university of groningen

University Medical Center Groningen

University of Groningen

Bacillus subtilis: sporulation, competence and the ability to take up fluorescently labelled DNA

Boonstra, Mirjam

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Boonstra, M. (2017). *Bacillus subtilis: sporulation, competence and the ability to take up fluorescently labelled DNA*. Rijksuniversiteit Groningen.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 21-06-2022

Chapter 3

Analyses of the competent and non-competent subpopulations of *B. subtilis* reveal *yhfW*, *yhxC* and ncRNAs as novel players in competence.

Mirjam Boonstra[1], Marc Schaffer[2], Ulrike Mäder[2], Petra Hildebrandt[2], Praveen Kumar Sappa[2], Anne de Jong[1], Joana Sousa[3], Michael Lalk[3], Uwe Völker[2], Oscar P. Kuipers[1*]

1. Molecular genetics group, Groningen Biomolecular Sciences and Biotechnology Institute (GBB) University of Groningen; *o.p.kuipers@rug.nl
2. Department of Functional Genomic, Interfaculty Institute of Genetics and Functional Genomics, Ernst-Moritz-Arndt University, Greifswald, Germany
3. Department of Cellular Biochemistry/ Metabolomics, Institute of Biochemistry, Ernst-Moritz-Arndt University, Greifswald, Germany

Abstract

One of the intriguing features of *Bacillus subtilis* is its ability to become naturally competent and take up exogenous DNA. Under competence stimulating conditions only part of population becomes competent. As yet little is known about differential expression of non-coding RNAs and differential protein levels of the competent and non competent subpopulations. It is also not entirely clear if ComK is capable of direct down regulation of genes or if it is solely a transcriptional activator. To address these questions, we separated the competent and non-competent subpopulations by use of FACS and used transcriptome sequencing (RNA-seq) and proteome analysis (ESI-MS) to compare gene expression and protein levels within the subpopulations. We found significant up-regulation of expression of non-coding RNAs during competence, many of which contain ComK binding sites (K-boxes) in their promoter region. Thus, antisense ncRNAs play a role in down-regulation of specific genes during competence. By investigating the proteome during competence we found that the important competence factors MinD, Noc, SbcC and SbcD are regulated at the post-transcriptional level. We also identified *yhfW* as a gene the absence of which leads to a decrease in the expression of *comG*. We found that *yhfW* has a modest but statistically significant effect on *comK* expression, and also changes expression of another important regulator gene, *srfA*. Metabolomic analysis of *ΔyhfW* under competence stimulating conditions reveals significant reductions in TCA cycle components and several amino acids. Interestingly in the *ΔyhfW* strain there is higher expression of the NAD synthesis genes *nadA, nadB* and *nadC*. Absence of YhfW causes a significant increase in the expression of Spo0A under sporulation conditions. Inactivation of the neighbouring gene of *yhfW, yhxC* decreases the total number of competent cells approximately by half. Inactivation of *yhxC* also effects the expression of *comK* and *srfA*. Together, our data provide a unique high resolution view of competence development in *B. subtilis* and reveal a number of novel factors involved in this process.

Introduction

Bacillus subtilis is a Gram positive soil bacterium capable of developing natural competence. During the competence state it can take up exogenous DNA from the environment. Competent bacteria no longer undergo cell division and are inhibited in their ability to synthesize new DNA (Briley Jr et al., 2011; Hahn et al., 2015; Haijema et al., 2001; Mirouze et al., 2015). Under nutrient limited conditions in the lab about 5-25 % of a *B. subtilis* 168 population becomes competent. The genes involved in competence are under control of the master regulator of competence, ComK. The competence state (K-state) of *B. subtilis* has previously been studied with microarray techniques (Berka et al., 2002; Hamoen et al., 2002; Ogura et al., 2002) and LacZ fusions (Ogura et al., 2002). To overcome the problem posed by the small fraction of competent cells, these studies compared *comK* and/or *mecA* deletion mutants with WT strains. In these studies *comK* was found to be a transcriptional activator with only a few potential, but not significantly, down-regulated genes found.
Despite the thorough investigations done in previous research, several questions still remain unanswered, *e.g.*: Is there down-regulation of genes in the competent subpopulation? Is there differential expression of non-coding RNAs during competence? Are there competence proteins for which regulation takes place primarily at the post transcriptional level?Here we use deep RNA sequencing, proteomics and metabolomics to address these questions. In contrast to previous studies we physically separated the two subpopulations using a competence specific GFP reporter (P_{comG}-*gfp*) in which the competent subpopulation expressing P_{comG} will be fluorescent (Smits et al., 2005). We quickly fixed the cells using sodium chloride to prevent degradation of RNA (Brown and Smith, 2009; Nilsson et al., 2014) and used FACS to separate the competent and non competent subpopulations. Separating and comparing the two subpopulations allowed us to study a more natural situation than is created when using knock-outs and allowed us to better determine whether significant down-regulation occurs in the competent sub-population.

Because of the higher sensitivity of RNA-seq compared to microarray analyses we could also gain a better insight in operon expression, sRNA occurrences and changes in weakly expressed genes that may not be detected by standard microarray technology. Recently, the expression and function of non-coding RNAs in *B. subtilis* has gained keen interest. Strain 168 is predicted to harbour approximately 100 potential small RNAs (Irnov et al., 2010). More non-coding RNAs have been found in the condition-depended transcriptome study by Nicolas (Nicolas et al., 2012).
However, little is known about the expression of ncRNAs in *B. subtilis* during competence and we therefore decided to examine if differential expression of non-coding RNAs occurs during competence. In the previous gene expression studies certain genes known to be involved in competence were not found differentially expressed at the RNA level. We therefore decided to also investigate protein levels between the two subpopulations in order to determine if regulation of some of the competence factors occurs primarily on a post-transcriptional level.
Under the conditions used by us, *B. subtilis* becomes competent after approximately 5 hours and transformability is highest during a two hour window after entrance into the competent state. In this time window we investigated changes in expression of the competent subpopulation and the non-competent subpopulations. We discovered several new genes differentially expressed during competence in our study, as well as ncRNAs responsible for down regulation of important genes, and we decided to investigate the role in competence of the most highly up regulated new gene *yhfW* and its neighbouring gene *yhxC* in more detail.

Results

Differential expression of protein encoding genes

To gain insight into the progression of the competence state, *B. subtilis* 168 p_{comG}-*gfp* was grown in competence medium and cells were sorted by FACS. Samples were taken early in the competence state at 5.5hrs and at a later stage at 6.5hrs in order to gain insight in the progression of competence.

We subsequently compared the transcriptome and proteome of the competent subpopulation with those of the non-competent subpopulation at both time points.
To prevent degradation of RNA the cells were fixed with 2M NaCl in PBS before FACS and sorted in to 4M NaCl in PBS (Brown and Smith, 2009; Nilsson et al., 2014). The NaCl fixation method was tested by microarray and RNA was isolated as described in (Nicolas et al., 2012). To exclude a difference in sporulation initiation under these conditions, we looked specifically at expression of sporulation genes. We did not find a significant difference between the two subpopulations with respect to entry into sporulation. After transcriptome analysis of the two subpopulations by T-REx (de Jong et al., 2015) a total of 135 genes were found differentially expressed between the competent and non-competent subpopulations at 5.5 hours (Table 1a) and 106 genes at 6.5 hours when using a cut off value of 2-fold and an adjusted P-value of 0.05. The fold-difference in gene expression at the RNA level was higher at the early time point, which was expected since the changes in gene expression would be greatest upon switching from one state to another. Our results were in accordance with previous studies with regard to the core ComK regulon. Some of the genes found previously were not found in our results. This was in part expected due to the use of different strains of *B. subtilis*, different media and possibly because of the use of knock out mutants used in previous studies.
Because of the high sensitivity of RNA-seq we also found several extremely high fold change genes with a very low expression. We did not investigate these genes further. In total we found 44 new genes for the first time point and 38 in the second. Newly found highly up-regulated genes in the first time point (>5fold) are *yhfW*, *phrH*, *ygaK*, *ccpB*, *yvqJ*, *rsoA*, *ydeB*, *clpE*, *maa*, *ybzI*, *sacB* and *yeeI*. Some of these genes such as *phrH*, *ccpB, maa* and *ybzI* are part of operons also found in previous studies. As in previous studies we found a large number of genes of unknown function as well as genes involved in metabolism. When only the protein coding genes are used in comparing the competent and non-competent subpopulations a total of 12 known regulatory genes were found differentially expressed.

For nine of these we also found other genes in their regulon up-regulated. Only for *mta*, *ftsR* and *licT* no other genes in their regulon were differentially expressed. Unlike previous studies, we found also significantly down-regulated genes, primarily at the first time point with *jag* being the only gene down regulated at both time points. Two of the down-regulated genes, i.e. *ywdK* and *degS* have not been previously identified. The detection of *degS* was likely due to the sensitivity of RNA-seq as the other gene in the operon *degU* has been previously detected. Four of the down-regulated genes in this study were found up-regulated by Berka et al. and two by Hamoen et al. These are *degU*, *sigA*, *jag* and *lipL* (*ywfL*). None of these genes contain a K-box in the promoter region. Deletion of jag, the only gene found down-regulated at both time-points, did not result in a change in competence.

