

# Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK

Randy M. Berka,<sup>1</sup> Jeanette Hahn,<sup>2</sup> Mark Albano,<sup>2</sup>  
Irena Draskovic,<sup>2</sup> Marjan Persuh,<sup>2</sup> Xianju Cui,<sup>1</sup>  
Alan Sloma,<sup>1</sup> William Widner<sup>1</sup> and David Dubnau<sup>2\*</sup>

<sup>1</sup>Novozymes Biotech, Inc., Davis, CA 95616, USA.

<sup>2</sup>Public Health Research Institute, 455 First Avenue,  
New York, NY 10016, USA.

## Summary

In *Bacillus subtilis*, the competence transcription factor ComK activates its own transcription as well as the transcription of genes that encode DNA transport proteins. ComK is expressed in about 10% of the cells in a culture grown to competence. Using DNA microarrays representing  $\approx 95\%$  of the protein-coding open reading frames in *B. subtilis*, we compared the expression profiles of wild-type and *comK* strains, as well as of a *mecA* mutant (which produces active ComK in all the cells of the population) and a *comK mecA* double mutant. In these comparisons, we identified at least 165 genes that are upregulated by ComK and relatively few that are downregulated. The use of reporter fusions has confirmed these results for several genes. Many of the ComK-regulated genes are organized in clusters or operons, and 23 of these clusters are preceded by apparent ComK-box promoter motifs. In addition to those required for DNA uptake, other genes that are upregulated in the presence of ComK are probably involved in DNA repair and in the uptake and utilization of nutritional sources. From this and previous work, we conclude that the ComK regulon defines a growth-arrested state, distinct from sporulation, of which competence for genetic transformation is but one notable feature. We suggest that this is a unique adaptation to stress and that it be termed the 'K-state'.

## Introduction

Competence is defined as a physiological state in which exogenous DNA can be internalized, leading to a genetic transformation event. In *Bacillus subtilis*, only 5–10% of the cells in a given population differentiate to this state,

and ComK and MecA were identified as key regulators of the competence response (Dubnau and Roggiani, 1990; Kong *et al.*, 1993; van Sinderen *et al.*, 1995). As cells grow exponentially, MecA binds to ComK, the competence transcription factor, targeting it for degradation by the ClpC/ClpP proteasome (Turgay *et al.*, 1998). When the competence quorum-sensing mechanism is activated late in exponential growth, the small protein, ComS, is synthesized (D'Souza *et al.*, 1994; Hamoen *et al.*, 1995; Solomon *et al.*, 1995). ComS binds to MecA, releasing ComK and thereby preventing its degradation. ComK activates a number of promoters, including its own, thereby establishing a positive autoregulatory circuit (van Sinderen and Venema, 1994). This results in a switch-like phenotype, in which the *comK* gene is expressed explosively in a few cells. Among the known ComK targets are several loci encoding proteins needed for the uptake of transforming DNA, as well as *recA* (Hamoen *et al.*, 1998). As expected, the mutational inactivation of *mecA* results in overproduction of ComK, as the latter is no longer targeted for degradation and is free to act positively on its own promoter. In *mecA* mutants, *comK* is transcribed in all the cells (Hahn *et al.*, 1995).

The C-terminal domain of MecA binds to ClpC, whereas the N-terminal domain binds to ComK and ComS (Persuh *et al.*, 1999). In this way, MecA targets the regulatory protein ComK for degradation by a proteolytic machine that is also an important part of the heat shock response of *B. subtilis*, presumably degrading misfolded and aggregated proteins (Krüger *et al.*, 2000). MecA therefore acts as an adapter, harnessing a general proteolytic mechanism for a specific regulatory response. ComK synthesis occurs in stationary phase cells, partly because the quorum-sensing system activates this synthesis in response to high population density. ComK-expressing cells are arrested with respect to growth and cell division and, upon dilution into fresh medium, retain the appearance of stationary phase cells for several hours after the majority population of non-competent cells has resumed growth (Haijema *et al.*, 2001).

Genome-wide expression profiling with DNA microarrays has rapidly emerged as a powerful method for studying differential gene expression. The value of this method is greatly enhanced when applied to organisms for which the complete genomic DNA sequence is avail-

Accepted 6 December, 2001. \*For correspondence. E-mail dubnau@phri.org; Tel. (+1) 212 578 0842; Fax (+1) 212 578 0804.

able. Among the many bacterial species whose genomes have been published, DNA microarrays have been deployed for genome-wide expression studies in several, including *B. subtilis* (Antelmann *et al.*, 2000; Fawcett *et al.*, 2000; Ye *et al.*, 2000; Helmann *et al.*, 2001; Kobayashi *et al.*, 2001; Lee *et al.*, 2001; Ogura *et al.*, 2001), *Escherichia coli* (Richmond *et al.*, 1999; Barbosa and Levy, 2000; Oh and Liao, 2000a,b; Wei *et al.*, 2001), *Helicobacter pylori* (Sillen *et al.*, 1999; Salama *et al.*, 2000), *Streptococcus pneumoniae* (Peterson *et al.*, 2000; de Saizieu *et al.*, 2000) and *Mycobacterium tuberculosis* (Wilson *et al.*, 1999; Talaat *et al.*, 2000).

We have used a DNA microarray to determine the set of genes that is up- or downregulated in a ComK-dependent manner. It is arguable that more is known concerning the regulatory networks of *B. subtilis* than of any other bacterial species, particularly those networks that affect post-exponential gene expression. This system is therefore an attractive subject for exploration using modern expression profiling methods, as internal controls abound, and much information is available to facilitate the biological interpretation of array-derived results.

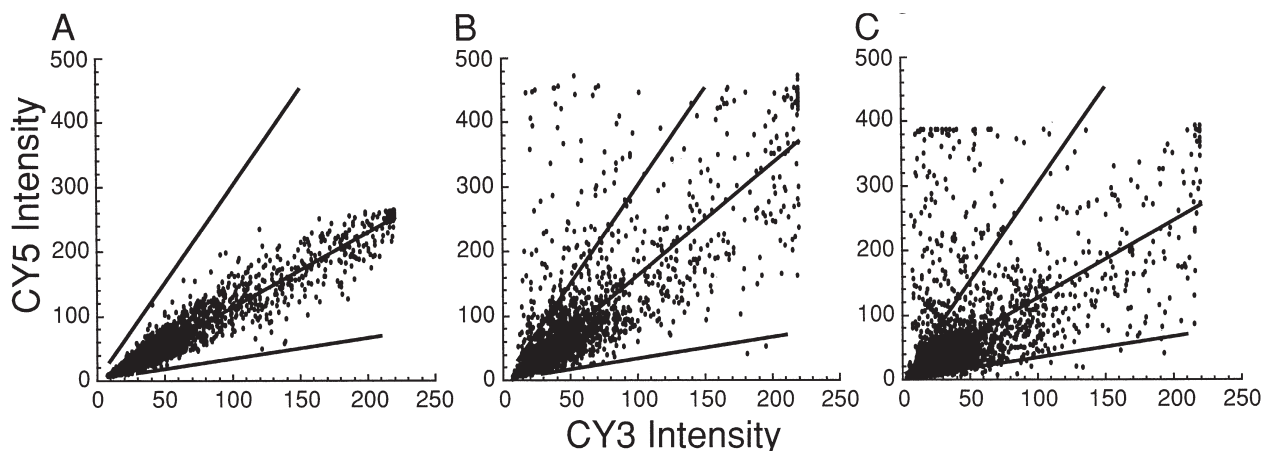
Unexpectedly, we have found that the expression of at least 165 genes is upregulated in the presence of ComK. These include open reading frames (ORFs) that were previously shown to be ComK dependent as well as many for which there was no prior evidence of ComK control. In several cases, validation of the microarray data was achieved through the use of promoter fusions. This profound alteration in the expression programme involves many genes that appear to have no role in transformation. We propose that competence as usually defined is but one feature of a differentiated, growth-arrested state, which we propose to call the K-state.

## Results

### *Bacillus subtilis* genome microarrays and experimental design

To explore the global expression changes orchestrated by ComK, we first amplified individual protein-encoding ORFs by polymerase chain reaction (PCR), using a commercially available primer set. We were successful with  $\approx 95\%$  of the *B. subtilis* ORFs. The PCR products were spotted onto poly L-lysine-coated glass slides. These microarrays were subsequently queried with fluorescent probes, labelled with Cy3 or Cy5, comprising first-strand cDNA made from randomly primed total cellular RNA (Eisen and Brown, 1999). Using such probes, derived from stationary phase cultures grown in rich media, we typically detected hybridization signals (greater than twice background in at least one channel) for about 40–60% of the genes that were spotted onto the microarrays.

The purpose of our experiments was to describe the set of genes that is up- or downregulated in the presence of ComK. To this end, one comparison was between RNA from wild-type and ComK mutant strains grown to competence (type 1 experiment). As ComK is synthesized in only about 10% of the cells in a competent culture, comparison of wild-type and *comK* mutant cultures would fail to detect repressed genes. In addition, if the transcription of a gene were merely augmented in the presence of ComK, but were not completely dependent on this factor, the comparison would be intrinsically insensitive, as a large transcriptional background would be present in the non-competent cells. When *mecA* is inactivated, ComK is produced in essentially all the cells (Hahn *et al.*, 1995). In this study, we have therefore also compared *mecA* with



**Fig. 1.** Scatter diagrams for representative data from *mecA* versus *mecA* (A), wild type versus *comK* (B) and *mecA* versus *mecA comK* comparisons.

B and C. The wild-type and *mecA* cDNAs, respectively, were labelled with Cy3. The lines of best fit are shown, together with lines that denote the three-fold limits.

**Table 1.** Summary of genes identified as ComK regulated by comparison of ratios and by SAM.

	Type 1 comparisons (wild type versus <i>comK</i> )						Type 2 comparisons ( <i>mecA</i> versus <i>mecA comK</i> )					
	RNA A		RNA B		Common in RNA A and B		RNA C		RNA D		Common in RNA C and D	
	+ <sup>a</sup>	- <sup>a</sup>	+ <sup>a</sup>	- <sup>a</sup>	+ <sup>a</sup>	- <sup>a</sup>	+ <sup>a</sup>	- <sup>a</sup>	+ <sup>a</sup>	- <sup>a</sup>	+ <sup>a</sup>	- <sup>a</sup>
Ratio	92	5	140	2	66	0	234	27	176	24	165	9
SAM	97	9	161	9	55	0	196	83	190	122	126	30
Common (ratio and SAM)	85	0	111	0	49	0	165	27	132	22	108	8

a. The + and – symbols refer to genes that are up- and downregulated, respectively, in the presence of ComK.

*mecA comK* cultures (type 2 experiment), thus obviating difficulties resulting from inhomogeneity in the competent cultures. However, it should be kept in mind that the type 2 comparison might identify genes that are regulated by ComK only in a *mecA* background.

Gene expression profiles were compared in two-colour microarray experiments (Eisen and Brown, 1999) as described above. The relative representation of a transcript between the two RNA pools was assayed by measuring the fluorescence intensities of the two dyes at a given target on the array. For each type of comparison, two independent sets of RNA samples were used (sample pairs A and B and sample pairs C and D for type 1 and 2 comparisons respectively). A total of 9–18 replications were carried out with each sample for a given comparison.

