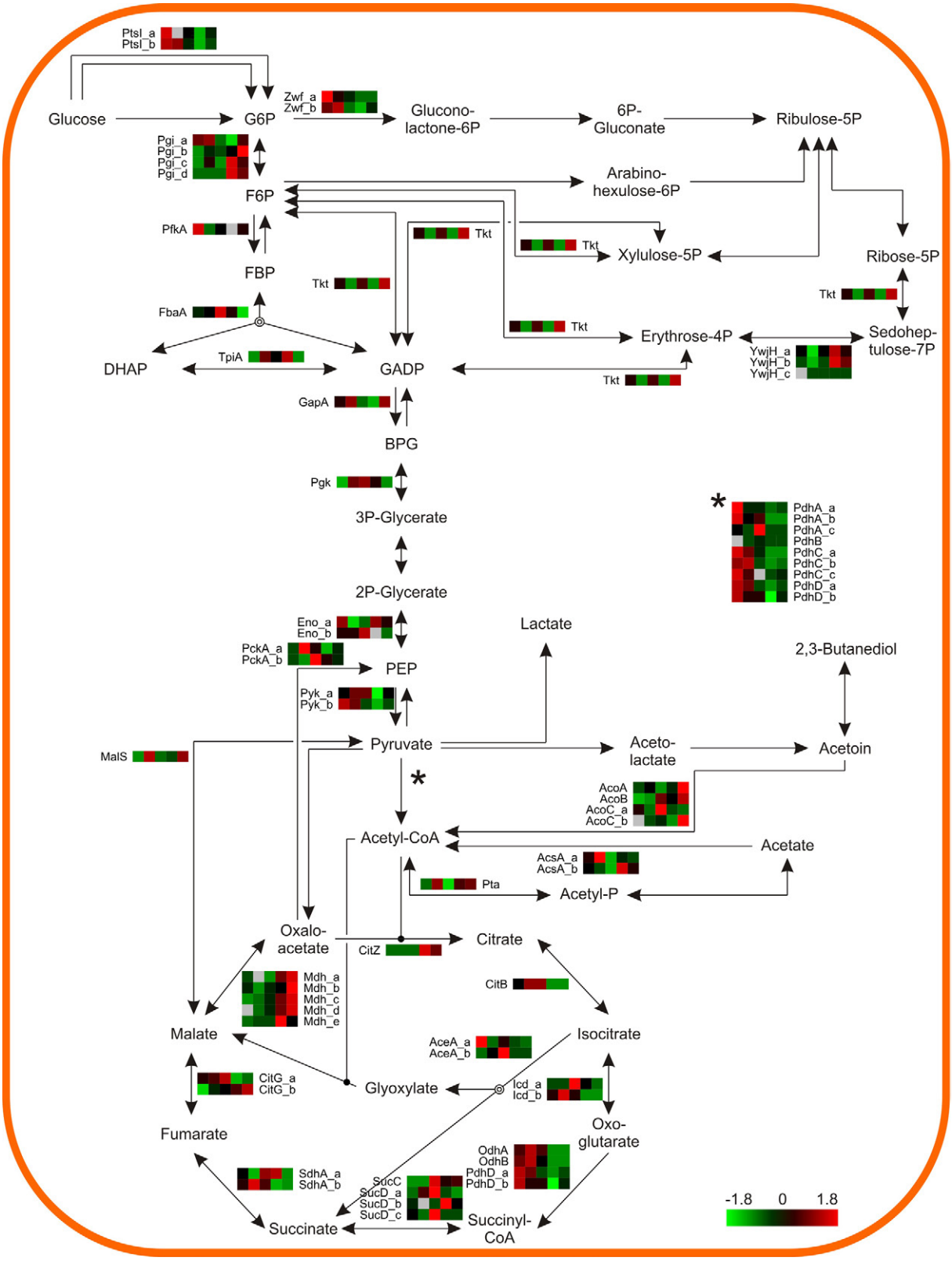
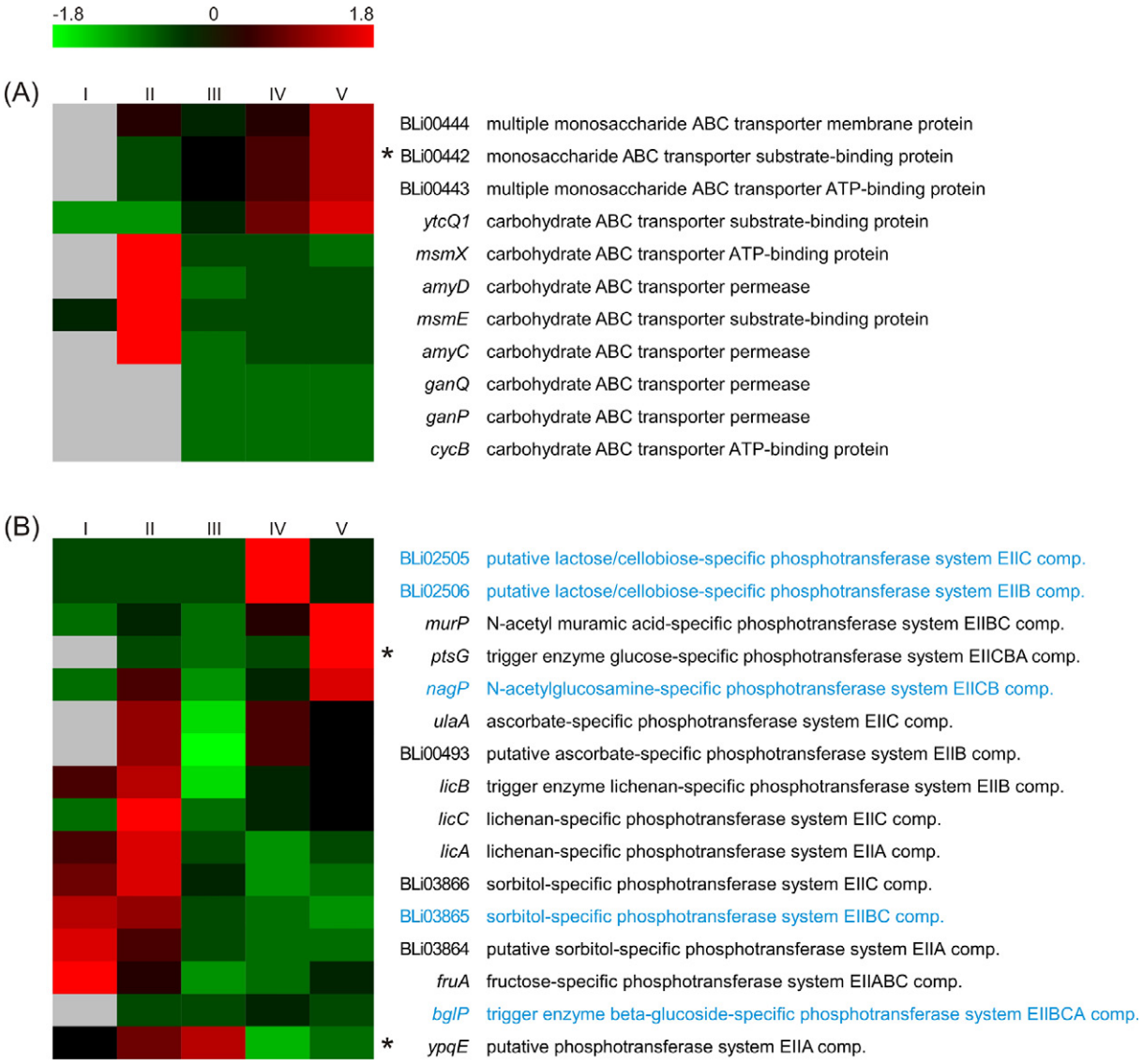


Figure S5 Proteome of the central carbon metabolism
Heat map representation of Z-score transformed protein spot volumes of proteins involved in central carbon metabolism. In cases where a specific protein is assigned to more than one spot, the particular spots are indicated by an ordering letter. Statistically not significant values are indicated by light grey boxes. For the corresponding transcriptome data please refer to Figure 7.


-1.8
0
1.8
(A)
I
II
III
IV
V
BLi00444 multiple monosaccharide ABC transporter membrane protein
* BLi00442 monosaccharide ABC transporter substrate-binding protein
BLi00443 multiple monosaccharide ABC transporter ATP-binding protein
ytcQ1 carbohydrate ABC transporter substrate-binding protein
msmX carbohydrate ABC transporter ATP-binding protein
amyD carbohydrate ABC transporter permease
msmE carbohydrate ABC transporter substrate-binding protein
amyC carbohydrate ABC transporter permease
ganQ carbohydrate ABC transporter permease
ganP carbohydrate ABC transporter permease
cycB carbohydrate ABC transporter ATP-binding protein
(B)
I
II
III
IV
V
BLi02505 putative lactose/cellobiose-specific phosphotransferase system EIIC comp.
BLi02506 putative lactose/cellobiose-specific phosphotransferase system EIIB comp.
murP N-acetyl muramic acid-specific phosphotransferase system EIIBC comp.
* ptsG trigger enzyme glucose-specific phosphotransferase system EIICBA comp.
nagP N-acetylglucosamine-specific phosphotransferase system EIICB comp.
ulaA ascorbate-specific phosphotransferase system EIIC comp.
BLi00493 putative ascorbate-specific phosphotransferase system EIIB comp.
licB trigger enzyme lichenan-specific phosphotransferase system EIIB comp.
licC lichenan-specific phosphotransferase system EIIC comp.
licA lichenan-specific phosphotransferase system EIIA comp.
BLi03866 sorbitol-specific phosphotransferase system EIIC comp.
BLi03865 sorbitol-specific phosphotransferase system EIIBC comp.
BLi03864 putative sorbitol-specific phosphotransferase system EIIA comp.
fruA fructose-specific phosphotransferase system EIIABC comp.
bglP trigger enzyme beta-glucoside-specific phosphotransferase system EIIBCA comp.
* ypqE putative phosphotransferase system EIIA comp.

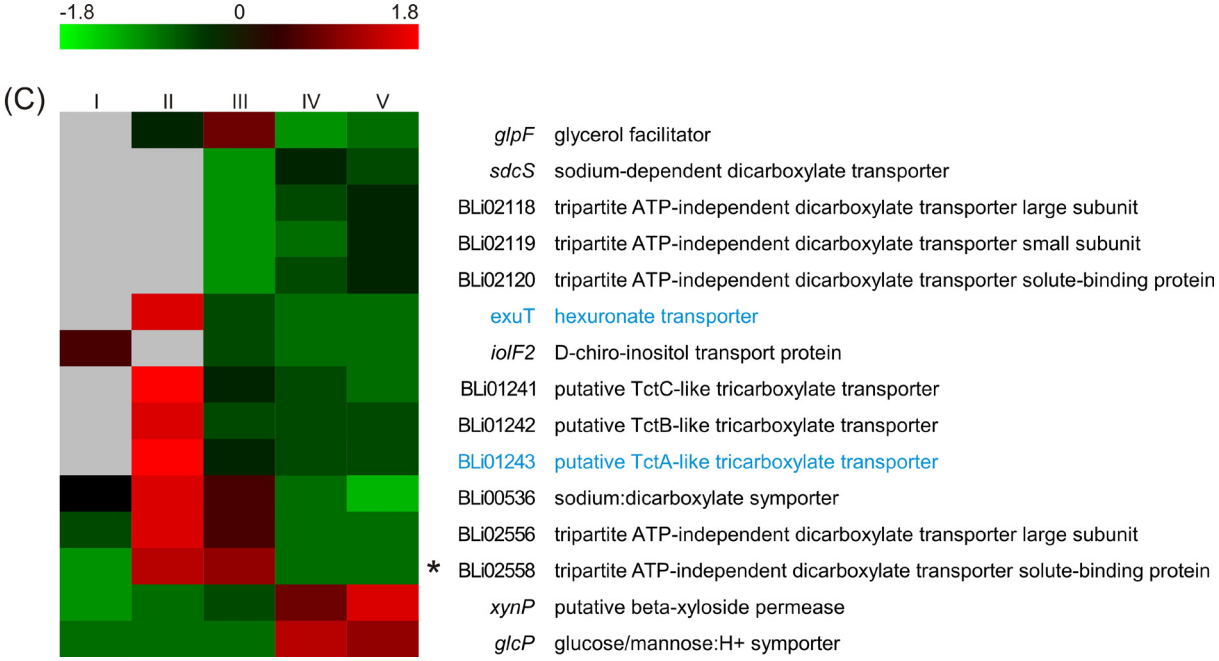


Figure S6 Carbohydrate transport
Heat map representation of Z-score transformed NPKM values. The depicted genes have been annotated as **(A)** carbohydrate ABC transporter components, **(B)** phosphotransferase system EII components and **(C)** further carbohydrate transporters. Please note that the figure does not give a complete list of genes involved in carbohydrate transport. Genes with an assigned antisense RNA (Chapter B) are marked in blue, asterisks indicate a detected protein spot for the respective gene (Table S3) and statistically not significant values are marked by grey boxes.
At sampling point II, the transcript abundances indicate the consumption of the previously synthesized acetate (Figure 7). These results are supported by RNA abundance shifts of associated transporters. The transcript of a tripartite ATP-independent periplasmic dicarboxylate transporter (TRAP) of the TAXI type (BLi02556 and BLi02558), which is suggested as capable of acetate transport (Bakermans et al., 2009), is highly abundant.
During the late production stages, transcripts of transporters of diverse sugars (GlcP, XynP, BglP, SdcS) and cell wall components (MurP, NagP) are abundant. It is likely that this reaction can be accounted to the availability of such compounds due to cell lysis, as for example described at the onset of sporulation in *B. subtilis* (González-Pastor, 2011), or shear effects in the fermenter.

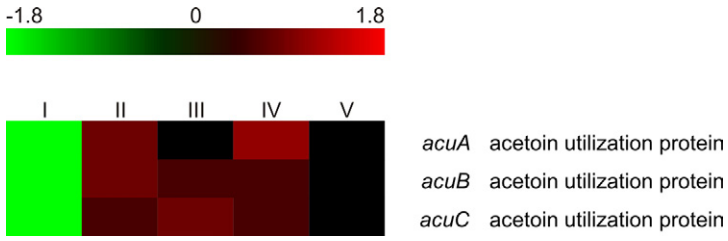


Figure S7 Acetoin utilization operon *acuABC*
Heat map representation of Z-score transformed NPKM values. The depicted genes are annotated as acetoin utilization operon *acuABC*.

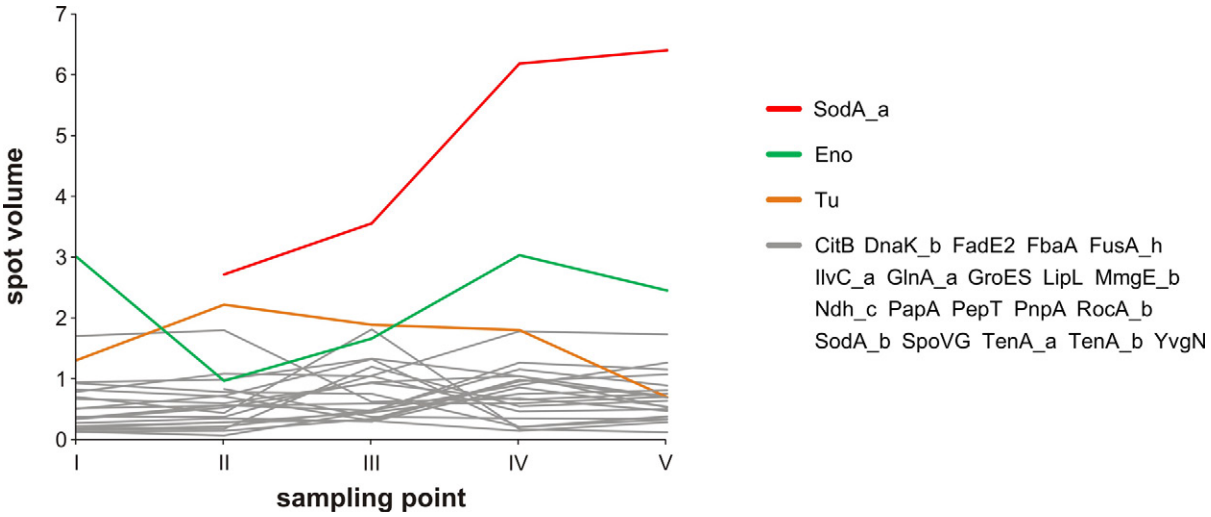


Figure S8 Most abundant proteins
Mean spot volumes of the most abundant proteins are plotted against sampling points (see also Table S3). Grey lines represent protein spot volumes either higher than 0.8% at one sampling point or higher than 0.5% at all sampling points. Lines colored green, red and orange indicate the three most abundant protein spots Eno, Tu and SodA_a. In case a specific protein can be assigned to more than one spot, the particular spot is indicated by an ordering letter. Please note that statistically not significant values are not shown.
Highly abundant transcripts (Figure 2) for which predominant proteins could also be observed are coding for SodA, Tu and SpoVG. Other strongly synthesized proteins are, for example, the enolase Eno and proteins involved in carbon and amino acid metabolism or cofactor synthesis. They also comprise peptidases, the heat shock proteins GroES and DnaK and elongation factor G. Protein spots corresponding to the strongly transcribed genes *lanA1* and *lanA2* could not be identified. This was expected, as these proteins are probably exported by ABC transporters (Dischinger et al., 2009) and thus cannot be detected by proteome analysis of cytoplasmic proteins. The findings of highly abundant proteins match the results for *B. licheniformis* shown by Voigt et al. (Voigt et al., 2004). The only major exception is the absence of the flagellin protein Hag, which is also not highly abundant on transcript level. This effect is due to repression of *hag* gene expression in the presence of amino acids (Bergara et al., 2003).

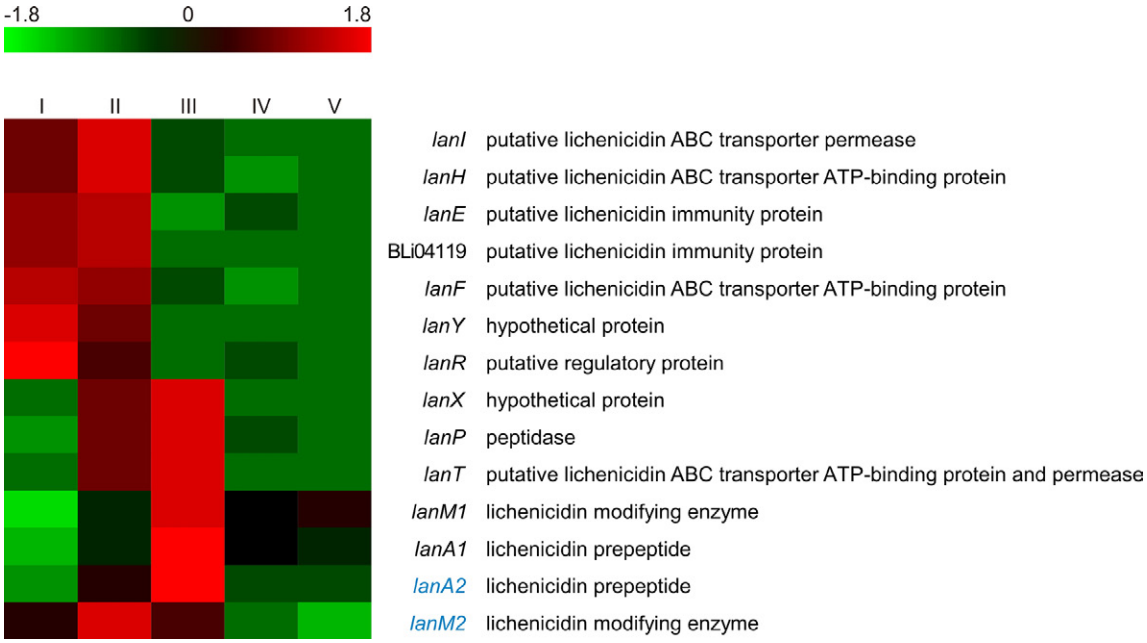


Figure S9 Lichenicidin gene cluster
Heat map representation of Z-score transformed NPKM values. The depicted genes have been identified as two-peptide lantibiotic lichenicidin-processing gene cluster Lan in *B. licheniformis* (Begley et al., 2009; Dischinger et al., 2009). Genes with an assigned antisense RNA (Chapter B) are marked in blue.

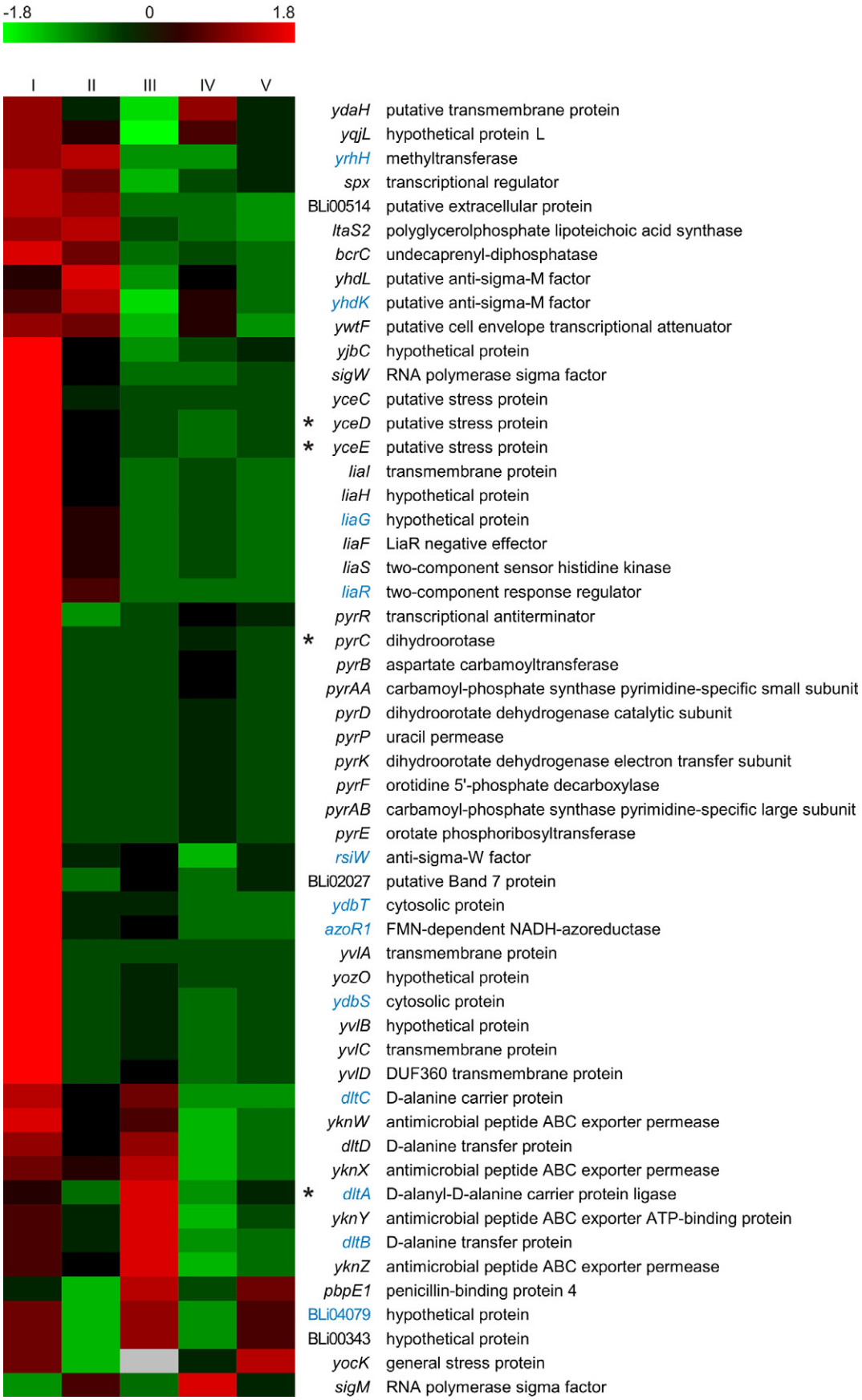
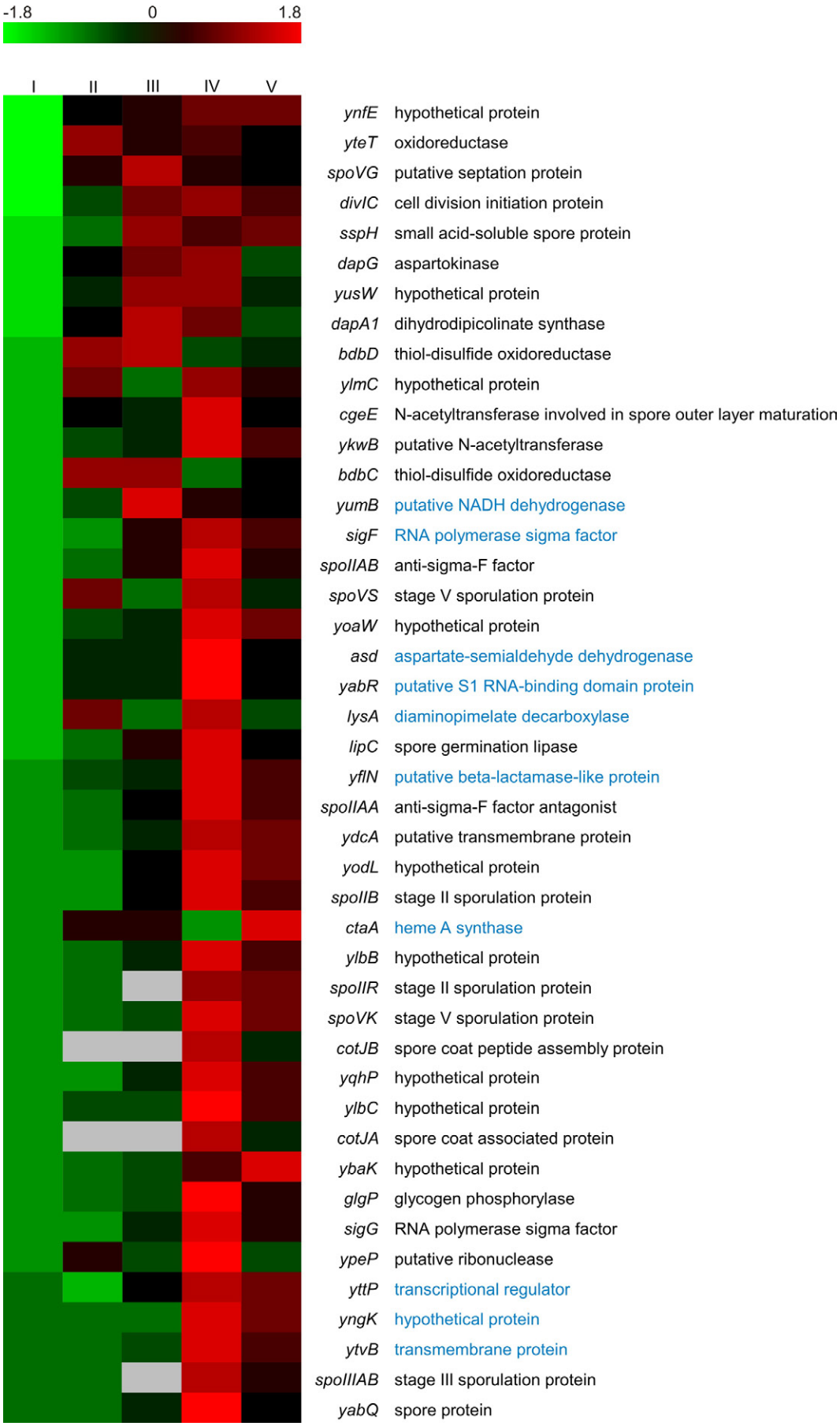
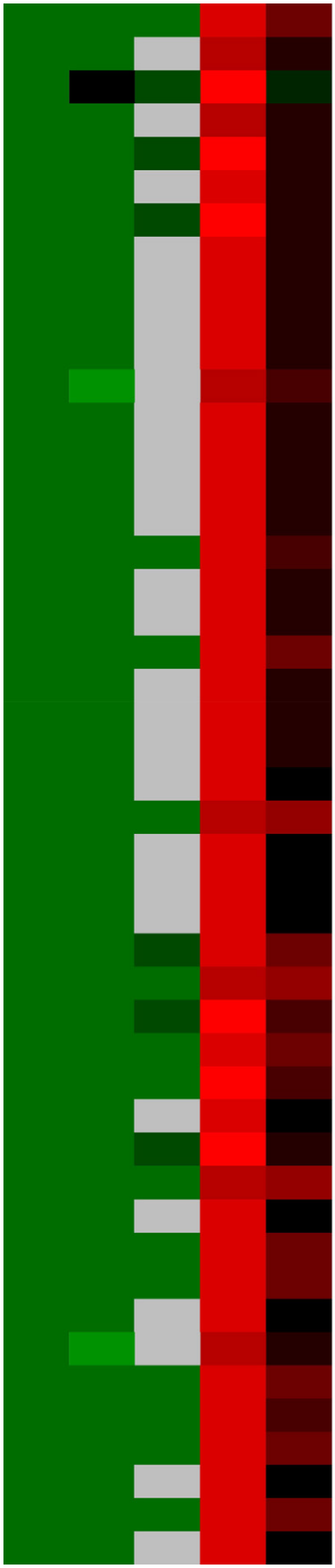
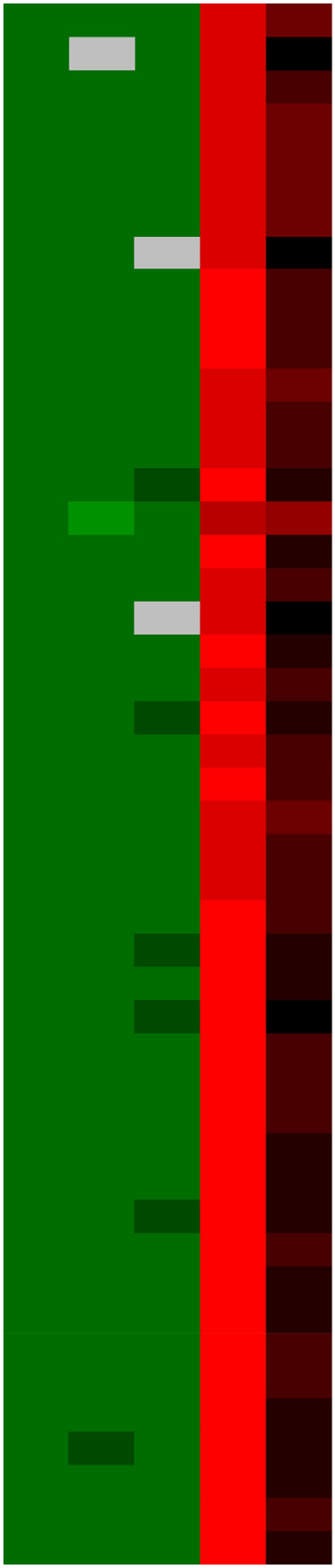

