We have developed an antisense oligonucleotide microarray for the study of gene expression and regulation in *Bacillus subtilis* by using Affymetrix technology. Quality control tests of the *B. subtilis* GeneChip were performed to ascertain the quality of the array. These tests included optimization of the labeling and hybridization conditions, determination of the linear dynamic range of gene expression levels, and assessment of differential gene expression patterns of known vitamin biosynthetic genes. In minimal medium, we detected transcripts for approximately 70% of the known open reading frames (ORFs). In addition, we were able to monitor the transcript level of known biosynthetic genes regulated by riboflavin, biotin, or thiamine. Moreover, novel transcripts were also detected within intergenic regions and on the opposite coding strand of known ORFs. Several of these novel transcripts were subsequently correlated to new coding regions.

**MATERIALS AND METHODS**

**Microarray design.** An “antisense” oligonucleotide array complementary to the *Bacillus subtilis* genome was custom designed by Affymetrix (Santa Clara, Calif.) using the published DNA sequence (GenBank accession no. NC_000964). A general description of Affymetrix GeneChip microarrays has been previously given (10, 11, 21, 30, 40). In the antisense format, the oligonucleotide sequence on the microarray is the same as the coding region sequence, and the labeled target sequence is complementary to the mRNA sequence. Each oligonucleotide probe is 25 nucleotides in length and is specifically selected using an Affymetrix proprietary algorithm. The probes are organized as probe pairs consisting of a perfect match probe (PM) and a mismatch probe (MM). Probe pairs are further organized into larger groups referred to as probe sets; one probe set is used to detect a single putative transcript. Probe sets were present for 4,112 predicted open reading frames (ORFs), as well as for 59 rRNA and 3 rRNA coding regions. More than 95% of these probe sets contained 20 probe pairs. Probe sets to most of the remaining genes contained 10 to 20 probe pairs; only a few had fewer than 10 probe pairs. A probe set to one gene (spuE) could not be designed. Probe sets to 590 selected genes were also duplicated and placed at different sites on the microarray. Of these, 550 were also included in the sense format (probe sets complementary to the mRNA sequence). Probe sets specific to both strands of 603 intergenic or interorneron (“gap”) regions larger than 250 nucleotides are represented. In some cases, larger gap regions contained two or more probe sets. In a few cases, a probe set to only one strand was made. Also included were 50 putative ORFs with *Bacillus*-like ribosome binding sites (RBS); 15 of these were located on the noncoding strand of known genes, and the other 35 were located within intergenic regions.