Figure 1 shows typical scatter plots for the type 1 and type 2 comparisons (Fig. 1B and C), as well as for a control experiment in which two different *mecA* RNA samples (Fig. A) were compared. Although the Cy3 and Cy5 signal intensities are grouped relatively closely around the line of best fit in the control graph, more scatter

is evident in the type 1 and type 2 graphs. To analyse the microarray data further, we have used two approaches, as described in *Experimental procedures*. The first approach defined genes as regulated by ComK if they were associated with normalized Cy5 to Cy3 fluorescence intensity ratios of  $\geq 3$  or  $\leq 0.33$ . In the second approach, we used significance analysis of microarrays (SAM) (Tusher *et al.*, 2001). A summary of the results with these two methods is presented in Table 1.

#### Expression profile changes affected by ComK (type 1 comparison)

Using the ratio method with the type 1 comparison, 92 and 140 genes were called as upregulated in the presence of ComK with RNA pairs A and B respectively. Of these genes, 66 were called in common using both RNA pairs (Tables 1 and 2). Using SAM with the type 1 comparison, 97 and 161 upregulated genes were called with RNA pairs A and B. Of these, 55 were called in common using the two RNA pairs, and 49 of these genes were

Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>
<i>comER</i>	22.5	<i>comEC</i>	11.7	<i>rpsF</i>	7.7	<i>gidB</i>	4.7
<i>nucA</i>	22.4	<i>comC</i>	11.6	<i>hisZ</i>	7.5	<i>yqeN</i>	4.6
<i>pta</i>	19.7	<i>ywfL</i>	11.3	<i>yviA</i>	7.3	<i>rpsR</i>	4.6
<i>comGA</i>	19.2	<i>maf</i>	11	<i>ywpH</i>	7.3	<i>mreB</i>	4.3
<i>comGB</i>	18.4	<i>smf</i>	10.9	<i>spolIJJ</i>	6.7	<i>jag</i>	4.3
<i>comGD</i>	17.4	<i>yckB</i>	10.6	<i>comEB</i>	6.7	<i>yndK</i>	4.2
<i>yyaF</i>	16.2	<i>radC</i>	10.3	<i>yndJ</i>	6.4	<i>mreC</i>	4.2
<i>comEA</i>	15.8	<i>hxIA</i>	10	<i>mtlR</i>	6.2	<i>ywpJ</i>	4.1
<i>nin</i>	15.3	<i>rnpA</i>	9.6	<i>yxaG</i>	6.2	<i>mreD</i>	4.1
<i>ywfK</i>	15.1	<i>comFB</i>	9.6	<i>ssb</i>	6	<i>yvrO</i>	3.9
<i>comGC</i>	14.2	<i>hisG</i>	9.4	<i>ycbO</i>	5.9	<i>bdbD</i>	3.9
<i>comGE</i>	14.1	<i>hxIB</i>	9	<i>oxdC</i>	5.9	<i>hxIR</i>	3.7
<i>yckC</i>	13.8	<i>comFC</i>	9	<i>ywhA</i>	5.3	<i>yvrN</i>	3.6
<i>comK</i>	13.1	<i>ywfM</i>	8.5	<i>tuaG</i>	5.2	<i>atpB</i>	3.5
<i>comGF</i>	13.1	<i>comFA</i>	8.3	<i>cwlJ</i>	5.1	<i>yhbB</i>	3.3
<i>comGG</i>	13	<i>yckA</i>	7.9	<i>ywfl</i>	4.9		
<i>yckD</i>	12.9	<i>ycbM</i>	7.7	<i>yurH</i>	4.8		

**Table 2.** Putative ComK-upregulated genes identified in type 1 comparisons.

a. This column lists the genes identified as upregulated in the type 1 comparisons, with ratios  $\geq 3.0$ , using both A and B RNA pairs. Among these genes, those also identified by SAM as upregulated are indicated in bold.

b. This column presents average ratios for the genes listed in the first column. The averages include the 18 repeats carried out with RNA pair A and the 12 repeats with RNA pair B.

called in common by both the ratio and SAM approaches and with both RNA sample pairs (Table 1). These are regarded as the most likely candidates for positive regulation (direct or indirect) by ComK detected by the type 1 comparisons, and are listed in bold in Table 2.

Only five and two genes were called as downregulated from RNA sample pairs A and B using the ratio method, and only nine genes were called by SAM with each RNA pair. None of the seven genes called by the ratio method or the 18 genes called by SAM as negatively regulated were called in common by both RNAs or by both the ratio criterion and SAM. As we do not expect to detect down-regulated genes from a type 1 comparison, and as no genes were called in common in these various comparisons, we regard these calls as false positives.

*Expression profile changes affected by ComK (type 2 comparison)*

With the type 2 comparisons, the ratio method called 234

and 176 positively regulated genes with RNA pairs C and D. Of these, 165 were called in common with the two RNA pairs (Table 1). These genes are listed in Table 3. SAM called similar numbers of positively regulated genes with both RNA pairs and, of these, 126 were called in common with RNA pairs C and D (Table 1). Table 3 lists the genes called in common as positively regulated by ComK with both RNA pairs using the ratio method. The genes also called by SAM are indicated in bold. Those genes called by both type 1 and type 2 comparisons and by both SAM and the ratio method are underlined in Table 3. Of the genes identified by the type 1 comparison, all but three (*atpB*, *yurH* and *ywfl*) were also identified by the type 2 comparison. In fact, *ywfl* was called by both ratio and SAM using RNA pair C. With RNA pair D, the ratio was 2.7, just below the cut-off. In the type 2 comparison, it is possible to detect genes that are downregulated in the presence of ComK. Nine genes are called using both RNA pairs by the ratio method (Table 4) and 30 using SAM. Of these, eight were called in common by both criteria.

**Table 3.** Putative ComK-upregulated genes identified in type 2 comparisons.

Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>	Gene <sup>a</sup>	Ratio <sup>b</sup>
<u>comGA</u>	37.4	<u>mtlR</u>	13.6	<u>tuaG</u>	8.4	<u>mreC</u>	5.4	<u>argJ</u>	4.1
<u>comGD</u>	29.7	<u>yckD</u>	13.2	<u>ywpE</u>	8.4	<u>cwlH</u>	5.4	<u>ywfG</u>	4
<u>comGE</u>	28.9	<u>yckC</u>	13	<u>yfhB</u>	8.4	<u>ywiB</u>	5.4	<u>usd</u>	4
<u>pta</u>	28.6	<u>rnpA</u>	12.8	<u>rapH</u>	8.3	<u>ycbL</u>	5.4	<u>ykuK</u>	4
<u>nucA</u>	28.1	<u>comZ</u>	12.7	<u>ycbQ</u>	8.2	<u>yhcl</u>	5.2	<u>yqeF</u>	4
<u>nin</u>	26.9	<u>hisZ</u>	12.6	<u>ywfM</u>	7.8	<u>mreD</u>	5.2	<u>exoA</u>	3.9
<u>comGB</u>	26.1	<u>yvyG</u>	11.9	<u>yddR</u>	7.5	<u>gidB</u>	5.2	<u>yocl</u>	3.9
<u>comGF</u>	25.7	<u>yviE</u>	11.9	<u>ywfF</u>	7.5	<u>yvrL</u>	5.2	<u>trpB</u>	3.9
<u>yvaY</u>	24.2	<u>xpt</u>	11.8	<u>aroC</u>	7.4	<u>dinB</u>	5.1	<u>ybdL</u>	3.9
<u>comER</u>	23.8	<u>yqeN</u>	11.6	<u>sacX</u>	7.4	<u>ycbP</u>	5.1	<u>ybxI</u>	3.9
<u>comFB</u>	23.7	<u>hisG</u>	11.4	<u>yhcE</u>	7.3	<u>jag</u>	5.1	<u>yoxD</u>	3.8
<u>smf</u>	22.5	<u>pbuX</u>	11.3	<u>yxaG</u>	7.1	<u>cwlJ</u>	5	<u>ywfB</u>	3.8
<u>comGC</u>	21.9	<u>yckA</u>	11.3	<u>yqzG</u>	7	<u>ywoH</u>	5	<u>ywhH</u>	3.7
<u>comEA</u>	21.9	<u>rpsR</u>	11.3	<u>ywtF</u>	6.9	<u>recA</u>	5	<u>yvrP</u>	3.7
<u>maf</u>	21.2	<u>yckB</u>	11.2	<u>yvrQ</u>	6.8	<u>ywoG</u>	4.9	<u>dnaG</u>	3.7
<u>yyaF</u>	20.8	<u>comEB</u>	11.1	<u>degS</u>	6.8	<u>yyaC</u>	4.9	<u>yyaH</u>	3.6
<u>radC</u>	20.4	<u>yhzc</u>	11.1	<u>yjzB</u>	6.6	<u>mreB</u>	4.8	<u>sboA</u>	3.6
<u>flgM</u>	18.2	<u>ssb</u>	11	<u>yviF</u>	6.5	<u>ywtE</u>	4.8	<u>yndG</u>	3.6
<u>comGG</u>	17.7	<u>rpsF</u>	10.9	<u>yqiX</u>	6.5	<u>yvtA</u>	4.8	<u>cypC</u>	3.6
<u>comFC</u>	17.6	<u>med</u>	10.7	<u>spolIIJ</u>	6.5	<u>sucD</u>	4.8	<u>yfhC</u>	3.5
<u>comC</u>	17.3	<u>yndK</u>	10.6	<u>yvrN</u>	6.4	<u>ydeC</u>	4.7	<u>groES</u>	3.5
<u>ywfK</u>	17	<u>yndH</u>	10.4	<u>ywpD</u>	6.3	<u>bdbC</u>	4.6	<u>minC</u>	3.4
<u>argC</u>	16.6	<u>yjbF</u>	10.4	<u>sigA</u>	6.3	<u>topA</u>	4.6	<u>ykuJ</u>	3.4
<u>comEC</u>	16.3	<u>ycbM</u>	10.2	<u>ywfE</u>	6.1	<u>yyaA</u>	4.4	<u>yqgP</u>	3.4
<u>ywpJ</u>	15.4	<u>flgL</u>	10.1	<u>flgK</u>	6	<u>csrA</u>	4.4	<u>ydeE</u>	3.4
<u>yndJ</u>	15.3	<u>yvyE</u>	10	<u>bdbD</u>	6	<u>yvaX</u>	4.3	<u>phrF</u>	3.4
<u>yviA</u>	15.3	<u>mbI</u>	10	<u>degU</u>	6	<u>ycgS</u>	4.3	<u>yjaX</u>	3.4
<u>yqzE</u>	14.9	<u>ywpH</u>	10	<u>malL</u>	5.9	<u>tuaF</u>	4.3	<u>yomI</u>	3.3
<u>comFA</u>	14.6	<u>yvhJ</u>	9.4	<u>oxdC</u>	5.7	<u>argD</u>	4.3	<u>yqiL</u>	3.3
<u>hxlA</u>	14.6	<u>ycbN</u>	9.3	<u>ywhA</u>	5.7	<u>yddQ</u>	4.3	<u>hipO</u>	3.3
<u>comK</u>	14.3	<u>hxlB</u>	9.2	<u>ywfC</u>	5.6	<u>yvyF</u>	4.3	<u>minD</u>	3.3
<u>spolIID</u>	13.6	<u>hxlR</u>	9.1	<u>yvrM</u>	5.5	<u>argB</u>	4.1	<u>yqxD</u>	3.2
<u>ywfL</u>	13.6	<u>ybdK</u>	9	<u>yhbB</u>	5.5	<u>yxaF</u>	4.1	<u>yknW</u>	3.1

a. This column lists the genes identified as upregulated (ratios  $\geq 3.0$ ) in the type 2 comparisons using both the C and D RNA pairs. Among these, the genes identified by SAM as upregulated are indicated in bold. The genes also identified as upregulated in the type 1 comparisons (Table 2) by both the ratio method and SAM are underlined.

b. This column presents average ratios for the genes listed in the first column. The averages include the 16 repeats carried out with RNA pair C and the nine repeats with RNA pair D.