Table 1. Differential expression of protein encoding genes at the first time point, after 5.5hrs growth in competence medium. The last column indicates if the gene was found differentially expressed in the previous transcriptomics studies B: Berka, H: Hamoen, O: Ogura (Berka et al., 2002; Hamoen et al., 2002; Ogura et al., 2002). Table continues on following pages

gene	fold	description	previous
ssbB	199.3	single-stranded DNA-binding protein SsbB	B,H,O
comGG	158.2	ComG operon protein 7	B,H,O
comEA	156.6	ComE operon protein 1	B,H,O
comGD	147.0	ComG operon protein 4	B,H,O
comC	139.0	type 4 prepilin-like proteins leader peptide-processing enzyme	B,H,O
comGA	129.7	ComG operon protein 1	B,H,O
comGE	129.2	ComG operon protein 5	B,H,O
comGF	119.5	ComG operon protein 6	B,H,O
comGB	119.1	ComG operon protein 2	B,H,O
comFA	116.2	ComF operon protein 1	B,H,O
comGC	101.4	ComG operon protein 3	B,H,O
yqzE	90.6	hypothetical protein	B,H
nucA	73.1	DNA-entry nuclease	B,H,O
comFB	63.2	ComF operon protein 2	B,H,O
comFC	58.1	ComF operon protein 3	B,H,O
ybdK	56.3	sensor histidine kinase	B,H,O
yhfW	52.5	rieske 2Fe-2S iron-sulfur protein YhfW	
rapH	51.9	response regulator aspartate phosphatase H	B,H,O
nin	51.9	DNA-entry nuclease inhibitor	B,H,O
comEB	51.2	ComE operon protein 2	B,H,O
phrH	48.0	inhibitor of regulatory cascade	

gene	fold	description	previous
comEC	47.8	ComE operon protein 3	B,H,O
sacX	46.4	negative regulator of SacY activity	B,H,O
radC	45.3	hypothetical protein	B,H,O
yyaF	44.4	GTP-dependent nucleic acid-binding protein EngD	B,H,O
maf	42.7	septum formation protein Maf	B,H,O
exoAA	39.9	exodeoxyribonuclease	B,H,O
comK	39.4	competence transcription factor	B,H,O
dprA	31.9	protein smf	B,H,O
yvyF	30.9	hypothetical protein	B,H,O
hxlR	30.0	HxlR family transcriptional regulator	B,H
spoIIB	29.1	stage II sporulation protein B	H
sacY	24.9	levansucrase and sucrase synthesis operon antiterminator	H,O
coiA	23.9	competence protein CoiA	H,B
ygaK	20.4	FAD-linked oxidoreductase YgaK	
rpsF	18.2	30S ribosomal protein S6	O
yhcE	17.7	hypothetical protein	B
yviE	15.9	hypothetical protein	B
ssbA	15.1	single-stranded DNA-binding protein	B
yhcF	14.0	GntR family transcriptional regulator	H,O
rpsR	13.8	30S ribosomal protein S18	B
yvrO	13.8	ABC transporter ATP-binding protein	B
recA	13.8	recombinase RecA	B,H,O
yckB	12.0	ABC transporter extracellular-binding protein YckB	B,H
glcR	11.6	HTH-type transcriptional repressor GlcR	H,O
ccpB	11.5	catabolite control protein B	
yckA	11.4	amino acid ABC transporter permease protein YckA	B,H
ycbP	11.0	hypothetical protein	B,H
hxlB	10.9	3-hexulose-6-phosphate isomerase	B,H
guaD	10.1	guanine deaminase	H
yvqJ	10.0	MFS transporter	
hxlA	9.9	3-hexulose-6-phosphate synthase	B,H
yviF	9.6	flagellar assembly factor FliW	B
groES	9.4	10 kDa chaperonin	B,H
flgM	9.3	negative regulator of flagellin synthesis	B,H
yvyG	9.2	hypothetical protein	B,H
ydeE	9.1	AraC family transcriptional regulator	B
ywpJ	9.0	phosphatase YwpJ	B.H
rsoA	8.7	sigma-O factor regulatory protein RsoA	
flgK	8.6	flagellar hook-associated protein 1	B
oxdC	8.5	oxalate decarboxylase OxdC	B
csrA	8.4	carbon storage regulator homolog	B
flgL	8.4	flagellar hook-associated protein 3	B
yhcG	8.1	spermidine/putrescine ABC transporter ATP-binding protein	O
med	7.8	transcriptional activator protein med	B,H
ydeB	7.6	transcription factor YdeB	

gene	fold	description	previous
groEL	7.5	60 kDa chaperonin	
yhcH	7.3	ABC transporter ATP-binding protein	H,O
clpE	7.1	ATP-dependent Clp protease ATP-binding subunit ClpE	
maa	7.0	maltose O-acetyltransferase	
yvhJ	6.8	transcriptional regulator YvhJ	B
trpB	6.8	tryptophan synthase beta chain	B,H
yhcI	6.6	hypothetical protein	B,H
yvyE	6.6	IMPACT family member YvyE	B
yvrN	6.5	ABC transporter permease	B
yvrP	6.2	efflux system protein YvrP	B,H,O
sucC	6.2	succinyl-CoA ligase [ADP-forming] subunit beta	H
ybzI	6.1	hypothetical protein	
gid	6.0	TrmFO	
sucD	5.8	succinyl-CoA ligase [ADP-forming] subunit alpha	B,H
sacB	5.6	levansucrase	
topA	5.6	DNA topoisomerase 1	B,H
comZ	5.5	ComG operon repressor	B,H
yeeI	5.1	transcriptional regulator	
ypzG	5.0	hypothetical protein	
ybdJ	5.0	transcriptional regulator	
spbC	4.7	killing factor SdpC	
yjcM	4.4	hypothetical protein	
yopL	4.4	hypothetical protein	
ydzE	4.2	permease	
radA	4.1	DNA repair protein RadA	
ymzE/2	3.9	Pseudogene	
yxiP	3.9	hypothetical protein	H
holA	3.9	hypothetical protein	
eglS	3.9	endoglucanase	
sigO	3.9	RNA polymerase sigma factor SigO	
yoqW	3.8	hypothetical protein	
yomL	3.7	hypothetical protein	B
yjiA	3.7	hypothetical protein	
yuaG	3.6	hypothetical protein	H
parA	3.6	sporulation initiation inhibitor protein Soj	
mta	3.5	HTH-type transcriptional activator mta	
yocI	3.4	ATP-dependent DNA helicase RecQ	
ywtF	3.4	transcriptional regulator YwtF	B
parB	3.4	stage 0 sporulation protein J	
gidB	3.4	ribosomal RNA small subunit methyltransferase G	B
ycgP	3.3	hypothetical protein	
ytzJ	3.3	hypothetical protein	
ftsR	3.2	LysR family transcriptional regulator	
ywfI	3.2	heme peroxidase	B,H
ykuK	3.1	hypothetical protein	B,H

gene	fold	description	previous
bdbC	3.1	disulfide bond formation protein C	B,H
dinB	3.0	protein DinB	B,H
hrcA	2.8	heat-inducible transcription repressor HrcA	
yeeK	2.8	spore coat protein YeeK	
bdbD	2.8	disulfide bond formation protein D	B,H
mcsA	2.7	hypothetical protein	
licT	2.7	transcription antiterminator LicT	
bpr	2.7	bacillopeptidase F	
trmE	2.6	tRNA modification GTPase MnmE	O
ywhH	2.5	hypothetical protein	B
gidA	2.5	MnmG	
mreB	2.4	rod shape-determining protein MreB	B
yfhB	2.4	isomerase	B
mcsB	2.4	ATP:guanido phosphotransferase YacI	
yddT	2.3	hypothetical protein	
yfhC	2.3	NAD(P)H nitroreductase YfhC	B
comN	2.2	post-transcriptional regulator	
aroD	2.2	3-dehydroquinate dehydratase	
degS	-2.3	signal transduction histidine-protein kinase/phosphatase DegS	(B up)
sigA	-2.4	RNA polymerase sigma factor RpoD	
ywdK	-2.5	hypothetical protein	(B up)
DegU	-2.7	transcriptional regulatory protein DegU	(B up)
jag	-2.8	protein jag	(B,H up)
ywfL	-3.0	octanoyl-[GcvH]:protein N-octanoyltransferase	(B,H up)

Expression patterns of non-coding RNAs

Little is known about differential expression within the subpopulations of non-coding RNAs during competence. We therefore decided to look at their expression patterns. We found at T1 a total of 36 elements (47 including high fold-change, but low expression) of which 17 are antisense RNAs (Table 2) and at T2, 25 differentially expressed ncRNAs. The previously found up-regulated genes *degU, sigA, jag* and *lipL* were found to have up-regulated anti-sense RNAs instead of up-regulated mRNAs, which caused the misinterpretation found in previous microarray analyses. The detection of these genes as being up-regulated in the previous studies was likely the result of the use of amplicon arrays in these studies. Because the probes are made from double stranded (dsDNA) it cannot distinguish between sense and antisense DNA.

Hamoen et al. already determined that *comER* was one of these false positives, and indeed we found up-regulation of the antisense *comER* RNA (S963), but not of *comER* itself. Hamoen et al. also found up-regulation of *lipL* and *ywfM* and Berka et al. up-regulation of *lipL*, *ywfM,* and *pta;* these genes form an operon consisting of *pta, cysl, lipL* and *ywfM*. In fact Berka et al. investigated expression of *pta* during competence using a promoter-*lacZ* fusion, but could only see a small effect on expression during competence, whereas one of their array comparisons suggested a robust effect. As with *comER* this operon is covered by a large antisense RNA called S1458, which we find up-regulated.
To determine whether expression of the ncRNAs could be controlled by ComK we looked at the presence of potential K-boxes in their respective promoter regions using genome2DTFBS. We found potential K-boxes for 7 of the ncRNAs within the first 100bp upstream region and 2 ncRNAs with K-boxes within the first 300bps (Table 2). 10 ncRNAs are preceded by competence genes with K-boxes in their respective promoter regions. The majority of the antisense RNAs are preceded by potential K-boxes. We did not find ncRNAs with a K-box at the second time point that were not present at the first time point. Although we found 17 antisense RNAs, only four of the up-regulated antisense RNAs have corresponding down-regulated genes. Down-regulation of *degU* could be a direct result of the up-regulation of its anti-sense RNA S1354 as DegU has been shown to directly regulate its own expression (Ogura and Tsukahara, 2010).
The down-regulated *lipL* has been shown to be essential for biosynthesis of lipoic acid, an enzyme co-factor (Christensen et al., 2011; Martin et al., 2011). One of the most important enzyme complexes requiring lipoic acid is the pyruvate dehydrogenase complex (Perham, 2000). Recently it has been shown that the E1α subunit of this complex is important for correct Z ring formation, and the complex is indicated as playing a role in the coordination of nutrient availability and cell division (Monahan et al., 2014). S1458, the antisense RNA to *lipL,* also contains a potential K-box in the promoter region. S1458 is a very large antisense RNA covering four genes (*pta, cysl, lipL* and *ywfM*). In *Escherichia coli* deletion of *pta* has been found to rescue cell division in *dnaA, dnaB, dnaE, dnaG* and *dnaN* mutants (Maciąg-Dorszyńska et al., 2012).

S1579*, i.e.* the *jag* and *spoIIIJ* antisense RNA, was also up-regulated in our data. Up-regulated S951 is antisense to *sigA* and partially overlaps *dnaG*. The only down-regulated gene not covered by an antisense RNA was *ywdK*. *ywdK* has previously been found to be negatively regulated by TnrA under Nitrogen limiting conditions (Yoshida et al., 2003).
The presence of up-regulated anti-sense RNA and the down-regulation of the genes confirm that the up-regulation of these genes found in the previous studies were false positives of the same nature as found for *comER*. Aside from the antisense RNAs there are also many 3'UTR and 5'UTR RNAs up-regulated. Some of these are at the 3’ or 5' end of up-regulated genes, but others are in front of or behind genes that have no known relation to competence. Interestingly, the competence gene *noc* that has not been found to be differentially expressed in any of the transcriptome studies is preceded by the ncRNA S1577, which is up-regulated in the competent subpopulation.