-1.8
0
1.8
I
II
III
IV
V
ydaH putative transmembrane protein
yqjL hypothetical protein L
yrhH methyltransferase
spx transcriptional regulator
BLi00514 putative extracellular protein
ltaS2 polyglycerolphosphate lipoteichoic acid synthase
bcrC undecaprenyl-diphosphatase
yhdL putative anti-sigma-M factor
yhdK putative anti-sigma-M factor
ywtF putative cell envelope transcriptional attenuator
yjbC hypothetical protein
sigW RNA polymerase sigma factor
yceC putative stress protein
* yceD putative stress protein
* yceE putative stress protein
liaI transmembrane protein
liaH hypothetical protein
liaG hypothetical protein
liaF LiaR negative effector
liaS two-component sensor histidine kinase
liaR two-component response regulator
pyrR transcriptional antiterminator
* pyrC dihydroorotase
pyrB aspartate carbamoyltransferase
pyrAA carbamoyl-phosphate synthase pyrimidine-specific small subunit
pyrD dihydroorotate dehydrogenase catalytic subunit
pyrP uracil permease
pyrK dihydroorotate dehydrogenase electron transfer subunit
pyrF orotidine 5'-phosphate decarboxylase
pyrAB carbamoyl-phosphate synthase pyrimidine-specific large subunit
pyrE orotate phosphoribosyltransferase
rsiW anti-sigma-W factor
BLi02027 putative Band 7 protein
ydbT cytosolic protein
azoR1 FMN-dependent NADH-azoreductase
yvlA transmembrane protein
yozO hypothetical protein
ydbS cytosolic protein
yvlB hypothetical protein
yvlC transmembrane protein
yvlD DUF360 transmembrane protein
dltC D-alanine carrier protein
yknW antimicrobial peptide ABC exporter permease
dltD D-alanine transfer protein
yknX antimicrobial peptide ABC exporter permease
* dltA D-alanyl-D-alanine carrier protein ligase
yknY antimicrobial peptide ABC exporter ATP-binding protein
dltB D-alanine transfer protein
yknZ antimicrobial peptide ABC exporter permease
pbpE1 penicillin-binding protein 4
BLi04079 hypothetical protein
BLi00343 hypothetical protein
yocK general stress protein
sigM RNA polymerase sigma factor

Figure S10 Cell envelope stress response
Heat map representation of Z-score transformed NPKM values. The depicted genes have been identified as marker genes for the *B. licheniformis* cell envelope stress response by Wecke et al. (2006). Genes with an assigned antisense RNA (Chapter B) are marked in blue, asterisks indicate a detected protein spot for the respective gene (Table S3) and statistically not significant values are marked by grey boxes. The genes *yvnB*, *sigY*, *pbpX* and *yxlCDEFG* are not shown due to lacking transcript abundances (NPKM values <10).

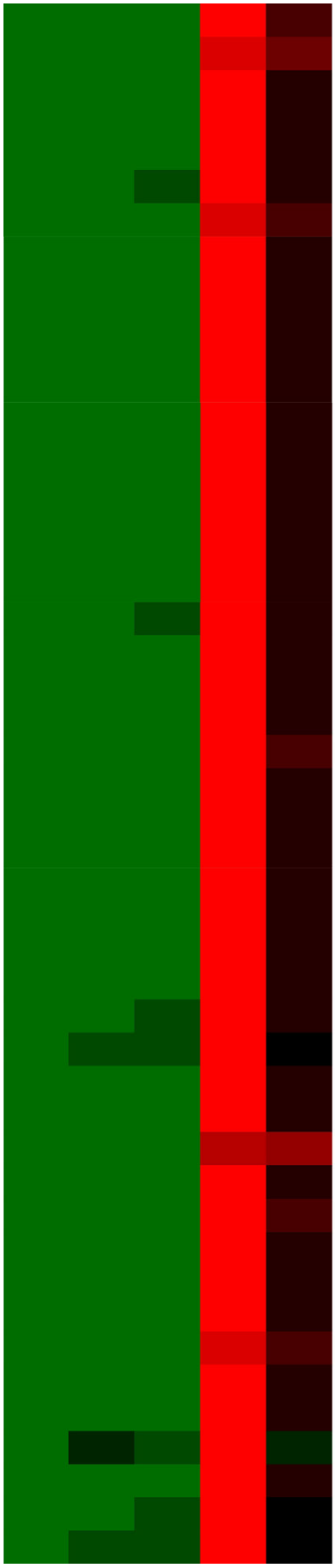

-1.8
0
1.8
I
II
III
IV
V
ynfE hypothetical protein
yteT oxidoreductase
spoVG putative septation protein
divIC cell division initiation protein
sspH small acid-soluble spore protein
dapG aspartokinase
yusW hypothetical protein
dapA1 dihydrodipicolinate synthase
bdbD thiol-disulfide oxidoreductase
ylmC hypothetical protein
cgeE N-acetyltransferase involved in spore outer layer maturation
ykwB putative N-acetyltransferase
bdbC thiol-disulfide oxidoreductase
yumB putative NADH dehydrogenase
sigF RNA polymerase sigma factor
spoIIAB anti-sigma-F factor
spoVS stage V sporulation protein
yoaW hypothetical protein
asd aspartate-semialdehyde dehydrogenase
yabR putative S1 RNA-binding domain protein
lysA diaminopimelate decarboxylase
lipC spore germination lipase
yflN putative beta-lactamase-like protein
spoIIAA anti-sigma-F factor antagonist
ydcA putative transmembrane protein
yodL hypothetical protein
spoIIB stage II sporulation protein
ctaA heme A synthase
ylbB hypothetical protein
spoIIR stage II sporulation protein
spoVK stage V sporulation protein
cotJB spore coat peptide assembly protein
yqhP hypothetical protein
ylbC hypothetical protein
cotJA spore coat associated protein
ybaK hypothetical protein
glgP glycogen phosphorylase
sigG RNA polymerase sigma factor
ypeP putative ribonuclease
yttP transcriptional regulator
yngK hypothetical protein
ytvB transmembrane protein
spoIIIAB stage III sporulation protein
yabQ spore protein



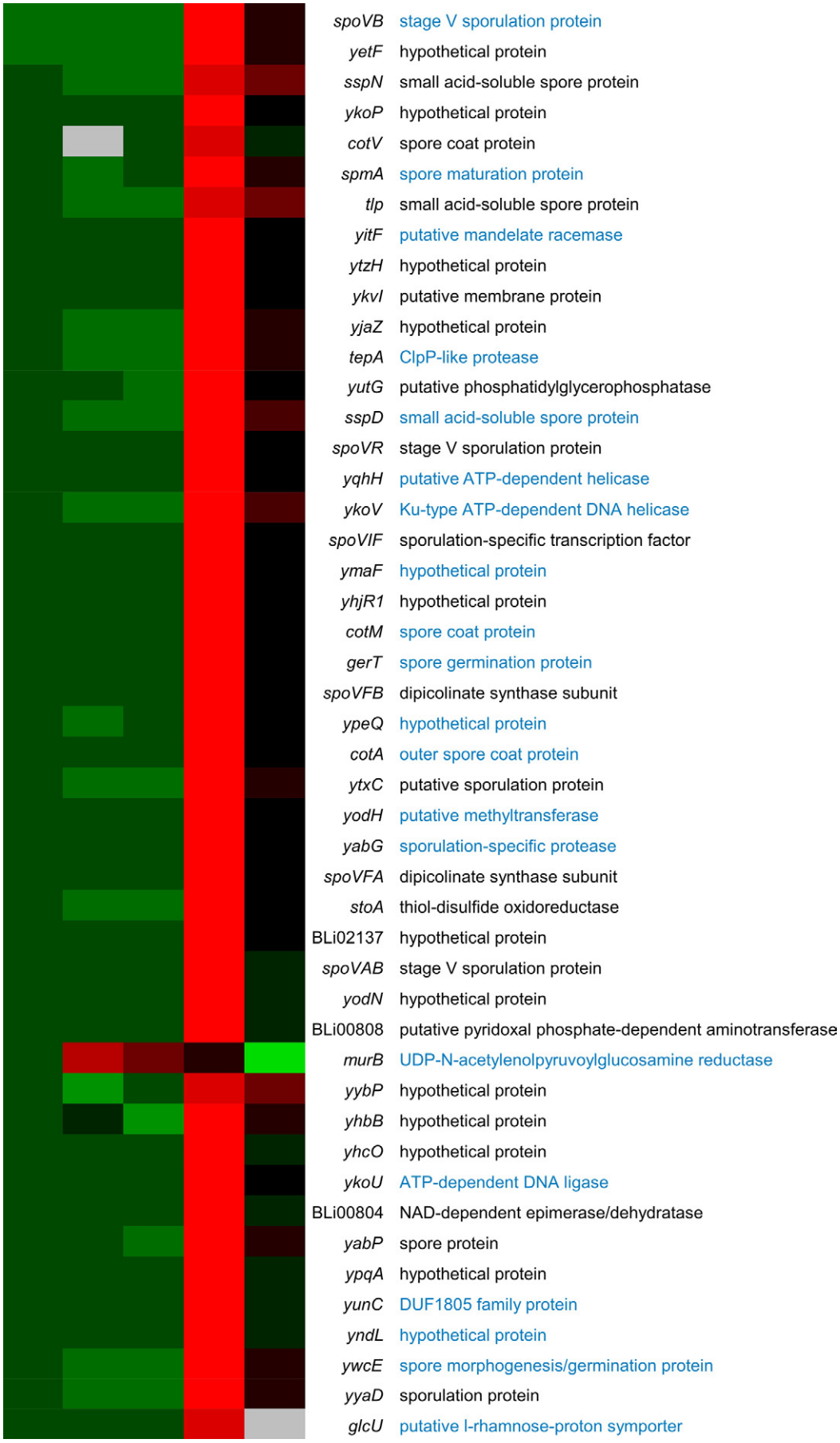
yhdB	hypothetical protein
spoIIID	stage III sporulation protein
ypfB	hypothetical protein
spoIVA	stage IV sporulation protein
spoVD	stage V sporulation protein
yyaC	sporulation protein
yqfC	sporulation protein
spoIIIAG	stage III sporulation protein
yabT	putative serine/threonine-protein kinase
yqfD	SigE-dependent sporulation gene
yabS	hypothetical protein
yozD	hypothetical protein
yqxA	hypothetical protein
yngHB	biotin/lipoyl attachment protein
yhxC	putative oxidoreductase
yaaH	glycoside hydrolase family protein
yhaL	hypothetical protein
spoIIIAH	stage III sporulation protein
yodQ	4-acetamidobutyryl-CoA deacetylase
yuzC	sporulation protein
ydcC	putative sporulation protein
yngE	methylcrotonoyl-CoA carboxylase
ysxE	positive modulator
sqhC	squalene--hopene cyclase
ymfJ	hypothetical protein
spoIIP	cell wall hydrolase
prkA	serine protein kinase
spoIIGA	sigma-E factor-processing peptidase
yqhO	putative acyl hydrolase/lysophospholipase
ytfI	DUF2953 family protein
spoIISA	stage II sporulation protein
spoVAC	stage V sporulation protein
yhaX	hypothetical protein
spoIIIAF	stage III sporulation protein
glgA	glycogen synthase
BLi03524	putative small acid-soluble spore protein
kamA	L-lysine 2,3-aminomutase
yhcN	forespore-specific protein
yhcV	CBS domain protein
yngG	3-hydroxy-3-methylglutaryl-coenzyme A lyase
yteV	sporulation protein
ykzD	hypothetical protein
yjbA	hypothetical protein
BLi01360	hypothetical protein
spoIIIAC	stage III sporulation protein
sspA	small acid-soluble spore protein
yngJ	putative acyl-CoA dehydrogenase



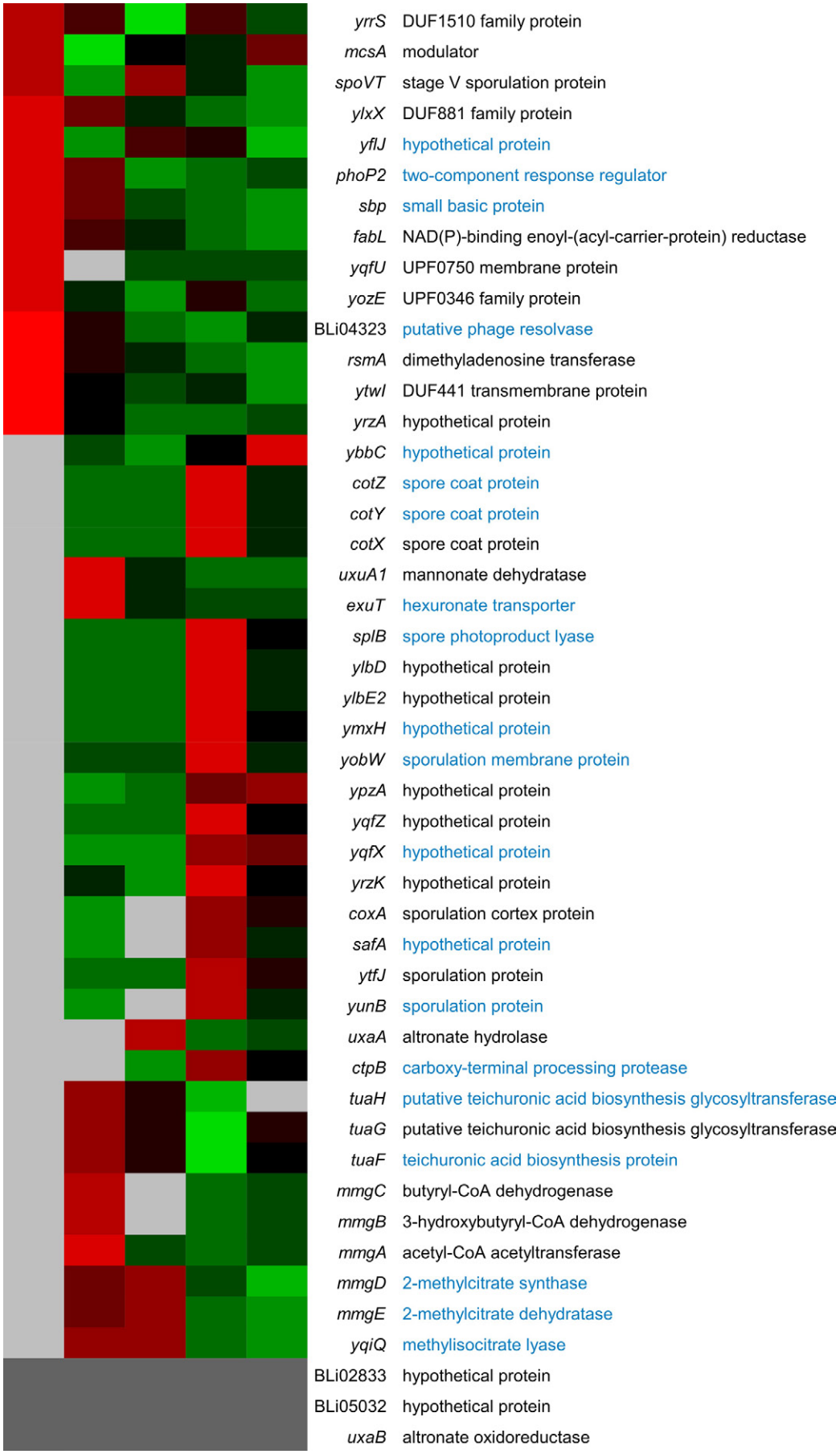
sspO	small acid-soluble spore protein
katX	catalase
sspI	small acid-soluble spore protein
BLi02230	hypothetical protein
sspP	small acid-soluble spore protein
BLi01951	hypothetical protein
yisY	AB hydrolase superfamily protein
yjbE	membrane protein
yngF	short chain enoyl-CoA hydratase
glgD	glycogen biosynthesis protein
spoIIQ	stage II sporulation protein
spoIVB	peptidase S
yqhR	hypothetical protein
yuzA	general stress protein
rsfA	putative transcriptional regulator
yqgO	hypothetical protein
yheD	hypothetical protein
yhcQ	hypothetical protein
yhbH	sporulation protein
yndM	hypothetical protein
yqhG	hypothetical protein
ywcB	DUF485 transmembrane protein
yrrD	forespore-specific sporulation protein
alr2	alanine racemase
azoR2	FMN-dependent NADH-azoreductase
cotO	spore coat morphogenetic protein
sspB	small acid-soluble spore protein
yqhV	hypothetical protein
sodF	putative superoxide dismutase
gerR	putative transcriptional regulator
yhfN	putative metalloprotease
yjaV	hypothetical protein
BLi00840	hypothetical protein
yhfW	putative Rieske 2Fe-2S iron-sulfur protein
dacF	D-alanyl-D-alanine carboxypeptidase
yrzE	transmembrane protein
accC1	acetyl-CoA carboxylase biotin carboxylase subunit
cwlC	N-acetylmuramoyl-L-alanine amidase
spoIIIAD	stage III sporulation protein
yodR	6-acetamido-3-oxohexanoate:acetyl-CoA CoA transferase beta subunit
glgB	1,4-alpha-glucan-branching enzyme
bofC	general stress protein
cotJC	spore coat peptide assembly protein
ywcA	Na+/solute symporter
ytlD	putative anion ABC transporter permease
lonB	Lon-like ATP-dependent protease
spoVAF1	stage V sporulation protein



yurZ	putative alkylhydroperoxidase
yoaR	hypothetical protein
spoIVFB	stage IV sporulation protein
sspK	small acid-soluble spore protein
ytrH	sporulation membrane protein
ytrI	sporulation membrane protein
dacB	D-alanyl-D-alanine carboxypeptidase
yutH	spore coat protein
ycgF	putative amino acid efflux protein
spoIIIAE	stage III sporulation protein
spoIIIAA	stage III sporulation protein
ylaJ	hypothetical protein
spoIVFA	stage IV sporulation protein
yusN	putative spore coat protein
spmB	spore maturation protein
yodT	N(6)-acetyl-beta-lysine transaminase precursor
yodP	beta-lysine acetyltransferase
ydhD	putative sporulation-specific glycosylase
pdaB	polysaccharide deacetylase
ytlC	putative anion ABC transporter ATP-binding protein
ylzJ	hypothetical protein
yheC	hypothetical protein
yuiC	sporulation protein
ytlA	putative anion ABC transporter substrate-binding protein
spoVAA	stage V sporulation protein
yodS	6-acetamido-3-oxohexanoate:acetyl-CoA CoA transferase alpha subunit
yknT	putative sporulation protein
asnO	asparagine synthase
cotF	spore coat protein
yqfT	hypothetical protein
glgC	glucose-1-phosphate adenylyltransferase
gin	forespore-specific protein
spoIIM	stage II sporulation protein
pbpG	penicillin-binding protein 2D
sspF	small acid-soluble spore protein
yraD	spore coat protein F-like protein
yteA	sporulation protein
yjcA	sporulation protein
tgl	protein-glutamine gamma-glutamyltransferase
ydfS	UPF0702 transmembrane protein
seaA	spore envelope assembly protein
yokU	hypothetical protein
yitG	putative major facilitator superfamily protein
yqhQ	hypothetical protein
yitD	phosphosulfolactate synthase
ylbJ	putative sporulation integral membrane protein
ytvI	sporulation integral membrane protein


spoVB stage V sporulation protein
yetF hypothetical protein
sspN small acid-soluble spore protein
ykoP hypothetical protein
cotV spore coat protein
spmA spore maturation protein
tlp small acid-soluble spore protein
yitF putative mandelate racemase
ytzH hypothetical protein
ykvI putative membrane protein
yjaZ hypothetical protein
tepA ClpP-like protease
yutG putative phosphatidylglycerophosphatase
sspD small acid-soluble spore protein
spoVR stage V sporulation protein
yqhH putative ATP-dependent helicase
ykoV Ku-type ATP-dependent DNA helicase
spoVIF sporulation-specific transcription factor
ymaF hypothetical protein
yhjR1 hypothetical protein
cotM spore coat protein
gerT spore germination protein
spoVFB dipicolinate synthase subunit
ypeQ hypothetical protein
cotA outer spore coat protein
ytxC putative sporulation protein
yodH putative methyltransferase
yabG sporulation-specific protease
spoVFA dipicolinate synthase subunit
stoA thiol-disulfide oxidoreductase
BLi02137 hypothetical protein
spoVAB stage V sporulation protein
yodN hypothetical protein
BLi00808 putative pyridoxal phosphate-dependent aminotransferase
murB UDP-N-acetylenolpyruvoylglucosamine reductase
yybP hypothetical protein
yhbB hypothetical protein
yhcO hypothetical protein
ykoU ATP-dependent DNA ligase
BLi00804 NAD-dependent epimerase/dehydratase
yabP spore protein
ypqA hypothetical protein
yunC DUF1805 family protein
yndL hypothetical protein
ywcE spore morphogenesis/germination protein
yyaD sporulation protein
glcU putative l-rhamnose-proton symporter

yitE UPF0750 membrane protein
ytzC DUF2524 family protein
lytH peptidoglycan hydrolase
yphA putative membrane protein
yueG spore germination protein
pbpI sporulation-specific penicillin-binding protein 4b
adhB alcohol dehydrogenase-like protein
phoR two-component sensor histidine kinase
cotE spore coat protein
yutC putative sporulation lipoprotein
ykzE hypothetical protein
spoIID stage II sporulation protein
ykuS UPF0180 family protein
cotH inner spore coat protein
ftsK DNA translocase
yhfA putative membrane protein
ypjB sporulation protein
ldt putative L,D-transpeptidase
clpC ATP-dependent Clp protease ATP-binding subunit
bofA pro-sigmaK processing inhibitor
ycgL hypothetical protein
jag hypothetical protein
cwlJ cell wall hydrolase
disA DNA integrity scanning protein
exuR HTH-type transcriptional repressor
ylxW DUF881 family protein
spoIIIJ membrane protein translocase
divIB division initiation protein
murG UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
spo0J stage 0 sporulation protein
ytkD nucleoside triphosphatase
yyxA putative serine protease
spoVE stage V sporulation protein
mbl cell-shape determining protein
ispG 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
BLi03803 NAD-dependent epimerase/dehydratase
soj sporulation initiation inhibitor protein
ykoY putative membrane protein
BLi03150 hypothetical protein
ftsY cell division protein
pbpF penicillin-binding protein 2C
yeaA hypothetical protein
splA transcriptional repressor
mcsB modulator
ytaF putative sporulation protein
nfo endonuclease 4
subA supressor


yrrS DUF1510 family protein
mcsA modulator
spoVT stage V sporulation protein
ylxX DUF881 family protein
yflJ hypothetical protein
phoP2 two-component response regulator
sbp small basic protein
fabL NAD(P)-binding enoyl-(acyl-carrier-protein) reductase
yqfU UPF0750 membrane protein
yozE UPF0346 family protein
BLi04323 putative phage resolvase
rsmA dimethyladenosine transferase
ytwI DUF441 transmembrane protein
yrzA hypothetical protein
ybbC hypothetical protein
cotZ spore coat protein
cotY spore coat protein
cotX spore coat protein
uxuA1 mannonate dehydratase
exuT hexuronate transporter
splB spore photoproduct lyase
ylbD hypothetical protein
ylbE2 hypothetical protein
ymxH hypothetical protein
yobW sporulation membrane protein
ypzA hypothetical protein
yqfZ hypothetical protein
yqfX hypothetical protein
yrzK hypothetical protein
coxA sporulation cortex protein
safA hypothetical protein
ytfJ sporulation protein
yunB sporulation protein
uxaA altronate hydrolase
ctpB carboxy-terminal processing protease
tuaH putative teichuronic acid biosynthesis glycosyltransferase
tuaG putative teichuronic acid biosynthesis glycosyltransferase
tuaF teichuronic acid biosynthesis protein
mmgC butyryl-CoA dehydrogenase
mmgB 3-hydroxybutyryl-CoA dehydrogenase
mmgA acetyl-CoA acetyltransferase
mmgD 2-methylcitrate synthase
mmgE 2-methylcitrate dehydratase
yqiQ methylisocitrate lyase
BLi02833 hypothetical protein
BLi05032 hypothetical protein
uxaB altronate oxidoreductase