**Table 4.** Putative ComK-downregulated genes identified in type 2 comparisons.

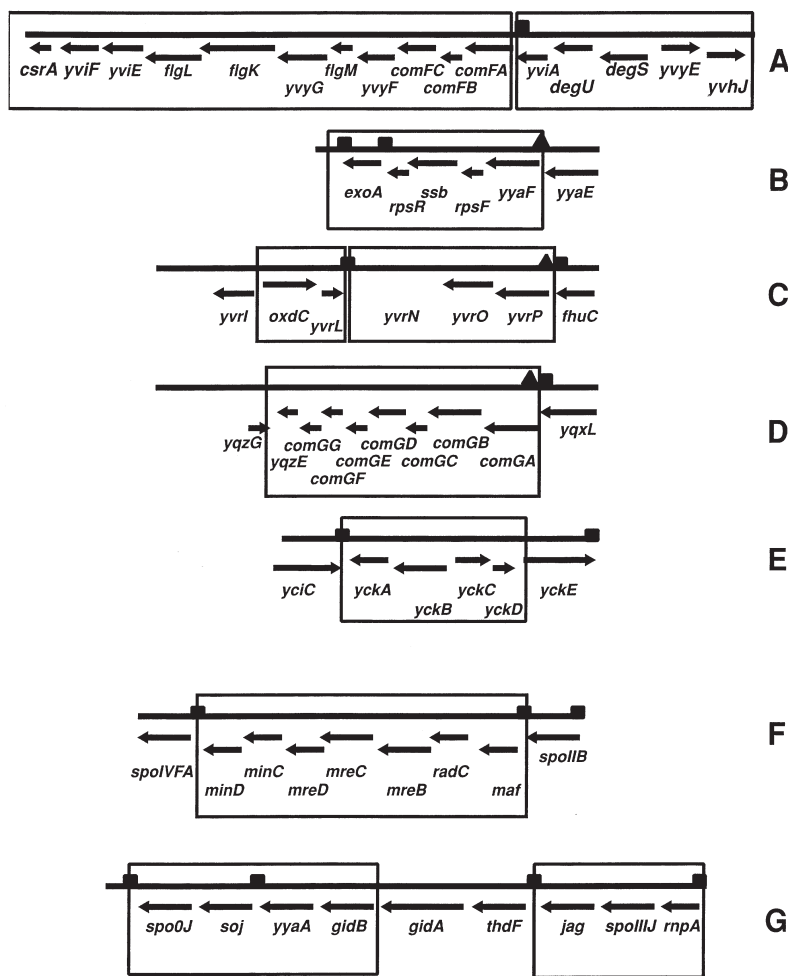
Gene <sup>a</sup>	Ratio <sup>b</sup>
<i>ywzA</i>	0.2
<i>nrgA</i>	0.24
<i>spo0A</i>	0.25
<i>csbA</i>	0.26
<i>grpE</i>	0.26
<i>mcpA</i>	0.29
<i>ureB</i>	0.29
<i>yjfB</i>	0.3
<i>yqfU</i>	0.32

**a.** The genes listed were identified as downregulated (ratio  $\leq 0.33$ ) in the type 2 comparisons with both the C and D RNA sample pairs. All but *ywzA* are in bold because they were also identified using SAM.  
**b.** These are grand average ratios computed from the 16 values from RNA pair C plus the nine values from RNA pair D.

For reasons to be discussed below, we regard the gene lists in Tables 2–4 as conservative. Nevertheless, we will largely restrict our discussion from this point to the genes listed in these tables. More complete data may be viewed at the website listed in *Experimental procedures*.

*Genomics of ComK-regulated genes*

Of the 66 genes listed in Table 2 that fall within the 3.0-fold cut-off with both the A and B RNA pairs, 48 are located in 14 clusters on the *B. subtilis* chromosome. Using this cut-off therefore identifies at least 32 ComK-dependent transcription units (66–48+14). The type 2 comparison identified 128 clustered genes in 39 clusters, corresponding to a total of 76 transcription units. Many of these ComK-regulated clusters undoubtedly represent operons, and several are shown in Fig. 2. In nearly every instance, the microarray data suggested differential gene expression for all the ORFs within a given cluster (e.g. *comFA–yviE* and *comGA–yqzE*; Fig. 2A and D). The ComK-dependent transcription of genes downstream from the *comF* operon (Fig. 2A) is consistent with the finding that readthrough transcription into the antisigma factor gene *flgM* plays a role in the downregulation of motility in competent cells (Liu and Zuber, 1998). The upregulation of *yqzE* downstream from the *comG* operon (Fig. 2D) and of *yqeN* downstream from the *comE* operon (Tables 2 and 3) may also represent readthrough from



**Fig. 2.** Examples of putative ComK-regulated operons and gene clusters. The boxes denote genes whose expression patterns from microarray analyses suggested co-expression driven by ComK. *spo0J* and *soj* are not listed in the tables but are apparently upregulated by ComK, as discussed in the text. Arrows denote direction of transcription; the solid rectangles indicate transcription terminators as annotated on the Subtilist website (<http://genolist.pasteur.fr/Subtilist/>), and ComK boxes are indicated by triangles.

known competence promoters. Neither of these downstream genes are required for competence (Albano *et al.*, 1989; Hahn *et al.*, 1993).

Not unexpectedly, putative ComK boxes precede several of the ComK-regulated gene clusters. The search for the ComK-binding motifs was based on the work of Hamoen *et al.* (1998), which identified AAAAN<sub>5</sub>TTTT as a dyad motif recognized by ComK. In the characterized ComK-driven promoters, a second such dyad element was always present, separated from the first by 8, 18 or 31 bp. Our search criteria permitted no more than two departures from consensus in a single dyad, and a total of no more than three mismatches in a single binding site, which are the limits observed with the promoters that have been studied in detail (Hamoen *et al.*, 1998). We have also imposed the conservative requirement that any putative ComK binding site be located within 500 bp of the first codon of a ComK-regulated gene cluster. Searching the entire genome with these criteria, using the SEARCH PATTERN utility (<http://genolist.pasteur.fr/Subtilist/>) generated hits for many of the ComK-dependent genes. Of the 76 upregulated transcription units identified in the type 2 comparison, 20 possess putative ComK-binding motifs (Table 5). Of the nine downregulated genes listed in Table 4, three (*yjbF*, *nrgA* and *ywzA*) are also preceded by putative ComK boxes (Table 5). Of the 23 ComK boxes presented in Table 5, 20 are newly identified. Several hundred putative ComK boxes were detected that are

apparently not associated with ComK-regulated genes. The significance of this finding is discussed below.

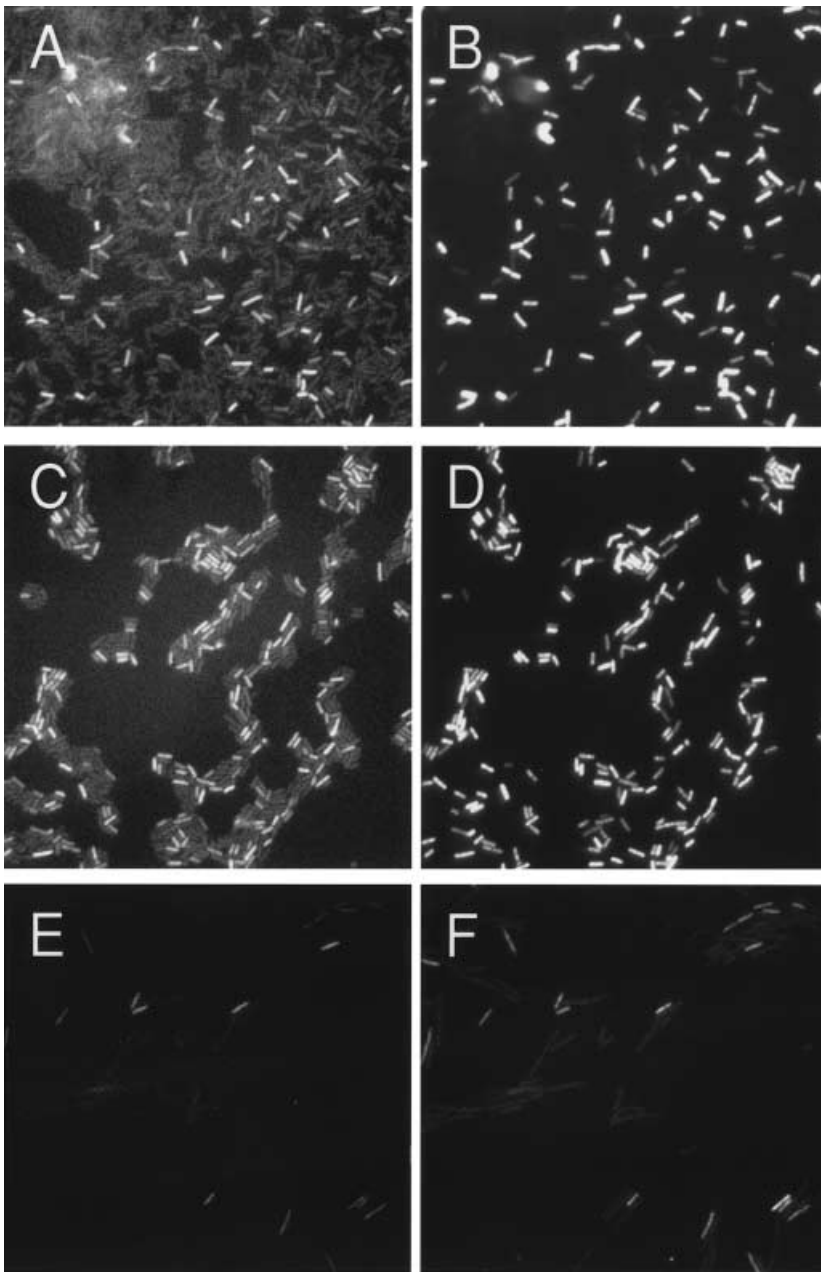
#### Verification of ComK-dependent genes