Table 2. Differential expression of ncRNAs at the first time point. The description is taken from Nicolas et al. 2012 The second last column indicates if the ncRNA has a K-box predicted by Genome2D TFBS. The type of K-box was manually determined according to the specifications used by Hamoen et al., 2002. The last column indicates the distance of the K-box to the start of the transcript, measured from the end of the K-box to the start codon.

name	fold	antisense	description	K-box	bp distance to start transcript
S963	184.6	*comER*	5'UTR of *comEA*	II-14	31
S962	173.6	*yqzM*	independent transcript	*comE*	
S1354	167.8	*degU*	independent transcript	I-13	65
S1458	166.4	*pta*	5'UTR of *hemQ*	I-15	29
S98	121.5	*cwlJ*	5'UTR of *ycbP*	II-14	0
S122	117.4	*bglC*	intergenic region	*nucA*	
S125	113.2	*tlpC*	5'UTR of hxlR	II-13	95
S1399	100.8		3'UTR of ssbB	*ssbB*	
S652	98.1	*yndK*	3' of S653	no	
S1579	96.6	*spoIIIJ*	independent transcript	II-15	5
S97	93	*ycbO*	3'UTR of *ycbP*	no	
S925	80.3	*yqzG*	3'UTR of *yqzE*	*comG*	
S245	43.4		intergenic region	*rapH*	

name	fold	antisense	description	K-box	bp distance to start transcript
S1357	32.3		5'UTR of *yvyE*	no	
S1575	27.9		5'UTR of *rpsF*	no	
S401	26	*yjzB*	intergenic region	*med*	
S1175	24.2		5'UTR of *mntA*	II-15	51
S1353	22.3		intergenic region	*comF*	
S366	22.1	*yhxD*	intergenic region	*comK*	
S655	21.5	*yndL*	5' of S653	*no*	
S367	17.3	*yhxD*	intergenic region	comK	
S951	16.1	*sigA*	independent transcript	no	
S876	11.3	*aroC*	3'UTR of *serA*	no	
S1278	10.6		5'UTR of *oxdC*	no	
S583	10.2		5'UTR of *topA*	I-13	275
S653	9.6		independent transcript	no	
S208	8.9		5'UTR of *groES*	no	
S209	8.3		3'UTR of *groEL*	no	
S967	5.8		3'UTR of *sda*	no	
S959	4.6		intergenic region	no	
S30	4		5'UTR of *sspF*	no	
S1577	3.2		intergenic region	*trmE*	256
S174	3.1		3'UTR of *yddM*	no	
S515	2.8		intergenic region	no	
S296	-2.9		5'UTR of *yfhP*	no	
S488	-5.4		5'UTR of *ykvA*	no	

Differential protein levels between the competent and non-competent subpopulations

B. subtilis 168 cells, sampled at 5.5 and 6.5hrs, were sorted by FACS onto a filter manifold system. The filters were collected and stored at -80 ^{o}C. Samples were digested and analysed by ESI-MS. At the first time point we found 53 proteins to be differentially expressed, six of which were downregulated in the competent subpopulation (Table 3). The second time point had 94 differentially expressed proteins, 20 of which were down-regulated in the competent fraction (Table 4)

The higher number of proteins found at the second time point can be explained by the maturation time of the proteins. 23 of the proteins found in the first time point and 20 of the proteins found in the second time point were also found in the RNA-seq data.
None of the genes found down-regulated in the RNA data were found down-regulated at the protein level. This is likely caused by the shorter half live of mRNA compared to that of proteins. None of the down-regulated genes found in the protein data were found in the RNA-seq data. Most of the down regulated proteins are involved in metabolism, with a few unknown genes at the second time point.
As we expected, some of the proteins involved in competence that were not found differentially expressed at the RNA level in our and the previous studies, were found in the proteomics data. For some of these proteins the gene is part of an operon, where other genes were differentially expressed at the RNA level. The gene of cell division inhibitor MinD lies in an operon with *mreB, radC* and *maf*. The gene *noc* is part of the *trmE* operon of which *thdF gidA* and *gidB* were also upregulated on a RNA level. The genes of SbcC, SbcD, however, do not reside in an operon of which the other genes were differentially expressed at the RNA level. Ogura et al. found differential expression of *sbcC* in one of their replicates, but none of the other studies found differential expression of these genes. This indicates that the regulation of SbcC and SbcD during competence takes place primarily on a post-transcriptional level. The only known competence genes that were not found differentially expressed in our study at either RNA or protein level were *addA, addB, hlpB, yhjB* and *yhjC.* With the exception of *yhjB* which has been found by Ogura et al, none of these genes have been found significantly up-regulated in any of the four studies. This strongly indicates that basal expression levels of these genes are sufficient for competence.
Other interesting proteins with higher levels in the competent subpopulation are the fatty acid biosynthesis proteins FabHA and FabF (5.5hrs), and FloT which is involved in regulation of membrane fluidity. Higher levels of these proteins may indicate a difference in fluidity and lipid composition of the membrane during competence.

In the same operon as the known competence gene *coiA* lies *pepF*, for which we found higher protein levels in the competent sub-population. Over-expression of *pepF* has been shown to inhibit sporulation initiation (Kanamaru et al., 2002).

Table 3. Difference in protein levels in the competent subpopulation at T1 Table continues on following pages

protein	logFC	description
ComEB	6.48	late competence protein required for DNA binding and uptake
NucA	6.24	catalyzes DNA cleavage during transformation
Nin	5.69	inhibitor of the DNA degrading activity of NucA
RecA	4.17	homologous recombination
SsbA	4.14	single-strand DNA-binding protein
YyaF	3.86	GTP-binding protein/ GTPase
FlgL	3.11	flagellar hook-associated protein 3 (HAP3)
FliW	2.78	checkpoint protein for hag expression, CsrA anatagonist
YdeE	2.64	similar to transcriptional regulator (AraC family)
YvrP	2.44	Unknown
TrmFO	2.35	tRNA:m(5)U-54 methyltransferase, glucose-inhibited division protein
Maa	1.96	maltose O-acetyltransferase
SucD	1.79	succinyl-CoA synthetase (alpha subunit)
SucC	1.70	succinyl-CoA synthetase (beta subunit)
YlbA	1.67	Unknown
FloT	1.59	involved in the control of membrane fluidity
TagT	1.57	phosphotransferase, attachment of anionic polymers to peptidoglycan
Noc	1.46	spatial regulator of cell division to protect the nucleoid
BdbD	1.41	required for the formation of thiol disulfide bonds in ComGC
Ffh	1.40	signal recognition particle (SRP) component
Spo0J	1.36	chromosome positioning near the pole, antagonist of Soj
SipW	1.25	signal peptidase I
GidA	1.24	glucose-inhibited division protein
ThdF	1.23	GTP-binding protein, putative tRNA modification GTPase
YckB	1.23	similar to amino acid ABC transporter (binding protein)
GrpE	1.21	heat-shock protein (activation of DnaK)
YfmM	1.17	similar to ABC transporter (ATP-binding protein)
YwfH	1.14	short chain reductase
SbcD	1.12	exonuclease SbcD homolog
MurB	1.10	UDP-N-acetylenolpyruvoylglucosamine reductase

protein	logFC	description
YdgI	1.05	similar to NADH dehydrogenase
YvbJ	1.01	Unknown
ClpY	1.01	two-component ATP-dependent protease, ATPase subunit
HemQ	0.99	heme-binding protein, essential for heme biosynthesis
FabHA	0.98	beta-ketoacyl-acyl carrier protein synthase III
ZosA	0.95	zinc transporter
HprT	0.93	hypoxanthine phosphoribosyltransferase
SwrC	0.91	similar to acriflavin resistance protein
GroEL	0.90	chaperonin and co-repressor for HrcA
FabF	0.89	Iinvolved in the control of membrane fluidity
YtsJ	0.83	malic enzyme
MinD	0.81	cell-division inhibitor (septum placement)
SbcC	0.79	DNA exonuclease
PepF	0.77	Oligoendopeptidase
DltC	0.76	D-alanine carrier protein
YtwF	0.70	Unknown
YqaP	0.68	Unknown
HisD	-0.80	histidinol dehydrogenase
PyrAA	-0.86	carbamoyl-phosphate synthetase (glutaminase subunit)
PheS	-0.99	phenylalanyl-tRNA synthetase (alpha subunit)
HisG	-1.12	ATP phosphoribosyltransferase
GudB	-1.23	trigger enzyme: glutamate dehydrogenase
AtpF	-0.83	ATP synthase (subunit b)

Table 4. Difference in protein levels in the competent subpopulation at T2 Table continues on following pages

protein	logfc	description
Nin	6.67	inhibitor of the DNA degrading activity of NucA
YyaF	4.45	GTP-binding protein/ GTPase
RecA	4.25	DNA repair/ recombination
FliW	3.16	checkpoint protein for hag expression, CsrA anatagonist
YvrP	3.06	unknown
ComGA	2.95	late competence gene, traffic ATPase
TrmFO	2.89	glucose-inhibited division protein
TopA	2.67	DNA topoisomerase I
RapA	2.40	dephosphorylates Spo0F-P, control of the phosphorelay
YwrO	2.35	similar to NAD(P)H oxidoreductase
Maa	2.31	maltose O-acetyltransferase
ThdF	2.10	GTP-binding protein, putative tRNA modification GTPase
Noc	2.06	spatial regulator of cell division to protect the nucleoid
Bcd	2.05	valine, isoleucine and L-leucine dehydrogenase
SucD	1.97	succinyl-CoA synthetase (alpha subunit)
TagV	1.97	attachment of anionic polymers to peptidoglycan
bdbD	1.94	required for the formation of thiol disulfide bonds in ComGC
SipW	1.87	signal peptidase I
FlgL	1.86	flagellar hook-associated protein 3 (HAP3)
YdeE	1.83	similar to transcriptional regulator (AraC family)
SucC	1.77	succinyl-CoA synthetase (beta subunit)
FabHA	1.75	beta-ketoacyl-acyl carrier protein synthase III
Spo0J	1.75	antagonist of Soj-dependent inhibition of sporulation initiation
InfC	1.70	translation initiation factor IF-3
SbcC	1.67	DNA exonuclease
Soj	1.63	negative regulation of sporulation initiation
RsmG	1.58	glucose-inhibited division protein
BacB	1.56	biosynthesis of the antibiotic bacilysin
YfhC	1.55	unknown
Ndk	1.50	nucleoside diphosphate kinase
YqhL	1.49	membrane protein
FloT	1.49	involved in the control of membrane fluidity
MetC	1.46	cystathionine beta-lyase
YpgR	1.45	unknown
YckB	1.43	similar to amino acid ABC transporter (binding protein)