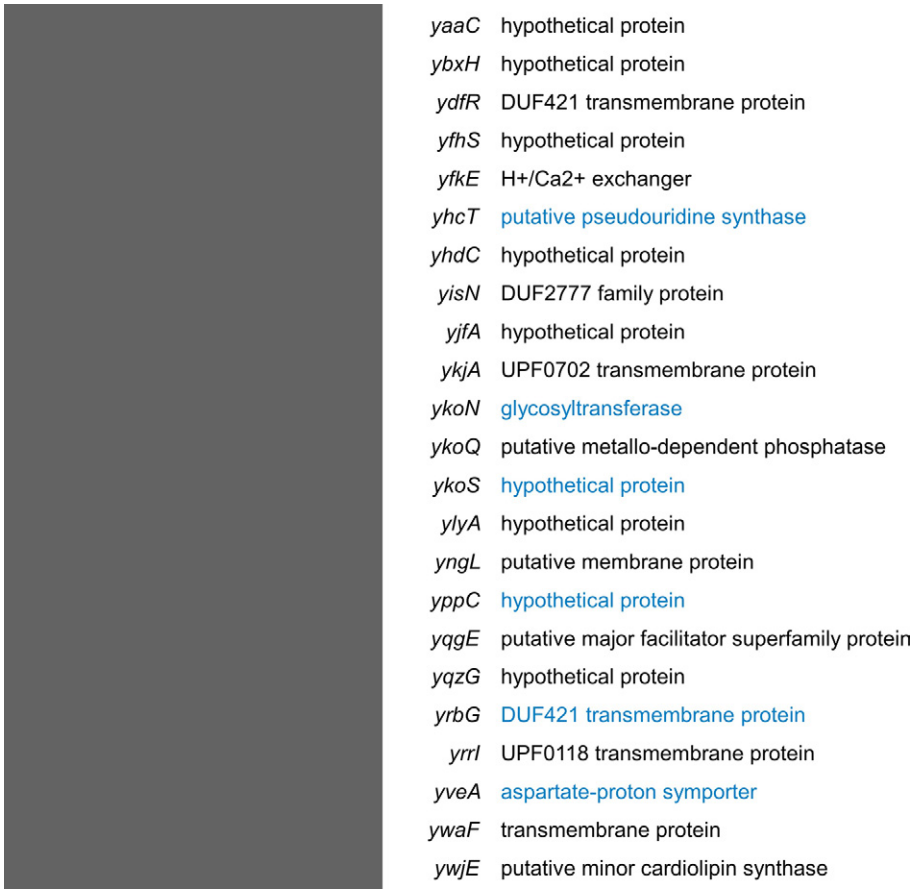


Figure S11 Sporulation
Heat map representation of Z-score transformed NPKM values. The depicted genes were identified as members of the sporulation cascade by reciprocal BLAST analysis (Lechner et al., 2011) between *B. licheniformis* genes (Chapter B) and *B. subtilis* genes assigned to sporulation (Mäder et al., 2012). Genes with an assigned antisense RNA (Chapter B) are marked in blue, genes with NPKM values <10 at all sampling points are indicated by dark grey boxes, and statistically not significant values are indicated by light grey boxes.

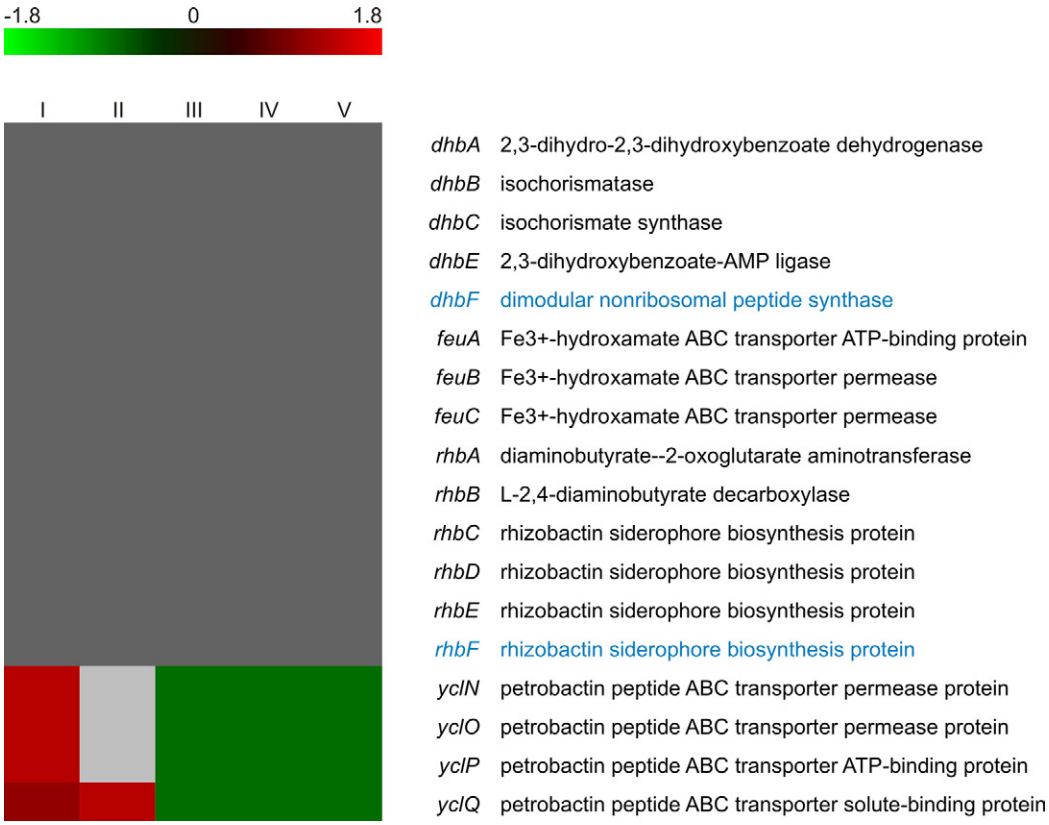


Figure S12 Iron starvation
Heat map representation of Z-score transformed NPKM values. The depicted genes have been identified as marker genes for *B. licheniformis* iron starvation by Nielsen et al. (2010). Genes with an assigned antisense RNA (Chapter B) are marked in blue, genes with NPKM values <20 at all sampling points are indicated by dark grey boxes and statistically not significant values are indicated by light grey boxes.
A well described indicator for iron starvation is the induction of siderophore anabolic genes (*dhbABCEF* and *rhbCDEF*) and siderophore importer genes (*feuABC* and *yclNOPQ*) (Nielsen et al., 2010), which are under control of the transcriptional regulator Fur (Ollinger et al., 2006). With exception of the Ycl operon, these transcripts of these genes are not abundant during the fermentation process. This behavior indicates that iron starvation does not occur in any of the examined fermentation phases and corresponds to the observation that the Fur regulon is only induced during growth in minimal medium with restricted iron supply but not in rich medium (Helmann et al., 2003; Mostertz et al., 2004; Nielsen et al., 2010), as used in this study.

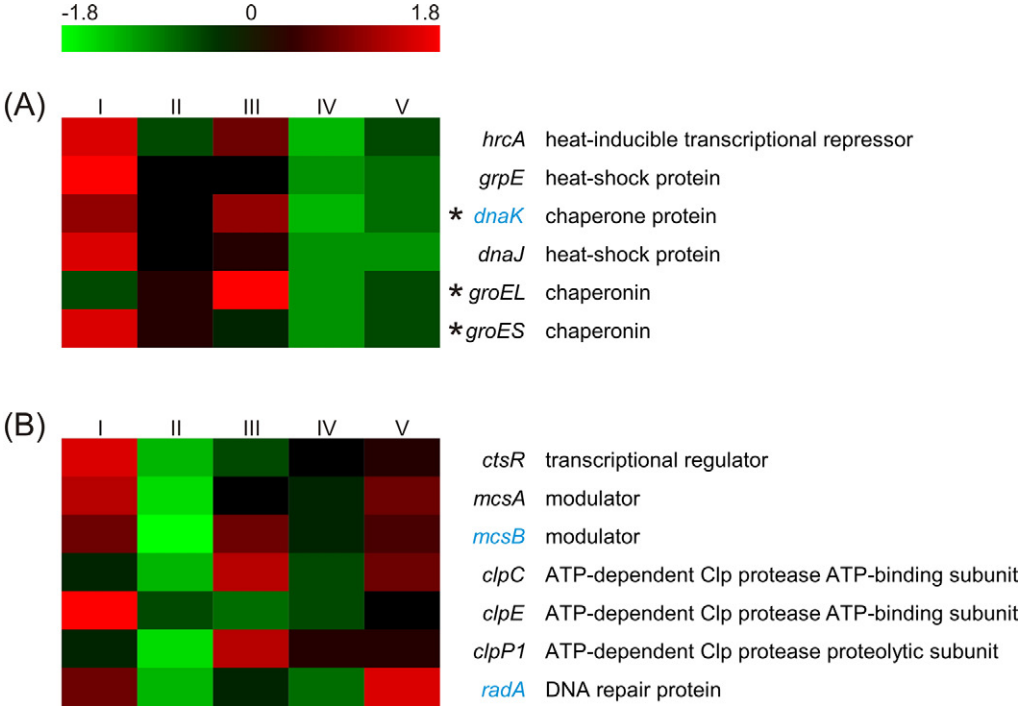


Figure S13 Heat shock response
Heat map representation of Z-score transformed NPKM values. The depicted **(A)** HrcA and **(B)** CtsR regulons have been identified as *B. licheniformis* heat stress markers by Nielsen et al. (2010). Genes with an assigned antisense RNA (Chapter B) are marked in blue, asterisks indicate a detected protein spot for the respective gene (Table S3).

In cases of heat shock in *B. licheniformis*, the HrcA regulon including the *dnaK* and the *groE* operons and the CtsR regulon have been shown to be induced (Nielsen et al., 2010). Furthermore, genes involved in iron and purine metabolism were upregulated, whereas the ABC transporter encoding *ytrABCEF* operon was repressed (Nielsen et al., 2010). Since the here analyzed fermentation was not performed under heat shock conditions, the data did not show a typical heat shock response. Nevertheless, a transcriptional reaction to the fermentation temperature of 39 °C cannot be excluded, as the responses between moderate and severe heat shock can differ remarkably (Barreiro et al., 2009). Despite the constant and moderate process temperature, differential expression of genes involved in a classic heat shock response (Helmann et al., 2001; Nielsen et al., 2010; Schumann, 2003) was observed. The ATP-dependent ClpCP protease shows changes in transcript abundance. However, the enzyme has been reported to be involved in proteolysis of misfolded proteins, which were caused by a variety of different stressors (Frees et al., 2007; Krüger et al., 2000) including for example the mentioned oxidative stress (Leichert et al., 2003; Mostertz et al., 2004; Schroeter et al., 2011). A reaction to non-heat stressors is also known for the chaperonins GroEL and GroES (Babu et al., 2011; Lee et al., 2011; Seydlová et al., 2012), which were transcribed with NPKM values >1000 at all sampling points or sampling point I, respectively. Although both chaperonins are encoded by the *groE* operon, the ratio of the transcript levels of the genes is highly variable between sampling points I and III – whereas *groES* declines over time, the *groEL* transcripts seem to be increasing. An explanation might be the processing of the bicistronic transcript into two monocistronic mRNAs. This effect has been described for *Agrobacterium tumefaciens* (Segal and Ron, 1995), in which the accumulation of *groEL* over time could also be shown. Furthermore, the determined GroEL protein amount (Table S3) does not correlate with the amount of *groEL* mRNA, but shows the strongest abundance of protein at sampling point I.

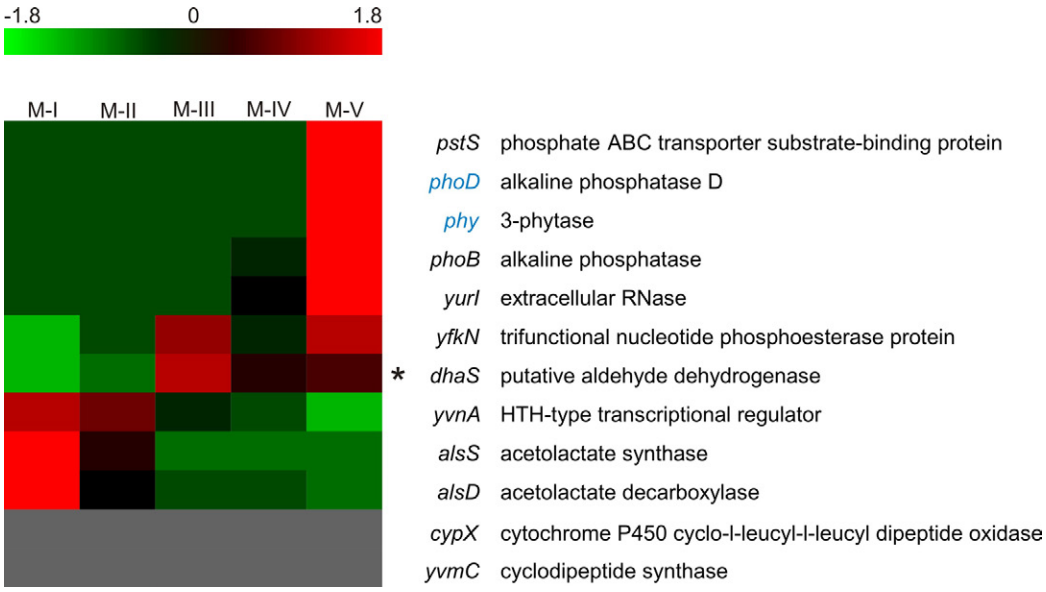


Figure S14 Phosphate starvation response
Heat map representation of Z-score transformed NPKM values. The depicted genes have been identified as marker genes for the *B. licheniformis* phosphate starvation response by Hoi et al. (2006). The figure shows the transcript abundances of those marker genes within the fermentation samples of one replicate (replicate M). Genes with an assigned antisense RNA (Chapter B) are marked in blue; asterisks indicate a detected protein spot for the respective gene Table S3. Genes with NPKM values <10 at all sampling points are indicated by dark grey boxes.

At sampling point V, transcripts of *pstS* and the downstream *pst* operon encoding a phosphate ABC transporter (see also Figure 3), but also of *phy*, *phoB*, *phoD*, and *yfnK* are highly abundant. All of the mentioned genes are members of the Pho regulon, also known to provide a specific phosphate starvation stress response in *B. subtilis* (Allenby et al., 2005; Antelmann et al., 2000). Furthermore, *yurI* and *dhaS* abundant as described by Hoi et al. (2006), supporting the inference of a phosphate shortage in this sample. However, no corresponding transcriptional response could be observed for the other fermentation samples (L and R) of sampling point V. This disparity coincides with an increased partial pressure of oxygen, starting ~1 h ahead of sampling point M-V (Figure S1), which indicates a decreased metabolic activity in this sample at this fermentation stage. At the same time, the genes *yvnA*, *alsD,* and *alsS* show no increased transcript abundance. This is in accordance with the results obtained by Hoi et al. (2006), who observed no induction of these genes in the early phases (~2 h) of declining metabolic activity after phosphate exhaustion. In addition, transcripts of genes involved in metabolic pathways, chemotaxis and especially teichoic acid synthesis were lacking, as also shown before (Hoi et al., 2006). The transcript abundance of the teichuronic acid synthesis operon (*tuaABCDEFG*), which enables the replacement of phosphate-rich teichoic acids, has been described as additional indicator of phosphate starvation in *B. subtilis* (Allenby et al., 2005; Liu and Hulett, 1998) and could also be observed in sample M-V (data not shown). In summary, these results indicate that fermentation M is under phosphate limitation during the latest stage of the process. This finding might be a target for bioprocess optimization aiming at the prolongation of the fermentation process.

Figure S15 Putative Tat signal peptide of Subtilisin Carlsberg
(A) Tat signal peptides contain a conserved twin-arginine motif, defined as SRRXFLK, in which the consecutive arginine residues are almost always invariant (Palmer and Berks, 2012). Nevertheless, one arginine residue (§), but not both, can be replaced by a lysine residue (DeLisa et al., 2002; Ize et al., 2002). The other motif residues occur with a frequency of >50%, whereas the amino acids at the # positions are usually hydrophobic and the amino acid at the X position is polar (Kouwen et al., 2009; Palmer and Berks, 2012). **(B)** The signal peptide of Subtilisin Carlsberg matches the Tat model motif with the exception of two hydrophilic residues in the middle positions, a condition not considered an exclusion criterion for a putative Tat signal peptide motif (Berks, 1996; van Dijl et al., 2002). In Sec signal peptides, helix breaking glycine or proline residues are often found in the middle of the hydrophobic domain following the N-terminal domain (de Vrije et al., 1990). These are not found in *B. subtilis* Tat signal peptides (van Dijl et al., 2002) and also not in the signal peptide of Subtilisin Carlsberg.

Table S1 *k-means* clustering of expression profiles

Please refer to “Chapter C_Additional information” on digital medium.

Table S2 GO term enrichment analysis of k-means clusters

Please refer to “Chapter C_Additional information” on digital medium.

Table S3 Proteome data

Please refer to “Chapter C_Additional information” on digital medium.

References

Allenby, N.E.E., Connor, N.O., Prágai, Z., Ward, C., Wipat, A., and Harwood, C.R. (2005). Genome-Wide Transcriptional Analysis of the Phosphate Starvation Stimulon of *Bacillus subtilis*. J Bacteriol *187*, 8063.

Antelmann, H., Scharf, C., and Hecker, M. (2000). Phosphate Starvation-Inducible Proteins of *Bacillus subtilis*: Proteomics and Transcriptional Analysis. J Bacteriol *182*, 4478.

Babu, M.M.G., Sridhar, J., and Gunasekaran, P. (2011). Global transcriptome analysis of *Bacillus cereus* ATCC 14579 in response to silver nitrate stress. J Nanobiotechnology *9*, 49.

Bakermans, C., Sloup, R.E., Zarka, D.G., Tiedje, J.M., and Thomashow, M.F. (2009). Development and use of genetic system to identify genes required for efficient low-temperature growth of *Psychrobacter arcticus* 273-4. Extremophiles *13*, 21–30.

Barreiro, C., Nakunst, D., Hüser, A.T., De Paz, H.D., Kalinowski, J., and Martín, J.F. (2009). Microarray studies reveal a "differential response" to moderate or severe heat shock of the HrcA- and HspR-dependent systems in *Corynebacterium glutamicum*. Microbiology+ *155*, 359–372.

Begley, M., Cotter, P.D., Hill, C., and Ross, R.P. (2009). Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. Appl Environ Microbiol *75*, 5451–5460.

Bergara, F., Ibarra, C., Iwamasa, J., Patarroyo, J.C., Aguilera, R., and Ma, L.M. (2003). CodY Is a Nutritional Repressor of Flagellar Gene Expression in *Bacillus subtilis*. J Bacteriol *185*, 3118–3126.

Berks, B.C. (1996). A common export pathway for proteins binding complex redox cofactors? Mol Microbiol *22*, 393–404.

DeLisa, M.P., Samuelson, P., Palmer, T., and Georgiou, G. (2002). Genetic analysis of the twin arginine translocator secretion pathway in bacteria. J Biol Chem *277*, 29825–29831.

Detsch, C., and Stülke, J. (2003). Ammonium utilization in *Bacillus subtilis*: transport and regulatory functions of NrgA and NrgB. Microbiology+ *149*, 3289–3297.

de Vrije, G.J., Batenburg, A.M., Killian, J.A., and De Kruijff, B. (1990). Lipid involvement in protein translocation in *Escherichia coli*. Mol Microbiol *4*, 143–150.

Dischinger, J., Josten, M., Szekat, C., Sahl, H.-G., and Bierbaum, G. (2009). Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13. PLoS One *4*, e6788.

Frees, D., Savijoki, K., Varmanen, P., and Ingmer, H. (2007). Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. Mol Microbiol *63*, 1285–1295.

González-Pastor, J.E. (2011). Cannibalism: a social behavior in sporulating *Bacillus subtilis*. FEMS Microbiol Rev *35*, 415–424.

Helmann, J.D., Wu, M.F.W., Kobel, P.A., Gamo, F.-J., Wilson, M., Morshedi, M.M., Navre, M., Paddon, C. (2001). Global Transcriptional Response of *Bacillus subtilis* to Heat Shock. J Bacteriol *183*, 7318.

Helmann, J.D., Wu, M.F.W., Kobel, P.A., Gamo, F.-J., Wilson, M., Morshedi, M.M., Navre, M., Paddon, C. (2003). The Global Transcriptional Response of *Bacillus subtilis* to Peroxide Stress Is Coordinated by Three Transcription Factors. J Bacteriol *185*, 243.

Hoi, L.T., Voigt, B., Jürgen, B., Ehrenreich, A., Gottschalk, G., Evers, S., Feesche, J., Maurer, K.-H., Hecker, M., and Schweder, T. (2006). The phosphate-starvation response of *Bacillus licheniformis*. Proteomics *6*, 3582–3601.

Ize, B., Gerard, F., Zhang, M., Chanal, A., Voulhoux, R., Palmer, T., Filloux, A., and Wu, L.-F. (2002). In vivo Dissection of the Tat Translocation Pathway in *Escherichia coli*. J Mol Biol *317*, 327–335.

Kouwen, T.R.H.M., van der Ploeg, R., Antelmann, H., Hecker, M., Homuth, G., Mäder, U., and van Dijl, J.M. (2009). Overflow of a hyper-produced secretory protein from the *Bacillus* Sec pathway into the Tat pathway for protein secretion as revealed by proteogenomics. Proteomics *9*, 1018–1032.

Krüger, E., Witt, E., Ohlmeier, S., Hanschke, R., and Hecker, M. (2000). The Clp proteases of *Bacillus subtilis* are directly involved in degradation of misfolded proteins. J Bacteriol *182*, 3259–3265.

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P.F., and Prohaska, S.J. (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics *12*, 124.

Lee, N., Yeo, I., Park, J., and Hahm, Y. (2011). Growth inhibition and induction of stress protein, GroEL, of *Bacillus cereus* exposed to antibacterial peptide isolated from *Bacillus subtilis* SC-8. Appl Biochem Biotechnol *165*, 235–242.

Leichert, L.I.O., Scharf, C., and Hecker, M. (2003). Global Characterization of Disulfide Stress in *Bacillus subtilis*. J Bacteriol *185*, 1967.

Liu, W., and Hulett, F.M. (1998). Comparison of PhoP binding to the *tuaA* promoter with PhoP binding to other Pho-regulon promoters establishes a *Bacillus subtilis* Pho core binding site. Microbiology+ *144*, 1443–1450.

Mäder, U., Schmeisky, A.G., Flórez, L.A., and Stülke, J. (2012). SubtiWiki--a comprehensive community resource for the model organism *Bacillus subtilis*. Nucleic Acids Res *40*, D1278–87.

Mostertz, J., Scharf, C., Hecker, M., and Homuth, G. (2004). Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress. Microbiology+ *150*, 497–512.

Nielsen, A.K., Breüner, A., Krzystanek, M., Andersen, J.T., Poulsen, T.A., Olsen, P.B., Mijakovic, I., and Rasmussen, M.D. (2010). Global transcriptional analysis of *Bacillus licheniformis* reveals an overlap between heat shock and iron limitation stimulon. J Mol Microbiol Biotechnol *18*, 162–173.

Ollinger, J., Song, K.-B., Antelmann, H., Hecker, M., and Helmann, J.D. (2006). Role of the Fur Regulon in Iron Transport in *Bacillus subtilis*. J Bacteriol *188*, 3664–3673.

Palmer, T., and Berks, B.C. (2012). The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol *10*, 483–496.