Several candidate ComK-regulated genes (*soj-spo0J*, *ywpH*, *yvrL*, *pta*, *smf* and *bdbDC*) were studied further to verify the microarray-derived results. In one type of experiment, we used fluorescence microscopy to determine whether candidate ComK-dependent genes were expressed preferentially in the 10% of cells in a given culture that also expressed *comK*. This approach uses the majority non-*comK*-expressing cells as internal controls and does not employ a *mecA* background, permitting the study of gene expression in a wild-type context. For these experiments, we first selected *ywpH* and *yvrL*. The *ywpH* gene was of interest because its predicted product resembles single-strand DNA-binding proteins, and it might therefore play a role in transformation. *yvrL* and the gene immediately upstream (*oxdC*) were both identified as upregulated (Fig. 2C). Fusions of *ywpH* and *yvrL* to *yfp* (the gene for the yellow fluorescent protein) were constructed. These fusion constructs were each inserted at their homologous chromosomal loci by single reciprocal recombination. This was done in a genetic background that also included a fusion of *comK* to *cfp*, the gene for the cyan fluorescent protein, as well as an intact functional copy of *comK*, so that these strains exhibited

**Table 5.** Putative ComK-binding motifs<sup>a</sup> preceding selected ComK-regulated genes.

Bp from start codon	Gene	Pattern sequence
-240	<b><i>nucA</i></b>	tgaaaa <del>caaat</del> at <del>tttc</del> gatctgct <del>aaaa</del> ttacgat <del>tttc</del> ctgc
-487	<i>bdbD</i>	tagaa <del>gct</del> aa <del>gct</del> <del>tttc</del> g <del>cct</del> agtg <del>aaaa</del> actca <del>tttc</del> ttgtc
-165	<i>med</i>	gc <del>aaaa</del> ca <del>aaaa</del> at <del>tat</del> cgag <del>tgg</del> aaa <del>gaca</del> at <del>tttc</del> catc
-462	<i>topA</i>	ccc <del>aaaa</del> at <del>ggc</del> at <del>ttc</del> cc <del>ct</del> atg <del>agaa</del> accg <del>tatt</del> atcagc
-106	<i>ycpL</i>	aaaa <del>aaa</del> ga <del>ata</del> at <del>tttc</del> tata <del>aaaa</del> caca <del>cac</del> g <del>tttc</del> ttcaat
-170	<i>yjbF</i>	gg <del>aaat</del> c <del>ggt</del> <del>tttc</del> ttat <del>gca</del> c <del>gaaa</del> at <del>gcca</del> ttttagag
-175	<i>yjfb</i>	ttg <del>aaaa</del> cag <del>ct</del> <del>tttc</del> g <del>tatt</del> g <del>ctc</del> aa <del>agg</del> t <del>ct</del> <del>tttc</del> catt
-90	<i>yndG</i>	gg <del>ata</del> aa <del>ag</del> ata <del>ct</del> <del>tttc</del> tt <del>gg</del> ag <del>agata</del> tacac <del>tttc</del> caaa
-167	<i>yqzG</i>	cg <del>aaat</del> c <del>gct</del> ca <del>tttc</del> gt <del>ctgt</del> tt <del>aaaa</del> ga <del>atg</del> <del>tttc</del> catc
-105	<b><i>comGA</i></b>	gaa <del>aga</del> at <del>tggt</del> <del>tttc</del> cag <del>cat</del> ata <del>aac</del> at <del>ctc</del> aca <del>aaa</del> at <del>cacg</del> <del>tttc</del> ccct
-114	<i>pta</i>	aaaa <del>ata</del> ag <del>att</del> <del>tttc</del> tat <del>gaa</del> ag <del>cgt</del> ata <del>aat</del> gaa <del>agt</del> tg <del>gct</del> gt <del>ttt</del> gaa
-84	<i>rapH</i>	tg <del>aaaa</del> aca <del>aat</del> <del>ttct</del> t <del>ctct</del> at <del>att</del> gt <del>tttc</del> aaa <del>at</del> tt <del>ggg</del> at <del>ttg</del> ata
-199	<i>ywfl</i>	tca <del>aaa</del> cag <del>cca</del> act <del>tttc</del> catt <del>atag</del> c <del>gtt</del> ct <del>ata</del> aaaa <del>aat</del> ct <del>tatt</del> tttc
-163	<i>ywpH</i>	tt <del>aaaa</del> gt <del>taca</del> t <del>attt</del> ct <del>ctc</del> aa <del>agg</del> aaaa <del>agc</del> aaa <del>gat</del> gt <del>tttc</del> tagct
-324	<i>ywzA</i>	tca <del>aaaa</del> at <del>ttc</del> at <del>tttc</del> gt <del>taac</del> ct <del>tct</del> at <del>agat</del> ag <del>gaa</del> t <del>tatt</del> <del>tttc</del> ctcg
-138	<i>yyaF</i>	ttg <del>aaaa</del> aa <del>gca</del> t <del>aat</del> caa <del>agg</del> gat <del>tgg</del> ca <del>agct</del> gaa <del>at</del> g <del>ct</del> <del>tttc</del> ttgtc
-147	<b><i>comK</i></b>	agt <del>aaaa</del> at <del>cgg</del> tt <del>tatt</del> act <del>agtc</del> at <del>ttt</del> agt <del>acc</del> at <del>taaat</del> at <del>cat</del> aaa <del>agat</del> g <del>attt</del> at
-429	<i>hipO</i>	aga <del>agaa</del> aga <del>agg</del> <del>tttc</del> gatt <del>ctt</del> tg <del>gag</del> ct <del>ttt</del> aa <del>caa</del> ag <del>caaat</del> ta <del>aaa</del> agg <del>ata</del> <del>tttc</del> gt
-84	<i>nrgA</i>	cat <del>gaaa</del> at <del>g</del> tt <del>tatt</del> catt <del>ctt</del> tt <del>ttct</del> ct <del>tata</del> aat <del>ga</del> aa <del>aat</del> ata <del>aat</del> att <del>gct</del> <del>ttt</del> ttat
-230	<i>ybdK</i>	cgg <del>aaa</del> at <del>tttc</del> tt <del>ttt</del> at <del>tatat</del> g <del>aaa</del> at <del>tag</del> agac <del>ag</del> tat <del>cct</del> gac <del>aaa</del> gg <del>aaca</del> <del>tttc</del> ctt
-132	<i>yfhB</i>	tata <del>aaa</del> ca <del>aga</del> at <del>ttct</del> ta <del>atc</del> gt <del>tac</del> ccc <del>at</del> tt <del>ttct</del> ta <del>aaaa</del> ag <del>aaa</del> at <del>ggg</del> <del>ttt</del> ttt
-202	<i>yhzC</i>	gat <del>aaa</del> at <del>cat</del> ct <del>tttc</del> aat <del>gat</del> at <del>tta</del> at <del>gg</del> tact <del>aaat</del> gact <del>ag</del> tata <del>aaa</del> acc <del>gat</del> <del>tttc</del> ac
-169	<i>yvrP</i>	tg <del>aaaa</del> aat <del>ccc</del> gg <del>ttt</del> ga <del>agc</del> cg <del>gg</del> at <del>ttt</del> tt <del>tatt</del> ct <del>cg</del> ct <del>ttg</del> aaa <del>at</del> taca <del>at</del> <del>ttt</del> ctc

a. No more than a total of three mismatches were allowed for the entire motif (AAAANNNTTTT < N8, N18 or N31 > AAAANNNTTTT) and no more than two mismatches per half. The search was restricted to regions within 500 bp upstream of start codons. The ComK box motifs are underlined. Genes with previously identified ComK-binding motifs are in bold, and one such characterized gene is included in each spacing class for purposes of illustration. All the genes listed are upregulated in the presence of ComK except for *yjfb*, *nrgA* and *ywzA*, which are downregulated.



**Fig. 3.** *yvrL*, *ywpH* and *spo0J* are preferentially expressed in competent cells. Strains carrying a *comK-cfp* fusion and *yvrL-yfp* (A and B), *ywpH-yfp* (C and D) or *spo0J-yfp* (E and F) were examined for CFP (A, C and E) or YFP (B, D and F) fluorescence. The CFP autofluorescence is obvious and may be used to identify the locations of cells that do not synthesize ComK-CFP.

normal expression of competence. The YFP and CFP fluorescence signals can be readily distinguished in the microscope using appropriate filter sets. The *comK-cfp* fusion permitted the microscopic identification of the ComK-expressing cells that were therefore competent (Haijema *et al.*, 2001). If *ywpH* and *yvrL* were indeed ComK dependent, we would expect their YFP fluorescence signals to be expressed preferentially in the cells that also expressed CFP. Figure 3 documents that this was indeed the case; YFP and CFP are co-expressed in 5–10% of the cells, whereas in the majority of cells, neither fluorescent signal was detected above back-

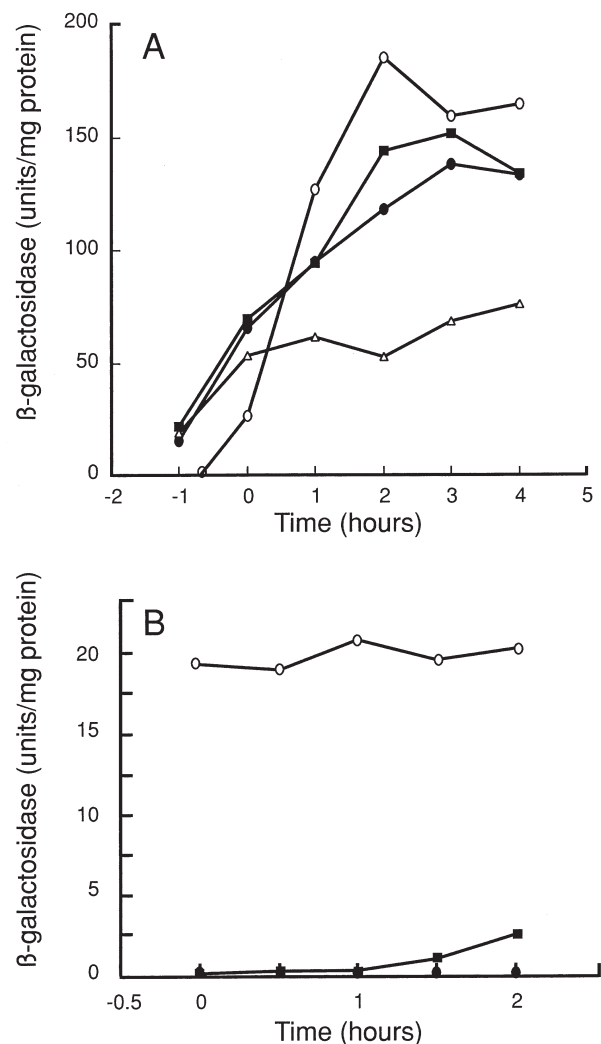
ground. We have reported previously an all-or-nothing pattern for *comK* expression (Haijema *et al.*, 2001), and the present results suggest that the transcription of *ywpH* and *yvrL* is greatly enhanced when ComK is active.

During the course of unrelated experiments, we observed enhanced fluorescence from a *spo0J-yfp* fusion in competence-expressing cells (Fig. 3). As expected, the Spo0J-Yfp signal is visible in all the cells, as this protein is present even in the absence of ComK. It should also be noted that, in these stationary phase cells, the Spo0J-YFP signals were usually not localized at the cell

poles as they are during exponential growth (Lin *et al.*, 1997). Although the *soj spo0J* operon is not listed in Tables 2 or 3, these genes fell just below our arbitrary criteria. For instance, in the type 1 comparison, *soj* and *spo0J* both exhibited ratios of 1.5 with RNA pair A and 1.7 with RNA pair B. In the type 2 comparison, the ratios for *soj* and *spo0J* were 4.2 and 3.8 with RNA pair C and 2.8 and 2.4 with RNA pair D respectively. The results obtained by fluorescence and with the microarrays have also been confirmed in Western blots using antisera raised against Soj and Spo0J proteins; markedly stronger signals were obtained with both proteins in the wild-type compared with a *comK* mutant strain (not shown). We conclude that the *soj spo0J* operon is positively regulated by ComK, as indicated in Fig. 2G, possibly by readthrough from the upstream *gidB* and *yjaA* genes.