**Bacterial strains and growth conditions.** The *B. subtilis* strains used in this study were PY79 (SPI* prototroph) (43) and BS0011 (*leuA8 metB5 polC*::neo) (N. Mouncey and H.-P. Hohmann, unpublished data) and isogenic derivatives BI421 (3) containing a biotin (biA) deregulatory mutation and BS0012 (Mouncey and Hohmann, unpublished) containing a riboflavin (rhC) deregulatory mutation. Cells were grown overnight in either Luria-Bertani (LB) medium (without glucose) or Spizizen’s minimal medium (SS) containing 0.1% sodium glutamate, 0.4% glucose, and trace amounts of micronutrients [CaCl₂, FeSO₄, MgSO₄, ZnSO₄, CoCl₂, (NH₄)₂HMo₇O₂₄, AlCl₃, and CuCl₂] in the presence or absence of biotin (0.1 μg/ml), thiamine (0.34 μg/ml), and riboflavin (20 μg/ml) (without glucose). Overnight cultures were diluted 50-fold into fresh medium and grown to exponential growth phase (optical density at 600 nm ~ 0.7). Cells from half of the culture were pelleted, and the total RNA was immediately extracted. In some experiments, the remaining culture was allowed to grow to early stationary phase before RNA extraction. Early stationary phase was determined to be 30 min after glucose exhaustion.
RNA isolation and cDNA labeling. Total RNA was extracted from cells as described by Wei et al. (38) except that the initial lysis buffer contained a mixture of glass beads (bead size, 106 μm), macaloid clay, phenol-chloroform, and sodium dodecyl sulfate (1). RNA isolation was performed at 4°C or on ice. Isolated RNA was treated with RNase-free DNase I to remove contaminating DNA and further purified by using the Qiagen RNEasy Midi kit according to the manufacturer’s instructions. Purified total RNA was stored at −20°C. Preparation of cDNA targets was based on a previously described method (10). Random hexamer primers (9 μg [Gibco-BRL]) and total RNA (30 μg) were added to a mixture containing 6.0 μl of deoxyribonucleotide triphosphate mix (10 mM dCTP, dGTP, and dTTP; 4 mM dATP) and 30.0 μl of biotin-labeled Bio-14 dATP (1 mM [NEN Life Sciences Products]) in a total volume of 60 μl. The mixture was incubated at 70°C for 5 min to denature the primers and RNA. The mixture was cooled on ice, and 24 μl of 5× Strand buffer (250 mM Tris-Cl [pH 8.3], 375 mM KCl, 15 mM MgCl2), 12 μl of 0.1 M diithiothreitol, and 25 μl of Superscript II (200 U/μl Gibco-BRL) was added to a final volume of 120 μl. cDNA synthesis was performed at room temperature for 10 min and then at 45°C for 2 to 4 h. The mixture was treated with NaOH to degrade the RNA strands, followed by neutralization with HCl and Tris-Cl (pH 7.0) (38). cDNA was further purified by two ethanol precipitations to remove unincorporated biotin-labeled dATP. Fragmentation was performed in One-Phor-All buffer (Amersham Pharmacia Biotech) containing 0.4 U of DNase I mix at 37°C. DNase I was inactivated by heating at 99°C for 10 min. Fragmented cDNA was recovered after overnight ethanol precipitation at −20°C.

Genomic DNA labeling. Genomic DNA was labeled by nick translation. Briefly, 5 to 10 μg of genomic DNA was added to a mixture of 50 μl of dCTP, dGTP, and dTTP (0.2 mM each), 30 μl of Bio-14 dATP (0.4 mM), and a combination of 15 U of DNA polymerase I and 12 μl of DNase I mix (Gibco-BRL). After incubation at 22°C for 2 h, the labeled DNA was then precipitated twice by ethanol to remove the unincorporated label.

Hybridization and staining procedures. Microarrays were first prehybridized for 10 min at 43°C in a buffer consisting of 100 mM MES (morpholineethanesulfonic acid), 0.01% Tween 20, bovine serum albumin at 1.6 mg/ml, and fragmented yeast RNA at 0.4 mg/ml. Then, 240 μl of the hybridized mixture containing 1 μl of “Checkboard” (biotin-labeled oligonucleotide [Affymetrix] used to outline the microarray boarder) and 22.5 μg of labeled cDNA was added to the microarray, and the mixture was hybridized overnight at the same temperature in a rotisserie oven. Washing and staining with GeneChip were performed according to Affymetrix’s standard protocol except that the stringent wash was performed at 45°C (9). No signal amplification was performed.

Data analysis. Microarrays were scanned twice at 570 nm at a 3-μm resolution with an Affymetrix scanner and analyzed as previously described (11, 21) by using the Affymetrix gene expression analysis suite. Data were converted to a text format and normalized according to the mean of all of the comparable experiments. Gene mining software Genespring (Silicon Genetics, Inc.) and Spotfire (Spotfire, Inc.) were also used to further analyze the expression data.

Real-time RT-PCR. Real-time RT-PCR using the SYBR green staining method was performed according to the protocol described by Wei et al. (39). PCR primer sequences (Gibco-BRL) to the following genes were as follows: bioA-forward (5′-CGCCGTTCCTCCAGATT-3′), and bioA-reverse (5′-CAA ATATCCTTCCGATCAC-3′), gap-forward (5′-CTCTAGTTCCGGCACA AAGA-3′) and gap-reverse (5′-GGTGGATGTGGAAGGTGATGTTTTCA-3′), mbiG-forward (5′-CAGAAGGGACAGACCGAATCCA) and mbiG-reverse (5′-GA CAATTTGGCAGCTTTAAGA-3′), and slaA-forward (5′-CGTGAATGGA TTATCCGCAAT-3′) and slaA-reverse (5′-TTTCAACGCGCCTGAATAATC G-3′).