protein	logfc	description
YydD	1.43	unknown
AnsB	1.38	L-aspartase
DltA	1.35	biosynthesis of teichoic acid
yjbL	1.34	unknown
GidA	1.31	glucose-inhibited division protein
YkuJ	1.29	unknown
YtxJ	1.28	general stress protein
SbcD	1.21	exonuclease SbcD homolog
YkvS	1.16	unknown
Ffh	1.13	signal recognition particle (SRP) component
McsB	1.11	modulator of CtsR-dependent repression
YvaK	1.10	general stress protein, similar to carboxylesterase
MurB	1.08	UDP-N-acetylenolpyruvoylglucosamine reductase
GroEL	1.08	chaperonin and co-repressor for HrcA
YeeI	1.07	unknown
HemQ	1.07	heme-binding protein, essential for heme biosynthesis
GroES	1.06	chaperonin, universally conserved protein
ClpY	0.99	two-component ATP-dependent protease, ATPase subunit
ThiC	0.95	biosynthesis of the pyrimidine moiety of thiamine
Rho	0.93	transcriptional termination protein
DapA	0.92	dihydrodipicolinate synthase
YkpA	0.90	similar to ABC transporter (ATP-binding protein)
EcsA	0.84	regulation of the secretion apparatus and of intra-membrane proteolysis
ClpC	0.83	ATP-dependent Clp protease, ATPase subunit
MinD	0.82	cell-division inhibitor (septum placement)
YvbJ	0.81	unknown
PepF	0.78	oligoendopeptidase, inhibits sporulation upon overexpression
HisS	0.77	histidyl-tRNA synthetase
YheA	0.76	unknown
YugI	0.72	unknown
YxkC	0.71	unknown
WalR	0.71	two-component response regulator, controls cell wall metabolism
AsnS	0.70	asparagyl-tRNA synthetase
DnaK	0.68	class I heat-shock protein (molecular chaperone)
YsaA	0.64	unknown
Zwf	0.62	glucose 6-phosphate dehydrogenase, pentose-phosphate pathway

protein	logfc	description
YorD	0.62	unknown
YwfH	0.61	short chain reductase
LeuD	-0.58	3-isopropylmalate dehydratase (small subunit)
MsrA	-0.59	peptide methionine sulfoxide reductase
PheT	-0.59	phenylalanyl-tRNA synthetase (beta subunit)
yxjH	-0.66	putative methionine synthase
YqzC	-0.71	unknown
AmhX	-0.72	amidohydrolase
YpsC	-0.75	similar to SAM-dependent 23S rRNA methyltransferase
YukE	-0.82	unknown
YxiE	-0.97	unknown
gpsA	-1.07	glycerol-3-phosphate dehydrogenase (NAD)
YmcB	-1.08	unknown
SdaAA	-1.12	L-serine deaminase
HisF	-1.23	cyclase-like protein
pyrB	-1.29	aspartate carbamoyltransferase
MetQ	-1.30	methionine ABC transporter (binding lipoprotein)
AlsS	-1.42	acetolactate synthase
PckA	-1.44	phosphoenolpyruvate carboxykinase
YvaQ	-2.09	membrane-bound chemotaxis receptor
YeeB	-2.23	unknown
GsaB	-2.56	glutamate-1-semialdehyde aminotransferase

Differential expression within the subpopulations at T1 and T2 at RNA and protein level

For both the competent and non-competent populations on the RNA level as well as the protein level The difference was primarily in genes and proteins involved in primary metabolism. More genes were significantly changed in expression between the two time points in the competent fraction compared to the non-competent fraction. At the first time point there was a higher expression of aminoacid biosynthesis genes, in particular arginine and histidine synthesis genes as well as increased expression of phosphate uptake genes. Protein levels of arginine, asparagine, leucine, methionine and threonine synthesis proteins were more abundant at T1.

Higher levels of amino-acid synthesis may be the result of increased demand for the production of the large competence machinery. There were also higher levels of proteins involved in the synthesis of purine and pyrimidine in both subpopulations at the first time point. Not only were there higher levels of amino acid synthesis proteins at the first time point, the levels of RpsJ and RpsL, i.e. the ribosomal s10 and s12 proteins were also higher. At T2 a significant down-regulation of *groES* and RNA binding protein *hfq* in the competent subpopulation at the RNA level was found. In the non-competent subpopulation there was also significant up-regulation of *rsoA, comGD, radC* and *ssbB* and down-regulation of *fliT* at T2. On the protein level at T2 the RNA polymerase beta subunit proteins RpoB and RpoC had increased levels in both subpopulations. There were decreased levels of SrfAA, SrfAC and SrfAD in both the competent and non-competent subpopulations at T2 whereas levels of MecA were increased in the second time point, which one might expect since competence is a transient state requiring removal of ComK to leave it. With regards to differential expression of ncRNAs between both time point there was very little difference with only. S963 that is antisense to *comER* was the only antisense RNA significantly changed in expression between the two time points in the non-competent subpopulation.

Investigation into yhfW and yhxC

Among the newly found genes in competence, *yhfW* was the highest up-regulated new gene. The only gene found down-regulated at both time points is *jag*. Deletion of *jag* did not result in a significant change in competence. As *yhfW* is up-regulated to a similar level as known competence genes we hypothesised that *yhfW* is involved in competence and that a deletion would lead to a reduction in competence. However, deletion of the *yhfW* gene did not lead to a strong decrease in the amount of competent cells, rather the expression of *comG* was significantly reduced (Fig. 1A). To determine how YhfW might be affecting competence we looked at the effect of inactivation of *yhfW* on the expression of known competence regulators.

We tested the expression of *comK*, *srfA, spooA* and *rok*. In the mutant the *comK* (Fig. 1B) expressing population was larger, but the intensity is slightly reduced when maximum competence is achieved.
This difference was statistically significant before full formation of the competent and non-competent sub-populations.

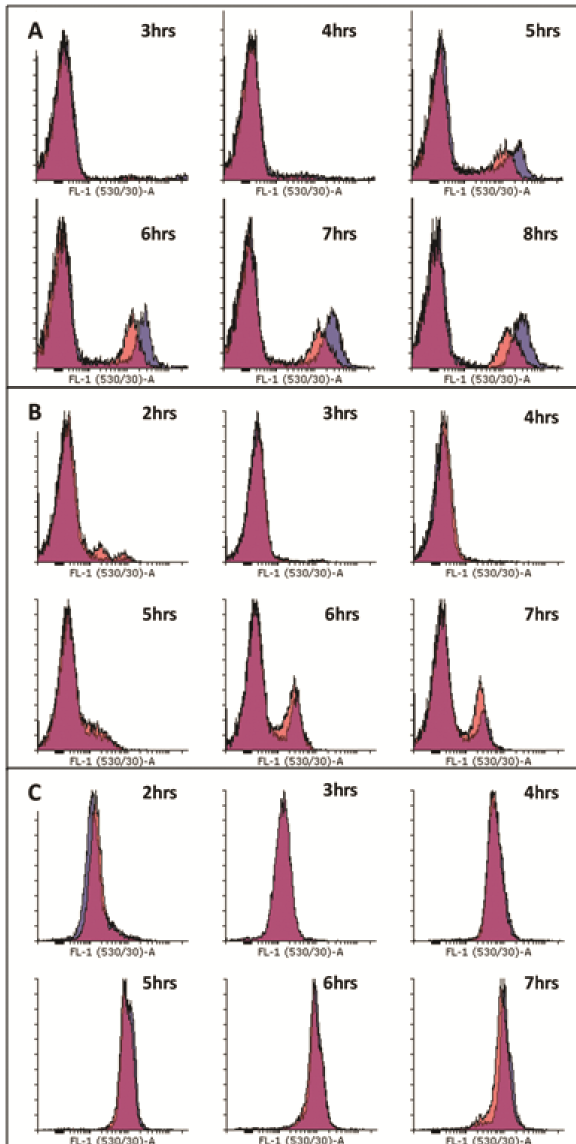

Fig.1. Differences in expression under competence stimulating conditions. **Blue:** control **Red:** BFA1698 (*ΔyhfW*)
A: expression of $P_{comG\text{-}gfp}$. The difference is statistically significant , P <0.04-0.001.
B: Expression of $P_{comK\text{-}gfp}$ The difference is statistically significant at 3-4hrs P<0.001.
C: Expression of $P_{srfA\text{-}gfp}$ statistically significant at 2-3hrs P0.002-0.019.

Statistics were performed using the Mann-Whitney Rank-Sum-test.

Expression of *srfA* (Fig. 1C) was also significantly increased in the mutant before on-set of competence and reduced later in the stationary phase, although not statistically significantly so. Expression of *spo0A* was lower in the mutant, but this effect is not statistically significant. There also was no statistically significant effect on the expression of *rok* under competence stimulating conditions.

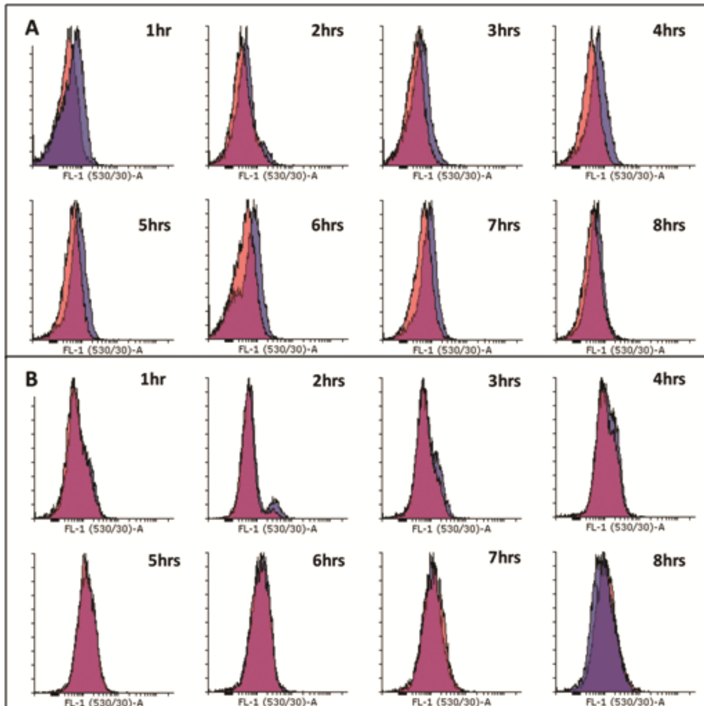

Fig.2. Differences in expression of regulator expression under competence stimulating conditions. **Blue:** control **Red:** BFA1698 (*ΔyhfW*)

A: Expression of $P_{spo0A\text{-}gfp}$ the difference is not statistically significant.
B: Expression of $P_{rok\text{-}gfp}$ the difference is not statistically significant.