We performed a similar experiment with *pta*, which encodes phosphotransacetylase and was predicted to be ComK dependent (Tables 2 and 3). This gene was chosen for study as it is involved in intermediary metabolism, is presumably not required for transformation and is most likely expressed even in non-competent cells. As expected, a strong YFP signal was detected in all the cells in the population (not shown), and no difference could be detected between the competent and non-competent cells. A transcriptional fusion of *pta* to *E. coli lacZ* was then obtained from T. Henkin, and  $\beta$ -galactosidase activity was determined in wild-type, *mecA*, *mecA comK* and *comK* backgrounds. Figure 4A demonstrates that the *pta-lacZ* fusion in the wild-type, *comK* and *mecA* backgrounds was growth stage regulated in the three strains, and that the fusion was expressed slightly more in the *mecA* background and slightly less in the *comK* background. These small effects would be reasonable for a gene that is ComK regulated but can also be expressed independently of ComK. In the *mecA comK* double mutant, however, the expression of the fusion was markedly reduced. Taken together, these data support the microarray results and suggest that *pta* transcription is ComK dependent, at least in a *mecA* background. We do not understand why the type 1 comparison suggests a robust effect of ComK on *pta* transcription (Table 2), whereas only a small effect was apparent using the *pta-lacZ* fusion (Fig. 4A). It is possible that the amount of *lacZ* mRNA produced from the *pta* fusion promoter does not always faithfully reflect that of the *pta* transcript.

*smf* is apparently ComK dependent (Tables 2 and 3) and is highly similar to genes from *Haemophilus influenzae*, *H. pylori* and *S. pneumoniae* that are required for full transformability in those organisms (Karudapuram *et al.*, 1995; Campbell *et al.*, 1998; Smeets *et al.*, 2000). Figure 4B confirms that the transcription of a *smf-lacZ* fusion is ComK dependent and is strongly induced when *mecA* is inactivated.



**Fig. 4.** *smf* and *pta* are regulated by ComK. Strains carrying *pta* (A) and *smf* (B) fusions to *lacZ* were grown in competence medium and sampled at the indicated times for assay of  $\beta$ -galactosidase. A. *pta-lacZ* (closed squares), *pta-lacZ mecA* (open circles), *pta-lacZ comK* (closed circles) *pta-lacZ mecA comK* (open triangles). B. *smf-lacZ* (closed squares), *smf-lacZ mecA* (open circles), *smf-lacZ comK* (closed circles).

A further confirmation of the microarray data was performed as part of another study (Meima *et al.*, 2001). The *bdbDC* operon encodes a pair of thiol-disulphide oxidoreductases that are required for transformation, apparently because they are needed for the correct folding of ComGC, an essential competence protein. The microarray data suggest that this operon is upregulated in the presence of ComK. *bdbD* satisfies our criteria in both type 1 and type 2 experiments. *bdbC* is called in the type 2 experiment but exhibits ratios of 2.3 with both RNA pairs in the type 1 comparison. The *comK* dependence of the *bdbDC* operon has been confirmed using a transcriptional fusion to *lacZ* (Meima *et al.*, 2001).

### Transformation phenotypes of new *ComK*-dependent genes

We have tested mutants of several of the newly identified *ComK*-dependent genes for transformability. Inactivation of *smf* resulted in a 100-fold decrease in transformation frequency when selection was for Leu prototrophy (not shown). Insertion of an intact copy of *smf* in the ectopic *amyE* locus, under *pSPAC* control, fully complemented the transformation phenotype of the *smf* knock-out. Also, inactivation of the *ssb* paralogue *ywpH* resulted in a 10-fold decrease in transformation (not shown). To demonstrate that the *ywpH* phenotype was not caused by polar effects on downstream genes (*glcR* and *ywpJ*), we constructed a double knock-out of these with no discernible effect on transformability. It is interesting that *ssb* expression is also elevated when *ComK* is present (Table 1). *ssb* could not be inactivated (not shown), and this gene is most probably essential for viability. Inactivation of *oxdC* (predicted to be polar on *yvrL*) or of *pta* had no effects on transformation. Finally, it should be noted that *yjbF* is an orthologue of *coiA*, which has been characterized as a competence gene in *S. pneumoniae* (Pestova and Morrison, 1998).

### Discussion

Expression profiling documents a profound alteration of the gene expression programme in response to the synthesis of *ComK*. This can be visualized in Fig. 1. Although we have emphasized the arbitrarily adopted expression ratio cut-offs of 3.0 to identify *ComK*-regulated genes, inspection of Fig. 1B and C conveys the impression that many additional genes that fall within these limits are affected by *ComK*. In fact, for several reasons, we regard the gene lists in Tables 2–4 as conservative. First, we have deliberately adjusted our criteria to exclude a small number of the known *ComK*-regulated genes (see *Experimental procedures*). Secondly, in several cases, a few genes in an operon are identified as *ComK* regulated, whereas others are not. The genes not identified often narrowly fail to satisfy our criteria. An example of this is provided by the gene cluster composed of *maf*, *radC*, *mreB*, *mreC*, *mreD*, *minC* and *minD* (Fig. 2F). In the type 1 comparison, all these genes were identified except for *minC* and *minD* (Table 2). *minC* exhibited average ratios of 2.7 and 3.5 using RNA pairs A and B, and *minD* exhibited average ratios of 2.0 and 2.5. In the type 2 comparison, both genes met the ratio criterion with both RNA pairs, but *minD* was called by SAM only with RNA pair C. These clustered genes therefore did not always meet our criteria, although it appears nearly certain that they are both actually upregulated in the presence of *ComK*. Additional examples of this kind were presented above in the validated cases of *bdbD*, *bdbC*, *soj* and

*spo0J*. These considerations lead us to conclude that the actual list of genes regulated by *ComK* is larger than indicated by Tables 2–4. It is noteworthy that we detected more genes upregulated by *comK* in the type 2 than in the type 1 comparisons. This was predicted, as only a minority of the cells in a *mecA*<sup>+</sup> culture express *comK*. In this connection, it is noteworthy that many of the genes called in the type 2 but not in the type 1 comparison are also expressed independently of *ComK*. Among these are *argB*, *argD*, *argJ*, *csrA*, *degS*, *degU*, *dnaG*, *minC*, *minD*, *sigA* and *topA*. As these genes will be expressed in the non-*comK*-expressing cells, the fold change in their expression in the presence and absence of *ComK* will be decreased.

One plausible explanation for the large number of genes apparently regulated by *ComK* is that the transcription of a distinct subset of affected genes is directly dependent on the binding of *ComK* to their promoters, and that the products of these genes have secondary effects on many others. It is notable in this connection that 28 genes, identified in this study as dependent on *ComK*, have been annotated as regulators of transcription or translation (*comZ*, *degS*, *degU*, *hisZ*, *hxlR*, *med*, *mtlR*, *phrF*, *rapH*, *rnpA*, *rpsF*, *rpsL*, *sacX*, *sigA*, *spolIID*, *spolIJJ*, *usd*, *ybdK*, *ycbL*, *ycbM*, *ydeC*, *ydeE*, *yvhJ*, *ywfK*, *ywoH*, *ywpD*, *ywtF* and *yxaF*). Of these, three (*med*, *rapH* and *ycbL*, in addition to *comK* itself) are associated with predicted *ComK* boxes, and any of them may be responsible for the *ComK*-activated transcription of additional genes. An additional 19 *ComK*-regulated genes appear to be involved in intermediary metabolism (*argB*, *argC*, *argD*, *argJ*, *aroC*, *fabHA*, *hxlB*, *hipO*, *hisG*, *mall*, *pta*, *sucD*, *trpB*, *ycgS*, *yddQ*, *ywpJ*, *ywfG*, *yoxD* and *xpt*), and their activities may easily affect the expression of downstream genes. In general, it is likely that homeostatic mechanisms will tend to dampen such ripple effects. However, these dampening mechanisms may be overpowered by regulatory proteins that directly activate or repress many genes. In the present case, it is clear that the set of genes directly dependent on *ComK* is large, as at least 24 such genes, organized in nine operons, have been described previously (Ogura *et al.*, 1997; Hamoen *et al.*, 1998; Ogura and Tanaka, 2000), and many of the genes identified in this study appear to be associated with *ComK* boxes (Table 5).

It is satisfying that many of the *ComK*-activated genes are organized in clusters, or putative operons, as if one gene in an operon is *ComK* dependent, one would expect this to be true, by and large, of the others. It is also satisfying that many of the *ComK*-dependent genes are preceded by apparent *ComK* boxes. This strongly suggests that the transcription of these genes may be directly enhanced (or repressed) by *ComK* binding. However, a cautionary note is sounded by the observation that

several hundred genes that are apparently not activated by ComK are also preceded by predicted ComK boxes (not shown). It is likely that our understanding of the rules defining functional ComK boxes is deficient, and that many of these sequences have been erroneously identified while others may have remained unpredicted. It will be important to confirm the direct activity of ComK at a sample of these predicted ComK boxes and to refine our understanding of what constitutes a ComK regulation site. However, it is also possible that many of these ComK boxes are in fact not overpredicted, but represent binding sites that do not function directly to modulate the transcription of adjacent genes. One such site was apparently identified in a previous study (Hamoen *et al.*, 1998). ComK-expressing cells contain about 90 000 monomers of this transcription factor, or about 23 000 active tetramers (Turgay *et al.*, 1998). This is highly unusual for a regulatory protein, suggesting that ComK may play a structural role, adjusting the folding of DNA in the nucleoid, in addition to its activity as a directly acting transcription factor.

An interesting pattern can be detected in five instances in which divergently transcribed genes are ComK dependent. These pairs of genes, identified among those listed in Tables 2 and 3, are *pta/ywfl*, *comK/yhzC*, *degS/yvyE*, *yckB/yckC* and *hxlA/hxlR*. At least in the cases of *comK/yhzC* and *pta/ywfl*, there are predicted (palindromic) ComK boxes between the divergent transcription units, within 500 bp of the respective start codons. This raises the possibility that ComK can activate transcription in both directions after binding to a single palindromic site. This arrangement may permit modulation of gene expression by global changes in supercoiling (Opel *et al.*, 2001).

Induction by ComK of several newly identified ComK-driven genes was verified with *yfp* or *lacZ* fusions. In the cases of *bdbDC*, *ywpH*, *yvrL*, *spo0J-soj* and *smf*, ComK-dependent transcription was confirmed (Figs 3 and 4) (Meima *et al.*, 2001). As noted above, it is possible that some of the ComK-activated genes that we have identified comparing *mecA comK* and *mecA* RNA may not be significantly dependent on ComK for their transcription in a *mecA*<sup>+</sup> background. However, the substantial overlap of the ComK-dependent genes identified in *mecA comK* versus *mecA* and in the *comK* versus wild-type comparisons (Tables 2 and 3) strongly suggests that most of the genes identified are truly ComK dependent even in the presence of a functional *mecA* gene.