RESULTS

Microarray quality. Control experiments were carried out with total RNA isolated from a prototroph strain (PY79) grown to either late exponential or early stationary stage in minimal medium. By using biotin-labeled cDNA targets prepared from the RNA, the optimal hybridization temperature range was determined to be between 42 and 45°C, with 43°C used in all subsequent experiments. To determine the linear range of gene transcript levels, independent microarrays were hybridized with undiluted and diluted (1:2, 1:5, and 1:10) labeled targets. Comparison of the data sets indicated that the transcript levels were linear over an average intensity difference range of 50 to 20,000. An average intensity difference value of 20 corresponded to background noise. Measurement of transcripts from highly expressed genes (i.e., genes with an average difference of >20,000) was inconsistent. The presence of such highly abundant transcripts (e.g., rRNA) often resulted in significant hybridization to the mismatch probe pairs, causing the average difference metric calculations to indicate that transcript levels were low or nonexistent. In these instances, RT-PCR could be used to measure transcript levels.

Additional analysis of the microarray data revealed that >70% of the known B. subtilis ORFs produced a detectable RNA transcript. This indicated that the sensitivity of the system is sufficiently high to detect even low-abundance transcripts. Within the group that did not produce a detectable transcript, eight genes (cca, hisC, ynzH, yobE, yoaI, yorV, ypjA, and pyjC) did not hybridize to labeled genomic DNA isolated from B. subtilis strains PY79 and BS0011 combined. In addition, 15 probe sets from intergenic regions also failed to generate a detectable signal.

Comparison of the duplicated probe sets within a single hybridization experiment also showed good reproducibility of the fluorescent signals with a Pearson correlation coefficient (r) of >0.99 (Fig. 1). This result indicated uniform hybridization of the microarray and provided additional data sets for statistical analysis.

Differential expression of vitamin-regulated genes. The Bacillus GeneChip microarray was also tested to detect differential gene expression patterns of known vitamin regulons for riboflavin, biotin, and thiamine. Control studies showed good reproducibility (r = 0.989) of transcript levels between duplicated bacterial cultures of B. subtilis PY79 grown to late exponential phase in minimal medium (Fig. 2). In contrast, significant and specific changes in transcript levels were detected when cells were grown with biotin (Fig. 3A), thiamine (Fig. 3B), or riboflavin (Fig. 3C) at levels known to repress tran-
scription of their respective biosynthetic genes. Biotin biosynthetic genes bioWAFDBIorf2 exhibited a 30- to >100-fold repression in transcript levels by biotin (Table 1). As expected, genes outside this cluster (ytaP and ytcP) did not show any change in transcript levels. Moreover, in an isogenic strain (BI421) containing a B. subtilis birA mutation that derepresses the bio operon (3), constitutive high-level transcription of the biosynthetic genes was observed. Transcription of birA remained unchanged in both strains.

Similarly, a 30- to 90-fold thiamine-specific repression was observed with transcripts from two major thiamine biosynthetic operons, thiA and tenAI-goxB-yjbSTUV. However, transcript levels of a third operon containing thiK and thiC remained unchanged. Other putative biosynthetic genes, ydiA, ytbJ, and yqiE, also showed no regulation (Table 1).

Transcription of the riboflavin operon (ribGBAH) also exhibited riboflavin-specific repression, but the change in average difference levels was only threefold for genes ribA, ribB, and ribH and twofold for ribG (Table 1). Similar results were obtained when a related strain, B. subtilis BS0011, was grown under similar conditions (data not shown). Moreover, in the presence of the ribC mutation (in B. subtilis 1012), which has been shown to derepress the rib operon (6, 22), constitutive, high-level transcription of the biosynthetic genes was observed (Table 1). Transcripts levels of ribGBAH were increased by >10-fold in the ribC-containing strain B. subtilis BS0012 compared to the isogenic parent B. subtilis BS0011 when both were grown in the presence of riboflavin. Transcription of the ribC gene remained unchanged; transcripts of ribR, a ribC paralog (31), were not detected. Collectively, these results were in good agreement with past Northern blot and lacZ reporter gene fusion studies (1, 26, 29, 44, 45; K. Eichler, S. Taylor, C. Vockler, Y. Zhang, V. Delague, T. P. Begley, and A. P. G. M. van Loon, unpublished results).