Schroeter, R., Voigt, B., Jürgen, B., Methling, K., Pöther, D.-C., Schäfer, H., Albrecht, D., Mostertz, J., Mäder, U., Evers, S., et al. (2011). The peroxide stress response of *Bacillus licheniformis*. Proteomics *11*, 2851–2866.

Schumann, W. (2003). The *Bacillus subtilis* heat shock stimulon. Cell Stress Chaperon *8*, 207–217.

Segal, G., and Ron, E. (1995). The *groESL* Operon of *Agrobacterium tumefaciens*: Evidence for Heat Shock-Dependent mRNA cleavage. J Bacteriol *177*, 750–757.

Seydlová, G., Halada, P., Fišer, R., Toman, O., Ulrych, A., and Svobodová, J. (2012). DnaK and GroEL chaperones are recruited to the *Bacillus subtilis* membrane after short-term ethanol stress. J Appl Microbiol *112*, 765–774.

van Dijl, J.M., Braun, P.G., Robinson, C., Quax, W.J., Antelmann, H., Hecker, M., Müller, J., Tjalsma, H., Bron, S., and Jongbloed, J.D.H. (2002). Functional genomic analysis of the *Bacillus subtilis* Tat pathway for protein secretion. J Biotechnol *98*, 243–254.

Voigt, B., Schweder, T., Becher, D., Ehrenreich, A., Gottschalk, G., Feesche, J., Maurer, K.-H., and Hecker, M. (2004). A proteomic view of cell physiology of *Bacillus licheniformis*. Proteomics *4*, 1465–1490.

Wecke, T., Veith, B., Ehrenreich, A., and Mascher, T. (2006). Cell envelope stress response in *Bacillus licheniformis*: integrating comparative genomics, transcriptional profiling, and regulon mining to decipher a complex regulatory network. J Bacteriol *188*, 7500–7511.

CHAPTER D

TRAV: A GENOME CONTEXT SENSITIVE TRANSCRIPTOME BROWSER

Sascha Dietrich, **Sandra Wiegand** and Heiko Liesegang

PLoS One (2014), 9(4):e93677

Authors' contributions to this work

Programmed TraV: SD
Performed benchmark tests: SD
Developed the prediction algorithms: SW, SD
Wrote paper: SD, HL, SW
Conceived and designed the experiments: HL

TraV: A Genome Context Sensitive Transcriptome Browser

Sascha Dietrich, Sandra Wiegand, Heiko Liesegang*

Abteilung für Angewandte und Genomische Mikrobiologie, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität Göttingen, Göttingen, Germany

Abstract

Next-generation sequencing (NGS) technologies like Illumina and ABI Solid enable the investigation of transcriptional activities of genomes. While read mapping tools have been continually improved to enable the processing of the increasing number of reads generated by NGS technologies, analysis and visualization tools are struggling with the amount of data they are presented with. Current tools are capable of handling at most two to three datasets simultaneously before they are limited by available memory or due to processing overhead. In order to process fifteen transcriptome sequencing experiments of *Bacillus licheniformis* DSM13 obtained in a previous study, we developed TraV, a RNA-Seq analysis and visualization tool. The analytical methods are designed for prokaryotic RNA-seq experiments. TraV calculates single nucleotide activities from the mapping information to visualize and analyze multiple transcriptome sequencing experiments. The use of nucleotide activities instead of single read mapping information is highly memory efficient without incurring a processing overhead. TraV is available at http://appmibio.uni-goettingen.de/index.php?sec = serv.

Citation: Dietrich S, Wiegand S, Liesegang H (2014) TraV: A Genome Context Sensitive Transcriptome Browser. PLoS ONE 9(4): e93677. doi:10.1371/journal.pone.0093677

Editor: Vinod Scaria, CSIR Institute of Genomics and Integrative Biology, India

Received October 7, 2013; **Accepted** March 10, 2014; **Published** April 7, 2014

Copyright: © 2014 Dietrich et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work has been funded by the Bundesministerium für Bildung und Forschung (FKZ-0315387). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: hlieseg@gwdg.de

Introduction

The possibility to sequence complete transcriptomes (RNA-Seq) opens a new level of quality of transcriptomics [1]. The development of next generation sequencing (NGS) methods [2] enabled the quantitative strand-specific investigation of transcriptional activities of genomes at single nucleotide resolution [3,4]. The discovery of new regulatory RNA features like small RNAs, riboswitches and antisense transcripts revealed the existence of an RNA based regulation layer in prokaryotic genomes [5,6]. A detailed analysis of highly resolved data on transcriptionally active genome loci should therefore enable the identification of these regulators on a whole genome scale. A comprehensive workflow in a microbial transcriptome experiment may comprise four steps: (i) production of sequence data from an RNA sample, (ii) filtering of bad quality reads and reads for rRNA and tRNA sequences, (iii) strand-specific mapping of the remaining RNA-derived sequences to the genome, and finally (iv) functional analysis of the transcriptionally active genomic regions in their physiological context. A deep sequencing experiment on a microbial genome with Illumina technology may result in several million reads per experiment [7–9]. This sheer amount of data is challenging with regards to read mapping as well as detailed analysis. Obviously, sophisticated bioinformatics tools are crucial to enable convenient and rapid RNA-Seq data analysis. Whereas NGS techniques like 454, Illumina and ABI Solid have continuously advanced over the years, approaches to improve or develop bioinformatics tools to handle the generated data have focused almost exclusively on mapping of the sequences to a genomic backbone, e.g. SSAHA2 [10], bowtie2 [11] and/or BWA [12]. Current visualization tools like Artemis [13], SAMSCOPE [14] or Integrative Genomics Viewer [15] focus on single or few parallel datasets and face performance issues due to the handling of single read mapping information. Therefore, the analysis of multiple datasets in parallel remains difficult, thus demanding further developments in the area of visualization and automated analysis. Here, we introduce TraV (**Tra**nscriptome **V**iewer), a freely available tool which provides support in organization and analysis of multiple transcriptome datasets in relation to the corresponding genomic context. TraV focuses on the identification of regions of transcriptional activity in correspondence to known genes like 5′ and 3′ untranslated regions (UTRs), transcripts that do not correspond to known genomic features, antisense transcripts and transcription start sites (TSS). TraV's ability to process many RNA-Seq data sets simultaneously enables the comparison of many different experimental conditions at the same time. The handling of multiple RNA-Seq datasets is based upon a data abstraction which transforms read mapping data into single base transcriptional activities of the genome. In case single read mapping information is required other tools have to be applied. Thereby, the tool facilitates the search for novel features based on comparative RNA-Seq analysis. TraV's capabilities make the tool an appropriate choice for the comparative analysis of multiple transcriptome experiments with focus on the transcriptional activities of corresponding genome loci from different experiments.

Materials and Methods

Calculation of base activity counts

TraV uses single base resolution coverage counts for both positive and negative strand as basis for all calculations and

graphical presentations of mapping information, a method firstly described by Wurtzel *et al.* [4]. The coverage counts are calculated from a SAM mapping file [16], obtainable from currently available single read mappers like e.g. bowtie2 [11]. The SAMtoTDS tool provided with TraV can be used to convert SAM files into TraV's innate TDS format (see below for details). For each successfully mapped read, the associated base coverage counts for regions covered by the read are increased by one. To constitute a successful mapping, a read has to be uniquely mapped, meaning that it has only one optimal mapping position in the genome. In case of multiple best mapping locations, a read is considered multi-mapped and will not be included in the base coverage calculation. After all mapped reads have been processed; the resulting base coverage strings serve as an abstracted representation of the mapping. This procedure greatly reduces the memory required to handle such transcriptome mapping data and thereby allows TraV to deal with multiple datasets simultaneously.

Data analysis

The analytical methods embodied in TraV are designed for the identification of regions of transcriptional activity fulfilling a set of constraints. These regions of transcriptional activity can be indicative of novel RNA features. The methods are implemented as background tools and do not interfere with the functionality of the display view. Each analytical method can be performed independently on loaded data sets. Results are provided either as tab-separated value files or General Feature Format (GFF version 3) formatted files. TraV's analytical methods work on a single nucleotide resolution by avoiding sliding window-based approaches. This allows TraV to make predictions accurate to a single base, giving the maximal possible precision of the analytical approaches like e.g. transcription start site (TSS) predictions. The comparison of RNA-Seq experiments from different conditions requires normalization of the mapped data. To achieve this goal, Mortazavi *et al.* [17] introduced the read oriented RPKM values. RPKMs represent the number of reads mapped for any transcriptionally active region normalized against the total number of reads per experimental condition. As TraV does not use single read information, RPKMs cannot be calculated. TraV instead uses nucleotide activities per kilobase of exon model per million mapped reads (NPKM) values (Equation 1 and [9]) to represent transcriptional activity of all identified regions of transcriptional activity in its analytical methods.

$$\textbf{NPKM value}: \; NKPKM(n,m) = 10^9 \frac{\sum_{i=1}^{m} f(i)}{\sum_{i=1}^{m} g(i)(m-n)} \qquad (1)$$

Where n and m are the start and stop of the region of interest, $f(i)$ is the base activity of base i on a specific strand and $g(i)$ is the sum of the activities of base i of positive and negative strands [9].

Implementation

TraV is implemented as a JAVA web application for Linux-based webservers capable of providing a java container. Runtime-critical and memory-limited procedures have been implemented in C++. For data storage and retrieval a PostgreSQL database (PostgreSQL version 8.4 or higher) is used. A dedicated user management has been established which consists of two administrative and one application level. The user management is implemented via a WWW interface with dedicated user accounts. TraV features three different levels of access: (i) the admin user who may create and delete user accounts and has the ability to create new projects, (ii) administrative users who may import, export and delete transcriptome data sets for their assigned projects and (iii) standard users who may view and analyze the transcriptome data sets of their projects. The webserver-based implementation with differing levels of access makes TraV a good solution for cooperative projects. TraV centralizes data storage and processing for workgroups and provides access to the server-based environment for workstations that do not provide the hardware to cope with the amount of data generated in whole genome transcriptome sequencing experiments. The secure user password restricted WWW access was designed to enable users to work from any reliable internet connection. TraV may use annotated genome information, either as EMBL- or as GenBank-formatted files, to generate the genomic context of transcriptionally active regions. It is possible to assign multiple EMBL or GenBank-formatted files to a single genome project, thus enabling projects that include strains with multiple replicons or draft genomes.

Graphical user interface. The display view represents the main working interface of TraV (Figure 1). The interface allows navigating, zooming, searching by specific genomic features, loading of additional transcriptome data sets for comparison (Figure 2) as well as accessing the analytical methods. All displayed plots are log scaled to enable the view of a wide range of possible transcriptional activities within the display. Strand-specificity is addressed by two different coverage plots: red and blue mark the transcriptional activity of the positive and negative strand, respectively, whereby positive and negative represent the orientation of the genome in relation to *dnaA*. TraV has been developed for RNA-Seq protocols which generate strand-specific data, thus the TraV default working mode is "strand-specific". However, data sets may not always contain strand-specific information. To enable the analysis of strand-unspecific data, a corresponding work mode has been implemented (Figure 3). Within the strand-unspecific mode, coverage counts from both strands are summed up. To gain the full value of the single nucleotide resolution of RNA-Seq data, TraV contains a magnification view, which displays sequence information in context to their transcriptional activity (Figure 4). For instance, this function proved suitable for manual promoter pattern searches guided by TSS. The graphical viewer of the transcriptome and the genomic features are interactive and display information on coverage and genome position upon mouse input events. All graphics comply with the Scalable Vector Graphics (SVG) standard. Should rasterized graphics be required, TraV can convert the SVG graphics to Portable Network Graphics (PNG) images. TraV supports user-provided annotations in the form of GFF3-formatted files. Each loaded annotation set is displayed by coloured arrows between the provided genome annotations. GFF-encoded loci may be used for feature-oriented navigation as well as input for annotation-dependent analytical methods. To enable comparisons of different experiments with different sequencing efficiencies within the interface, TraV contains a normalization method for data sets depending on their mapped read count. The normalization factor x is calculated specifically for each loaded dataset (Equation 2: Normalization factor: $x = \frac{m}{\max(m)}$). The factor x exclusively scales the graphical representation of activity plots to the data set with the highest read number. Currently, TraV does not offer a normalization method to account for sequencer bias such as Illumina read biases deriving from random hexamer priming, as described by Risso *et al.* [18] and Hansen *et al.* [19].

$$\textbf{Normalization factor}: \; x = \frac{m}{\max(m)} \qquad (2)$$

Where x is the normalization factor for the current dataset applied to each base activity, m is the number of mapped reads in the current dataset and max(m) is the greatest amount of mapped reads of all currently loaded data sets.

Analytical methods. TraV has been designed for the identification of regions of transcriptional activity such as 5′ and 3′ untranslated regions [20], transcriptional activities of non-annotated genome loci (which may encode sRNAs or hitherto unknown protein genes) and antisense transcripts. Furthermore, transcription pattern search for sharp increases in transcription activity can be used to identify transcription start sites. To represent transcriptional activities, TraV can calculate NPKM values for all identified regions as well as already annotated genomic or user-provided features.

TraV offers five analytical methods to accomplish the detection of RNA (Figure 5). Since some analytic procedures can consume a considerable amount of computing power, it is possible to choose which data sets should be included for further analysis. In case multiple RNA-Seq data sets are selected, one merged dataset accumulating all available data is generated from these data sets. TraV adds the base transcriptional activities of each selected dataset for each base position to create the merged dataset. This approach allows replicates to close gaps in the sequencing coverage and therefore improves the accuracy of the predictions. Merging data sets originating from different growth phases or environmental conditions enables a combined search for overlapping as well as for differentially expressed features. However, this approach comes with the potential drawback that features like alternative transcription start sites or alternative termination sites become less obvious. Therefore, in cases where the merging of the data sets might obfuscate interesting features, only single data sets should be used for analysis. Prior to merging different data sets, TraV performs a noise filtering on each dataset by removing singularly mapped reads which never intersect with at least one other mapped read from the dataset. This is accomplished by searching for regions of transcriptional activity going from zero to one and back to zero activity without ever encountering an activity greater than one. TraV does not check for a minimum or maximum length of such a region since the different sequencing methods are producing reads of varying length. Subsequently, TraV performs the described merging process and applies the selected analytical method to the merged dataset. For each region identified by an analytical method, TraV calculates the NPKMs from the original data sets instead of the merged set. This allows a comparison of the transcriptional activity of a region between the different data sets. All analytical methods, with the exception of "Transcription Start Site prediction" can use either the replicon provided annotations or user-provided annotations in form of imported GFFs.

(i) **Calculate NPKM Values** determines NPKMs for annotated regions, which may also include any GFF-definable features (i.e. genes, operons or phages). NPKMs enable impromptu comparisons of the transcriptional expression strength of features. These comparisons may indicate candidate genes for more sophisticated, statistical methods like e.g. baySeq [21].

(ii) **3′ and 5′UTR Search** identifies regions of uninterrupted transcriptional activity entering or leaving an annotation. The algorithm searches regions of transcriptional activity flanking annotated features. The start and stop of transcriptional activity has to lie outside of an annotated feature on the same strand.

(iii) **Free Transcript Search** scans the transcriptome for regions of transcriptional activity that do not intersect with any annotated genomic features on either positive or negative strand. Therefore this method is suitable to scan for genes which might correspond to not yet annotated sRNAs or protein-coding genes.

(iv) **Antisense Transcript Search** is functionally similar to "Free Transcript Search". In contrast to the constraints of a free transcript, an antisense transcript requires an annotation on the opposite strand of its genomic location. Since the method requires strand-specific transcriptome data, it is unavailable in TraV's strand-unspecific mode.

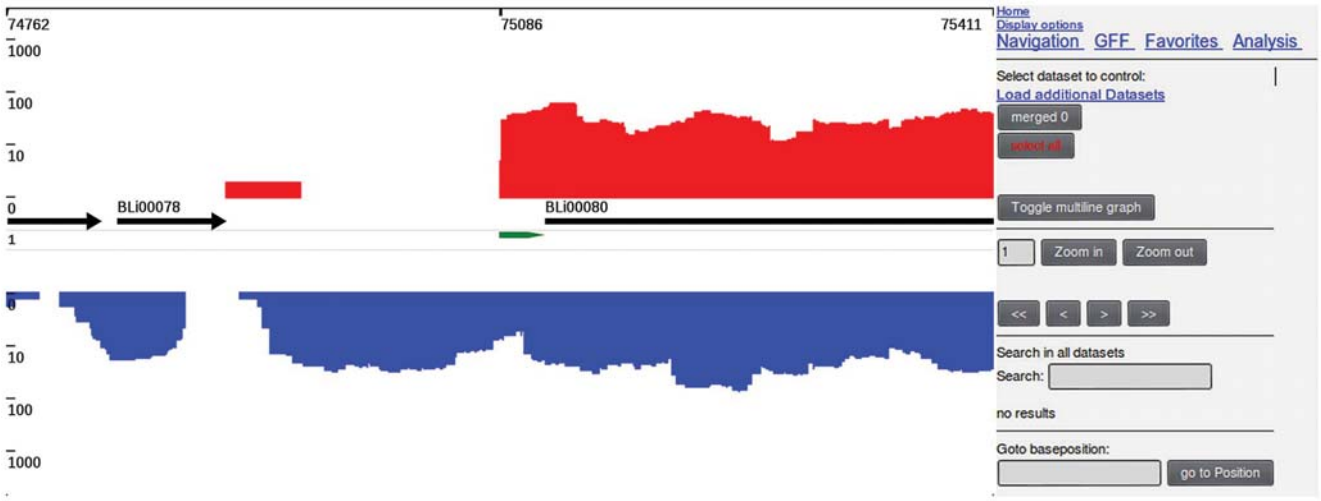


Figure 1. Main display of TraV. Transcriptional activity is depicted by graphs (red graph representing positive strand, blue representing negative strand). Annotated genome information is represented by black arrows. User-provided GFF-based annotations are represented in tracks between the original genome information by coloured arrows (green arrow in this figure).
doi:10.1371/journal.pone.0093677.g001

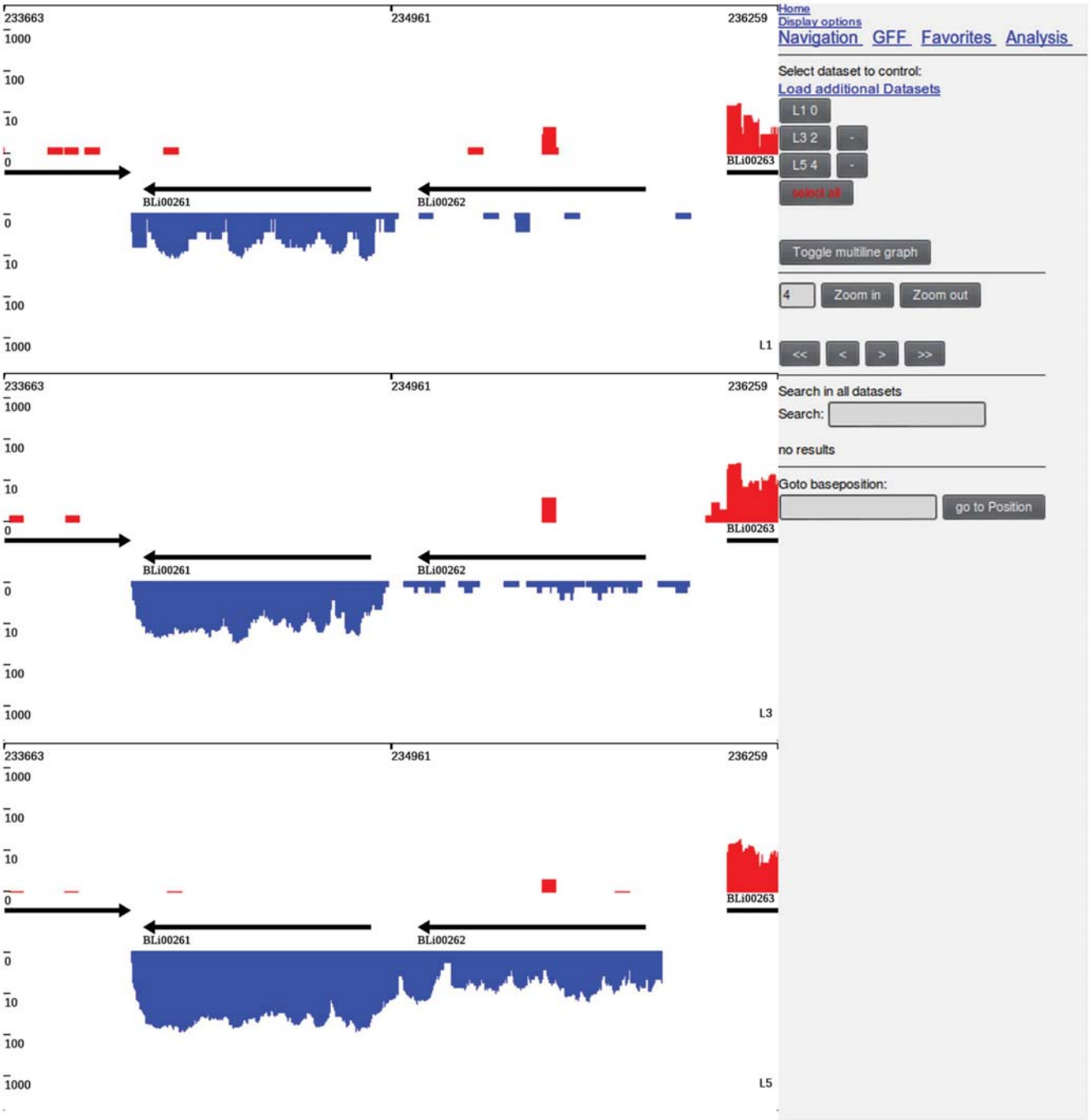


Figure 2. Differential expression. Differential expression of genes BLi00261 and BLi00262 at selected time points of a fermentative process. Dataset L-I shows the transcriptional activity of *B. licheniformis* DSM13 at the early exponential growth phase, L-III shows the transcriptional activity at the end of the exponential growth phase and L-V shows the transcriptional activity at the late stationary phase.
doi:10.1371/journal.pone.0093677.g002

(v) **Transcription Start Site Prediction** predicts candidates for TSS based on the identification of strong increases in transcriptional activity over a very short distance. This is done by calculating the slope of the transcriptional activity graph at any given position of the genome. This method is controlled by two values: the minimal height of the increase in transcriptional activity that constitutes the minimum height to define a positive hit and the slope distance which defines how many bases the slope should include. TraV offers a function to suggest an appropriate slope for a dataset. This function does a 5′UTR search and checks the distance necessary to obtain a positive hit with the current minimum height. The suggested standard value for the minimal height of three proved to be a good compromise between accuracy and noise in our test cases. Obviously, this method is dependent on the coverage of the sequencing experiment and insufficient coverage information may lead to gaps in

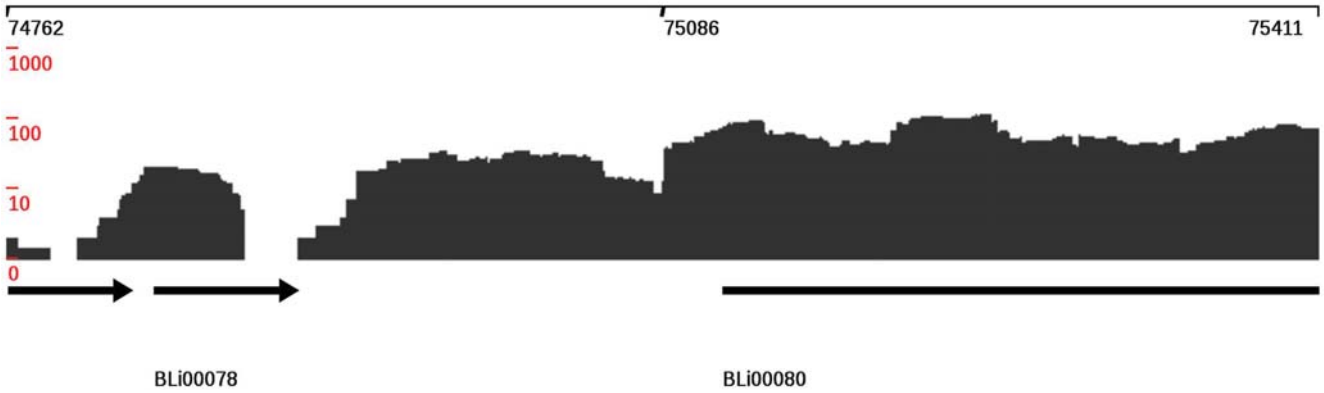


Figure 3. Strand unspecific mode of TraV. In case that the available mapped read data does not contain strand specificity information TraV can be switched into the strand unspecific mode in which all mapped reads are summed up to an general activity of a genome locus.
doi:10.1371/journal.pone.0093677.g003

transcript coverage to be misinterpreted as TSS candidates.

Please note that, depending on operon structure and regulation of transcription units, the different analytical methods may identify regions that belong to another feature class. For instance, the "Free Transcript Search" method may identify features that are actually *cis*-regulatory elements inhibiting downstream transcription which are therefore not connected to the respective gene. Therefore each region identified by the analytical methods has to be considered as a candidate for a RNA feature and the biological relevance has to be evaluated carefully.

TDS format and SAMtoTDS converter. TraV employs the generic TDS flat file data format to describe and save transcriptome mappings. The format consists of the following three sections: (i) a count data section which stores the number of mapped and unmapped reads and the total activity count of the RNA-Seq experiment. The mapped reads count is used by TraV for normalization of the graphical representation of different RNA-Seq experiments. The total activity count is used for the calculation of NPKM values. (ii) A coverage section, introduced by the string \\COV. In this section, the mapping is represented by transcriptional activity values per base. Every line in this section represents the coverage data of one base. (iii) An optional \\READS section can be used to provide the single read mapping information. TraV does not use single read mappings, however, as downstream analysis tools like mpileup [16] or GATK [22] require single read mapping information, this data can be included. TraV provides a command line conversion tool for SAM-formatted [16] mappings to generate its TDS import data format. This tool uses the CIGAR and bit flag information of the SAM format to determine successful read mappings. SAMtoTDS can use the CIGAR string information to apply a sequence similarity filter. This function is essential since some mappers do not apply a whole sequence length similarity filter. SAMtoTDS has been successfully used on SAM files generated by bowtie2, BWA and SSAHA2. This conversion step from SAM to TDS is done outside of TraV to conserve bandwidth of the server as the resulting TDS files are much smaller than the original SAM files.