Statistics were performed using the Mann-Whitney Rank-Sum-test.

The expression pattern of *yhfW* is nearly identical to that of its neighbour *yhxC,* which is transcribed in the opposite direction (Nicolas et al., 2012). *yhfW* has been shown to be regulated by SigF and *yhxC* by SigE (Arrieta-Ortiz et al., 2015; Wang et al., 2006). However there is not a significant difference in sporulation regulators and their regulons between the two subpopulations during competence, so we can exclude effects of subpopulations going differently into sporulation. Both genes also share a number of predicted regulator binding sites (Table 5). We therefore decided to also investigate the effect of inactivation of *yhxC* on competence. In the absence of *yhxC* the total amount of competent cells is significantly reduced by approximately a factor of two (Fig3. A). In contrast to Δ*yhfW* the expression of *comK* is reduced in the mutant, and again this difference is only significant before maximum competence is achieved (Fig. 3B).As for Δ*yhfW* the expression of *srfA* is significantly increased in Δ*yhxC* before the on-set of competence (Fig. 3C). The expression of *spo0A* is slightly lower, but as for *yhfW* not statistically significant (Fig. 4A). As for Δ*yhfW,* inactivation of *yhxC* does not significantly affect the expression of *rok* (Fig. 4B). The transformability of the Δ*yhfW* strain is 1.4X lower than the control, however this difference is not statistically significant. The transformability of Δ*yhxC* is 2 times lower and is statistically significant.

predicted regulator	***yhfW***	***yhxC***
CcpC	•	
CitT	•	
CtsR	•	
DegU	•	•
GltR	•	•
RocR	•	•
Xre		•
Zur		•

Table 5. Predicted sigma factor and regulator binding sites

Regulator binding sites in the 300bp promoter region of *yhfW* and *yhxC*. All were predicted by Genome2D TFBS search.

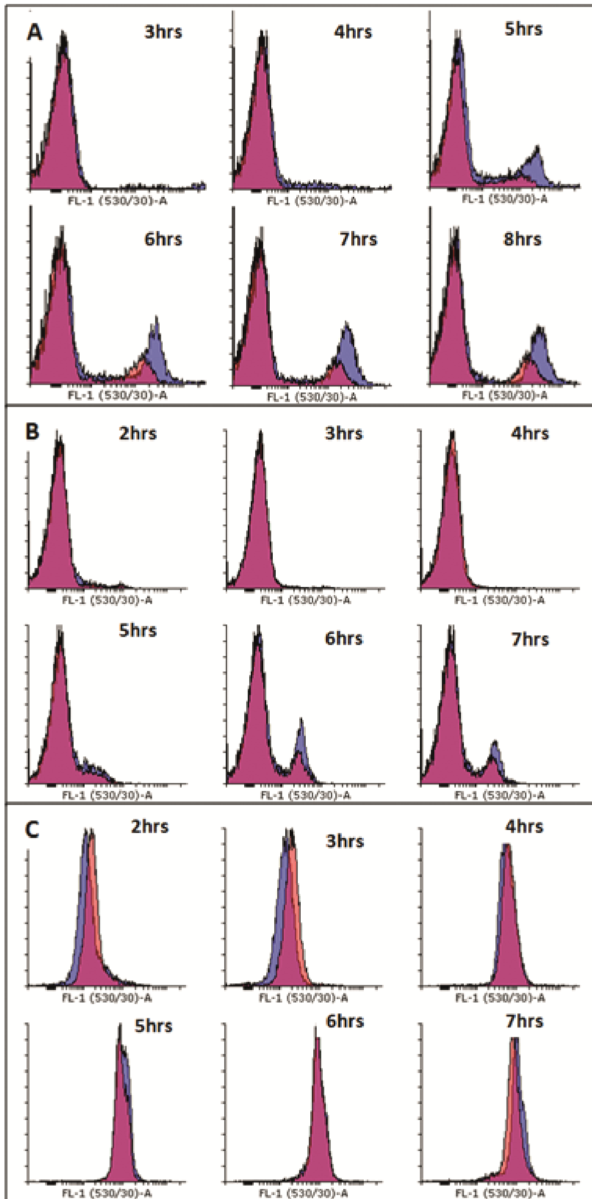

Fig. 3. Differences in expression under competence stimulating conditions. **Blue:** control **Red:** BFA1701 (*ΔyhxC*)

A. expression of $P_{comG\text{-}gfp}$. The difference is statistically significant P <0.001
B. Expression of $P_{comK\text{-}gfp}$ The difference is statistically significant at 2hrs P0.008
C. Expression of $P_{srfA\text{-}gfp}$ statistically significant at 2 and 4hrs P<0.001-0.014

Statistics were performed using the Mann-Whitney Rank-Sum-test.

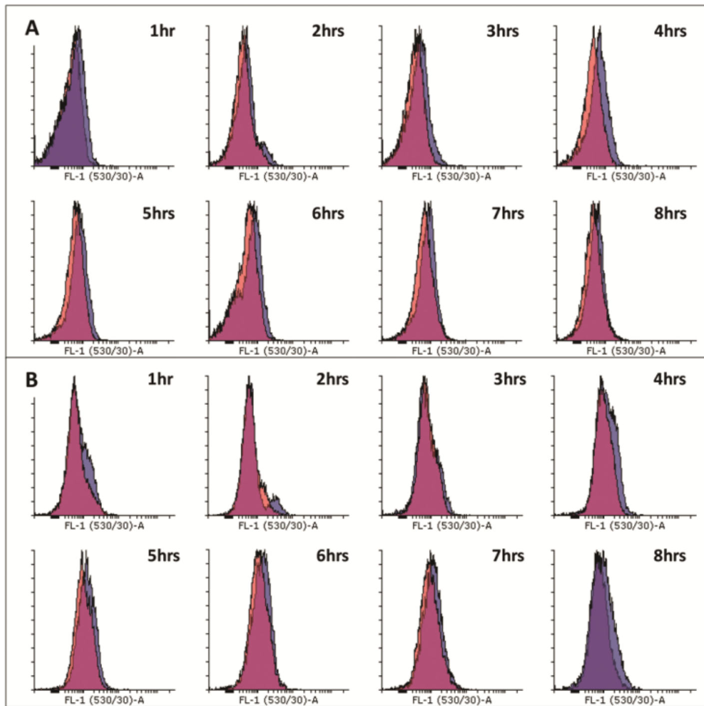

Fig.4. Differences in expression under competence stimulating conditions. **Blue:** control **Red:** BFA1701 (*ΔyhxC*)

A. Expression of P $_{spo0A\text{-}gfp}$ the difference is not statistically significant.
B. Expression of P$_{rok\text{-}gfp}$ the difference is not statistically significant.

Statistics were performed using the Mann-Whitney Rank-Sum-test.

Effect of yhfW inactivation on the metabolome

Both YhfW and YhxC are predicted oxidoreductases of unknown function. YhfW is predicted to be a FAD-linked oxidoreductase and contains a Rieske 2Fe-2S domain at the C-terminus as indicated by Interpro. YhxC belongs to the Short Chain Dehydrogenase (SDR_c1) family of proteins and shows similarity to FabG and 3-oxo-ACP reductase domains (NCBI-pBLAST). Proteins similar to YhfW and YhxC (Blast-P) are present in strains of many species of *Bacillus* closely related to *B. subtilis*. Most of these *Bacillus* species also contain ComK, Spo0A, and SrfA. Proteins with high similarity to YhfW and YhxC are also found in *Paenibacillus*, *Geobacillus, Anoxybacillus, Fictibacillus, Brevibacillus* species and several other members of the Bacillaceae genus. Interestingly, when firmicutes were excluded from the YhfW Blast, similar proteins are found in Archaea and Cyanobacterial species, but not in Gram negative bacteria (100 species cut-off), but in contrast YhxC like proteins are found in gram negative bacteria.
Because YhfW is predicted to be an enzyme we decided to determine if inactivation of this gene has an effect on the metabolome under competence conditions. A growth curve was determined to inspect possible differences in growth between the mutant and the control, which are shown to be nearly identical (S10). Samples were taken when maximum *comG-gfp* expression was achieved, for this experiment 6-7 hrs. The intracellular metabolome revealed differences in metabolite levels between control and Δ*yhfW* (Fig.5). The promoter region of *yhfW* contains (predicted) binding sites for regulators of the citric acid cycle and citrate uptake CcpC and CitT (Table 5). It was therefore interesting to see significant changes in several TCA cycle metabolites at 6 hrs, such as fumarate, 2-oxoglutarate, and citrate. There were also significant changes in amino acids and amino acid intermediates such as L-threonine, 2-oxoglutarate, phenylpyruvate, L-methionine, L-tryptophan, L-aspartate and L-glutamate. Other significant changes were found in dCTP an dTTP as well as the cell-wall metabolite N-acetyl muramoyl-Ala. At 7 hrs fewer differences in metabolites were found. Significantly changed were N-acetyl muramoyl-Ala, UDP-MurNac, GDP and FAD.

The changes in TCA cycle are interesting as mutaions in TCA cycle genes can lead to defects in sporulation. (Craig et al., 1997; Fortnagel and Freese, 1968, 1968; Ireton et al., 1995). Mutations have a negative effect on phosphorelay activation resulting in lower levels of Spo0A~P (Craig et al., 1997; Ireton et al., 1995). The TCA cycle is linked to amino acid metabolism and it is possible that the reduction of aminoacid levels is a result of defects in the TCA cycle.

Transcriptomic analysis of BFA1698 (ΔyhfW)

To determine if there are changes in amino acid biosynthesis and TCA cycle genes in the mutant we performed RNA-seq on samples harvested at the same time in the same experiment as those used for the metabolomics experiment. Although there are quite a few metabolites significantly changed we only found 17 differentially expressed genes in the RNA-seq data (Table 6). None of the genes found are known amino acid or TCA cycle genes. Interestingly we did find up-regulation of NAD biosynthesis genes *nadA* ,*nadB* and *nadC*. Biosynthesis of NAD in *B. subtilis* occurs from aspartate and uses fumarate or oxygen as electron acceptor for FAD reoxidation (Marinoni et al., 2008). The expression of the three NAD synthesis genes is low under competence conditions in WT *B. subtilis* (Nicolas et al., 2012). It is possible that the changes in TCA cycle and a possible resulting defect in NAD/NADH homeostasis is responsible for up-regulation of NAD synthesis genes, or the up-regulation of NAD synthesis disrupts NAD/NADH homeostasis. Interestingly we do not see a significant increase in the levels of NAD nor in the levels of NADP in the metabolomics data. It is also possible that NAD is processed further in the Nicotinate and Nicotinamide pathway. As we do not find significant changes in the expression levels of amino acid synthesis genes it seems likely that the reduction in the levels of amino acid synthesis intermediates and amino acids are the result of a disruption in the TCA cycle. We also found up-regulation of the Na+/H+ antiporter NhaC, which may be the result of internal pH disruptions due to the lower levels of amino acids and intermediates such as fumarate, 2-oxoglutarate, aspartate, glutamate and citrate. NhaC has been found to be involved in pH homeostasis and the uptake of Na+ (Prágai et al., 2001).