Real-time RT-PCR experiments were also performed to assess the accuracy of the GeneChip microarray to detect changes in transcript levels. By using primers complementary to the 5' and 3' ends of the bioA, thiA, and ribG genes, transcript levels were increased 33-, 8-, and 3-fold, respectively, in bacteria grown in minimal medium compared to cells grown in the same medium supplemented with their respective vitamin (Table 2). In a parallel control experiment, the transcript level of a housekeeping gene, gap, showed no significant change.

FIG. 2. Comparison of transcript levels from duplicate bacterial cultures. A scatter plot of average intensity difference values from a hybridization experiment with cDNA prepared from duplicate cultures of B. subtilis PY79 grown to exponential growth phase in minimal medium is shown. Average intensity differences from 50 to >20,000 U are plotted. r is the Pearson correlation coefficient.

FIG. 3. Identification of biotin-, thiamine-, and riboflavin-regulated genes. Scatter plots of average intensity difference values from hybridization experiments with cDNA prepared from cultures of B. subtilis PY79 grown to exponential growth phase in minimal medium in the presence (x axis) or absence (y axis) of biotin (0.1 μg/ml) (A), thiamine (0.34 μg/ml) (B), or riboflavin (200 μg/ml) (C) are shown. r is the Pearson correlation coefficient. The diagonal lines indicate a 10-fold change in transcript levels. Known and potential biotin- or thiamine-regulated genes are circled.
under these growth conditions. Qualitatively, these results were in good agreement with microarray data, although in some cases the exact fold change in gene expression differed. Based on these results and standard deviation criteria, average intensity difference ratios of 3 or higher were considered significant.

Analysis of the transcript levels of selected biosynthetic genes showed that the data conformed to known operon structures and regulation mechanisms. As examples, results with the biotin and riboflavin biosynthetic operons are illustrated in Fig. 4. Transcription of the biotin operon is tightly regulated by a repressor or operator mechanism involving biotin and the B. subtilis BirA-like repressor (3, 4). Location of a rho-independent transcription terminator between the fifth and sixth genes results in the synthesis of two polycistronic transcripts: a full-length transcript of 7.2 kb covering the entire operon (bioWAFDBI-ytbQ) and a shorter 5.2-kb mRNA species that cover the first five genes (bioWAFDB) (26). Northern blots indicated that abundance of the shorter transcript is ~8-fold higher than the full-length transcript. Microarray data showed that, in cells grown in the absence of biotin, bioWAFDB transcript levels were threefold greater than those of bioI and ytbQ (Fig. 4). In the presence of biotin, however, no transcripts were detected. Interestingly, no readthrough transcription from the biotin operon into ytcP was detected, indicating strong transcription termination at the 3' terminator.

The riboflavin operon is one of several well-studied operons that are regulated by a transcription termination-antitermination mechanism (27). The genes are transcribed from three vegetative promoters, two of which (ribF1 and ribF2) are regulated by FMN (flavin mononucleotide) and B. subtilis RibC (6, 22). A proximal rho-independent transcription terminator located between a vegetative promoter and the first structural gene is the key element in the regulatory mechanism that limits expression of the operon under excess FMN conditions. A probe set to the 5' RFN leader region (15) easily detected the