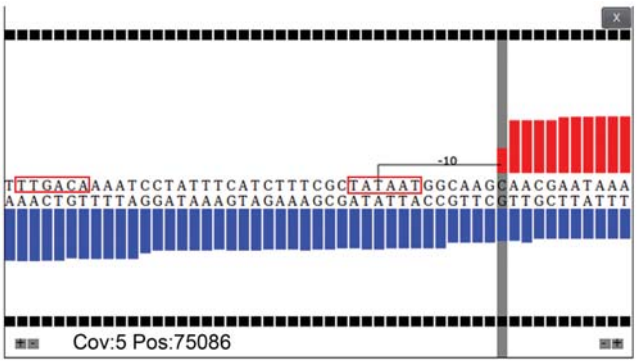


Figure 4. Magnification view of TraV. Transcriptional activities are displayed at single nucleotide resolution. The interactive graphic allows access to coverage information as well as to distinct base distances between the selected base (grey marker) and any other base of the displayed sequence. In the example perfect −10 and −35 boxes of a sigma70 promoter are visible.
doi:10.1371/journal.pone.0093677.g004

Results and Discussion

TraV was developed and extensively tested for the analysis of RNA-Seq experiments on *Bacillus licheniformis* DSM13 during an industrial fermentation process [9]. The tested RNA-Seq data was sequenced using an Illumina HiSeq 2000 machine with a read length of 50 nucleotides and is available in the Sequence Read Archive (SRA) under accession no. SRP018744. Mapping was performed using the Wurtzel *et al.* [4] BLAST-based method with a minimum sequence similarity of 98% for the complete read length. TraV was furthermore used to successfully visualize eukaryotic RNA-Seq data from *Schizosaccharomyces pombe* (SRA accession no. PRJEB3065). Please note: the analytical methods are currently only applicable to prokaryotic data since TraV does not yet have methods to deal with splicing events.

Transcriptionally active region evaluation

Predicted UTR and Free/Antisense transcript candidates were checked with the Rfam [23] covariance models for known structured RNAs. The results of this analysis are presented in Table 1. These predictions were used as foundation for a detailed manual curation and characterization process of regulatory RNA candidates [9]. Among the identified candidates are expected RNA genes like the 6S RNAs, the tmRNA as well as orthologous riboswitches described in the Rfam database. However, beside the elements with an Rfam assignment, a great number of transcribed RNA features without known function with a size greater than 100 nt have been identified which may contain interesting candidates for novel regulatory RNA features.

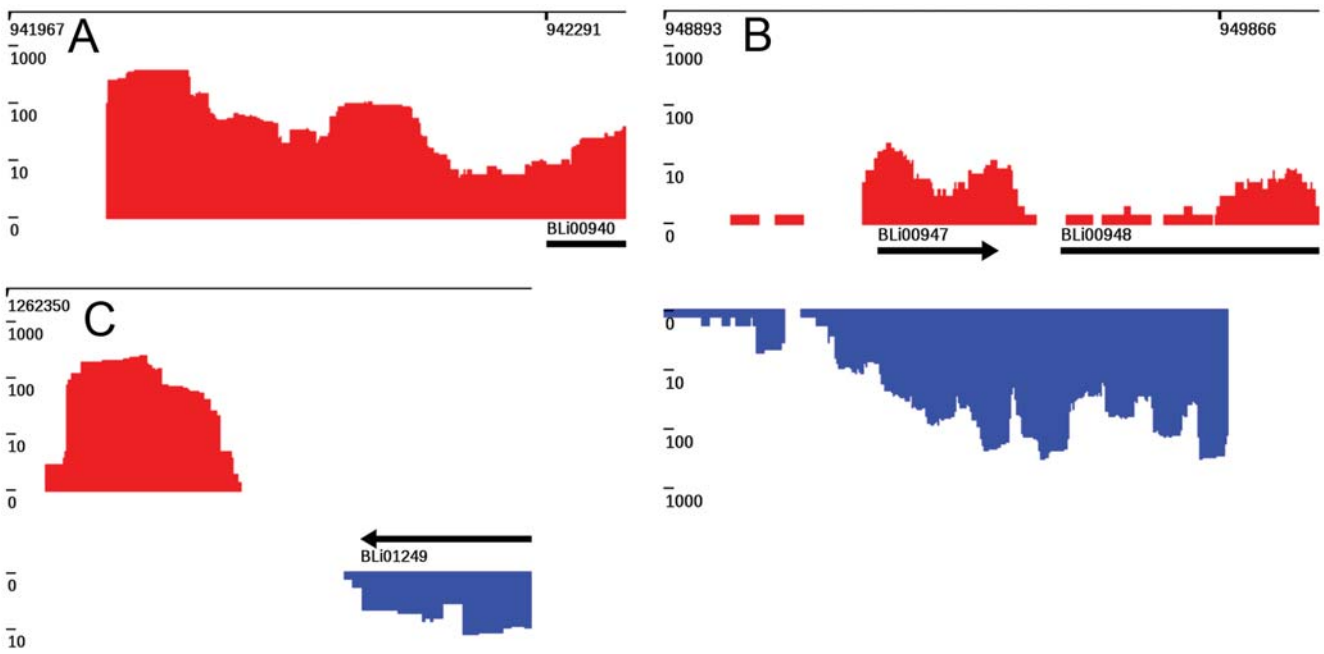


Figure 5. Patterns visualized by TraV. (**A**) 5′UTR of annotated gene feature. (**B**) Antisense transcript. (**C**) Transcriptional activity at a non-annotated region (free transcripts). A, B and C also show strong increases in transcriptional activity, which TraV predicts as transcription start sites. All graphs shown represent data from RNA-Seq experiments performed on *B. licheniformis* DSM13 [9] (indicated by red boxes). All graphs shown represent data from RNA-Seq experiments performed on *B. licheniformis* DSM13 [9].
doi:10.1371/journal.pone.0093677.g005

Visualization and Memory Management

Tools like T-ACE [24], Tablet [25] and SAMSCOPE [14] use indexed single reads and display the mapping information based on these single reads. Tools like T-ACE [24] suggest upper limits to their input data due to these limitations (60,000 bp contig size with a suggested maximum of 100,000 mapped reads). This reduces their usability with increasing contig sizes commonly used in whole genome RNA-Seq experiments. With the application of high coverage generating NGS techniques they are inevitably challenged by memory consumption issues. Programs like Tablet, SAMSCOPE or Artemis circumvent this issue by using the indexing information of BAM [16] to stream the loading and unloading of read information based on the displayed genomic region. This approach reduces the memory footprint but causes computational overhead due to the streamed loading mechanism leading to bad responsiveness when dealing with multiple datasets at the same time and to possible memory limitations at big window sizes. In contrast, TraV loads the whole data set and keeps it available in memory; therefore the tool does not incur stream loading congestions. The memory requirements on our test data resulted in approximately 50 Mbyte per data set with TraV when loaded in the working environment. In our tests TraV still works with 15 simultaneously loaded data sets in a comparative analysis on a workstation with 6 Gbyte RAM. Since TraV sacrifices single read information, it is unsuitable for data analysis dependent on information derived from base differences of single reads like single nucleotide polymorphism (SNP) studies, phase variation analysis or methylome studies by bisulfate sequencing.

Transcription start site prediction

RNA-Seq-based transcription start site prediction has been shown to be efficient by Sharma *et al.* [26,27]. The method described therein generates region lists within a 500 nt sliding window based on the comparison of transcriptional activities from corresponding RNA-Seq and differential RNA-Seq data sets, which were then manually evaluated. Differential RNA-Seq is a specialized version of RNA-Seq which is designed to identify microbial transcriptions start sites [28]. It is based on a selective digestion of all transcripts which do not represent primary transcripts. In contrast, the TSS prediction by TraV generates TSS candidate lists of single nucleotide resolution on single data sets from high coverage RNA-Seq or preferably differential RNA-Seq data sets, which have to be evaluated by a biological expert. Furthermore, the TraV approach is not dependent on the existence of corresponding RNA-Seq and differential RNA-Seq data sets and can thus be applied to a single deep sequencing

Table 1. Identified UTRs and Free Transcripts in *B. licheniformis* DSM13.

	3′UTR	5′UTR	Free Transcripts	Antisense Transcripts
Identified with TraV	1396	1404	476	3777
Candidates >100 nt	581	446	124	1933
Candidates with Rfam hits	78	27	10	212

doi:10.1371/journal.pone.0093677.t001

experiment. However, TraV's TSS prediction is directly dependent on the coverage; therefore an application of the data of Sharma *et al.* resulted in an accelerated error rate of artefact TSSs due to mapping gaps within transcriptionally active regions. Therefore TraV is not applicable for data sets with an insufficient coverage.

Conclusion

The hereby introduced RNA-Seq analysis software package TraV includes an import function for mapping data from SAM-formatted mappings, five prediction tools to identify RNA features and a memory efficient transcriptome visualization engine addressing the typical work flow of a comparative RNA-Seq analysis. TraV has been successfully applied to Illumina generated data sets consisting of fifteen RNA-Seq and five differential RNA-Seq records [9]. The analytical tools of TraV identified previously missing protein genes and RNA-based regulatory features like e.g. riboswitches or regulatory RNAs. TraV is inapplicable for single read based data as necessary in SNP analysis, phase variations or bisulfate sequencing. TraV is positioned as a tool focused on comparative high coverage transcriptome mappings on well-polished and annotated microbial genomes. TraV operates simultaneously on multiple RNA-Seq mappings and provides analysis functions that focus on whole contigs rather than limited areas. The program's primary purpose is the global identification of RNA-based regulators within a genome and providing researchers with automatically generated candidate lists and interactive evaluation tools.

Acknowledgments

The authors would like to thank Stephan Waack for expert review and valuable discussion as well as Rolf Daniel for valuable discussion and hosting our group at the Georg-August University. We thank Stefania Neuber for intensive proof reading and substantial improvement of the manuscript. We are grateful for extensive testing and valuable feedback by Anja Poehlein.

Author Contributions

Conceived and designed the experiments: HL SD. Performed the experiments: SD SW. Analyzed the data: SD SW HL. Contributed reagents/materials/analysis tools: SD SW. Wrote the paper: SD SW HL.

References

1. Malone JH, Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 9: 34. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 3104486&tool = pmcentrez&rendertype = abstract. Accessed 22 July 2012.
2. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE (2011) Landscape of next-generation sequencing technologies. Anal Chem 83: 4327–4341. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 3437308&tool = pmcentrez&rendertype = abstract. Accessed 24 July 2012.
3. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57–63. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2949280&tool = pmcentrez&rendertype = abstract. Accessed 13 July 2012.
4. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons B, et al. (2010) A single-base resolution map of an archaeal transcriptome. Genome Res 20: 133–141. Available: http://www.ncbi.nlm.nih.gov/pubmed/19884261.
5. De Lay N, Schu DJ, Gottesman S (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. J Biol Chem 288: 7996–8003. Available: http://www.ncbi.nlm.nih.gov/pubmed/23362267. Accessed 13 August 2013.
6. Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152: 17–24. Available: http://www.ncbi.nlm.nih.gov/pubmed/23332744. Accessed 15 August 2013.
7. Filiatrault MJ, Stodghill PV, Myers CR, Bronstein PA, Butcher BG, et al. (2011) Genome-wide identification of transcriptional start sites in the plant pathogen Pseudomonas syringae pv. tomato str. DC3000. PLoS One 6: e29335. Available: http://dx.plos.org/10.1371/journal.pone.0029335. Accessed 23 May 2012.
8. Wang Y, Li X, Mao Y, Blaschek HP (2012) Genome-wide dynamic transcriptional profiling in Clostridium beijerinckii NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics 13: 102. Available: http://www.biomedcentral.com/1471-2164/13/102. Accessed 12 March 2013.
9. Wiegand S, Dietrich S, Hertel R, Bongaerts J, Evers S, et al. (2013) RNA-Seq of Bacillus licheniformis: active regulatory RNA features expressed within a productive fermentation. BMC Genomics 14: 667. Available: http://www.biomedcentral.com/1471-2164/14/667. Accessed 1 October 2013.
10. Ning Z, Cox AJ, Mullikin JC (2001) SSAHA: a fast search method for large DNA databases. Genome Res 11: 1725–1729. Available: http://genome.cshlp.org/content/11/10/1725. Accessed 29 January 2013.
11. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359. Available: http://www.ncbi.nlm.nih.gov/pubmed/22388286. Accessed 4 October 2012.
12. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760. Available: http://bioinformatics.oxfordjournals.org/content/25/14/1754.long. Accessed 28 January 2013.
13. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28: 464–469. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 3278759&tool = pmcentrez&rendertype = abstract. Accessed 23 September 2013.
14. Popendorf K, Sakakibara Y (2012) SAMSCOPE: An OpenGL based real-time interactive scale-free SAM viewer. Bioinformatics: bts122–. doi:10.1093/bioinformatics/bts122.
15. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14: 178–192. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 3603213&tool = pmcentrez&rendertype = abstract. Accessed 17 September 2013.
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. Available: http://bioinformatics.oxfordjournals.org/content/25/16/2078.long. Accessed 28 January 2013.
17. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621–628. Available: http://dx.doi.org/10.1038/nmeth.1226. Accessed 10 July 2011.
18. Risso D, Schwartz K, Sherlock G, Dudoit S (2011) GC-Content Normalization for RNA-Seq Data. BMC Bioinformatics 12: 480. Available: http://www.biomedcentral.com/1471-2105/12/480. Accessed 19 December 2011.
19. Hansen KD, Brenner SE, Dudoit S (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res 38: e131. Available: http://nar.oxfordjournals.org/content/38/12/e131.long. Accessed 4 March 2013.
20. Breaker RR (2011) Prospects for riboswitch discovery and analysis. Mol Cell 43: 867–879. Available: http://www.ncbi.nlm.nih.gov/pubmed/21925376. Accessed 13 August 2013.
21. Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11: 422. Available: http://www.biomedcentral.com/1471-2105/11/422. Accessed 8 March 2012.
22. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297–1303. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2928508&tool = pmcentrez&rendertype = abstract. Accessed 17 September 2013.
23. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, et al. (2005) Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 33: D121–4. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 540035&tool = pmcentrez&rendertype = abstract. Accessed 9 August 2012.
24. Philipp EER, Kraemer L, Mountfort D, Schilhabel M, Schreiber S, et al. (2012) The Transcriptome Analysis and Comparison Explorer - T-ACE: a platform-independent, graphical tool to process large RNAseq data sets of non-model organisms. Bioinformatics 28: 777–783. Available: http://bioinformatics.oxfordjournals.org/cgi/content/abstract/28/6/777. Accessed 19 March 2012.
25. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, et al. (2012) Using Tablet for visual exploration of second-generation sequencing data. Brief Bioinform: bbs012–. Available: http://bib.oxfordjournals.org/cgi/content/abstract/bbs012v1. Accessed 25 March 2012.
26. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, et al. (2010) The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464: 250–255. Available: http://dx.doi.org/10.1038/nature08756. Accessed 31 October 2012.
27. Schmidtke C, Findeiss S, Sharma CM, Kuhfuss J, Hoffmann S, et al. (2012) Genome-wide transcriptome analysis of the plant pathogen Xanthomonas identifies sRNAs with putative virulence functions. Nucleic Acids Res 40: 2020–

2031. Available: http://nar.oxfordjournals.org/content/40/5/2020.long. Accessed 14 November 2012.
28. Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, et al. (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803. PNAS 108: 1–6. doi:10.1073/pnas.1015154108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1015154108.

CHAPTER E

COMPLETE GENOME SEQUENCE OF *GEOBACILLUS* SP. GHH01, A THERMOPHILIC LIPASE-SECRETING BACTERIUM

Sandra Wiegand, Ulrich Rabausch, Jennifer Chow,
Rolf Daniel, Wolfgang R. Streit and Heiko Liesegang

Genome Announcements (2013), Vol. 1, Issue 2, e00092-13

Authors' contributions to this work

Isolated strain and performed experiments: UR, JC
Performed sequencing and annotation: SW
Wrote the paper: SW
Conceived and designed the experiments: HL, WRS

genomeA Journals.ASM.org

Complete Genome Sequence of *Geobacillus* sp. Strain GHH01, a Thermophilic Lipase-Secreting Bacterium

Sandra Wiegand,[a] Ulrich Rabausch,[b] Jennifer Chow,[b] Rolf Daniel,[a] Wolfgang R. Streit,[b] Heiko Liesegang[a]

Department of Genomic and Applied Microbiology, Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität, Göttingen, Germany[a]; Abteilung für Mikrobiologie und Biotechnologie, Biozentrum Klein Flottbek, Universität Hamburg, Hamburg, Germany[b]

***Geobacillus* sp. strain GHH01 was isolated during a screening for producers of extracellular thermostable lipases. The completely sequenced and annotated 3.6-Mb genome encodes 3,478 proteins. The strain is genetically equipped to utilize a broad range of different substrates and might develop natural competence.**

Received 8 February 2013 **Accepted** 13 March 2013 **Published** 25 April 2013

Citation Wiegand S, Rabausch U, Chow J, Daniel R, Streit WR, Liesegang H. 2013. Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. Genome Announc. 1(2):e00092-13. doi:10.1128/genomeA.00092-13.

Copyright © 2013 Wiegand et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Heiko Liesegang, hlieseg@gwdg.de.

The genus *Geobacillus* contains thermophilic strains, which produce a variety of thermostable hydrolytic extracellular enzymes, such as proteases, amylases, and lipases. These features are interesting for future production platforms used in industrial applications (1).

Here, we present the complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase producer. The strain was isolated from an enrichment culture originally sampled at Botanischer Garten, University of Hamburg, Germany, and was cultivated at 60°C with 1.5% native olive oil as the sole carbon source. Recombinant expression in *Escherichia coli* revealed that the *Geobacillus* sp. GHH01 lipase (locus tag GHH_c20570) is highly active but only moderately thermostable.

The genome sequence of *Geobacillus* sp. GHH01 was determined by a combined approach of 454 GS-FLX Titanium XL paired-end sequencing (454 Life Sciences, Branford, CT) and Genome Analyzer II single-read sequencing (TruSeq Chemistry, Illumina, San Diego, CA), resulting in average coverages of 10.91-fold and 33.03-fold, respectively. The assembly employing the MIRA v3.4.1.1 software (2) yielded 84 contigs >3 kbp. Gap closure and quality improvement were performed by PCR-based techniques and subsequent Sanger sequencing (ABI 3730xl, Life Technologies, Carlsbad, CA). Initial gene prediction was performed with IMG/ER (3), followed by manual curation based on comparisons to the Swiss-Prot, TrEMBL (4), and InterPro (5) databases. For the identification of rRNA and tRNA genes, RNAmmer v1.2 and tRNAscan-SE v1.4 (6, 7) were used, respectively.

The complete genome consists of a 3,582,992-bp chromosome with a G+C content of 52.3%. In total, 3,597 genes were identified, including 10 rRNA gene clusters and 88 tRNA genes. The annotation resulted in 2,724 protein-encoding genes with assigned functions.

16S rRNA gene phylogenetic analysis confirmed the affiliation of *Geobacillus* sp. GHH01 to the genus *Geobacillus*, whereas an assignment to a described species was not possible. We determined average nucleotide identities (8) of approximately 96% between the *Geobacillus* sp. GHH01 genome and the genomes of *Geobacillus kaustophilus* HTA426 (9) and *Geobacillus thermoleovorans* CCB_US3_UF5 (10). The recently mentioned (10) high synteny between *G. thermoleovorans* CCB_US3_UF5 and *G. kaustophilus* HTA426 (97.94%) calls into question their assignment to distinct species. Hence, a sequence similarity-based assignment of *Geobacillus* sp. GHH01 to a distinct species could not be employed.

Geobacillus sp. GHH01 is predicted to secrete 139 enzymes by the Sec-dependent pathway (11, 12), including the identified lipase, diverse peptidases, proteinases, an amylopullanase (GHH_c32620), an alpha-amylase (GHH_c32630), and an alkaline phosphatase (GHH_c27900). Several substrate-binding proteins of ABC transporters indicate the potential for utilization of a broad range of substrates. The ability to take up extracellular DNA is a crucial mechanism for strain development. Eighteen out of 25 main competence-related structural genes identified for *Bacillus subtilis* (13) were detected, featuring a possible mechanism of DNA uptake.

Genome comparisons revealed seven distinct GHH01-specific genomic islands (14). Furthermore, 123 putative transposases, five clustered regularly interspaced short palindromic repeat (CRISPR) regions, and nine CRISPR-associated genes of subtype III-B (15) could be detected.

Nucleotide sequence accession number. The genome sequence of *Geobacillus* sp. GHH01 has been deposited in GenBank under accession no. CP004008. The strain is available upon request at the Bacillus Genetic Stock Center (BGSC).

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) within the network grants GenoMik-Plus and GenoMik-Transfer.

We thank Anja Poehlein for support during sequencing, annotation, and submission and Stefanie Offschanka for technical assistance.

REFERENCES

1. **McMullan G, Christie JM, Rahman TJ, Banat IM, Ternan NG, Marchant R.** 2004. Habitat, applications and genomics of the aerobic, thermophilic genus *Geobacillus*. Biochem. Soc. Trans. **32**:214–217.
2. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer Science and Biology: Proceedings of the German Conference on Bioinformatics. CiteSeer, Pennsylvania State University, University Park, PA.
3. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.
4. **UniProt Consortium.** 2010. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res. **37**:D169–D174.
5. **Zdobnov EM, Apweiler R.** 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**: 847–848.
6. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
8. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. **106**: 19126–19131.
9. **Takami H, Horikoshi K.** 2000. Analysis of the genome of an alkaliphilic *Bacillus* strain from an industrial point of view. Extremophiles **4**:99–108.
10. **Muhd Sakaff MKL, Abdul Rahman AY, Saito JA, Hou S, Alam M.** 2012. Complete genome sequence of the thermophilic bacterium *Geobacillus thermoleovorans* CCB_US3_UF5. J. Bacteriol. **194**:1239.
11. **Petersen TN, Brunak S, Von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.
12. **Sonnhammer EL, Von Heijne G, Krogh A.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6**:175–182.
13. **Kovács AT, Smits WK, Mirończuk AM, Kuipers OP.** 2009. Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries. Environ. Microbiol. **11**:1911–1922.
14. **Langille MG, Brinkman FS.** 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics **25**:664–665.
15. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, Van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. **9**:467–477.

CHAPTER F

DISCUSSION AND OUTLOOK

F.1 Transcriptome complexity

The advent of genome-wide transcriptome analysis techniques like RNA-Seq and tiling arrays has revealed many unexpected findings, including a plethora of antisense and small RNAs, versatile untranslated regions and alternative operon structures (Sorek and Cossart, 2010). With the exception of sRNAs, which were thoroughly discussed in Chapter B, this unforeseen, domain-spanning transcriptome complexity will be presented in this chapter.

F.1.1 Untranslated regions

The RNA-Seq-based analysis of 5'UTRs in *B. licheniformis* revealed an unexpected size range of the leader regions. For example, the 5'UTR of the sporulation inhibitor-encoding gene *kapD* can reach 2226 nt in length. Although this 5'UTR is exceptionally long, still 4% of all detected leader regions are longer than 400 nt, as shown in Figure 1. Altogether, 33% of the detected 5'UTRs are more than 100 nt in size and might therefore bear a regulatory function (Güell et al., 2011; Chapter B). A GO term-based enrichment analysis of the genes downstream of 5'UTRs >400 nt revealed an enrichment of cation-binding protein genes, which might indicate yet unknown *cis*-regulatory RNAs in this region. However, in *Photobacterium profundum*, a significant enrichment of genes involved in energy, coenzyme and lipid metabolism has been observed, suggesting that some classes of genes might preferably be regulated by mechanisms intrinsic to long 5'UTRs (Campanaro et al., 2012). Comparisons of 5'UTR lengths in different species (Figure 1), with special consideration of the respective median values, point out that 5'UTR sizes are not only very diverse within one organism, but are also highly variable between species. Especially the two archaea *Methanosarcina mazei* and *Sulfolobus solfataricus* have either very long or very short 5'UTRs,

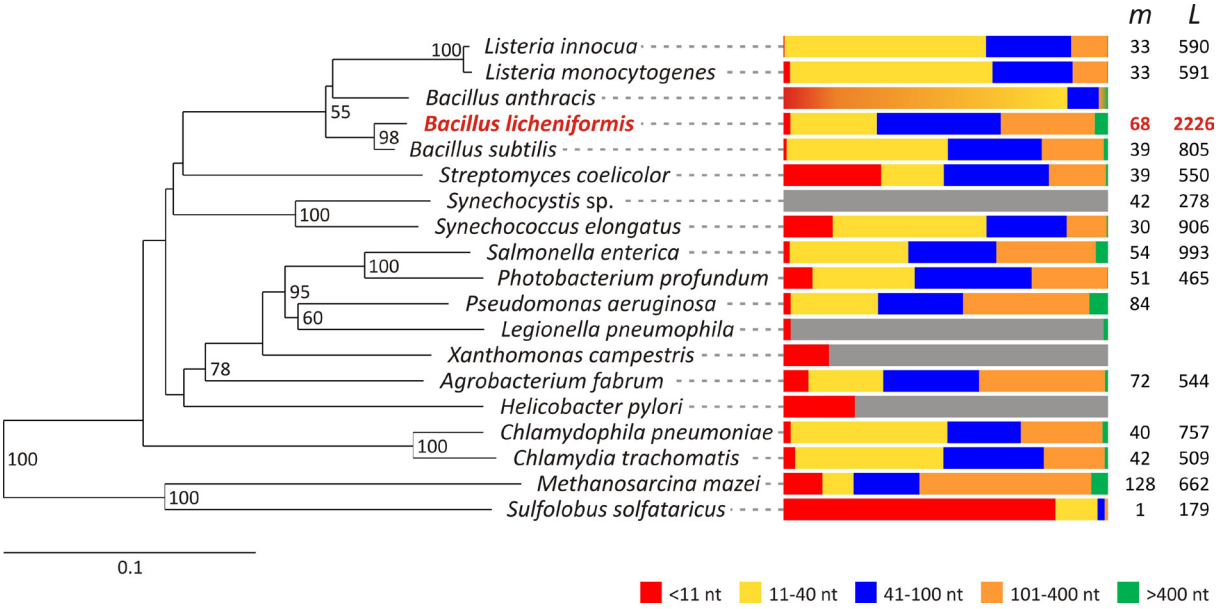


Figure 1 Distribution of 5'UTR lengths
The 16S rRNA gene sequence neighbor-joining tree was calculated with ARB (Ludwig et al., 2004) for those organisms with available 5'UTR length data. A detailed register of strain designations and corresponding studies is given in Table S1. *B. licheniformis* DSM13, examined in this study, is indicated in red. Only bootstrap values above 50% are given at the respective nodes.
Colored bars indicate the percentage of identified 5'UTRs with a certain length, as outlined in the legend. On the right hand side, the median (*m*) of the determined 5'UTR lengths and the longest identified 5'UTR length (*L*) are given.

respectively. It has been shown that bacteria have widely different average intergenic region sizes (Molina and van Nimwegen, 2008). These appear to correlate slightly with the lengths of 5'UTRs, supporting the hypothesis that large intergenic regions have a high regulatory potential (Campanaro et al., 2012). When comparing the varying 5'UTR lengths of closely related species like *B. subtilis* and *B. licheniformis*, where high similarities are expected (Wurtzel et al., 2012), it becomes obvious that the different approaches to determine the sizes of 5'UTR bear a high potential for methodical biases. Hence, comparative analyses not relying on data generated with a comparable method are prone to errors and should be considered with care.