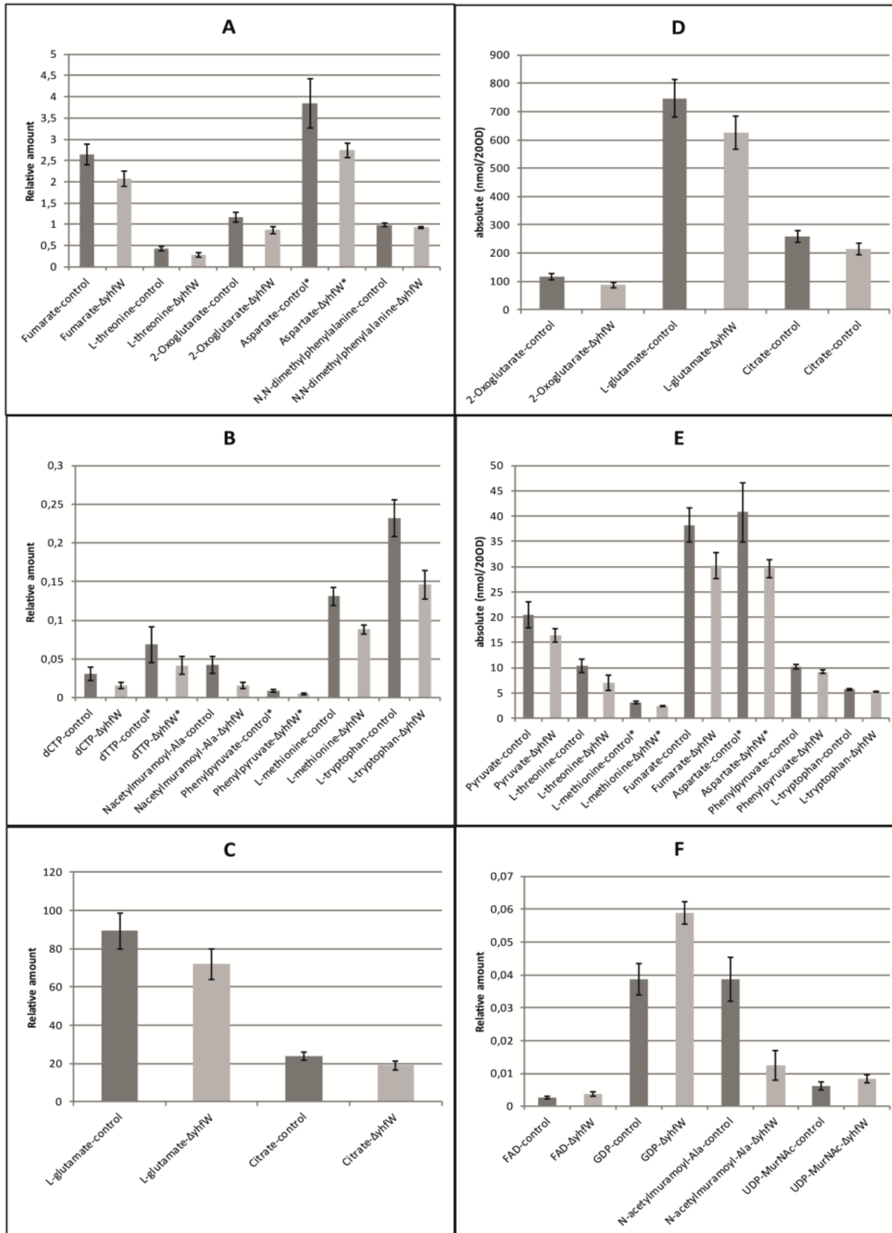

Fig. 5 Significantly changed metabolites. (A, B, C). relative amounts detected by LC-MS at 6hrs. Statistics two. **(D, E).** Absolute amounts detected by GC-MS at 6hrs. Statistics T-test. **(F).** relative amounts detected by LC-MS at 7hrs. Statistics were done using a 2 tail T-test, for samples indicated by asterisk a rank-sum test was performed. Error bars respresent the standard d.eviation

The majority of the down-regulated genes have no known function, but the expression pattern of *yxeD* and *sspD* is very similar to that of *yhfW* (Nicolas et al., 2012) (subtiwiki). Because the down-regulated genes have no known function in either *B. subtilis* or other species further research is required to determine what role they have in the metabololomic changes.

gene	fold	description
nadB	39.1	L-aspartate oxidase
nadC	35.5	nicotinate-nucleotide diphosphorylase (carboxylating)
nadA	29.2	quinolinate synthetase
lip	11.7	extracellular lipase
trnY-Phe	7.3	transfer RNA-Phe
nhaC	5.5	Na^+/H^+ antiporter
tyrS	5.2	tyrosyl-tRNA synthetase
yrzI	4.3	Unkown
opuCB	4	glycine betaine/carnitine/choline ABC transporter
ykzN	-3.7	Unkown
corA	-6.1	Unkown
ywjC	-8.7	Unkown
ywqJ	-11.9	Unkown
yosF	-42	Unkown
sspP	-79.5	probable small acid-soluble spore protein
yxeD	-204.3	Unkown
ywqI	-334.7	Unkown

Table 6. Differential gene expression of BFA1698 under competence stimulating conditions.

Effects of yhfW deletion on sporulation

Although the effect of deletion of *yhfW* on *spo0A* expression is not statistically significant under competence conditions there is a clear difference. As *yhfW* is primarily regulated by SigF, we decided to determine whether absence of *yhfW* could lead to a significant difference in *spo0A* expression under sporulation conditions.

BFA1698 (Δ*yhfW*) was grown in chemically defined sporulation medium (CDSM) containing alanine. In contrast to the competence stimulating conditions growth in CDSM significantly affects the expression of *spo0A*.
Interestingly the expression of *spo0A* is higher in the mutant compared to the control, whereas the expression of *spo0A* was lower in the mutant under competence stimulating conditions (Fig. 6). We also looked at spore outgrowth. The spore crops were diluted to the same OD and heated for 10 minutes at 80°C and plated on LB agar and grown overnight at 37°C. There is no difference in the number of colonies formed between the control and Δ*yhfW*, but strikingly the mutant contains a significantly higher number of large colonies (Fig. 7). The larger size of the colonies could indicate faster germination or a different response to the heat treatment. Sub-lethal heat exposure has been shown to affect germination (Smelt et al., 2008). Different methods of spore isolation and investigations into heat resistance of the spores should lead to more insight into the effect of YhfW on sporulation and spore properties. Investigations into the physical spore properties may lead more insight in the apparent increased germination as YhfW has been found as a putative spore coat protein by (Abhyankar et al., 2015).
High levels of citrate as a result of mutations in TCA cycle genes has been indicated as being responsible for defects in sporulation (Craig et al., 1997; Matsuno et al., 1999). We find under competence conditions lower levels of citrate. As we found higher levels of *spo0A* expression under sporulation conditions, it would be interesting to determine citrate levels under sporulation conditions.

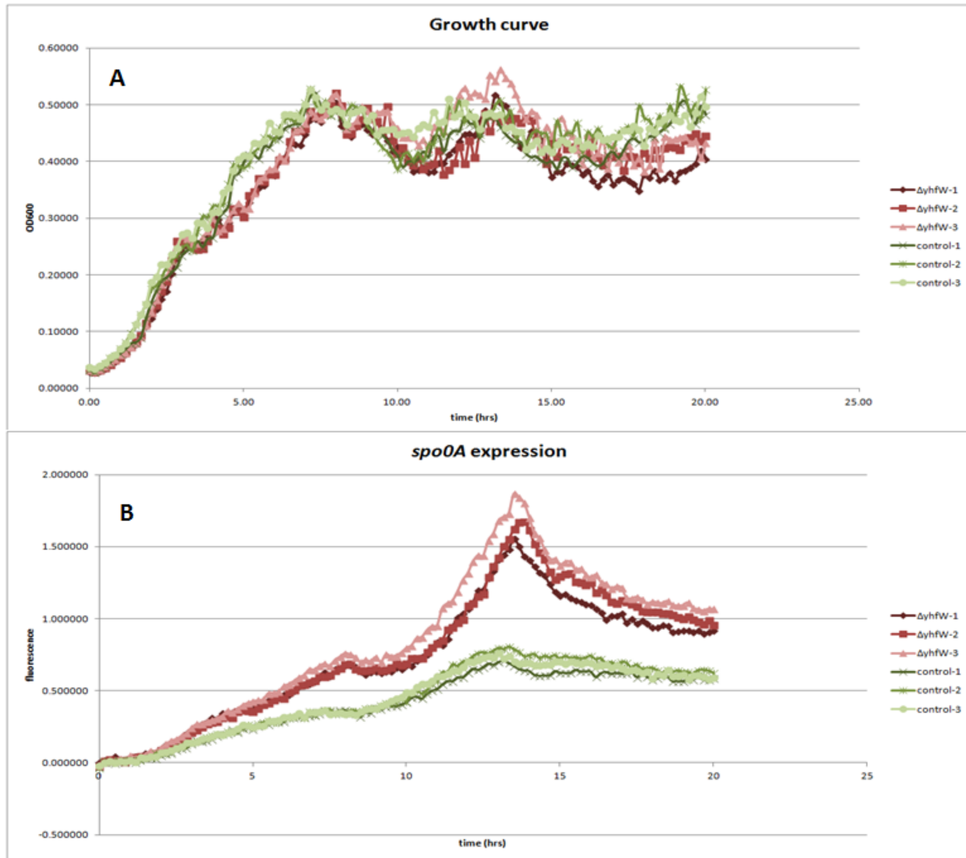

Fig. 6(A). Growth of BFA1698_P$_{spo0A}$ -*gfp* (red) and P$_{spo0A}$ -*gfp* (control , green) in CDSM + alanine. **(B).** Expression of *spo0A* in CDSM + alanine red BFA1698 green control.There is a statistically significant difference in both the growth P<0.001 and *spo0A* expression P 0.0001 between BFA1698 and the control (Mann-Whitney *U*-test)