### TABLE 1. Transcript levels of genes involved in biosynthesis of riboflavin, biotin, and thiamine

<table>
<thead>
<tr>
<th>Biosynthesis type and gene</th>
<th>Enzyme and/or function</th>
<th>Ratio&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>wt-/wt+  deg+/wt+</td>
</tr>
<tr>
<td><strong>Riboflavin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ribG</td>
<td>Riboflavin-specific deaminase or reductase</td>
<td>2.0</td>
</tr>
<tr>
<td>ribB</td>
<td>Riboflavin synthase (α subunit)</td>
<td>2.5</td>
</tr>
<tr>
<td>ribA</td>
<td>GTP cyclohydrolase IV 3,4-dihydroxy-2-butanoic 4-phosphate synthase</td>
<td>2.9</td>
</tr>
<tr>
<td>ribH</td>
<td>Lumazine synthase (β subunit)</td>
<td>2.7</td>
</tr>
<tr>
<td>ribT</td>
<td>Unknown (not required for riboflavin synthesis)</td>
<td>NC</td>
</tr>
<tr>
<td>ribC</td>
<td>Riboflavin kinase or FAD synthase</td>
<td>NC</td>
</tr>
<tr>
<td>ribR</td>
<td>Cryptic riboflavin kinase</td>
<td>NT</td>
</tr>
<tr>
<td>ynaA</td>
<td>Unknown transmembrane protein</td>
<td>NC</td>
</tr>
<tr>
<td><strong>Biotin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bioW</td>
<td>Pimeloyl-CoA synthase</td>
<td>89</td>
</tr>
<tr>
<td>bioA</td>
<td>DAPA aminotransferase</td>
<td>&gt;100</td>
</tr>
<tr>
<td>bioF</td>
<td>7-KAP synthase</td>
<td>52</td>
</tr>
<tr>
<td>bioD</td>
<td>DTB synthetase</td>
<td>48</td>
</tr>
<tr>
<td>bioB</td>
<td>Biotin synthase</td>
<td>94</td>
</tr>
<tr>
<td>bioI</td>
<td>Biotin cytochrome P-450</td>
<td>33</td>
</tr>
<tr>
<td>ytbQ</td>
<td>Unknown and/or not required for biotin synthesis</td>
<td>62</td>
</tr>
<tr>
<td>birA</td>
<td>Biotin operon repressor/biotin-protein ligase</td>
<td>NC</td>
</tr>
<tr>
<td>yuiG</td>
<td>Unknown function or homologous to B. sphaericus bioY</td>
<td>21</td>
</tr>
<tr>
<td>yhfU</td>
<td>Unknown function or homologous to B. sphaericus bioY</td>
<td>11</td>
</tr>
<tr>
<td><strong>Thiamine</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thiA</td>
<td>Biosynthesis of hydroxymethylpyrimidine phosphoribosyltransferase</td>
<td>62</td>
</tr>
</tbody>
</table>
the addition, the transcription of both genes was derepressed in higher, respectively, in wild-type cells grown in minimal me-
difference levels of yuiG, yfhS (regulated transcription pattern, suggesting that these four
B. sphaericus bioY/growth conditions. Transcription from the unregulated
ribP3 last gene of the operon, ribT than from ribP1
avin (200 RNA isolated from cells grown to exponential phase in minimal medium in the
last transcript levels could be detected between exponential and
stationary growth phases or between minimal medium-
transcript levels were used to determine the onset of stationary phase. Reproducibility improved if a biochemical indicator was used to
to indicate the stationary phase. We arbitrarily chose 30 min
after glucose exhaustion as an indicator of cells entering sta-
tionary phase. By this method, the transcript levels of 439
exponentially expressed genes decreased threefold or more
when cells entered stationary phase. Conversely, transcription of 230 genes poorly or not expressed during exponential phase
were increased threefold or more in stationary phase. Table 3
lists examples of the mostly highly expressed genes from both
classes. Of notable interest, genes acoA, acoC, gvhC, gapB, licB, licC, pckA, and tdh and the rbs gene cluster (rbsA and rbsd as
examples) have been previously shown to be glucose repressed (22, 39). Others are involved in the cold shock response (cspB and
cspD), carbohydrate (gap, kbl, lctE, pdhA, pdhB, yidE, and
yedP) and lipid metabolisms (yüd), transport (lctP), and ribo-
somal function (rplJ and rpsD).
Identification of new transcripts. Selinger et al. (30) have
recently reported on the use of an E. coli oligonucleotide
genome microarray to detect new transcripts within intergenic
regions and located on the noncoding strand of known coding
regions (i.e., antisense RNA). Inclusion of probe sets to the
Bacillus microarray, which were complementary to intergenic
regions of 250 bp or larger, allowed a similar screening for
unidentified transcripts. Approximately 33% of these probe
sets detected a transcript from cells grown in either minimal
medium or LB broth. For many of these signals, we suspect
that they represented either 3' readthrough transcripts or un-
translated 5' leader transcripts from existing genes. To deter-
mine whether any of these transcripts originated from coding
regions, a subclass of transcripts were analyzed that were ori-
ented in the opposite direction relative to transcripts from
known flanking genes. All 35 such transcripts hybridized to a
specific cluster of probe pairs, indicating the presence of a
novel transcript. Subsequent placement of the oligonucleotide
probe sequence onto the DNA sequence identified 20 ORFs
with a B. subtilis-type RBS. In many cases, known Bacillus
promoter sequences and rho-independent transcription termina-
tion sites could also be identified from the DNA sequence.
Several examples of coding regions identified by this method
are listed in Table 4. In most of these examples, changes in the
transcript levels could be detected between exponential and
stationary growth phases or between minimal medium-and LB medium-grown cells. Moreover, in all six examples tested (GAP74-R, GAP123-F, GAP163-2-R, GAP206A-F, GAP240-F, and GAP342-F), these gene expression changes
could be confirmed qualitatively by real-time RT-PCR (data
not shown). For two of the intergenic (GAP66D-R and
GAP163-2-R) regions, two coding regions were identified. The
* Triple real-time RT-PCRs were performed on cDNA prepared from total
RNA isolated from cells grown to exponential phase in minimal medium in the
presence or absence of riboflavin (200 μg/ml), biotin (0.1 μg/ml), or thiamine
(0.34 μg/ml) as described in Materials and Methods. Values presented were
calculated by dividing the average fold change from untreated cells by those from
vitamin-treated cells. The fold change values determined by microarray
analysis were taken from Table 1 for comparison. NC, no change in real-time RT-PCR
values.