To date, 65 *cis*-regulatory elements, comprising mainly riboswitches and T-boxes, were identified in *B. licheniformis* by covariance models (Burge et al., 2013; Markowitz et al., 2012). All of them showed transcriptional activity during the monitored fermentation process (Chapter B). Analyses of the frequency of common *cis*-regulatory elements in different bacterial species (Figure 2) show that *cis*-regulatory elements are evenly distributed within the *Bacillus subtilis* as well as within the *Bacillus cereus* group, whereas the latter appears to contain considerably more of these elements. It is noteworthy that riboswitches and T-boxes

seem to be highly abundant in *Firmicutes*. It has been proposed previously that members of this phylum seem to make more extensive use of riboswitches and T-boxes than other bacterial phyla, Archaea or Eukarya (Barrick and Breaker, 2007; Vitreschak et al., 2008). Interestingly, riboswitches tend to affect transcription in Gram-positive bacteria whereas translation is more frequently regulated in Gram-negative bacteria (Nudler and Mironov, 2004). This effect might be due to the fact that Gram-positive bacteria contain larger biosynthetic operons, hence more resources can be saved by tight transcriptional regulation (Nudler and Mironov, 2004; Waters and Storz, 2009).

Knowledge of the length of 5'UTRs can be helpful for the prediction of unidentified *cis*-regulatory elements. For example, three previously undetected T-box structures could be identified in unexpectedly long 5'UTRs by covariance analyses in this study (Chapter B). *In silico* analyses for the prediction of yet unknown regulatory motifs can also benefit from this knowledge, as they presently often rely on assumed 5'UTR lengths (Bastet et al., 2011; Wein-

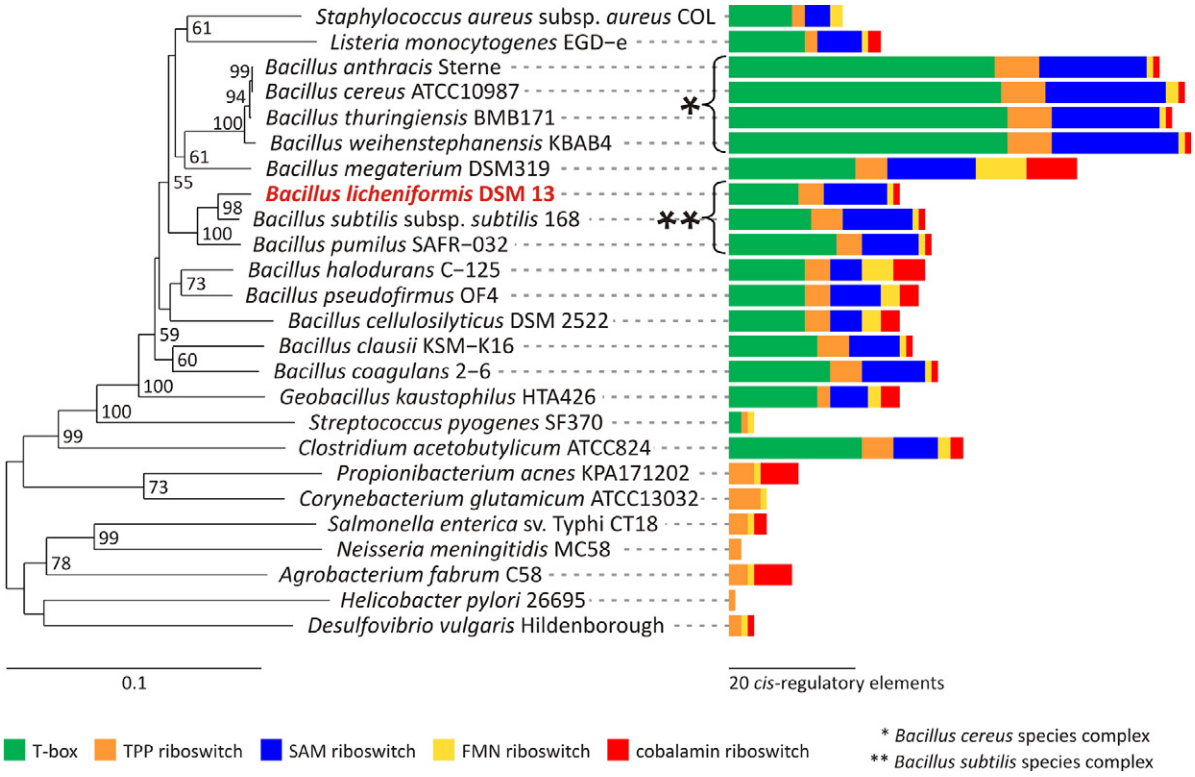


Figure 2 Distribution of five common *cis*-regulatory elements
The 16S rRNA gene sequence neighbor-joining tree was calculated with ARB (Ludwig et al., 2004) for selected organisms. *Sulfolobus solfataricus* P2 was used as outgroup. *B. licheniformis* DSM13, which was examined in this study, is indicated in red. Only bootstrap values above 50% are given at the respective nodes.
Colored bars depict the frequency of the following *cis*-regulatory elements, chosen due to their most common distribution within the bacterial domain (Barrick and Breaker, 2007), as outlined in the legend. The length of the bars corresponds to the number of Rfam predictions (Burge et al., 2013). A length scale for 20 *cis*-regulatory elements is given.

berg et al., 2007). Moreover, the identification of varying transcriptional activities of a 5'UTR and its following gene can help to elucidate putative regulatory events originating from regulatory elements in the untranslated region (personal communication, Robert Hertel).
Riboswitches do not only increase the complexity of the transcriptome by their presence, but also provide the opportunity to artificially increase the complexity by introducing (synthetic) riboswitch variants into the genome (Weigand and Suess, 2009). This approach has led to engineered riboswitches that allow conditional gene expression in bacteria (Topp et al., 2010) and might enable the fine-tuning of complex, metabolic engineered processes in future applications (Wittmann and Suess, 2012).

In general, translation initiation in prokaryotes is achieved by interaction of the Shine-Dalgarno (SD) sequence with a 16S rRNA, which triggers the anchoring of the 30S ribosomal subunit to the start codon. Nonetheless, two additional mechanisms of translation initiation not requiring a SD sequence were identified: one mechanism is dependent on the ribosomal protein S1 and at least a short 5'UTR (Chang et al., 2006; Malys and McCarthy, 2011); the other accomplishes the binding of the 70S ribosomal subunit directly to the start codon of mRNAs not possessing a 5'UTR (Brock et al., 2008; Giliberti et al., 2012; Moll et al., 2002). In *B. licheniformis*, 33 genes were found to be led by 5'UTRs with lengths between 0 and 10 nucleotides, which might be targets of SD-free, so-called leaderless translation initiation. RNA-Seq studies of other prokaryotes have assigned 1.3% to 81.4% of the determined 5'UTRs to be leaderless (Figure 1). This broad range of leaderless transcription fits the results of genome-based predictions of SD-free transcription by Zheng et al. (2011), who have predicted over 70% leaderless genes in some archaea. On the other hand, they found *α*-, *γ*- and *ε-Proteobacteria* to be amongst the bacteria with the lowest amount of leaderless transcripts. *Firmicutes* were calculated to harbor 2.4 to 16.6% leaderless mRNAs, which is close to the 2.1% determined for *B. licheniformis*. This rather low proportions of leaderless transcripts in *Bacilli* was also predicted by Nakagawa et al. (2010) by proposing a high amount of SD-containing genes in *Firmicutes*.

Next to 5'UTRs, high size variability was also shown for the generally less regarded 3'UTRs (Brenneis et al., 2007; Campanaro et al., 2012). In *B. licheniformis*, 1365 3'UTRs with a median length of 73 nt were determined. 19% of these 3'UTRs exceed a length of 400 nt and range up to a maximal length of 4508 nt (Chapter B).

Both 5'UTR as well as 3'UTR transcripts enhance the transcriptome complexity also by their extensive localization antisense to adjacent genes on the opposite strand (overlapping UTRs), which applies for 4% and 25% of the total 5' and 3'UTRs, respectively (Chapter B). In fact, 13 and 104 of these overlapping UTRs are longer than 1000 nt and span up to six genes. Examples for overlapping UTRs are also known from *B. subtilis* and *Anabaena* sp. PCC 7120 (Hernández et al., 2006; Nicolas et al., 2012; Rasmussen et al., 2009). The comparison of the overlapping 5'UTRs identified in *B. licheniformis* (53) to those identified by tiling arrays in *B. subtilis* (90; Nicolas et al., 2012), showed only 12 of these 5'UTRs to be length-conserved in both species. Furthermore, less than 6% of the overlapping 3'UTRs >1000 nt seem to be conserved. The variability among the detected UTRs can be partially explained by the different analytical methods, but manual inspection also gives the impression that overlapping UTRs might rather occur in the unconserved regions of the chromosome. Certainly, this impression will have to be evaluated by the comparison of related, uniformly analyzed genomes.

The phenomenon of overlapping bacterial UTRs has been analyzed in more detail in *Listeria monocytogenes* and *Listeria innocua* (Sesto et al., 2013; Toledo-Arana et al., 2009; Wurtzel et al., 2012). One example is the long 5'UTR of a conserved putative ABC transporter permease component operon, which overlaps a gene coding for a multidrug efflux pump (Wurtzel et al., 2012). Two promoters were identified for this operon. The promoter upstream of the long 5'UTR is dependent of the σ^B transcription factor: upon deletion of σ^B, the long 5'UTR shows no expression, but the overlapped gene is upregulated, confirming a negative effect of the antisense transcript (Wurtzel et al., 2012). It was also shown that the σ^B-dependent transcript contributes to the expression level of the downstream operon, in addition to the second, σ^B-independent promoter (Wurtzel et al., 2012). The finding of this and three further examples of adjacent genes with opposed or related function regulated by overlapping 5'UTRs in *L. monocytogenes* and *L. innocua* led to the introduction of the term *excludon* to describe this regulatory elements as expression-excluding operons (Sesto et al., 2013; Wurtzel et al., 2012). As excludon mRNAs are detected as multiple fragments with different sizes (Toledo-Arana et al., 2009; Wurtzel et al., 2012), it is likely that the mechanism behind these regulatory events relies on degradation of the double-stranded RNA resulting from hybridization of the sense and the antisense transcript (Sesto et al., 2013). In *B. licheniformis,* the transcriptional activity of the GltC regulon shows similarities to this regulatory mechanism (Chapter B): the gene of the associated transcriptional activator is encoded on the opposite strand of the glutamate synthase operon and overlapped by the long 5' UTR of the operon genes.

Taken together, these findings promote the idea that the arrangement of bacterial genes on the chromosome is not random (Lasa et al., 2012; Lawrence, 2003): in addition to the clustering of protein-encoding genes of the same metabolic pathway into operons, 5' and 3' overlapping UTRs seem to introduce a further regulatory level for the coordinated expression of adjacent genes (Lasa et al., 2012).

F.1.2 Non-coding antisense transcripts

Next to antisense transcripts deriving from overlapping 5' and 3' UTRs, many asRNAs are expressed independently from protein-coding transcripts. In *B. licheniformis*, 408 of such non-coding asRNAs were identified, comprising 89% of all identified non-coding RNA transcripts and 48% of all antisense transcripts (Chapter B). In total, 15% of all genes are co-located with an independently transcribed antisense transcript. In contrast to the occasional occurrence of bacterial antisense transcription anticipated in the past, it has now become clear that asRNAs are abundant in all investigated species (Sorek and Cossart, 2010) and are therefore expected to be ubiquitously distributed throughout the bacterial, but also the archaeal and eukaryotic domain (Georg and Hess, 2011; Raghavan et al., 2012). It has been

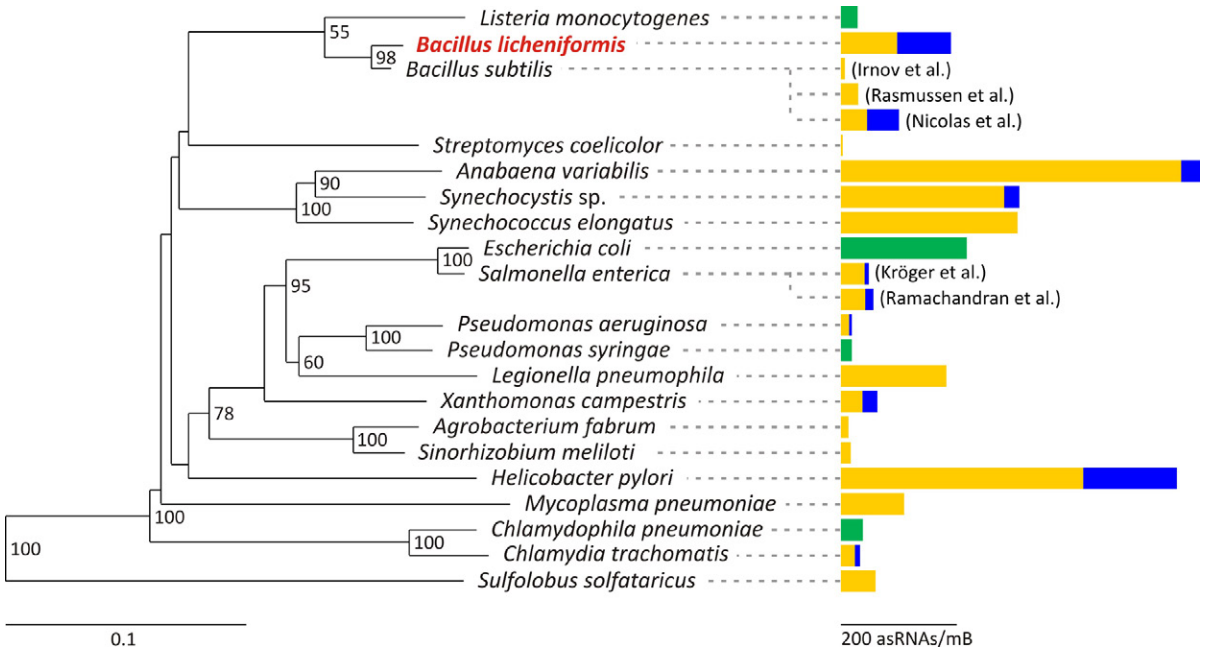


Figure 3 Distribution of antisense transcription
The 16S rRNA gene sequence neighbor-joining tree was calculated with ARB (Ludwig et al., 2004) for those bacteria and archaea with available antisense transcription data. A detailed register of strain designations and corresponding studies is given in Table S2. *B. licheniformis* DSM13, which was examined in this study, is indicated in red. Only bootstrap values above 50% are given at the respective nodes.
Colored bars indicate the numbers of identified antisense transcripts/mB: (*yellow*) non-coding asRNAs; (*blue*) antisense transcripts deriving from overlapping UTRs; (*green*) unspecified antisense transcripts. The length of the bars corresponds to the number of identified asRNAs. A length scale for 200 asRNAs is given.

speculated previously that the frequency of asRNAs depends on the taxonomic allocation of the investigated organism (Qiu et al., 2010). However, with the growing number of explored species (Figure 3), it is now assumed that antisense transcripts can be found for ~10-20% of the bacterial genes, and that smaller numbers observed are a result of low-sensitive detection methods (Güell et al., 2011). Figure 3 illustrates the differences of identified asRNAs for those prokaryotes with obtainable data sets. It becomes obvious that the so far available data, derived from different experimental approaches, are not readily comparable. This observation is especially pinpointed by the varying amounts of asRNAs identified in *B. subtilis*. However, high variability of independently transcribed asRNAs was suggested after comparison of in-depth analyses of *B. subtilis* and *B. licheniformis* (Chapter B). This raises the question, whether asRNAs are conserved between closely related species or, despite their costly production, derive from transcriptional noise due to spurious transcriptional events (Nicolas et al., 2012; Raghavan et al., 2012). A study by Raghavan et al. (2012) addressing this question in enteric bacteria revealed that only 14% of the detected asRNAs are abundant in *E. coli* as well as the related *Salmonella enterica* serovar Typhimurium. The authors showed furthermore, that promoters enabling antisense transcription do not seem to be under the same evolutionary pressure as promoters of protein-coding genes. Therefore, Raghavan et al. conclude that a large fraction of the observed antisense expression is non-functional and derives from promoter-like sequences under weak selection. This hypothesis is also prioritized by Nicolas et al. (2012), who found that many asRNAs are products of spurious transcription from alternative promoter sequences under low evolutionarily pressure.

Nevertheless, asRNAs with a clear impact on the sense transcript have been identified. In *B. subtilis*, the analysis of the toxin/antitoxin system of overlapping *bsrG*/SR4 showed severe growth defects of the organism upon depletion of the *sr4* gene (Jahn et al., 2012). This effect is prompted by the increased half-life of the *bsrG* mRNA when not degraded by RNases as a result of duplex formation with SR4 (Jahn et al., 2012). Two sense/antisense pairs encoding BsrG-like peptides and forming RNAs with SR4-like stem loops could be identified in *B. licheniformis* (Chapter B). An effect of an asRNA located strictly within the boundaries of the respective sense RNA was investigated in *Synechocystis* sp. PCC6803: the *isiA* gene encodes a protein involved in photosynthesis and is a member of the iron stress response regulon. The constitutively expressed IsiR asRNA is located antisense to this gene (Dühring et al., 2006). When both RNAs are transcribed at the same time, a rapidly degraded RNA duplex is formed; hence, the protein cannot be synthesized until the *isiA* mRNA levels are high enough to titrate the IsiR asRNAs (Dühring et al., 2006; Georg and Hess, 2011). An

interesting study by Dornenburg et al. (2010) showed the effect of the studied antisense transcripts by knocking out only the respective asRNA promoters. In consequence, the sense transcript of one investigated sRNA/asRNA pair showed higher transcript levels, pointing out the negative effect of the asRNA, whereas no altered transcript level could be observed for another pair (Dornenburg et al., 2010). These findings complement the above mentioned hypothesis, that not all of the identified asRNAs might have a regulatory potential.

F.1.3 The operon concept

In addition to the discussed features enhancing the complexity of the transcriptome, the operon structure itself contributes to raise the diversity of transcripts due to multiple possible transcriptional variants.

Until recently, the identification of operons was either based on computational predictions or analytical methods focusing on few specific targets. The advent of RNA-Seq and tiling arrays now enables the experimental determination of whole genome operon maps for bacteria (Campanaro et al., 2012; Güell et al., 2009; Qiu et al., 2010; Ramachandran et al., 2012; Sahr et al., 2012; Toledo-Arana et al., 2009) and archaea (Wurtzel et al., 2010). Especially the knowledge of transcription start sites (TSS) derived by dRNA-Seq aids the assignment of operons by pointing out those TSS not recognizable by standard RNA-Seq analyses (Sharma et al., 2010). The operon map compiled for *B. licheniformis* (Chapter B) contains 1677 monocistronic and 833 polycistronic operons, the latter of which comprise 3.15 genes in average. In comparison, a bioinformatic prediction available at the DOOR database lists 1629 monocistronic and 877 polycistronic operons with a mean of 3.03 genes per polycistronic operon (Mao et al., 2009). These results are in good accordance with the results achieved during operon map generation for other prokaryotic organisms (Campanaro et al., 2012; Güell et al., 2009; Qiu et al., 2010; Ramachandran et al., 2012; Sahr et al., 2012; Toledo-Arana et al., 2009), for which also no remarkable variations between the numbers of predicted and validated operons could be determined. Nevertheless, despite the similar operon counts, only 65% of the predicted *B. licheniformis* polycistronic operons have a perfect match to the experimentally verified operons, whereas the other operons are either extended or shortened. According to this observation, only 60% of predicted operons in *Salmonella enterica* serovar Typhimurium match the obtained transcriptome data (Ramachandran et al., 2012). These high rates of introduced changes show that computational approaches can give a good estimation

of the operon map, but are presently not capable of replacing the accuracy of experimental validation (Brouwer et al., 2008; Ramachandran et al., 2012).

A comprehensive, tiling array-based study of *B. subtilis* grown at 104 different conditions has lately revealed that 46% of all annotated CDS are transcribed from more than one promoter under the tested conditions (Nicolas et al., 2012). A similar result of 42% was found for *Mycoplasma pneumoniae* studied under 62 independent conditions (Güell et al., 2009). In *Helicobacter pylori*, grown under only five different conditions, 192 alternative transcripts complementing the 337 primary operons were predicted (Sharma et al., 2010). The analysis of TSS identified in *B. licheniformis* (Chapter B) is in accordance with these results. Secondary or tertiary starting sites, which suggest the presence of additional promoters, were revealed upstream (<500 nt) of 72 annotated genes. Furthermore, 125 TSS were identified upstream (<500 nt) of genes localized within an operon, also pointing out putative alternative promoter sites. In total, 43 mono- and 146 polycistronic operons, which putatively form alternative transcripts, were detected in the evaluated time-course experiment under only one condition. Notably, 267 further gene-internal TSS were detected; these might derive from promoters more distal to the annotated start codon and may therefore also contribute to the formation of alternative transcripts.

Another known effect is naturally occurring operon polarity (Adhya, 2003), describing a decay behavior within cistrons whereby promoter-distal genes are less expressed than genes in close proximity to the promoter. In *M. pneumoniae*, most of the observed operons show natural polarity, which appears to be condition-dependent in most cases (Güell et al., 2009). The results of a manual inspection of the *B. licheniformis* operons also match this observation. Specific analyses of natural polarity have revealed that there is not one global mechanism triggering this effect, but that many different components are involved in dependence of the respective operon (Laing et al., 2006). These components range from sRNAs and riboswitches to RNA polymerase binding factors to Rho-dependent internal terminators (Adhya, 2003; Güell et al., 2011). The effect of differential termination has also been found to be affected by transcription initiation from different promoters (Lee et al., 2008). Furthermore, a termination-inducing coupling of transcription and translation has been proposed, based on affectation of the RNA polymerase binding stability by the following ribosome (Bonekamp et al., 1984; Güell et al., 2011). Of course, another main influential effect is RNase-mediated mRNA degradation (Li and Altman, 2004).

F.2 Process monitoring and optimization

In addition to the identification of RNA-based regulators and the evaluation of *B. licheniformis* transcriptome complexity, a further aim of this thesis was the physiological monitoring of the fermentation process and the assignment of putative optimization targets, as described and discussed in the following passages.

F.2.1 Process monitoring by RNA-Seq

The initial steps in RNA-Seq data analysis are bioinformatic approaches to transfer the vast amount of sequencing data into interpretable units representing, for example, gene expression or transcript boundaries. First, the gained sequencing reads had to be aligned to a reference sequence. In this study, $1.5x10^6$ to $4.5x10^6$ reads per sample were mapped to the *B. licheniformis* DSM13 genome (Veith et al., 2004) utilizing the BLAST algorithm (Altschup et al., 1990; Chapter B). The results of this strategy proved to be equally exact as, but to require more computation time than, results gained by commonly used mapping tools like BWA or Bowtie 2 (personal communication, Sascha Dietrich; Langmead and Salzberg, 2012; Li and Durbin, 2009; Mutz et al., 2013). Hence, the usage of these sophisticated read alignment tools is proposed in Chapter D. One obstacle that had to be overcome during the mapping process was the handling of reads mapping to multiple locations due to high sequence similarities. A conservative approach, used in this study and described in Chapter B, is the elimination of such sequences from the data set (Mäder et al., 2011), whereas an alternative method proportionally allocates the multi-mapping reads based on the number of reads mapped to their neighboring unique sequences (Wang et al., 2009).

The next step of data analysis was the identification of RNA features at single base-resolution as described comprehensively in Chapter B and D. To allow comparison and illustration, the expression strength of each identified feature and every annotated gene was calculated as normalized NPKM value (Chapters B and D).

Furthermore, the RNA-Seq data were utilized for differential expression analysis of protein-encoding genes (Chapter C). In average, $2.8x10^6$ sequencing reads per sample were employed for this analysis. This amount of input data has been shown to be appropriate for differential expression analysis in bacteria; in fact, considerably higher read counts primarily increase the rate of false positives (Haas et al., 2012; Tarazona et al., 2011). At the moment, there is an ongoing debate on which statistical methods and analytical tools are best suited for the identification of differentially expressed genes in RNA-Seq experiments (Fang et al., 2012;

Oshlack et al., 2010; Trapnell et al., 2013). In this study the baySeq tool was employed, as it was the only tool with reliable performance for comparison of more than two different conditions (Hardcastle and Kelly, 2010; Kvam et al., 2012; Vijay et al., 2013) available at the time (autumn 2011) the analyses were performed.

F.2.2 Strategies for process monitoring and optimization

The monitoring of global expression profiles of cells at different time points throughout a fermentation process enables the exploration of physiological and genetic bottlenecks of the strain of interest (Jürgen et al., 2005a). Therefore, the generated data of expression strength (NPKM) and differential gene expression (Chapter B) were analyzed with regard to critical process parameters like carbon and nitrogen supply, oxidative, cell envelope and phosphate stress, as well as sporulation and secretion capacity (Chapter C; Schweder, 2011). This approach allowed the identification of several putative optimization targets, which will be discussed below (E.2.3).