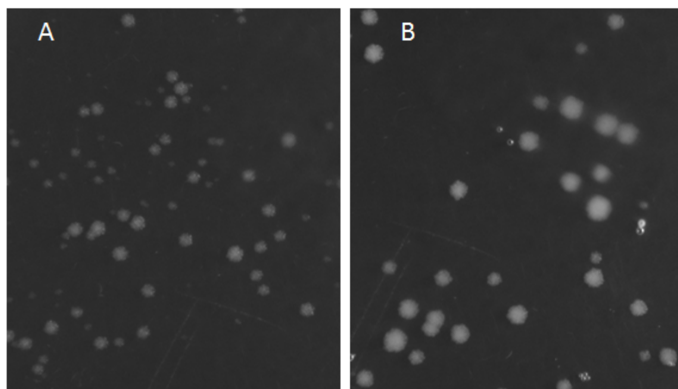

Fig. 7. Growth on LB of colonies germinated from spores. **(A).** Control. **(B)** BFA1698
There is no significant difference in the number of colonies, however there is a significant difference in the size of the colonies P<0.001 (Mann-Whitney *U*-test)

Discussion

In this study we attempted to elucidate some of the remaining questions regarding the K-state of *B. subtilis*. Is there differential expression of non-coding RNAs during competence? Does ComK directly down-regulate genes in the competent subpopulation? Are there competence proteins that are primarily regulated at the post transcriptional level? Our results are largely in accordance with previous studies with regard to the core ComK regulon. Unlike previous studies, we did find six genes that were significantly down regulated in the competent subpopulation, four of which have a corresponding up-regulated antisense RNAs. These are *degU, jag, sigA,* and *lipL*. The up-regulation of antisense RNAs and the presence of K-boxes in their promoter region indicates that down-regulation of the target genes by ComK is mostly indirect and may primarily occur through antisense RNA. Further studies are required to confirm direct regulation of the ncRNAs by ComK. The recently discovered Kre is so far the only gene found that may be directly inhibited by ComK as its expression is repressed in competent cells, and it contains several ComK binding sites (Gamba et al., 2015). Expression of antisense RNA may be a method of fine tuning expression.

During competence there may be a reduced demand for SigA regulated genes as cell division and replication are halted. DegU is a regulator of competence as well as other processes. Down-regulation of LipL may be related to the lack of division by competent cells. LipL is essential for lipoic acid formation, which is necessary for the pyruvate dehydrogenase complex of which one subunit affects Z-ring formation (Christensen et al., 2011; Martin et al., 2011; Monahan et al., 2014). S1458 also covers *pta* which has been found to affect cell-division in *E. coli* (Maciąg-Dorszyńska et al., 2012). The up-regulation of their antisense RNA to these genes during competence in *B. subtilis* may indicate a role in regulating cell division during competence. Known genes affecting cell division during competence are *maf*, *noc* and *minD,* unlike the gene for competence cell division inhibitor *maf*; *noc* and *minD* are not differentially expressed at the RNA level. They do however show increased protein levels in the competent subpopulation. These results indicate that regulation of cell division for MinD and Noc may for a large part take place on a post-transcriptional level. Aside from the before mentioned MinD and Noc the competence proteins SbcC and SbcD also have increased protein levels in the competent subpopulation but are not differentially expressed at the RNA level. Our results, combined with previous research, show that the important competence factors MinD, Noc, SbcC and SbcD are regulated at the post-transcriptional level.

Only five known competence genes were not significantly up-regulated either at the RNA or protein level; these are *addA, addB, hlpB, yhjB* and *yhjC*. This strongly indicates that their basal levels are sufficient for competence. Differences in gene expression in the competent and non-competent subpopulations at the two time points primarily involves amino acid biosynthesis genes. For both the expression of amino-acid synthesis genes is higher at the first time point. The competent sub-population may require higher amino acid production in order to make the proteins necessary for the rather large competence machinery. As the non-competent sub-population enters the stationary state the demand for protein synthesis likely decreases as the growth rate is reduced. The decrease of Srf proteins and the increase of MecA levels at T2 are expected as competence is a transient state requiring removal of ComK.

One of our goals was to determine if there are genes involved in competence that were not found in the previous transcriptomic studies. We did indeed find up-regulation of several genes that have not been found previously, but most notably we found strong up-regulation of *yhfW*, encoding a protein with an unknown function. Deletion of *yhfW* leads reduction in *comG* expression and causes a change in expression under competence conditions of the important *B. subtilis* regulators *comK,* and *srfA*. Its neighbouring gene *yhxC* which has a nearly identical expression profile to *yhxC* also affects expression of *comG, comK,* and *srfA*. In contrast to *yhfW* deletion of *yhxC* results in a strong decrease in the number of competent cells.
YhfW and YhxC are conserved within sporulating bacteria closely related to *B. subtilis,* the majority of which also contain *comK, srfA* and *spo0A*. absence of YhfW under competence conditions results in a significant decrease of several TCA cycle metabolites, and aminoacids. Aside from its effect on competence deletion of *yhfW* significantly increases expression of *spo0A* under sporulation conditions. Interestingly it may also affect the speed of germination resulting in larger colonies. The probable increased speed of spore outgrowth is particularly interesting with regards to the results of (Abhyankar et al., 2015) who indicated YhfW as a putative spore coat protein. Abhyankar et al. also found YhxC in the spore coat. The exact effect of YhfW on germination and spore properties merits further investigation.
YhfW deserves further investigation because absence of YhfW affects TCA cycle components and citrate mutations in TCA cycle genes have been show to affect sporulation (Craig et al., 1997; Fortnagel and Freese, 1968, 1968; Ireton et al., 1995). Ireton et al. proposed that accumulation of citrate in a *citB* mutant strain could be responsible for the sporulation defect, as citrate can chelate divalent cations required for phosphorelay activity (Ireton et al., 1995). We find lower levels of citrate during competence, but interestingly Spo0A expression levels are lower during this condition, although not statistically significantly lower. In contrast, there is a statistically significant increase in Spo0A expression under sporulation conditions and it would therefore be very interesting to determine citrate levels under sporulation conditions.

Although *yhfW* is confirmed to be regulated by SigF no other SigF-regulated genes are differentially expressed between the two subpopulations, nor is there a difference in expression of *sigF*. Further investigation into the other predicted regulators is therefore of interest as it is likely that the up-regulation under competence conditions is the result of one or more of the other predicted regulators.

Together with our data, the presence of these genes in *Bacillus* species that likely are capable of sporulation, competence and surfactin production indicates involvement of these genes in these adaptive strategies. These two genes are interesting because they are predicted to be enzymes and may represent a novel way of affecting differentiation in *Bacillus*. In the case of YhfW this regulation may occur through an effect on the TCA cycle. YhfW could have a pleiotropic metabolic function that is important for both competence and sporulation. Both competence and sporulation have metabolic demands that are different from those during vegetative growth.

To conclude, our data confirm that ComK is primarily a transcriptional activator and that down-regulation by ComK is indirect and occurs through specific ncRNAs. A small number of the known competence genes, in particular those involved in halting cell division, are primarily regulated at the protein level rather than at the transcriptional level. The high sensitivity of RNA-seq did indeed lead to the identification of a new and important gene, *yhfW,* which together with *yhxC* may play an important role in the adaptive lifestyles of *Bacillus*. For further research it would be interesting to study the ncRNAs differentially expressed during competence and confirm their regulation by ComK. Of particular interest is the role of *yhfW* during sporulation, as this is when its highest expression occurs.

Experimental procedures

Growth conditions

Unless otherwise indicated, we used the following competence medium adapted from (Spizizen 1958) and (Konkol et al., 2013): 18ml demi water, 2ml 10X competence medium stock (0.615M K_2HPO_4 . $3H_2O$, 0.385M KH_2PO_4, 20% glucose, 10ml 300mM Tri-Na-citrate, 1ml 2% ferric NH_4 citrate, 1g casein hydrolysate (Oxoid), 2g potassium glutamate) 100µl 2mg/ml tryptophan, 67µl 1M $MgSO_4$ Strains were streaked out from -80 stocks on Luria Bertani (LB) agar plates with antibiotics and grown overnight at 37°C. A single colonies (sc) was diluted 1000X in PBS or 1X Spizizen solution 100µl of the sc colony solution was added to 20ml medium in 100ml Erlenmeyer flasks and grown at 37°C 220rpm Exponential/early stationary overnight cultures were diluted to an OD600 of 0.05 in 20ml medium without antibiotics. Antibiotic concentrations chloramphenicol (cm) 5µg/ml, spectinomycin (sp) 50µg/ml, erythromicin (ery) 0.5 µg/ml, lincomycin 12.5 µg/ml. Growth conditions in CDSM (Hageman et al., 1984; Vasantha and Freese, 1980) + alanine (10mM)+tryptophan 1mM. Single colonies in triplicate were grown overnight at 37°C on LB agar + chloramphenicol (control) or chloramphenicol + erythromycin (BFA1698) were diluted and incubated in 2ml LB 37°C 220rpm in test tubes overnight cultures were mid exponential after overnight growth. Cultures were diluted to OD600 0.05 in 2 CDSM+ alanine +tryptophan and chloramphenicol (control) or chloramphenicol + erythromycin (BFA1698) in test tubes and grown to mid-exponential growth at 37°C 220rpm cultures were diluted to OD600 0.1 in 100µl CDSM+alanine+tryptophan without antibiotics in a 96 wells plate and grown at 37°C, 240rpm, 10min measuring interval for 20hrs in a Thermo Fisher Varioskan Lux . The remainder of the cultures was grown for 24hrs after which the cultures were kept in the dark at 4°C without shaking for 4 days. Spores were harvested by centrifugation at 10000g and washing 3X with double distilled water. The spore crops were diluted to the same OD and heated for 10 minutes at 80°C and dilutions were plated on LB agar with chloramphenicol and grown overnight at 37°C.

Growth conditions for RNA-seq and proteomics

B. subtilis 168 $p_{comG-gfp}$ chloramphenicol resistant variant was created by Jan Willem Veening. *B. subtilis* 168 p_{comG}-gfp was grown in 1X Competence medium, adapted from (Spizizen 1958) and (Konkol et al., 2013) containing 18ml demi water, 2ml 10X competence medium stock (0.615M K_2HPO_4 . $3H_2O$, 0.385M KH_2PO_4, 20% glucose, 10ml 300mM Tri-Na-citrate, 1ml 2% ferric NH_4 citrate, 1g casein hydrolysate (Oxoid), 2g potassium glutamate) 100µl 2mg/ml tryptophan, 67µl 1M $MgSO_4$. Cultures were grown overnight Cultures should be late exponential early stationary after overnight growth. Samples for protein analysis and RNA-seq analysis were taken at 5.5hrs and 6.5hrs One hour of sorting through FACS yields approximately 3x10^7 GFP-negative (non-competent cells) and 1.5x10^7 GFP-positive (competent) cells.