<table>
<thead>
<tr>
<th>Expt and gene</th>
<th>Fold change in transcript levelsa as determined by:</th>
<th>RT-PCR</th>
<th>Microarray</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/− riboflavin treatment</td>
<td>ribG</td>
<td>3</td>
<td>2.0</td>
</tr>
<tr>
<td>thiA</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>bioA</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>gap</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>+/− biotin treatment</td>
<td>ribG</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>thiA</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>bioA</td>
<td>8</td>
<td>&gt;100</td>
<td></td>
</tr>
<tr>
<td>gap</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>+/- thiamine treatment</td>
<td>ribG</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>thiA</td>
<td>33</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>bioA</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
<tr>
<td>gap</td>
<td>NC</td>
<td>NC</td>
<td></td>
</tr>
</tbody>
</table>

attenuator transcripts, which has been observed in previous
Northern studies. These transcripts were approximately seven-
fold higher than transcripts of the first four rib genes under
derepressing growth conditions. Transcript levels of ribA and
ribH were found to be similar to those of ribG and ribB, indi-
cating that any increase in transcript levels contributed by the
internal ribP2 promoter was not detected by the ribA and ribH
probe sets. The fact that transcription from ribP2 is weaker
than from ribP1 may account for this observation (28, 29). The
last gene of the operon, ribT, was highly expressed under both
growth conditions. Transcription from the unregulated ribP2
promoter might account for this result and “mask” transcripts
from ribP2, and ribP3.