The assessed data only provide insight into five discrete temporary states of the production process, but nevertheless allow exploitation for future applications. Generally, industrial fermentation processes are in demand of on-line analysis techniques to gain information about process-critical physiological parameters (Pioch et al., 2007; Schweder, 2011). Therefore, direct monitoring of the transcriptome by measuring distinct gene markers as indicators of unfavorable growth conditions would be most helpfully (Jürgen et al., 2005b; Pioch et al., 2007). The thereby obtained information could allow an enhanced control of the process, leading to better reproducibility, increased efficiency, cycle time reduction, and finally a diminished employment of resources (Rautio, 2007; Schweder, 2011). Whereas current state of the art is off-line RT-qPCR, more direct at-line methods based on mRNA hybridization techniques have been proposed (Jürgen et al., 2005b; Pioch et al., 2007; Rautio et al., 2006; Schweder, 2011). Next to necessary improvements regarding analysis time, these methods could also be advanced by the identification of additional markers derived from protein genes but also from sRNAs or asRNAs, as identified in Chapter B.

However, an important factor that has to be kept in mind when developing transcriptome monitoring systems is the distinct stability of the observed mRNAs. The decay of mRNA transcripts plays an important role in the cellular control of gene expression, which is highly regulated by the use of sRNAs, RNA-binding proteins, internal cleavage sites and diverse ribonucleases (Anderson and Dunman, 2009; Arraiano et al., 2010; Condon, 2007). For

example, mRNA half-lives from around 1 min to more than 15 min could be determined in *B. subtilis* (Hambraeus et al., 2003). Additionally, mRNA turnover has been shown to be condition dependent in other members of the division *Firmicutes* (Redon et al., 2005). Briefly, mRNA levels are determined by the momentary transcriptional activity of the respective gene as well as by the rate of regulatory mRNA decay (Condon and Bechhofer, 2011; Evguenieva-Hackenberg and Klug, 2011). Hence, when carrying out mRNA quantification experiments like hybridization assays or RNA-Seq, it is important to keep in mind that the determined mRNA amount does not exactly represent the actual gene expression, but only the mere presence of a transcript. Furthermore, it is not assignable whether the detected mRNA had been inactivated at the sampled time point (Evguenieva-Hackenberg and Klug, 2011).

To gain deeper insight into metabolic activities, the proteome of the analyzed fermentation process was explored as described in Chapter C. By the employed 2D gel-based approach 367 protein spots representing 260 different proteins could be identified and were assessed to confirm the insights gained by RNA-Seq. Similar, though microarray-based, experiments, which addressed the *B. licheniformis* stress and starvation responses, have been performed previously (Hoi et al., 2006; Schroeter et al., 2011; Voigt et al., 2007) and allowed thorough analysis of the production process evaluated in this study (Chapter C). To explore the obtained proteomic and transcriptomic data on a more global scale, integration of the different layers of information with system biological approaches will be necessary (Maier et al., 2011). However, as adjustment of the deployed methodologies is a major challenge (De Keersmaecker et al., 2006; Zhang et al., 2010), data integration will be part of future work embedded in the encompassing project (personal communication, Heiko Liesegang).

Systems biology is not only a powerful tool to monitor processes, but also to optimize bioprocesses by metabolic and process engineering based on the gained models (Carrondo et al., 2012). Whereas systems biology-based process engineering focuses on the optimized supply with nutrients and oxygen, metabolic engineering conventionally targets the rate-limiting steps of the given process identified by the developed process model (Deckwer et al., 2006; Vemuri and Aristidou, 2005). However, it might also be reasonable to keep in mind further strategies that do not aim at the optimization of an existing fermentation, but at the establishment of a whole new process. This might, for example, be the complete reorganization of an existing pathway or the insertion of a previously absent pathway by metabolic engineering (Park et al., 2007; Ro et al., 2006). Moreover, the employment of new

production platforms, either derived by methods of synthetic biology (Purnick and Weiss, 2009) or originating from new, but promising isolates like high-temperature production strains (Chapter E) might also contribute to the optimization of enzyme yields.

F.2.3 Putative optimization targets

In addition to the mentioned approaches towards bioprocess improvement (Chapter A, E.2.2), the RNA-Seq data obtained in this study also allowed the identification of putative optimization targets.

As described in Chapter B, antisense transcripts with a putative impact on protease production could be found located opposite to *apr* coding for Subtilisin Carlsberg, and opposite to the two-component regulatory system operons *degSU* and *cssRS,* which have roles in cell differentiation and the secretion stress response, respectively. Especially the antisense transcript addressing the alkaline protease Subtilisin Carlsberg is of high interest, as this RNA feature has no ortholog in the transcriptome of *B. subtilis* (Nicolas et al., 2012). The first step to elucidate the impact of the asRNA on the synthesis of the protease could, for example, be the removal of the asRNA promoter sequence (personal communication, Robert Hertel). This approach would show whether the asRNA has a stabilizing or degradative effect on the sense mRNA, effecting the production of the protease. In case of a stabilizing effect, it might also be promising to overexpress the asRNA to increase its impact. Subsequently, the results from the experiments targeting the chromosomally encoded *apr* gene might then be transferred to the plasmid-encoded industrial Subtilisin gene to eventually optimize protease production. This would be a new approach, as previous efforts of bacterial production strain engineering by modulation of antisense transcription have relied on introduction of artificial asRNAs into the cell to destabilize the sense target (Desai and Papoutsakis, 1999; Tummala et al., 2003).

The analysis of gene expression presented in Chapter C pointed out some reactions to changes in nutrient and oxygen supply, for example phosphate limitation and superoxide stress, which could be overcome by bioprocess engineering approaches. The data suggest furthermore, that the secretion of the produced protease might not be solely accomplished by the Sec protein translocation pathway, but may also involve the twin-arginine translocation (Tat) pathway. This finding could be further investigated by the creation of deletion mutants targeting either the Tat- or the Sec-dependent pathway. If a Tat-directed secretion could be confirmed, an improvement of the production process might be achieved by the optimization of the Subtilisin signal peptide (Degering et al., 2010).

In terms of bioprocess development, it is also interesting to consider the impact of multicellularity (Chapter A) on the capacity of protease secretion. Although differentiated cell states were not considered separately in this study, information can be gained from exploration of the whole transcriptome data. In order to give a brief overview on the influence of the three master regulators of cell differentiation, their mode of action will be described at first.

ComA-dependent differentiation. In *B. subtilis*, the genetic cascades regulated by ComA induce surfactin production as well as natural competence (Figure 4; Dubnau and Losick, 2006; Hoffmann et al., 2010; López and Kolter, 2010). The induction of these pathways starts with the production of the peptide pheromone ComX. Accumulation of the pheromone is sensed by the sensor histidine kinase ComP, which then phosphorylates ComA. Subsequently, ComA~P activates the expression of the surfactin operon and, via the transcriptional activator ComK, the formation of genetic competence. In contrast to *B. subtilis* 168 and mutant strains of *B. licheniformis* ATCC 9945A, ComP is found inactive in *B. licheniformis* DSM13 due to an insertion element in *comP* (Rey et al., 2004; Veith et al., 2004).

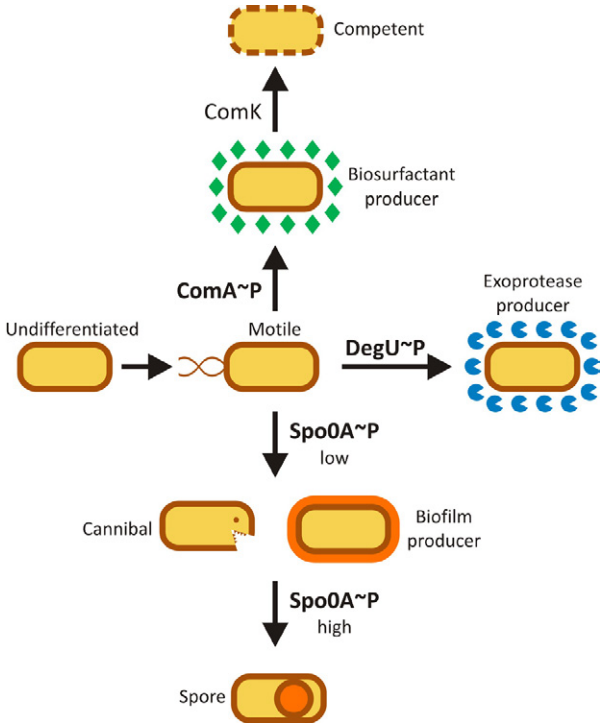


Figure 4 Multicellularity of *B. subtilis*
Schematic representation of the cell types in a *B. subtilis* community that differentiate beginning from motile cells, under the influence of the phosphorylated states of the three master regulators DegU, Spo0A and ComA, and the transcriptional activator ComK. The figure was modified from López and Kolter (2010).

DegU-dependent differentiation. The regulatory effect of DegU depends fairly on its phosphorylation state. Whereas non-phosphorylated DegU is necessary to support the development of competence, low levels of DegU~P facilitate swarming motility (Verhamme et al., 2007). Furthermore, moderate levels of DegU~P support colony architecture and finally high amounts of DegU~P are required for the activation of exoprotease production (Figure 4; Veening et al., 2008; Verhamme et al., 2007).

Spo0A-dependent differentiation. In *B. subtilis,* low levels of phosphorylated Spo0A trigger the secretion of the two main components of biofilms: exopolymeric polysaccharides (EPS) and the structural protein TasA (López et al., 2010). In the same subpopulation, two peptide toxins are expressed to facilitate the lysis of members of other subpopulations (López et al., 2009). The coupling of cannibalistic behavior and biofilm production is thought to be a mechanism to delay sporulation by providing nutrients originating from lysed cells under nutrient-limited conditions. (González-Pastor, 2011; López et al., 2009). High levels of Spo0~P, which emerge in response to starvation, activate the sporulation of a fraction of the matrix producer/ cannibal subpopulation (Figure 4; González-Pastor, 2011; López and Kolter, 2010).

The expression of the *fla/che*-operon, as marker for flagellum-based motility (Amati et al., 2004), is repressed at all sampling points. Biofilm formation, indicated by the *eps* operon and *tasA*, is highly induced at the early sampling points I to III whereas sporulation seems to occur mainly at the later sampling point IV. Additionally, enhanced expression of exoprotease-encoding genes is monitored throughout the sampling points with exception to sampling point I. Because the first sample was taken ten hours after onset of the fermentation process at a time point showing already high activation of biofilm formation, it can be concluded that a low phosphorylation level of Spo0A~P is given, inducing biofilm formation as well as repressing swimming motility (López and Kolter, 2010; Lopez et al., 2009). With the entry to carbohydrate limitation at sampling point IV (Chapter C), the level of Spo0A~P seems to rise to levels sufficient for the triggering of sporulation in a subset of cells (Schultz et al., 2009; Vlamakis et al., 2008). Meanwhile high levels of DegU~P appear to be preserved throughout the productive stages of the fermentation process in order to allow the formation of an exoprotease-producing subpopulation (Ogura et al., 2003; Verhamme et al., 2007). It is fairly natural to assume that this subpopulation is likewise suited for the production and secretion of the overexpressed protease of interest and, hence, the maintenance of this cell state is an important aspect of the industrial production process. Furthermore, neither the lichenysin-encoding operon nor the operon coding for the late competence genes show activation at any sampling point, implying that subpopulations of biosurfactant producing and competent cells (Chapter A) are not formed in the monitored process. As both multicellular states are expected to be induced in response to ComA~P (Hoffmann et al., 2010a; López and Kolter, 2010), the lack of operon activity might be a direct consequence of the missing functionality of the ComA-phosphorylating kinase ComP. Although the introduction of a functional copy of

ComP could not restore the genetic competence of *B. licheniformis DSM13*, a result not totally understood by now (Hoffmann et al., 2010), an impact on lichenysin production cannot be excluded. It is not difficult to imagine, that the lacking cell differentiation itself is not only a drawback in terms of genetic accessibility (Rachinger, 2010; Waschkau et al., 2008), but offers a crucial advantage over other biotechnological-relevant bacteria (Schallmey et al., 2004) by releasing anabolic and catabolic resources.

Some questions arise with regard to the existing subpopulations: Is the population of exoprotease-producing cells in fact the most suitable subpopulation for the industrial production of a plasmid-encoded protease? And, if an optimal cell state for this purpose could be identified, would it be possible to shift more cells to this subpopulation in order to increase the productivity of the process?

Both questions have been addressed by different approaches in order to raise the amount of available DegU~P in the cells by the Meinhardt group. They could increase the proteolytic activity of *B. licheniformis* by deleting the operon encoding genes for polyglutamate production, which is, like the production of exoproteases, regulated by DegU~P (Hoffmann et al., 2010b; Stanley and Lazazzera, 2005). Although the underlying reasons for this result remain unresolved, it is proposed that the deletion of competing molecular targets of DegU~P releases capacities to enhance production of other targets of this activator (Hoffmann et aAl., 2010b). In contrast, the deletion of the DegU-phosphorylation repressor RapG did not show any effect on exoprotease secretion (Borgmeier et al., 2012), which is contrary to the results of an according experiment carried out in *B. subtilis* (Ogura et al., 2003). In a different attempt, the Meinhardt group found that enhanced production of chromosomally encoded exoproteases can also be achieved in *B. licheniformis* by introduction of the *degU32* allele, known to cause hypersecretion in *B. subtilis* (Borgmeier et al., 2012). Furthermore, it was shown that a chromosomal copy of *degU32* has a positive effect on the production of a plasmid-encoded exoprotease in *B. amyloliquefaciens* (Borgmeier et al., 2011). Nevertheless, since this exoprotease was under the control of a DegU~P-responsive promoter (Borgmeier et al., 2011), it is not yet known if there is an impact of enhanced DegU~P availability on exoprotease production induced by a constitutive, DegU-independent promoter.

A further approach to optimize the productivity of the fermentation might be to diminish the influence of other abundant cell states, especially the spore-forming subpopulation (Chapter C). Moderate levels of Spo0A~P are not only necessary to induce biofilm formation, but also to establish the exoprotease-producing subpopulations, by repressing the

transcriptional repressor AbrB (Davidson et al., 2012; Schultz et al., 2009). Instead, high levels of phosphorylated Spo0A induce sporulation (Davidson et al., 2012; Schultz et al., 2009). Hindering Spo0A~P to accumulate could inhibit the formation of spores and could therefore contribute to the establishment of optimized multicellularity. Ways to address this goal might be the overexpression of the RapA phosphatase (Perego et al., 1996), the deletion of its repressor PhrA (Perego et al., 1996) or a reduced expression of the lichenicidin cluster, as described in Chapter C. However, sporulation itself might as well be a factor contributing to the bioprocess, for instance by providing nutrients through cannibalism-induced cell lysis.

It is obvious that manipulation of cellular pathways tightly controlled by the described master regulators and a plethora of additional activators and repressors is labor-intensive and highly speculative. Therefore, techniques to determine the effective cell state regarding protease production and general metabolic activity would be of high benefit. Some recent methods allowing single-cell analyses with the most prominent omics technologies could be used for these purposes. For example, developments in mass spectrometry now enable the detection of metabolites deriving from a single bacterial cell (Heinemann and Zenobi, 2011; Svatoš, 2011). Additionally, the proteome of cellular subpopulations sorted by flow cytometric methods can also be analyzed by mass spectrometry (Bernhardt et al., 2013). Other approaches allow protein analysis on single cell-resolution by imaging of fluorescent reporter proteins coupled to the promoter or protein of interest (De Jong et al., 2012; Taniguchi et al., 2010). Another imaging technique, fluorescence *in situ* hybridization, can also be applied for the identification of mRNA transcripts (Taniguchi et al., 2010). Finally, transcriptome analysis of single bacterial cells was performed successfully (Kang et al., 2011). Although the method was originally established for microarray detection, it should be applicable to RNA-Seq approaches, as has already been demonstrated for eukaryotic cells (Tang et al., 2009). This technique is expected to make future contributions to many leading questions concerning the differences between the heterogeneous cell populations (Filiatrault, 2011; Mäder et al., 2011), particularly when coupled with third-generation sequencing techniques.

References

Adhya, S. (2003). Suboperonic regulatory signals. Sci STKE *185*, pe22.

Altschup, S.F., Gish, W., Miller, W., Myers, E.W., and J, L.D. (1990). Basic Local Alignment Search Tool. J Mol Biol *215*, 403–410.

Amati, G., Bisicchia, P., and Galizzi, A. (2004). DegU-P Represses Expression of the Motility *fla-che* Operon in *Bacillus subtilis*. J Bacteriol *186*, 6003–6014.

Anderson, K., and Dunman, P. (2009). Messenger RNA Turnover Processes in *Escherichia coli*, *Bacillus subtilis*, and Emerging Studies in *Staphylococcus aureus*. Int J Microbiol *2009*, 525491.

Arraiano, C.M., Andrade, J.M., Domingues, S., Guinote, I.B., Malecki, M., Matos, R.G., Moreira, R.N., Pobre, V., Reis, F.P., Saramago, M., et al. (2010). The critical role of RNA processing and degradation in the control of gene expression. FEMS Microbiol Rev *34*, 883–923.

Barrick, J.E., and Breaker, R.R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol *8*, R239.

Bastet, L., Dubé, A., Massé, E., and Lafontaine, D.A. (2011). New insights into riboswitch regulation mechanisms. Mol Microbiol *80*, 1148–1154.

Bernhardt, J., Michalik, S., Wollscheid, B., Völker, U., and Schmidt, F. (2013). Proteomics approaches for the analysis of enriched microbial subpopulations and visualization of complex functional information. Curr Opin Biotechnol *24*, 112–119.

Bonekamp, F., Clemmesen, K., Karlstrom, O., and Jensen, K.F. (1984). Mechanism UTP-modulated attenuation at the *pyrE* gene of *Escherichia coli*: an example of operon polarity control through the coupling of translation and transcription. EMBO J *3*, 2857–2861.

Borgmeier, C., Biedendieck, R., Hoffmann, K., Jahn, D., and Meinhardt, F. (2011). Transcriptome profiling of *degU* expression reveals unexpected regulatory patterns in *Bacillus megaterium* and discloses new targets for optimizing expression. Appl Microbiol Biotechnol *92*, 583–596.

Borgmeier, C., Bongaerts, J., and Meinhardt, F. (2012). Genetic analysis of the *Bacillus licheniformis degSU* operon and the impact of regulatory mutations on protease production. J Biotechnol *159*, 12–20.

Brenneis, M., Hering, O., Lange, C., and Soppa, J. (2007). Experimental characterization of *Cis*-acting elements important for translation and transcription in halophilic archaea. PLoS Genet *3*, e229.

Brock, J.E., Pourshahian, S., Giliberti, J., Limbach, P.A., and Janssen, G.R. (2008). Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 59-terminal AUG. RNA *14*, 2159–2169.

Brouwer, R.W.W., Kuipers, O.P., and van Hijum, S.A.F.T. (2008). The relative value of operon predictions. Brief Bioinform *9*, 367–375.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., and Bateman, A. (2013). Rfam 11.0: 10 years of RNA families. Nucleic Acids Res *41*, D226–32.

Campanaro, S., Pascale, F. De, Telatin, A., Schiavon, R., Bartlett, D.H., and Valle, G. (2012). The transcriptional landscape of the deep-sea bacterium *Photobacterium profundum* in both a *toxR* mutant and its parental strain. BMC Genomics *13*, 567.

Carrondo, M.J.T., Alves, P.M., Carinhas, N., Glassey, J., Hesse, F., Merten, O.-W., Micheletti, M., Noll, T., Oliveira, R., Reichl, U., et al. (2012). How can measurement, monitoring, modeling and control advance cell culture in industrial biotechnology? Biotechnol J *7*, 1522–1529.

Chang, B., Halgamuge, S., and Tang, S.-L. (2006). Analysis of SD sequences in completed microbial genomes: Non-SD-led genes are as common as SD-led genes. Gene *373*, 90–99.

Condon, C. (2007). Maturation and degradation of RNA in bacteria. Curr Opin Microbiol *10*, 271–278.

Condon, C., and Bechhofer, D.H. (2011). Regulated RNA stability in the Gram positives. Curr Opin Microbiol *14*, 148–154.

Davidson, F., Seon-Yi, C., and Stanley-Wall, N. (2012). Selective Heterogeneity in Exoprotease Production by *Bacillus subtilis*. PLoS One *7*, e38574.

Deckwer, W.-D., Jahn, D., Hempel, D., and Zeng, A.-P. (2006). Systems Biology Approaches to Bioprocess Development. Eng Life Sci *6*, 455–469.

Degering, C., Eggert, T., Puls, M., Bongaerts, J., Evers, S., Maurer, K.-H., and Jaeger, K.-E. (2010). Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. Appl Environ Microbiol *76*, 6370–6376.

Desai, R.P., and Papoutsakis, E.T. (1999). Antisense RNA Strategies for Metabolic Engineering of *Clostridium acetobutylicum*. Appl Environ Microbiol *65*, 936.

Dornenburg, J.E., Devita, A.M., Palumbo, M.J., and Wade, J.T. (2010). Widespread Antisense Transcription in *Escherichia coli*. Mbio *1*, e00024.

Dubnau, D., and Losick, R. (2006). Bistability in bacteria. Mol Microbiol *61*, 564–572.

Dühring, U., Axmann, I.M., Hess, W.R., and Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. Proc Natl Acad Sci USA *103*, 7054–7058.

Evguenieva-Hackenberg, E., and Klug, G. (2011). New aspects of RNA processing in prokaryotes. Curr Opin Microbiol *14*, 587–592.

Fang, Z., Martin, J., and Wang, Z. (2012). Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. Cell Biosci *2*, 26.

Filiatrault, M.J. (2011). Progress in prokaryotic transcriptomics. Curr Opin Microbiol *14*, 579–586.

Georg, J., and Hess, W.R. (2011). *cis*-antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol R *75*, 286–300.

Giliberti, J., O'Donnell, S., van Etten, W.J., and Janssen, G.R. (2012). A 5'-terminal phosphate is required for stable ternary complex formation and translation of leaderless mRNA in *Escherichia coli*. RNA *18*, 508–518.

Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. Science *326*, 1268–1271.

Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. (2011). Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nat Rev Microbiol *9*, 658–669.

González-Pastor, J.E. (2011). Cannibalism: a social behavior in sporulating *Bacillus subtilis*. FEMS Microbiol Rev *35*, 415–424.

Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., and Livny, J. (2012). How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics *13*, 734.

Hambraeus, G., Von Wachenfeldt, C., and Hederstedt, L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. Mol Genet Genomics *269*, 706–714.

Hardcastle, T.J., and Kelly, K.A. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics *11*, 422.

Heinemann, M., and Zenobi, R. (2011). Single cell metabolomics. Curr Opin Biotechnol *22*, 26–31.

Hernández, J.A., Muro-Pastor, A.M., Flores, E., Bes, M.T., Peleato, M.L., and Fillat, M.F. (2006). Identification of a *furA* cis antisense RNA in the cyanobacterium *Anabaena* sp. PCC 7120. J Mol Biol *355*, 325–334.

Hoffmann, K., Wollherr, A., Larsen, M., Rachinger, M., Liesegang, H., Ehrenreich, A., and Meinhardt, F. (2010a). Facilitation of direct conditional knockout of essential genes in *Bacillus licheniformis* DSM13 by comparative genetic analysis and manipulation of genetic competence. Appl Environ Microbiol *76*, 5046–5057.

Hoffmann, K., Daum, G., Koster, M., Kulicke, W.-M., Meyer-Rammes, H., Bisping, B., and Meinhardt, F. (2010b). Genetically improved *Bacillus licheniformis* strains for efficient deproteinization of shrimp shells and production of high-molecular-mass chitin and chitosan. Appl Environ Microbiol *76*, 8211–8221.

Hoi, L.T., Voigt, B., Jürgen, B., Ehrenreich, A., Gottschalk, G., Evers, S., Feesche, J., Maurer, K.-H., Hecker, M., and Schweder, T. (2006). The phosphate-starvation response of *Bacillus licheniformis*. Proteomics *6*, 3582–3601.

Jahn, N., Preis, H., Wiedemann, C., and Brantl, S. (2012). BsrG/SR4 from *Bacillus subtilis*--the first temperature-dependent type I toxin-antitoxin system. Mol Microbiol *83*, 579–598.

De Jong, I.G., Veening, J.-W., and Kuipers, O.P. (2012). Single cell analysis of gene expression patterns during carbon starvation in *Bacillus subtilis* reveals large phenotypic variation. Environ Microbiol *14*, 3110–3121.

Jürgen, B., Tobisch, S., Wümpelmann, M., Gördes, D., Koch, A., Thurow, K., Albrecht, D., Hecker, M., and Schweder, T. (2005a). Global expression profiling of *Bacillus subtilis* cells during industrial-close fed-batch fermentations with different nitrogen sources. Biotechnol Bioeng *92*, 277–298.

Jürgen, B., Barken, K.B., Tobisch, S., Pioch, D., Wümpelmann, M., Hecker, M., and Schweder, T. (2005b). Application of an electric DNA-chip for the expression analysis of bioprocess-relevant marker genes of *Bacillus subtilis*. Biotechnol Bioeng *92*, 299–307.

Kang, Y., Norris, M.H., Zarzycki-Siek, J., Nierman, W.C., Donachie, S.P., and Hoang, T.T. (2011). Transcript amplification from single bacterium for transcriptome analysis. Genome Res *21*, 925–935.

De Keersmaecker, S., Thijs, I., Vanderleyden, J., and Marchal, K. (2006). Integration of omics data: how well does it work for bacteria? Mol Microbiol *62*, 1239–1250.

Kvam, V.M., Liu, P., and Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot *99*, 248–256.

Laing, E., Mersinias, V., Smith, C.P., and Hubbard, S.J. (2006). Analysis of gene expression in operons of *Streptomyces coelicolor*. Genome Biol *7*, R46.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

Lasa, I., Toledo-Arana, A., and Gingeras, T.R. (2012). An effort to make sense of antisense transcription in bacteria. RNA Biol *9*, 1039–1044.

Lawrence, J.G. (2003). Gene organization: selection, selfishness, and serendipity. Annu Rev Microbiol *57*, 419–440.