The 174 up- or downregulated genes that meet the three-fold cut-off criterion in the *mecA* versus *mecA comK* experiment fall into several categories. Although many are genes of unknown function, in many cases, they have been annotated in the database as exhibiting similarities to genes for which a function has been established. Based on these database annotations

(<http://genolist.pasteur.fr/SubtiList/>) and additional similarity searches, we have classified many of the activated genes as involved in transformation (25), sporulation (six), cell shape determination (including peptidoglycan synthesis) and cell division (11), transcriptional regulation (23), transport (11), intermediary metabolism (22), protein synthesis (five), DNA metabolism (10) and detoxification and the stress response (four). It is clear that the ComK regulon affects disparate cellular functions.

The competence-associated transcription of DNA repair genes has been reported previously (Love *et al.*, 1985). Among the transformation-related genes are several newly identified late competence gene candidates: *ywpH* (an *ssb* paralogue); *yjbF* (which resembles *coiA* of *S. pneumoniae*; Pestova and Morrison (1998) and *smf*. It was not expected that sporulation genes would be transcribed in response to ComK, as it is generally believed that sporulation and competence are mutually exclusive. It is interesting in this connection that *spo0A* was found to be downregulated by ComK (Table 4). However, all six ComK upregulated sporulation genes act at or after stage III, and their induction in the absence of functional early spore gene expression would not presage entry into the sporulation pathway.

It is certain that many (probably most) of the newly identified ComK-dependent genes are not required for competence, originally defined as receptivity to transformation (Lerman and Tolmach, 1957), nor for the recombination and recovery steps that follow DNA uptake. We have demonstrated that this is the case for *pta* and *oxdC/yvrL*, and it is certainly true for many of the 29 intermediary metabolism genes, the six sporulation genes and many of the newly identified ComK-dependent transcriptional regulators. It is therefore no longer appropriate to refer to the ComK-determined physiological state as 'competence', as more is involved than transformability. We propose to refer to this instead as the 'K-state', a neutral term with no functional connotation.

The cell shape and cell division genes are of particular interest, as the K-state is associated with inhibition of cell elongation and division (Hahn *et al.*, 1995; Hajjema *et al.*, 2001). The competence gene *comGA* plays a role in the inhibition of these two processes. One ComK-dependent gene cluster (Fig. 2) includes the genes for Maf (an inhibitor of cell division that has also been implicated in DNA repair; Butler *et al.*, 1993; Minasov *et al.*, 2000), MreB, MreC, MreD (shape determining factors; Jones *et al.*, 2001), MinC and D (inhibitors of cell division; Levin *et al.*, 1992) and RadC, a probable DNA repair protein. In addition to this cluster, *mbi*, *tuaF*, *tuaG*, *cwlH* and *cwlJ* are activated. Mbi plays a role in cell shape determination (Henriques *et al.*, 1998; Jones *et al.*, 2001), TuaG and F are required for the synthesis of teichuronic acid (Soldo *et al.*, 1999), and CwlH and J are cell wall hydrolases.

It appears that the K-state is accompanied by a reprogramming of cell shape, cell division and cell wall synthesis genes.

A minority of the cells in a given population reach the K-state, and these cells are arrested in cell division and growth (Haijema *et al.*, 2001). The reversal of this growth inhibition requires at least the degradation of ComK (Turgay *et al.*, 1998). In this sense, the K-state appears to be in some respects a resting state and is associated with the induction of a number of genes (*exoA*, *radC*, *recA*, *ssb*, *topA*, *maf* and *dinB*) that are likely to be involved in DNA repair. The arrest of cell division and growth may be an advantage if the K-state has evolved in part to deal with DNA damage, or if DNA repair is required after transformation, as growth in the presence of DNA lesions may be detrimental. In *E. coli*, the Sula protein is induced as part of the SOS regulon and inhibits cell division (Bi and Lutkenhaus, 1993), presumably until DNA damage has been repaired. Several genes that are activated in the K-state might facilitate the assimilation of novel nutritional sources. These include *malL*, *sucD*, *yoxD* and *ycgS* and the putative transport genes *pbuX*, *yckA*, *yckB*, *yckN*, *ywfF*, *yvrO*, *yvrN*, *yvrM*, *yqiX*, *ywoG* and *yvrP*. Several of these transport proteins, in particular *ywfF* and *ywoG*, might function instead as detoxifying efflux pumps. In this connection, it is worth mentioning some additional ComK-dependent genes. *oxdC* encodes an acid-induced oxalate dehydrogenase (Tanner and Bornemann, 2000), which has been suggested to play a role in pH homeostasis in response to acid stress. *hxlA* and *hxlB* encode enzymes of the ribulose monophosphate pathway (Yasueda *et al.*, 1999), and *hxlR* encodes a positive activator of their expression. It was suggested that this pathway is involved in the detoxification of formaldehyde. ComK induces all three of these genes. Additional stress response genes that are apparently upregulated in response to ComK include *groES* and possibly *yqxD*. Finally, two genes are likely to be involved in the synthesis of antibiotics (*sboA* and *cypC*; Hosono and Suzuki, 1983; Matsunaga *et al.*, 1999), which may serve to eliminate competitors.

In conclusion, we propose that the K-state is a global adaptation to stress, distinct from sporulation, which enables the cell to repair DNA damage, to acquire new fitness-enhancing genes by transformation, to use novel substrates (possibly including DNA; Redfield, 1993; Finkel and Kolter, 2001) and to detoxify environmental poisons. This view of the K-state suggests a reason for its expression in only a fraction of the cells in a given population. The K-state represents a specialized strategy for dealing with danger, but also carries with it inherent risks. Transient arrest of growth and cell division confer vulnerability to overgrowth by competing populations, and transformability opens the cell to invasion by foreign DNA. The

genome may therefore activate alternative systems in subpopulations to deal with adversity, and the K-state may be one such system. This strategy maximizes the probability that the genome will survive when faced with changing environments, a valuable capability for a soil-dwelling organism.

## Experimental procedures

### Bacillus subtilis strains and strain construction

All strains used were derivatives of *B. subtilis* strain 168, and all comparisons were made between isogenic pairs of strains. The following strains were used as sources of RNA for the microarray comparisons: BD630 (*his leu met*); BD2121 (*his leu met comK::kan*); BD2103 (*his leu met mecA::spc*); and BD2125 (*his leu met mecA::spc comK::kan*). The *pta* knock-out and *pta-lacZ* strains were gifts from T. Henkin.

The *oxdC* knock-out was constructed as follows. An internal 630 bp fragment of *oxdC* was amplified from the chromosome using the primers YVRK1 (5'-GGCGATATGCCCCG GGAAGTG) and YVRK2 (5'-GCTCTAGATCGGGTGCAGT GCAG). Underlined letters represent the sites for *SmaI* and *XbaI* respectively. The polymerase chain reaction (PCR) product was digested with these enzymes and cloned into the similarly digested vector pUC18CM. Plasmid DNA from the resultant clone was transformed into BD630 where it integrated by single reciprocal recombination, inserting the entire plasmid in *oxdC* and creating BD3226.

To disrupt the *ywpH* gene, a 1428 bp region bearing the *ywpH* gene was obtained by PCR amplification of BD630 chromosomal DNA using the primers YWPH-FP(#19)-*EcoRI* (5'-CGGAATTCCTGCAATGTTGGGGC) and YWPH-RP(#19)-*SphI* (5'-ACTGCGGCATGCGGTCTCCATCATGAA TAAG). This fragment was cloned into pCR2.1-TOPO (Invitrogen). The resulting plasmid, pED395, was digested with *MfeI*, and a spectinomycin resistance cassette obtained by *EcoRI* digestion of pIC156 was inserted in *ywpH*. We used this plasmid, pED396, to transform *B. subtilis*, generating BD3056. The disruption of *ywpH* by replacement recombination was confirmed by PCR. To disrupt *glcR* and *ywpJ*, a 1912 bp fragment was amplified by PCR from BD630 chromosomal DNA using the primers GLCR-FP(#50)-*BamHI* (5'-CGGGATCCAAAAAGTTCTCTCGTC) and GLCR-RP(#50)-*EcoRI* (5'-CGGAATTCCTCCGAGCCTCCAGC). The product was digested with *EcoRI* and *BamHI* and cloned into pUS19, a pUC19 derivative with a spectinomycin resistance cassette inserted between the *NdeI* and *NarI* sites. The resulting plasmid, pED480, was digested with *NarI*, and a kanamycin resistance cassette, obtained by *Clal* digestion of pDG782 (Guerout-Fleury *et al.*, 1995), was inserted. This plasmid, pED485, was transformed into BD630, giving rise to BD3194. In this mutant strain, the kanamycin resistance cassette replaced the last 612 bp of *glcR* and the first 196 bp of *ywpJ*.

A fusion of *comK* to the *cfp* gene was constructed as follows. A 383 bp *comK* DNA fragment, containing the first nine codons and the regulatory region of *comK*, was amplified from the chromosome using the primers: YFP-*EcoRI* (5'-GACATCGAATTCCTTTTGT) and YFP-*XhoI* (5'-CCGCT CGAGTAAAGGTGCGTCTGTTTTCTG). The fragment was

digested with *EcoRI* and *XhoI* and used to replace the *spo0J* fragment carried by the *spo0J-yfp* fusion vector pKL184. This plasmid was a gift from K. Lemon and A. Grossman. The *comK-yfp* vector was digested with *EcoRI* and *NcoI*, and a fragment containing the *comK* sequences, fused to the first 191 bp of the Yfp protein, was isolated and cloned into pFG10 generating a *comK-cfp* fusion plasmid. pFG10 was a gift from F. Gueiros and R. Losick. The resulting plasmid was then used to transform BD630 by single reciprocal recombination, integrating into the chromosomal *comK* gene to create BD2961.

From the *comK-yfp* fusion, *ywpH-yfp* and *yvrL-yfp* fusions were prepared by excising the *comK* sequences by cutting with *EcoRI* and *XhoI* and replacing this fragment with PCR products containing the *ywpH* or *yvrL* promoter regions. For *ywpH*, the primers were YWPH1 (5'-CGGAATTCATTACATACATACAACCGC) and YWPH2 (5'-CCGCTCGAGATCAGCAGCTTTTCCCGGGG). For *yvrL*, the primers were YVRL1 (5'-CGGAATTCGCACTCGTTCTCTGCAAGCCT) and YVRL2 (5'-CCGCTCGAGGCTGGGAAGAGCCGAGTGCA). The *EcoRI* and *XhoI* sites are underlined. The *ywpH-yfp* and *yvrL-yfp* fusion constructs, marked by chloramphenicol resistance, were combined with a *comK::kan* mutation to create BD2990 and BD3010 respectively.

To construct an *smf-lacZ* fusion, the C-terminus of *smf* was amplified from BD630 chromosomal DNA with PCR using the primers SMF-FP(#31)-*EcoRI* (CGGAATTCGTGATCAAATAAAGCAGC) and SMF-RP(#31)-*BamHI* (CGGGATCCCTTAAAGGTTCCGTATATTGAAC). The product was cloned into pMutin2 (Vagner *et al.*, 1998) using the underlined *EcoRI* and *BamHI* sites. The resulting plasmid, pED455, was integrated into the BD630 chromosome by single cross-over recombination, leading to *smf-lacZ* transcriptional fusion without the disruption of *smf* (BD3148). The chromosomal DNA of *comK::spc* (BD2259) was transformed into BD3148 to give BD3150. BD3158 was obtained by transforming BD3148 with chromosomal DNA from *mecA::kan* (BD2090).