Protein sample preparation and analysis

A non-sorted control of 4.10^6 cells was taken. A total of 4 biological replicates were used for the protein analysis. Samples were sorted onto a vacuum manifold filter system. Proteins were isolated and prepared for ESI-MS details regarding the digestion and ESI-MS settings can be found in Supplementary material S5.

Sample preparation for RNA-seq

Samples were harvested at 5.5 and 6.5hrs diluted in 2M NaCL 1XPBS and run through BDFACS Aria at 4ºC samples were sorted into 4M NaCl 1X PBS. Samples were filtered using a syringe and 13mm 0.22µm filter and washed using TE + 20mM sodium azide and put to liquid nitrogen. The cells on the filter were homogenized in a bead mill , and RNA was extracted as described in (Nicolas et al., 2012a). Two biological replicates were sent for sequencing by Primbio on a proton pI chip without ribosomal RNA depletion. Results were analysed using T-REx (http://genome2d.molgenrug.nl) (EdgeR trimmed-median mean method (TMM) normalization).

Comparisons were made between competent vs non-competent cells at T1 (5.5hrs), competent vs non-competent cells at T2 (6.5hrs), competent T1 vs competent cells T2, non-competent T1 vs non-competent cells T2. Samples for the RNA-seq analysis of BFA1698 were harvested and extracted as described in Nicolas et al.

Strain Construction

BFA1698 (Δ*yhfW*) and BFA1701(Δ*yhxC*) were made using pMUTIN4 by Dr. Rob Meima. BFA1698 and BFA1701 were transformed with genomic DNA from *B.subtilis* 168 P*comg-gfp, B.subtilis* 168 P_{comK}-*gfp, B.subtilis* 168 P_{spooA}-*gfp, B.subtilis* 168 P_{rok}-*gfp, B. subtilis* 168 P_{srfA}-*gfp*. Strain list supplementary material S6

FACS analysis of regulators in BFA1698 ΔyhfW and BFA1701 ΔyhxC background

3 single colony replicates were inoculated and grown as described under growth conditions. Samples were analysed every hour on a BD FACSCanto machine. Data were analysed using Flowing Software 2.5.1. Statistics were performed in Sigma plot using a Rank Sum test.
Supplementary material is available upon request.

References

Abhyankar, W., Pandey, R., Ter Beek, A., Brul, S., de Koning, L.J., and de Koster, C.G. (2015). Reinforcement of Bacillus subtilis spores by cross-linking of outer coat proteins during maturation. Food Microbiol. 45, Part A, 54–62.

Arrieta-Ortiz, M.L., Hafemeister, C., Bate, A.R., Chu, T., Greenfield, A., Shuster, B., Barry, **S.N., Gallitto, M**., Liu, B., Kacmarczyk, T., et al. (2015). An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network. Mol. Syst. Biol. 11.

Berka, R.M., Hahn, J., Albano, M., Draskovic, I., Persuh, M., Cui, X., Sloma, A., Widner, W., and Dubnau, D. (2002). Microarray analysis of the Bacillus subtilis K-state: genome-wide expression changes dependent on ComK. Mol. Microbiol. 43, 1331–1345.

Briley Jr, K., Prepiak, P., Dias, M.J., Hahn, J., and Dubnau, D. (2011). Maf acts downstream of ComGA to arrest cell division in competent cells of B. subtilis. Mol. Microbiol. 81, 23–39.

Brown, A.L., and Smith, D.W. (2009). Improved RNA preservation for immunolabeling and laser microdissection. RNA 15, 2364–2374.

Christensen, Q.H., Martin, N., Mansilla, M.C., de Mendoza, D., and Cronan, J.E. (2011). A Novel Amidotransferase Required for Lipoic Acid Cofactor Assembly in Bacillus subtilis. Mol. Microbiol. 80, 350–363.

Craig, J.E., Ford, M.J., Blaydon, D.C., and Sonenshein, A.L. (1997). A null mutation in the Bacillus subtilis aconitase gene causes a block in Spo0A-phosphate-dependent gene expression. J. Bacteriol. 179, 7351–7359.

Fortnagel, P., and Freese, E. (1968). Analysis of Sporulation Mutants II. Mutants Blocked in the Citric Acid Cycle. J. Bacteriol. 95, 1431–1438.

Gamba, P., Jonker, M.J., and Hamoen, L.W. (2015). A Novel Feedback Loop That Controls Bimodal Expression of Genetic Competence. PLoS Genet. 11.

Hageman, J.H., Shankweiler, G.W., Wall, P.R., Franich, K., McCowan, G.W., Cauble, S.M., Grajeda, J., and Quinones, C. (1984). Single, chemically defined sporulation medium for Bacillus subtilis: growth, sporulation, and extracellular protease production. J. Bacteriol. 160, 438–441.

Hahn, J., Tanner, A.W., Carabetta, V.J., Cristea, I.M., and Dubnau, D. (2015). ComGA-RelA interaction and persistence in the Bacillus subtilis K-state. Mol. Microbiol. 97, 454–471.

Haijema, B.-J., Hahn, J., Haynes, J., and Dubnau, D. (2001). A ComGA-dependent checkpoint limits growth during the escape from competence. Mol. Microbiol. 40, 52–64.

Hamoen, L.W., Smits, W.K., de Jong, A., Holsappel, S., and Kuipers, O.P. (2002). Improving the predictive value of the competence transcription factor (ComK) binding site in Bacillus subtilis using a genomic approach. Nucleic Acids Res. 30, 5517–5528.

Ireton, K., Jin, S., Grossman, A.D., and Sonenshein, A.L. (1995). Krebs cycle function is required for activation of the Spo0A transcription factor in Bacillus subtilis. Proc. Natl. Acad. Sci. 92, 2845–2849.

Irnov, I., Sharma, C.M., Vogel, J., and Winkler, W.C. (2010). Identification of regulatory RNAs in Bacillus subtilis. Nucleic Acids Res. 38, 6637–6651.

de Jong, A., van der Meulen, S., Kuipers, O.P., and Kok, J. (2015). T-REx: Transcriptome analysis webserver for RNA-seq Expression data. BMC Genomics 16, 663.

Kanamaru, K., Stephenson, S., and Perego, M. (2002). Overexpression of the PepF oligopeptidase inhibits sporulation initiation in Bacillus subtilis. J. Bacteriol. 184, 43–50.

Konkol, M.A., Blair, K.M., and Kearns, D.B. (2013). Plasmid-Encoded ComI Inhibits Competence in the Ancestral 3610 Strain of Bacillus subtilis. J. Bacteriol. 195, 4085–4093.

Maciąg-Dorszyńska, M., Ignatowska, M., Jannière, L., Węgrzyn, G., and Szalewska-Pałasz, A. (2012). Mutations in central carbon metabolism genes suppress defects in nucleoid position and cell division of replication mutants in Escherichia coli. Gene 503, 31–35.

Marinoni, I., Nonnis, S., Monteferrante, C., Heathcote, P., Härtig, E., Böttger, L.H., Trautwein, A.X., Negri, A., Albertini, A.M., and Tedeschi, G. (2008). Characterization of l-aspartate oxidase and quinolinate synthase from Bacillus subtilis. FEBS J. 275, 5090–5107

Martin, N., Christensen, Q.H., Mansilla, M.C., Cronan, J.E., and de Mendoza, D. (2011). A Novel Two-Gene Requirement for the Octanoyltransfer Reaction of Bacillus subtilis Lipoic Acid Biosynthesis. Mol. Microbiol. 80, 335–349.

Matsuno, K., Blais, T., Serio, A.W., Conway, T., Henkin, T.M., and Sonenshein, A.L. (1999). Metabolic imbalance and sporulation in an isocitrate dehydrogenase mutant of Bacillus subtilis. J. Bacteriol. 181, 3382–3391

Mirouze, N., Ferret, C., Yao, Z., Chastanet, A., and Carballido-López, R. (2015). MreB-Dependent Inhibition of Cell Elongation during the Escape from Competence in Bacillus subtilis. PLoS Genet 11, e1005299.

Monahan, L.G., Hajduk, I.V., Blaber, S.P., Charles, I.G., and Harry, E.J. (2014). Coordinating Bacterial Cell Division with Nutrient Availability: a Role for Glycolysis. MBio 5.

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.

Nilsson, H., Krawczyk, K.M., and Johansson., M.E. (2014). High salt buffer improves integrity of RNA after fluorescence-activated cell sorting of intracellular labeled cells. J. Biotechnol. 192, Part A, 62–65.

Ogura, M., and Tsukahara, K. (2010). Autoregulation of the Bacillus subtilis response regulator gene degU is coupled with the proteolysis of DegU-P by ClpCP. Mol. Microbiol. 75, 1244–1259.

Ogura, M., Yamaguchi, H., Kobayashi, K., Ogasawara, N., Fujita, Y., and Tanaka, T. (2002). Whole-Genome Analysis of Genes Regulated by the Bacillus subtilis Competence Transcription Factor ComK. J. Bacteriol. 184, 2344–2351.

Prágai, Z., Eschevins, C., Bron, S., and Harwood, C.R. (2001). Bacillus subtilis NhaC, an Na+/H+ Antiporter, Influences Expression of the phoPR Operon and Production of Alkaline Phosphatases. J. Bacteriol. 183, 2505–2515.

Smelt, J.P.P.M., Bos, A.P., Kort, R., and Brul, S. (2008). Modelling the effect of sub(lethal) heat treatment of Bacillus subtilis spores on germination rate and outgrowth to exponentially growing vegetative cells. Int. J. Food Microbiol. 128, 34–40.

Smits, W.K., Eschevins, C.C., Susanna, K.A., Bron, S., Kuipers, O.P., and Hamoen, L.W. (2005). Stripping Bacillus: ComK auto-stimulation is responsible for the bistable response in competence development. Mol. Microbiol. 56, 604–614.

Spizizen, J. (1958). Transformation of biochemically deficient strains of Bacillus subtilis by deoxyribonucleate. Proceedings of the National Academy of Sciences, 44(10), 1072-1078.

Vasantha, N., and Freese, E. (1980). Enzyme changes during Bacillus subtilis sporulation caused by deprivation of guanine nucleotides. J. Bacteriol. 144, 1119–1125

Wang, S.T., Setlow, B., Conlon, E.M., Lyon, J.L., Imamura, D., Sato, T., Setlow, P., Losick, R., and Eichenberger, P. (2006). The Forespore Line of Gene Expression in Bacillus subtilis. J. Mol. Biol. 358, 16–37.

Yoshida, K., Yamaguchi, H., Kinehara, M., Ohki, Y., Nakaura, Y., and Fujita, Y. (2003). Identification of additional TnrA-regulated genes of Bacillus subtilis associated with a TnrA box. Mol. Microbiol. 49, 157–165.