Lee, H.J., Jeon, H.J., Ji, S.C., Yun, S.H., and Lim, H.M. (2008). Establishment of an mRNA gradient depends on the promoter: an investigation of polarity in gene expression. J Mol Biol *378*, 318–327.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, Y., and Altman, S. (2004). Polarity effects in the lactose operon of *Escherichia coli*. J Mol Biol *339*, 31–39.

Lopez, D., Vlamakis, H., and Kolter, R. (2009). Generation of multiple cell types in *Bacillus subtilis*. FEMS Microbiol Rev *33*, 152–163.

López, D., and Kolter, R. (2010). Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. FEMS Microbiol Rev *34*, 134–149.

López, D., Vlamakis, H., Losick, R., and Kolter, R. (2009). Cannibalism enhances biofilm development in *Bacillus subtilis*. Mol Microbiol *74*, 609–618.

López, D., Vlamakis, H., and Kolter, R. (2010). Biofilms. Cold Spring Harb Perspect Biol *2*, a000398.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., et al. (2004). ARB: a software environment for sequence data. Nucleic Acids Res *32*, 1363–1371.

Mäder, U., Nicolas, P., Richard, H., Bessières, P., and Aymerich, S. (2011). Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. Curr Opin Biotechnol *22*, 32–41.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R., and Serrano, L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol Syst Biol *7*, 511.

Malys, N., and McCarthy, J.E.G. (2011). Translation initiation: variations in the mechanism can be anticipated. Cell Mol Life Sci *68*, 991–1003.

Mao, F., Dam, P., Chou, J., Olman, V., and Xu, Y. (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Res *37*, D459–63.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., et al. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res *40*, D115–22.

Molina, N., and van Nimwegen, E. (2008). Universal patterns of purifying selection at noncoding positions in bacteria. Genome Res *18*, 148–160.

Moll, I., Grill, S., Gualerzi, C.O., and Bläsi, U. (2002). Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. Mol Microbiol *43*, 239–246.

Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol *24*, 22–30.

Nakagawa, S., Niimura, Y., Miura, K., and Gojobori, T. (2010). Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc Natl Acad Sci USA *107*, 6382–6387.

Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., et al. (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. Science *335*, 1103–1106.

Nudler, E., and Mironov, A. (2004). The riboswitch control of bacterial metabolism. Trends Biochem Sci *29*, 11–17.

Ogura, M., Shimane, K., Asai, K., Ogasawara, N., and Tanaka, T. (2003). Binding of response regulator DegU to the *aprE* promoter is inhibited by RapG, which is counteracted by extracellular PhrG in *Bacillus subtilis*. Mol Microbiol *49*, 1685–1697.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. Genome Biol *11*, 220.

Park, J.H., Lee, K.H., Kim, T.Y., and Lee, S.Y. (2007). Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci USA *104*, 7797–7802.

Perego, M., Glaser, P., and Hoch, J.A. (1996). Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. Mol Microbiol *19*, 1151–1157.

Pioch, D., Jürgen, B., Evers, S., Maurer, K.-H., Hecker, M., and Schweder, T. (2007). At-line Monitoring of Bioprocess-Relevant Marker Genes. Eng Life Sci *7*, 373–379.

Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Bio *10*, 410–422.

Qiu, Y., Cho, B.-K., Park, Y.S., Lovley, D., Palsson, B.Ø., and Zengler, K. (2010). Structural and operational complexity of the *Geobacter sulfurreducens* genome. Genome Res *20*, 1304–1311.

Rachinger, M. (2010). Stammdesign in *B. licheniformis*. *PhD thesis*. Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, Germany.

Raghavan, R., Sloan, D.B., and Ochman, H. (2012). Antisense Transcription Is Pervasive but Rarely Conserved in Enteric Bacteria. mBio *3*, e00156–12.

Ramachandran, V.K., Shearer, N., Jacob, J.J., Sharma, C.M., and Thompson, A. (2012). The architecture and ppGpp-dependent expression of the primary transcriptome of *Salmonella Typhimurium* during invasion gene expression. BMC Genomics *13*, 25.

Rasmussen, S., Nielsen, H., and Jarmer, H. (2009). The Transcriptionally Active Regions in the Genome of *Bacillus subtilis*. Mol Microbiol *73*, 1043–1057.

Rautio, J. (2007). Development of rapid gene expression analysis and its application to bioprocess monitoring. *PhD thesis*. Faculty of Technology, University of Oulu, Finland.

Rautio, J.J., Kataja, K., Satokari, R., Penttilä, M., Söderlund, H., and Saloheimo, M. (2006). Rapid and multiplexed transcript analysis of microbial cultures using capillary electrophoresis-detectable oligonucleotide probe pools. J Microbiol Meth *65*, 404–416.

Redon, E., Loubière, P., and Cocaign-Bousquet, M. (2005). Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. J Biol Chem *280*, 36380–36385.

Rey, M., Ramaiya, P., Nelson, N., and Brody-Karpin, S. (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. Genome Biol *5*, R77.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940–943.

Sahr, T., Rusniok, C., Dervins, Ravault, D., Sismeiro, O., J-Y, C., and Buchrieser, C. (2012). Deep sequencing defines the transcriptional map of L. pneumophila and identifies growth phase-dependent regulated ncRNAs implicated in virulence. RNA Biol *9*, 503–519.

Schallmey, M., Singh, A., and Ward, O. (2004). Developments in the use of *Bacillus* species for industrial production. Can J Microbiol *50*, 1–17.

Schroeter, R., Voigt, B., Jürgen, B., Methling, K., Pöther, D.-C., Schäfer, H., Albrecht, D., Mostertz, J., Mäder, U., Evers, S., et al. (2011). The peroxide stress response of *Bacillus licheniformis*. Proteomics *11*, 2851–2866.

Schultz, D., Wolynes, P.G., Ben Jacob, E., and Onuchic, J.N. (2009). Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis*. Proc Natl Acad Sci USA *106*, 21027–21034.

Schweder, T. (2011). Bioprocess monitoring by marker gene analysis. Biotechnol J *6*, 926–933.

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat Rev Microbiol *11*, 75–82.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature *464*, 250–255.

Sorek, R., and Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet *11*, 9–16.

Stanley, N.R., and Lazazzera, B.A. (2005). Defining the genetic differences between wild and domestic strains of *Bacillus subtilis* that affect poly-gamma-dl-glutamic acid production and biofilm formation. Mol Microbiol *57*, 1143–1158.

Svatoš, A. (2011). Single-cell metabolomics comes of age: new developments in mass spectrometry profiling and imaging. Anal Chem *83*, 5037–5044.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377–384.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Tarazona, S., Garcıa-Alcade, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. Genome Res *21*, 2213–2223.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., et al. (2009). The *Listeria* transcriptional landscape from saprophytism to virulence. Nature *459*, 950–956.

Topp, S., Reynoso, C.M.K., Seeliger, J.C., Goldlust, I.S., Desai, S.K., Murat, D., Shen, A., Puri, A.W., Komeili, A., Bertozzi, C.R., et al. (2010). Synthetic riboswitches that induce gene expression in diverse bacterial species. Appl Environ Microbiol *76*, 7881–7884.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol *31*, 46–53.

Tummala, S.B., Junne, S.G., and Papoutsakis, E.T. (2003). Antisense RNA Downregulation of Coenzyme A Transferase Combined with Alcohol-Aldehyde Dehydrogenase Overexpression Leads to Predominantly Alcohologenic *Clostridium acetobutylicum* Fermentations. J Bacteriol *185*, 3644–3653.

Veening, J.-W., Igoshin, O.A., Eijlander, R.T., Nijland, R., Hamoen, L.W., and Kuipers, O.P. (2008). Transient heterogeneity in extracellular protease production by *Bacillus subtilis*. Mol Syst Biol *4*, 184.

Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K.-H., Ehrenreich, P., Bäumer, S., Henne, A., Liesegang, H., Merkl, R., et al. (2004). The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. J Mol Microbiol Biotechnol *7*, 204–211.

Vemuri, G., and Aristidou, A. (2005). Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. Microbiol Mol Biol R *69*, 197–216.

Verhamme, D.T., Kiley, T.B., and Stanley-Wall, N.R. (2007). DegU co-ordinates multicellular behaviour exhibited by *Bacillus subtilis*. Mol Microbiol *65*, 554–568.

Vijay, N., Poelstra, J.W., Künstner, A., and Wolf, J.B.W. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Mol Ecol *22*, 620–634.

Vitreschak, A.G., Mironov, A.A., Lyubetsky, V.A., and Gelfand, M.S. (2008). Comparative genomic analysis of T-box regulatory systems in bacteria. RNA *14*, 717–735.

Vlamakis, H., Aguilar, C., Losick, R., and Kolter, R. (2008). Control of cell fate by the formation of an architecturally complex bacterial community. Gene Dev *22*, 945–953.

Voigt, B., Hoi, L.T., Jürgen, B., Albrecht, D., Ehrenreich, A., Veith, B., Evers, S., Maurer, K.-H., Hecker, M., and Schweder, T. (2007). The glucose and nitrogen starvation response of *Bacillus licheniformis*. Proteomics *7*, 413–423.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57–63.

Waschkau, B., Waldeck, J., Wieland, S., Eichstädt, R., and Meinhardt, F. (2008). Generation of readily transformable *Bacillus licheniformis* mutants. Appl Microbiol Biotechnol *78*, 181–188.

Waters, L., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell *136*, 615–628.

Weigand, J., and Suess, B. (2009). Aptamers and riboswitches: perspectives in biotechnology. Appl Microbiol Biotechnol 229–236.

Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., et al. (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res *35*, 4809–4819.

Wittmann, A., and Suess, B. (2012). Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. FEBS Letters *586*, 2076–2083.

Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. Genome Res *20*, 133–141.

Wurtzel, O., Sesto, N., Mellin, J.R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., and Sorek, R. (2012). Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. Mol Syst Biol *8*, 1–14.

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple "omics" analysis for microbial biology: application and methodologies. Microbiology+ *156*, 287–301.

Zheng, X., Hu, G.-Q., She, Z.-S., and Zhu, H. (2011). Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. BMC Genomics *12*, 361.

Additional information

Table S1 Strains and studies depicted in Figure 1

Species	Strain	Study
Agrobacterium fabrum	C58	Wilms, I., Overlöper, A., Nowrousian, M., Sharma, C.M., and Narberhaus, F. (2012). Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. RNA Biol *9*, 446–457.
Anabaena variabilis	ATCC 29413	Mitschke, J., Vioque, A., Haas, F., and Hess, W.R. (2011a). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. P Natl Acad Sci USA *108*, 20130–20135.
Bacillus licheniformis	DSM13	Chapter B
Bacillus subtilis	subsp. *subtilis* 168	Irnov, I., Sharma, C., Vogel, J., and Winkler, W. (2010). Identification of regulatory RNAs in *Bacillus subtilis*. Nucleic Acids Res *38*, 6637–6651.
		Rasmussen, S., Nielsen, H., and Jarmer, H. (2009). The Transcriptionally Active Regions in the Genome of *Bacillus subtilis*. Mol Microbiol *73*, 1043–1057.
		Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., et al. (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. Science *335*, 1103–1106.
Chlamydia trachomatis	L2b	Albrecht, M., Sharma, C., Reinhardt, R., Vogel, J., and Rudel, T. (2010). Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res *38*, 868–877.
Chlamydophila pneumoniae	CWL029	Albrecht, M., Sharma, C.M., Dittrich, M.T., Müller, T., Reinhardt, R., Vogel, J., and Rudel, T. (2011). The transcriptional landscape of *Chlamydia pneumoniae*. Genome Biol *12*, R98.
Escherichia coli	K-12 MG1655	Dornenburg, J.E., Devita, A.M., Palumbo, M.J., and Wade, J.T. (2010). Widespread Antisense Transcription in *Escherichia coli*. Mbio *1*, e00024.
Helicobacter pylori	26695	Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature *464*, 250–255.
Legionella pneumophila	Paris	Sahr, T., Rusniok, C., Dervins, Ravault, D., Sismeiro, O., J-Y, C., and Buchrieser, C. (2012). Deep sequencing defines the transcriptional map of *L. pneumophila* and identifies growth phase-dependent regulated ncRNAs implicated in virulence. RNA Biol *9*, 503–519.
Listeria monocytogenes	EGD-e	Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., et al. (2009). The *Listeria* transcriptional landscape from saprophytism to virulence. Nature *459, 950-956*.
Mycoplasma pneumoniae	M129	Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. Science *326*, 1268–1271.
Pseudomonas aeruginosa	UCBPP-PA14	Dötsch, A., Eckweiler, D., Schniederjans, M., Zimmermann, A., Jensen, V., Scharfe, M., Geffers, R., and Häussler, S. (2012). The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. PloS One *7*, e31092.
Pseudomonas syringae	pv. tomato DC3000	Filiatrault, M.J., Stodghill, P.V, Bronstein, P.A., Moll, S., Lindeberg, M., Grills, G., Schweitzer, P., Wang, W., Schroth, G.P., Luo, S., et al. (2010). Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. J Bac *192*, 2359–2372.
Salmonella enterica	subsp. *enterica*sv. Typhimurium SL1344	Kröger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., et al. (2012). The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. P Natl Acad Sci USA *109*, E1277–86.

Table S1 Continued

Species	Strain	Study
		Ramachandran, V.K., Shearer, N., Jacob, J.J., Sharma, C.M., and Thompson, A. (2012). The architecture and ppGpp-dependent expression of the primary transcriptome of *Salmonella* Typhimurium during invasion gene expression. BMC Genomics *13*, 25.
Sinorhizobium meliloti	2011	Schlüter, J.-P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Jänicke, S., Becker, J.D., Giegerich, R., and Becker, A. (2010). A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. BMC Genomics *11*, 245.
Streptomyces coelicolor	A3(2)	Vockenhuber, M.-P., Sharma, C.M., Statt, M.G., Schmidt, D., Xu, Z., Dietrich, S., Liesegang, H., Mathews, D.H., and Suess, B. (2011). Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. RNA Biol *8*, 468–477.
Sulfolobus solfataricus	P2	Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. Genome Res *20*, 133–141.
Synechococcus elongatus	PCC7942	Vijayan, V., Jain, I.H., and O'Shea, E.K. (2011). A high resolution map of a cyanobacterial transcriptome. Genome Biol *12*, R47.
Synechocystis sp.	PCC 6803	Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voß, B., Steglich, C., Wilde, A., Vogel, J., et al. (2011b). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. P Natl Acad Sci USA *108*, 1–6.
Xanthomonas campestris	pv. vesicatoria 85-10	Schmidtke, C., Findeiss, S., Sharma, C.M., Kuhfuss, J., Hoffmann, S., Vogel, J., Stadler, P.F., and Bonas, U. (2012). Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. Nucleic Acids Res *40*, 2020–2031.

Table S2 Strains and studies depicted in Figure 3

Species	Strain	Study
Agrobacterium fabrum	C58	Wilms, I., Overlöper, A., Nowrousian, M., Sharma, C.M., and Narberhaus, F. (2012). Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. RNA Biol *9*, 446–457.
Bacillus anthracis	Sterne	Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., and Bergman, N.H. (2009). Structure and complexity of a bacterial transcriptome. J Bacteriol *191*, 3203–3211.
Bacillus licheniformis	DSM13	Chapter B
Bacillus subtilis	subsp. *subtilis* 168	Irnov, I., Sharma, C., Vogel, J., and Winkler, W. (2010). Identification of regulatory RNAs in *Bacillus subtilis*. Nucleic Acids Res *38*, 6637–6651.
Chlamydia trachomatis	L2b	Albrecht, M., Sharma, C., Reinhardt, R., Vogel, J., and Rudel, T. (2010). Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res *38*, 868–877.
Chlamydophila pneumoniae	CWL029	Albrecht, M., Sharma, C.M., Dittrich, M.T., Müller, T., Reinhardt, R., Vogel, J., and Rudel, T. (2011). The transcriptional landscape of *Chlamydia pneumoniae*. Genome Biol *12*, R98.
Helicobacter pylori	26695	Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature *464*, 250–255.

Table S2 Continued

Species	Strain	Study
Legionella pneumophila	Paris	Sahr, T., Rusniok, C., Dervins, Ravault, D., Sismeiro, O., J-Y, C., and Buchrieser, C. (2012). Deep sequencing defines the transcriptional map of *L. pneumophila* and identifies growth phase-dependent regulated ncRNAs implicated in virulence. RNA Biol *9*, 503–519.
Listeria innocua	Clip11262	Wurtzel, O., Sesto, N., Mellin, J.R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., and Sorek, R. (2012). Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. Mol Syst Biol *8*, 1–14.
Listeria monocytogenes	EGD-e	Wurtzel, O., Sesto, N., Mellin, J.R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., and Sorek, R. (2012). Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. MolSystBiol *8*, 1–14.
Methanosarcina mazei	Gö1	Jäger, D., Sharma, C., Thomsen, J., Ehlers, C., Vogel, J., and Schmitz, R. (2009). Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. P Natl Acad Sci USA *106*, 21878–21882.
Photobacterium profundum	SS9	Campanaro, S., Pascale, F. De, Telatin, A., Schiavon, R., Bartlett, D.H., and Valle, G. (2012). The transcriptional landscape of the deep-sea bacterium *Photobacterium profundum* in both a *toxR* mutant and its parental strain. BMC Genomics *13*, 567.
Pseudomonas aeruginosa	UCBPP-PA14	Dötsch, A., Eckweiler, D., Schniederjans, M., Zimmermann, A., Jensen, V., Scharfe, M., Geffers, R., and Häussler, S. (2012). The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. PloS One *7*, e31092.
Salmonella enterica	subsp. *enterica*sv. Typhimurium SL1344	Ramachandran, V.K., Shearer, N., Jacob, J.J., Sharma, C.M., and Thompson, A. (2012). The architecture and ppGpp-dependent expression of the primary transcriptome of *Salmonella* Typhimurium during invasion gene expression. BMC Genomics *13*, 25.
Streptomyces coelicolor	A3(2)	Vockenhuber, M.-P., Sharma, C.M., Statt, M.G., Schmidt, D., Xu, Z., Dietrich, S., Liesegang, H., Mathews, D.H., and Suess, B. (2011). Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. RNA Biol *8*, 468–477.
Sulfolobus solfataricus	P2	Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. Genome Res *20*, 133–141.
Synechococcus elongatus	PCC7942	Vijayan, V., Jain, I.H., and O'Shea, E.K. (2011). A high resolution map of a cyanobacterial transcriptome. Genome Biol *12*, R47.
Synechocystis sp.	PCC 6803	Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voß, B., Steglich, C., Wilde, A., Vogel, J., et al. (2011b). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. P Natl Acad Sci USA *108*, 1–6.
Xanthomonas campestris	pv. vesicatoria 85-10	Schmidtke, C., Findeiss, S., Sharma, C.M., Kuhfuss, J., Hoffmann, S., Vogel, J., Stadler, P.F., and Bonas, U. (2012). Genome-wide transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative virulence functions. Nucleic Acids Res *40*, 2020–2031.

SUMMARY

Bacillus licheniformis has been employed as industrial workhorse since several decades, especially for the production of alkaline serine proteases like Subtilisin Carlsberg, which is used as additive inhousehold detergents. Since the progress in second-generation sequencing technologies facilitated the determination of whole bacterial transcriptomes by RNA-Seq approaches, in-depth analyses of different stages of a fermentation process are now possible and allow new insights into bacterial regulation to accomplish optimization of production strains and bioprocesses.

In total, 15 samples of *B. licheniformis* DSM13 were derived from industrial-oriented fermentations, aiming at the production of Subtilisin Carlsberg. The samples were taken at five different time points during three independent fermentation processes. Whole transcriptome analysis yielded $2.4x10^7$ to $4.3x10^7$ RNA-Seq reads per sample. Analysis of these results and the utilization of public databases enabled the complete reannotation of the genome of *B. licheniformis*. Subsequently, the development of suitable prediction algorithms, led to the identification of 2798 untranslated regions and 461 non-coding RNAs in the generated transcriptome data sets. Thorough examination showed that 14% and 89% of these RNA features, respectively, are located in antisense orientation to a gene on the opposite strand, revealing a previously not expected wealth of putative antisense transcription-based regulation. Furthermore, the analyses allowed the confirmation and new identification of *cis*-regulatory elements, and the determination of several new RNAs, which are either expressed independently or located in intergenic regions. Cluster analysis of the transcriptome data allowed the assignment of differentially expressed protein and ncRNA genes to distinct expression patterns.

Moreover, differential RNA-Seq analysis of five of the aforementioned samples allowed the identification of 1500 transcription start sites. 408 of these are not located in a common promoter region, pinpointing additional regulatory sites. The integration of dRNA-Seq and

whole transcriptome RNA-Seq data made it possible to generate the first experimentally verified operon map of *B. licheniformis*.

Proteome analysis of the fermentation samples, based on 2D gel fractionation and subsequent assignment by mass spectrometry, yielded 367 protein spots representing 260 different proteins. Together with the obtained gene expression values, these data enabled the thorough examination of carbon and nitrogen metabolism, stress responses, sporulation, secretion capacities and cell differentiation, in order to provide an overview on the complex dynamics of *B. licheniformis* during an industry-oriented fermentation process.

The comprehensive analysis of all generated data facilitated the identification of putative targets for bioprocess and strain optimization approaches. These targets comprise, amongst others, the Tat-secretory pathway, the cell differentiation cascade and the antisense transcript located opposite to *apr*, which encodes the native Subtilisin Carlsberg preprotein.

Finally, the accomplished complete sequencing and annotation of the 3.6 Mb genome of *Geobacillus* sp. GHH01, a thermophilic, lipase-secreting member of the family *Bacillaceae*, grants access to a new possible industrial production platform for high-temperature applications.

ACKNOWLEDGEMENTS

Mein Dank gilt meinem Doktorvater Prof. Rolf Daniel für die Übernahme des Referates und seine oft kurzen, aber prägnanten Kommentare, die halfen diese Arbeit und die damit verbundenen Veröffentlichungen auf den richtigen Weg zu bringen. Desweiteren möchte ich meinem Korreferenten Prof. em. Gerhard Gottschalk, der es stets schafft mir einen guten Rat mit auf den Weg zu geben, herzlich für sein Interesse und seine Zeit danken. Den Mitgliedern der Prüfungskommission Prof. Pöggeler, Jun.-Prof. Heimel, PD Dr. Hoppert und PD Dr. Kramer möchte ich für ihre Bereitschaft diese Aufgabe zu übernehmen ebenfalls herzlich danken.

Ein ganz spezieller Dank gilt vor allem Dr. Heiko Liesegang für die Ermöglichung dieser Arbeit im weitesten Sinne. Dazu gehörte nicht nur die Konzipierung des Projektes, sondern auch eine intensive Betreuung, verbunden mit zahllosen Diskussionen und Vorschlägen. Danke auch für die zahlreichen Lektionen in positivem Denken und angewandter Diplomatie.

Weiterhin bedanken möchte ich mich bei Dr. Johannes Bongaerts, Stefan Evers und Ayhan Aydemir, die mir die Probenahme bei der Firma Henkel ermöglichten. Dr. Henning Hellmuth danke ich für die schnelle und unkomplizierte Unterstützung bei der Freigabe der erstellten Schriften. Bei Dr. Birgit Voigt, Antje Fengler, Stefan Handtke, Dr. Dirk Albrecht und Prof. Dr. Michael Hecker von der Universität Greifswald möchte ich mich für die Möglichkeit die Proteomanalysen in etabliertem Umfeld durchzuführen, sowie für die fortwährende Unterstützung und eine schöne Zeit bedanken. Desweiteren gilt mein Dank Prof. Dr. Wolfang Streit, Ulrich Rabausch und Dr. Jennifer Chow von der Universität Hamburg für die Überlassung des *Geobacillus*-Stammes zur Sequenzierung. Außerdem möchte ich Dr. Andrea Thürmer, Frauke-Dorothee Meyer, Maik Schlieper und im Besonderen Stefanie Offschanka für ihre Unterstützung der Arbeiten hier in Göttingen danken.

Ganz besonders herzlich bedanken möchte ich mich bei Dr. Anja Poehlein und Dr. Sonja Volland, die mir beide stets mit fachlicher Hilfe, aber vor allem auch mit moralischer Unterstützung, weiterhalfen. Danke, dass ihr auf mich aufgepasst habt.

Desweiteren gilt mein Dank meinen beiden Projekt-Mitstreitern Sascha Dietrich und Robert Hertel, die es mit ihrer Herzlichkeit schafften jede Schrulligkeit vergessen zu machen und mir zu wertvollen und liebgewonnenen Kollegen geworden sind.

Zuletzt, dafür aber besonders innig, möchte ich mich bei Marvin Djukic, Dr. Katrin Hartwich, Andreas Leimbach und John Vollmers dafür bedanken, mit mir geduldig jedes noch so kleine Problem gelöst und am Ende des Tages immer noch einen Grund zum Lachen, Weinen oder Biertrinken gefunden zu haben.

PROMOVIERENDEN-ERKLÄRUNG
DER GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Name: Wiegand, Sandra
Anschrift: Nonnenstieg 51, 37075 Göttingen

Ich beabsichtige, eine Dissertation zum Thema „RNA-Seq and proteomics based analysis of regulatory RNA features and gene expression in *Bacillus licheniformis*“ an der Georg-August-Universität Göttingen anzufertigen. Dabei werde ich von Herrn Prof. Rolf Daniel betreut.

Ich gebe folgende Erklärung ab:

1. Die Gelegenheit zum vorliegenden Promotionsvorhaben ist mir nicht kommerziell vermittelt worden. Insbesondere habe ich keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.
2. Hilfe Dritter wurde bis jetzt und wird auch künftig nur in wissenschaftlich vertretbarem und prüfungsrechtlich zulässigem Ausmaß in Anspruch genommen. Insbesondere werden alle Teile der Dissertation selbst angefertigt; unzulässige fremde Hilfe habe ich dazu weder unentgeltlich noch entgeltlich entgegengenommen und werde dies auch zukünftig so halten.
3. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen werden von mir beachtet.
4. Eine entsprechende Promotion wurde an keiner anderen Hochschule im In- oder Ausland beantragt; die eingereichte Dissertation oder Teile von ihr wurden nicht für ein anderes Promotionsvorhaben verwendet.

Mir ist bekannt, dass unrichtige Angaben die Zulassung zur Promotion ausschließen bzw. später zum Verfahrensabbruch oder zur Rücknahme des erlangten Grades führen.

Göttingen, den 29.08.2013