Interestingly, several new vitamin-regulated genes were
identified by using an algorithm that groups genes according to
their transcription patterns (GeneSpring). For example, tran-
scription of ypaA, a putative transport gene shown to contain
the riboflavin RFN regulatory region (15), was increased >10-
fold in a ribC mutant compared to wild-type cells when both
were grown in minimal medium containing riboflavin (Table 1).
These results are in good agreement with recent ypaA-lacZ
fusion studies (Mouncey and Hohmann, unpublished). Other
examples included two biotin-regulated transcripts, yuiG and
the yhfU. Both yuiG and yhfU encode proteins with strong similarity to the B. sphaericus bioY gene (16). The average difference
levels of yuiG and the yhfU were 20- and 10-fold higher,
respectively, in wild-type cells grown in minimal me-
dium than in cells grown in the presence of biotin (Table 1).
In addition, the transcription of both genes was derepressed in the
B. subtilis bioA mutant. Gene transcripts of three genes
(yhfT, yhfS, and yhfH) adjacent to yhfU showed a similar biotin-
regulated transcription pattern, suggesting that these four
genes are organized as a single operon. Moreover, inspection of the predicted S' leader region of yuiG and yhfU revealed
putative vegetative (r) promoter regions, followed by short
DNA segments with strong sequence homology to the “regu-
lar” site of the B. subtilis bio operon (Fig. 5). The presence
of these DNA elements is consistent with the structure of known
biotin-regulated genes.

Differential expression during transition from exponential
to stationary growth. Control studies initially showed poor
reproducibility (r < 0.9) of transcript levels between duplicated
bacterial cultures of PY79 when standard optical measure-
ments were used to determine the onset of stationary phase. Reproducibility improved if a biochemical indicator was used
to indicate the stationary phase. We arbitrarily chose 30 min
after glucose exhaustion as an indicator of cells entering sta-
tionary phase. By this method, the transcript levels of 439
exponentially expressed genes decreased threefold or more
when cells entered stationary phase. Conversely, transcription of 230 genes poorly or not expressed during exponential phase
were increased threefold or more in stationary phase.
recently updated Subtilist database now indicates that the GAP68D-R region contains a single gene, \textit{ygzB}, which encompasses the two coding regions detected by the microarray. The deduced amino acid sequence of several coding regions showed significant homology to known \textit{B. subtilis} genes (\textit{yobB}) or to related species (e.g., \textit{Bacillus halodurans} and \textit{Bacillus stearothermophilus}). In two cases, the detected genes have been recently identified and characterized (ynxB and \textit{sda}).

However, in coding regions from four intergenic regions, GAP96-R, GAP169A-R, GAP206A, and GAP342-F, no similarity was detected. Most of the coding regions listed in Table 4 are also present in a new public \textit{B. subtilis} genome database developed by Integrated Genomics, Inc. (2, 25). Importantly, coding regions from four intergenic regions, GAP74-R,
TABLE 3. Change in transcript levels of several highly expressed genes from exponential to stationary growth phase

<table>
<thead>
<tr>
<th>Gene</th>
<th>Enzyme and/or functions</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>aceA</td>
<td>Acetoic dehydrogenase E1 component (alpha subunit)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>aceC</td>
<td>Acetoic dehydrogenase E2 component</td>
<td>0.01</td>
</tr>
<tr>
<td>gapB</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>0.013</td>
</tr>
<tr>
<td>glvC</td>
<td>PTS arbutin-like enzyme IIBC component</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>kbl</td>
<td>2-Amino-3-ketobutyrate CoA ligase</td>
<td>0.014</td>
</tr>
<tr>
<td>licB</td>
<td>PTS lichen-specie enzyme IIB component</td>
<td>0.015</td>
</tr>
<tr>
<td>licC</td>
<td>PTS lichen-specie enzyme IIC component</td>
<td>0.013</td>
</tr>
<tr>
<td>pekA</td>
<td>Phosphoenolpyruvate carboxykinase</td>
<td>0.04</td>
</tr>
<tr>
<td>rbsA</td>
<td>Ribose ABC transporter (ATP-binding protein)</td>
<td>0.016</td>
</tr>
<tr>
<td>rbsD</td>
<td>Ribose ABC transporter (membrane protein)</td>
<td>0.019</td>
</tr>
<tr>
<td>tdiD</td>
<td>Threonine 3-dehydrogenase</td>
<td>0.038</td>
</tr>
<tr>
<td>ydeD</td>
<td>Unknown; similar to mannose-6-phosphate isomerase</td>
<td>0.023</td>
</tr>
<tr>
<td>yufD</td>
<td>Unknown; similar to glucan 1,4-alpha-maltohydrolase</td>
<td>0.011</td>
</tr>
<tr>
<td>cspB</td>
<td>Cold shock protein</td>
<td>13.8</td>
</tr>
<tr>
<td>cspD</td>
<td>Cold shock protein</td>
<td>9.7</td>
</tr>
<tr>
<td>gap</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>66.3</td>
</tr>
<tr>
<td>lctE</td>
<td>t-Lactate dehydrogenase</td>
<td>9.4</td>
</tr>
<tr>
<td>lctF</td>
<td>t-Lactate permease</td>
<td>&gt;100</td>
</tr>
<tr>
<td>pdhA</td>
<td>Pyruvate dehydrogenase (E1 α subunit)</td>
<td>14.3</td>
</tr>
<tr>
<td>pdhB</td>
<td>Pyruvate dehydrogenase (E1 β subunit)</td>
<td>18.3</td>
</tr>
<tr>
<td>pufD</td>
<td>Ribosomal protein L10 (BLS)</td>
<td>91.5</td>
</tr>
<tr>
<td>rspD</td>
<td>Ribosomal protein S4 (BS4)</td>
<td>72.0</td>
</tr>
</tbody>
</table>