To disrupt the *smf* gene, an 1800 bp fragment, bearing the *smf* gene, was obtained by PCR amplification of chromosomal DNA with SMF-FP(#20)-*EcoRI* (CGGAATTC AACATACG AAGCGGTGC) and SMF-RP(#20)-*SphI* (ACTGCGGCATGCGTTTTCAGCTCTTTTAATACC). This fragment was digested with *SphI* and cloned into pUK19, a pUC19 derivative with a kanamycin resistance cassette inserted into the *NarI* site. The resulting plasmid, pED398, was digested with *MfeI*, and the spectinomycin resistance cassette obtained by *EcoRI* digestion of pIC156 (Steinmetz and Richter, 1994) was inserted. pED398 was used to transform BD630 (*his leu met*) for spectinomycin resistance to produce BD3024. A spectinomycin-resistant and kanamycin-sensitive transformant obtained from this cross had a disrupted *smf* produced by replacement recombination, as verified by PCR.

#### DNA microarrays

A complete set of *B. subtilis* ORF-specific PCR primers was purchased from Eurogentec and used to amplify the protein-coding ORFs from *B. subtilis* 168 genomic DNA. Chromosomal DNA was prepared using the method described by Pitcher *et al.* (1989). The following PCR components were combined in 96-well plates: 25 pmol of each primer, 1× *Taq*

polymerase buffer (PE Applied Biosystems), 0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM TTP, 2 mM MgCl<sub>2</sub>, 0.1 µg of *B. subtilis* genomic DNA and 1.75 U of *Taq* polymerase (PE Applied Biosystems). An MJ Research PTC-225 thermocycler was programmed to incubate the reactions for 36 cycles of 95°C (30 s), 56°C (45 s) and 72°C (3 min 30 s). An aliquot of each completed reaction was analysed by agarose gel electrophoresis to ensure that a PCR product of the correct size and adequate yield was obtained. The success rate for a single-pass amplification of the 4107 ORFs was ≈ 92%. Failed reactions, most often occurring for genes with long ORFs (≈ 3–15 kb), were repeated using *Expand* polymerase mixture (Roche Molecular Biochemicals), and the reaction products were verified by agarose gel electrophoresis. Collectively, the PCR product pools generated in two rounds of amplifications represented ≈ 95% of the *B. subtilis* genome. The amplified *B. subtilis* ORFs were precipitated with isopropanol, resuspended in 15 µl of 3× SSC, and 5 µl aliquots were stored at -20°C in 384-well microplates (Eisen and Brown, 1999). From these plates, the ORFs were spotted onto poly L-lysine-coated glass microscope slides using the equipment and methods that are described on the website of P. O. Brown of Stanford University (<http://cmgm.stanford.edu/pbrown/protocols>).

#### Preparation of probes and hybridization

Fluorescent probes were prepared by reverse transcription of 25 µg of total RNA from *B. subtilis* to incorporate aminoallyl-dUTP into first-strand cDNA. The amino-cDNA products were subsequently labelled by direct coupling to either Cy3 or Cy5 monofunctional reactive dyes (Amersham Pharmacia Biotech). The details of this protocol can be found at <http://cmgm.stanford.edu/pbrown/protocols>. Cy3- and Cy5-labelled probes were combined and purified using Qiaquick PCR spin columns (Qiagen). The purified probes were dried under vacuum in a SpeedVac (Savant Instruments), resuspended in 15.5 µl of water and combined with the following: 3.6 µl of 20× SSC, 2.5 µl of 250 mM HEPES (pH 7.0), 1.8 µl of poly dA (500 µg ml<sup>-1</sup>; Amersham Pharmacia Biotech) and 0.54 µl of 10% SDS. Before hybridization, the solution was filtered with a 0.22 µm Ultrafree-MC microcentrifuge filter (Millipore), boiled for 2 min and cooled to room temperature. The probe was then applied to the microarray under a coverglass, placed in a humidified chamber and incubated at 63°C overnight. Before scanning, the arrays were washed consecutively in 1× SSC with 0.03% SDS, 0.2× SSC and 0.05× SSC and centrifuged for 2 min at 500 r.p.m. to remove excess liquid. Lastly, the slides were imaged using a custom-built confocal laser microscope (Eisen and Brown, 1999). For each type of comparison, two independent sets of RNA samples were used. For the type 1 comparisons (the wild-type BD630 compared with the *comK* mutant BD2121), we used RNA sample pairs A and B. For the type 2 comparisons (*mecA* strain BD2103 compared with the *mecA comK* strain BD2125), we used RNA sample pairs C and D. Each comparison with a given RNA pair was repeated a number of times. The type 1 experiments were repeated 18 and 12 times with RNAs A and B respectively. The type 2 experiments were repeated 16 and nine times with RNAs C and D respectively.

### Treatment of data

Relative fluorescence values for microarray spots were quantified using QUANTARRAY software from GSI Lumonics. The raw Cy3 and Cy5 intensity values were normalized in order to equalize the median values of the two data sets.

Two approaches to the analysis of data were used. In the first approach, we analysed ratios derived from normalized fluorescence intensity data. We calculated the average ratio of wild type to *comK* (type 1 comparison) or *mecA* to *mecA comK* (type 2 comparison) using normalized fluorescence intensities. We then selected a threshold ratio, above which we regarded a gene as ComK dependent for its transcription. To arrive at this threshold, we took advantage of the considerable published knowledge base concerning the ComK regulon. It has been shown, using reporter gene analysis, that the transcription of the following 24 genes is dependent on ComK: *addA*, *addB*, *comC*, *comEA*, *comEB*, *comEC*, *comER*, *comFA*, *comFB*, *comFC*, *comGA*, *comGB*, *comGC*, *comGD*, *comGE*, *comGF*, *comGG*, *comK*, *flgM*, *med*, *nin*, *nucA*, *recA* and *comZ* (Ogura *et al.*, 1997; Hamoen *et al.*, 1998; Liu and Zuber, 1998; Ogura and Tanaka, 2000). We then selected a threshold ratio that called nearly all these genes as ComK dependent. We deliberately set this ratio so that a small number of the test genes were not called, in order to limit the numbers of false positives. Using a threshold of three, 18 and 21 test genes were called with RNA samples A and B in the type 1 comparisons, and 22 genes were called with each of the two RNA samples in the type 2 comparisons.

The normalized fluorescence intensity data were also analysed using the significance analysis of microarrays (SAM) method (Tusher *et al.*, 2001). SAM calculates a *t*-like statistic ( $d_i$ ) and then uses a bootstrapping procedure to determine the significance of this statistic for each gene. SAM permits the adjustment of a significance parameter  $\Delta$ . Genes with  $d_i$  values exceeding  $\Delta$  or below  $-\Delta$  are called as significant. Although in both type 1 and 2 comparisons, our test genes were called as significant by a wide range of  $\Delta$ -values, we have selected  $\Delta$  for each comparison, using our test set of 24 genes as described above. Again, we adopted a conservative approach by selecting  $\Delta$ s that excluded a small number of the test genes. With the chosen values of  $\Delta$ , the type 1 experiments called 17 and 16 of the test genes with RNA samples A and B ( $\Delta = 1.75$  and 10 respectively). In the type 2 comparisons, 21 and 18 test genes were correctly called as ComK dependent with RNA samples C and D ( $\Delta = 9$  and 6 respectively).

As a control, we compared two independent samples of *mecA* (BD2103) RNA (from sample pairs C and D). This comparison was repeated 12 times. No genes exhibited average ratios greater than or equal to 3 or less than or equal to 0.33 (Fig. 1A). When analysed using SAM, no genes were called as either up- or downregulated.

The normalized data for the complete gene sets, together with the calculated ratios, can be viewed or downloaded from <http://www.phri.nyu.edu/dubnaud/dubnaudbio.htm>. The raw data are also available (contact R. Berka at [rambo@novozymesbiotech.com](mailto:rambo@novozymesbiotech.com)).

### Culture conditions and isolation of total RNA

Cultures (250 ml) were grown in competence medium

(Albano *et al.*, 1987) and harvested 2 h after the transition to stationary phase, at which time the wild type (BD630) reaches maximum competence. The cultures were chilled rapidly by the addition of ice, and the cells were collected by centrifugation at 4°C. The pellets were resuspended in 2.5 ml of ice-cold sterile water. RNA was prepared using the FastRNA blue kit (BIO101), according to the manufacturer's instructions, with minor modifications. Two hundred microlitres of cell suspension was added to each FastPrep blue tube containing the lysing matrix as well as the other required reagents. These mixtures were shaken in the FastPrep instrument at a speed rating of 6, twice for 45 s and then for an additional period of 30 s. The rest of the procedure was as described (BIO 101). The final RNA pellets were dissolved in 50  $\mu$ l of SAFE buffer (provided in the FastRNA kit) and pooled. Typically, 3–4 mg of RNA was obtained from 12 tubes. Each RNA sample was analysed on agarose gels to assess quality. In all cases, 23S rRNA signals were detected with greater intensity than the 16S rRNA bands.

### Fluorescence microscopy

Cultures were grown in competence medium (Albano *et al.*, 1987) to 2 h past the transition from the exponential to the stationary phase of growth. Cells were processed for microscopy as described previously (Hajjema *et al.*, 2001). Images were collected with the OPENLAB software package (Improvision) and then exported to ADOBE PHOTOSHOP where they were prepared for publication. We used a Zeiss Axiovert 135M microscope equipped with an Orca digital camera (Hamamatsu), and a Zeiss 1.3 NA Plan Neo-Fluar 100 $\times$  oil immersion objective. For detecting fluorescent images, the following optical filter sets were used: CFP, D480/20m (emitter), D436/20 (exciter), 455dclp (dichroic), all from ChromaTechnology; YFP, XF3082 (emitter), XF1068 (exciter), XF2030 (dichroic), all from Omega Optical.

### Acknowledgements

This study was funded by NIH grant GM57720 from the National Institutes of Health. We gratefully acknowledge the assistance of Michael W. Rey of Novozymes Biotech, Inc. for assistance in developing a database for storage of microarray data. We also thank Joe DeRisi, Vishy Iyer, Michael Eisen and Arkady Khodursky for assistance in constructing our arrayer and scanner and for helpful discussions. Finally, we gratefully acknowledge T. Henkin, F. Gueiros, R. Losick, K. Lemon and A. Grossman for donating strains and antisera, and all the members of our respective laboratories for helpful discussions.

### References

- Albano, M., Hahn, J., and Dubnau, D. (1987) Expression of competence genes in *Bacillus subtilis*. *J Bacteriol* **169**: 3110–3117.
- Albano, M., Breitling, R., and Dubnau, D.A. (1989) Nucleotide sequence and genetic organization of the *Bacillus subtilis comG* operon. *J Bacteriol* **171**: 5386–5404.

- Antelmann, H., Scharf, C., and Hecker, M. (2000) Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis. *J Bacteriol* **182**: 4478–4490.
- Barbosa, T.M., and Levy, S.B. (2000) Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *J Bacteriol* **182**: 3467–3474.
- Bi, E., and Lutkenhaus, J. (1993) Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *J Bacteriol* **175**: 1118–1125.
- Butler, Y.X., Abhayawardhane, Y., and Stewart, G.C. (1993) Amplification of the *Bacillus subtilis* *maf* gene results in arrested septum formation. *J Bacteriol* **175**: 3139–3145.
- Campbell, E.A., Choi, S.Y., and Masure, H.R. (1998) A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol Microbiol* **27**: 929–939.
- D'Souza, C., Nakano, M.M., and Zuber, P. (1994) Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **91**: 9397–9401.
- Dubnau, D., and Roggiani, M. (1990) Growth medium-independent genetic competence mutants of *Bacillus subtilis*. *J Bacteriol* **172**: 4048–4055.
- Eisen, M.B., and Brown, P.O. (1999) DNA arrays for analysis of gene expression. *Methods Enzymol* **303**: 179–205.
- Fawcett, P., Eichenberger, P., Losick, R., and Youngman, P. (2000) The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **97**: 8063–8068.
- Finkel, S.E., and Kolter, R. (2001) Dna as a nutrient: novel role for bacterial competence gene homologs. *J Bacteriol* **183**: 6288–6293.
- Guerout-Fleury, A.M., Shazand, K., Frandsen, N., and Stragier, P. (1995) Antibiotic-resistance cassettes for *Bacillus subtilis*. *Gene* **167**: 335–336.
- Hahn, J., Inamine, G., Kozlov, Y., and Dubnau, D. (1993) Characterization of *comE*, a late competence operon of *Bacillus subtilis* required for the binding and uptake of transforming DNA. *Mol Microbiol* **10**: 99–111.
- Hahn, J., Bylund, J., Haines, M., Higgins, M., and Dubnau, D. (1995) Inactivation of *mecA* prevents recovery from the competent state and interferes with cell division and the partitioning of nucleoids in *Bacillus subtilis*. *Mol Microbiol* **18**: 755–767.
- 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., Eshuis, H., Jongbloed, J., Venema, G., and van Sinderen, D. (1995) A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Mol Microbiol* **15**: 55–63.
- Hamoen, L.W., Van Werkhoven, A.F., Bijlsma, J.J.E., Dubnau, D., and Venema, G. (1998) The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev* **12**: 1539–1550.
- Helmann, J.D., Wu, M.F., Kobel, P.A., Gamo, F.J., Wilson, M., Morshedi, M.M., et al. (2001) Global transcriptional response of *Bacillus subtilis* to heat shock. *J Bacteriol* **183**: 7318–7328.
- Henriques, A.O., Glaser, P., Piggot, P.J., and Moran, C.P. Jr (1998) Control of cell shape and elongation by the *rodA* gene in *Bacillus subtilis*. *Mol Microbiol* **28**: 235–247.
- Hosono, K., and Suzuki, H. (1983) Acylpeptides, the inhibitors of cyclic adenosine 3',5'-monophosphate phosphodiesterase. I. Purification, physicochemical properties and structures of fatty acid residues. *J Antibiot (Tokyo)* **36**: 667–673.
- Jones, L.J., Carballido-Lopez, R., and Errington, J. (2001) Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell* **104**: 913–922.
- Karudapuram, S., Zhao, X., and Barcak, G.J. (1995) DNA sequence and characterization of *Haemophilus influenzae* *dprA+*, a gene required for chromosomal but not plasmid DNA transformation. *J Bacteriol* **177**: 3235–3240.
- Kobayashi, K., Ogura, M., Yamaguchi, H., Yoshida, K., Ogasawara, N., Tanaka, T., and Fujita, Y. (2001) Comprehensive DNA microarray analysis of *Bacillus subtilis* two-component regulatory systems. *J Bacteriol* **183**: 7365–7370.
- Kong, L., Siranosian, K.J., Grossman, A.D., and Dubnau, D. (1993) Sequence and properties of *mecA*, a negative regulator of genetic competence in *Bacillus subtilis*. *Mol Microbiol* **9**: 365–373.
- 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.
- Lee, J.M., Zhang, S., Saha, S., Santa Anna, S., Jiang, C., and Perkins, J. (2001) RNA expression analysis using an antisense *Bacillus subtilis* genome array. *J Bacteriol* **183**: 7371–7380.
- Lerman, L.S., and Tolmach, L.J. (1957) Genetic transformation I. Cellular incorporation of DNA accompanying transformation in *Pneumococcus*. *Biochim Biophys Acta* **26**: 68.
- Levin, P.A., Margolis, P.S., Setlow, P., Losick, R., and Sun, D. (1992) Identification of *Bacillus subtilis* genes for septum placement and shape determination. *J Bacteriol* **174**: 6717–6728.
- Lin, D.C.H., Levin, P.A., and Grossman, A.D. (1997) Bipolar localization of a chromosome partition protein in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **94**: 4721–4726.
- Liu, J., and Zuber, P. (1998) A molecular switch controlling competence and motility: competence regulatory factors ComS, MecA, and ComK control sigmaD-dependent gene expression in *Bacillus subtilis*. *J Bacteriol* **180**: 4243–4251.
- Love, P.E., Lyle, M.J., and Yasbin, R.E. (1985) DNA-damage-inducible (*din*) loci are transcriptionally activated in competent *Bacillus subtilis*. *Proc Natl Acad Sci USA* **82**: 6201–6205.
- Matsunaga, I., Ueda, A., Fujiwara, N., Sumimoto, T., and Ichihara, K. (1999) Characterization of the *ybdT* gene product of *Bacillus subtilis*: novel fatty acid beta-hydroxylating cytochrome P450. *Lipids* **34**: 841–846.
- Meima, R., Eschevins, C., Fillinger, S., Bolhuis, A., Hamoen, L.W., Dorenbos, R., et al. (2001) *bdbDC* operon of *Bacillus subtilis* encodes thiol-disulphide oxidoreductases required for competence development. *J Biol Chem*: in press.
- Minasov, G., Teplova, M., Stewart, G.C., Koonin, E.V., Anderson, W.F., and Egli, M. (2000) Functional implications from crystal structures of the conserved *Bacillus subtilis*

- protein Maf with and without dUTP. *Proc Natl Acad Sci USA* **97**: 6328–6333.
- Ogura, M., and Tanaka, T. (2000) *Bacillus subtilis comZ* (*yjzA*) negatively affects expression of *comG* but not *comK*. *J Bacteriol* **182**: 4992–4994.
- Ogura, M., Ohshiro, Y., Hirao, S., and Tanaka, T. (1997) A new *Bacillus subtilis* gene, *med*, encodes a positive regulator of *comK*. *J Bacteriol* **179**: 6244–6253.
- Ogura, M., Yamaguchi, H., Yoshida, K., Fujita, Y., and Tanaka, T. (2001) DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B. subtilis* two-component regulatory systems. *Nucleic Acids Res* **29**: 3804–3813.
- Oh, M.K., and Liao, J.C. (2000a) Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol Prog* **16**: 278–286.
- Oh, M.K., and Liao, J.C. (2000b) DNA microarray detection of metabolic responses to protein overproduction in *Escherichia coli*. *Metab Eng* **2**: 201–209.
- Opel, M.L., Arfin, S.M., and Hatfield, G.W. (2001) The effects of DNA supercoiling on the expression of operons of the *ilv* regulon of *Escherichia coli* suggest a physiological rationale for divergently transcribed operons. *Mol Microbiol* **39**: 1109–1115.
- Persuh, M., Turgay, K., Mandic-Mulec, I., and Dubnau, D. (1999) The N- and C-terminal domains of MecA recognize different partners in the competence molecular switch. *Mol Microbiol* **33**: 886–894.
- Pestova, E.V., and Morrison, D.A. (1998) Isolation and characterization of three *Streptococcus pneumoniae* transformation-specific loci by use of a *lacZ* reporter insertion vector. *J Bacteriol* **180**: 2701–2710.
- Peterson, S., Cline, R.T., Tettelin, H., Sharov, V., and Morrison, D.A. (2000) Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. *J Bacteriol* **182**: 6192–6202.
- Pitcher, D.G., Saunders, N.A., and Owen, R.J. (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett Appl Microbiol* **8**: 151–156.
- Redfield, R.J. (1993) Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered* **84**: 400–404.
- Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., and Blattner, F.R. (1999) Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* **27**: 3821–3835.
- de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., et al. (2000) Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J Bacteriol* **182**: 4696–4703.
- Salama, N., Guillemin, K., McDaniel, T.K., Sherlock, G., Tompkins, L., and Falkow, S. (2000) A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA* **97**: 14668–14673.
- Sillen, A., Dieckmann, A., Alzhanov, D., and Engstrand, L. (1999) Microarray for genetic analysis of the Cag pathogenicity island in *H. pylori*. *Gut* **45** (Suppl. 3): A21.
- van Sinderen, D., and Venema, G. (1994) *comK* acts as an autoregulatory control switch in the signal transduction route to competence in *Bacillus subtilis*. *J Bacteriol* **176**: 5762–5770.
- van Sinderen, D., Luttinger, A., Kong, L., Dubnau, D., Venema, G., and Hamoen, L. (1995) *comK* encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis*. *Mol Microbiol* **15**: 455–462.
- Smeets, L.C., Bijlsma, J.J., Kuipers, E.J., Vandenbroucke-Grauls, C.M., and Kusters, J.G. (2000) The *dprA* gene is required for natural transformation of *Helicobacter pylori*. *FEMS Immunol Med Microbiol* **27**: 99–102.
- Soldo, B., Lazarevic, V., Pagni, M., and Karamata, D. (1999) Teichuronic acid operon of *Bacillus subtilis* 168. *Mol Microbiol* **31**: 795–805.
- Solomon, J., Magnuson, R., Srivastava, A., and Grossman, A.D. (1995) Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes Dev* **9**: 547–558.
- Steinmetz, M., and Richter, R. (1994) Easy cloning of mini-Tn10 insertions from the *Bacillus subtilis* chromosome. *J Bacteriol* **176**: 1761–1763.
- Talaat, A.M., Hunter, P., and Johnston, S.A. (2000) Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nat Biotechnol* **18**: 679–682.
- Tanner, A., and Bornemann, S. (2000) *Bacillus subtilis* YvrK is an acid-induced oxalate decarboxylase. *J Bacteriol* **182**: 5271–5273.
- Turgay, K., Hahn, J., Burghoorn, J., and Dubnau, D. (1998) Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J* **17**: 6730–6738.
- Tusher, V.G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**: 5116–5121.
- Vagner, V., Dervyn, E., and Ehrlich, S.D. (1998) A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology* **144**: 3097–3104.
- Wei, Y., Lee, J.M., Richmond, C., Blattner, F.R., Rafalski, J.A., and LaRossa, R.A. (2001) High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol* **183**: 545–556.
- Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O., and Schoolnik, G.K. (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* **96**: 12833–12838.
- Yasueda, H., Kawahara, Y., and Sugimoto, S. (1999) *Bacillus subtilis yckG* and *yckF* encode two key enzymes of the ribulose monophosphate pathway used by methylotrophs, and *yckH* is required for their expression. *J Bacteriol* **181**: 7154–7160.
- Ye, R.W., Tao, W., Bedzyk, L., Young, T., Chen, M., and Li, L. (2000) Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J Bacteriol* **182**: 4458–4465.