* CoA, coenzyme A.

b The repression or derepression ratios were calculated by dividing the average difference values (after normalization) from duplicate hybridization experiments of cells grown to exponential phase in minimal medium without vitamin treatment by those grown to stationary phase. See text for details. For most genes (indicated in boldface), average difference values were obtained from duplicate probe sets per microarray.

gap169A-R, gap206A, and gap240-F, are not. Two of these did not show any homology to known proteins.

The presence of probe sets complementary to the noncoding strand of known coding regions also allowed a screening for “antisense” transcripts. In samples from cells grown in minimal or LB medium, hybridization signals exceeding noise by a factor of 5 or greater were observed for 18% (102 of 565) of the known genes tested (see Materials and Methods). In some cases, the transcript levels were observed to change under specific growth conditions or stages. The biological significance of these antisense transcripts is not known. The hybridization data of two such probe sets complementary to the noncoding strand of gapkA (diacetyl glycerol kinase) and yocS (unknown function) are illustrated in Fig. 6. In minimal medium, the yocS antisense transcript was eightfold higher in the exponential stage than in the stationary phase, with a maximum average intensity difference of 3.600. This transcript could be the result of readthrough transcription from odhB, which is upstream from yocS and is oriented in the same direction as this anti-sense transcript. However, a putative rho-independent transcription terminator is located between odhB and yocC that could block readthrough transcription. Moreover, the expression profile of odhB is significantly different, showing high constitutive levels between exponential and stationary growth phases (average intensity difference of between 6,000 and 7,000). The gapkA antisense transcript, alternately, was observed to increase three- to fourfold as cells entered exponential growth phase, with a maximum average intensity difference of 1,600. This gene is located within the middle of a six-gene operon. An antisense transcript from the adjacent gene cdd was not detected, indicating that the gapkA antisense transcript was not caused by readthrough transcription.

DISCUSSION

Genome-wide profiling of B. subtilis has been performed previously with nylon filters (microarrays) or microscope glass slides (microarrays) containing spotted PCR products that cor-
respond to the annotated ORFs. Here we report on the development and testing of an oligonucleotide-based antisense microarray for \textit{B. subtilis} using Affymetrix technology. The array was found to monitor gene expression changes in response to transition from exponential to stationary growth phase and exposure to three different vitamins that repress expression of biosynthetic genes. More importantly, these results were consistent with our current understanding of gene expression and regulation in \textit{B. subtilis}. We anticipate that this array can be applied to assess other gene expression changes such as sporulation, heat shock, stress, competence, and chemotaxis with equal accuracy.

The \textit{Bacillus} antisense microarray was found to detect transcripts for $>70\%$ of the known coding regions. This sensitivity level is in good agreement with gene detection levels reported for \textit{S. pneumoniae}, \textit{H. influenzae}, and \textit{E. coli} antisense Affymetrix oligonucleotide microarrays (10, 17, 30). Caldwell et al. (R. Caldwell, R. Sapolsky, W. Weyler, R. R. Maile, S. Causey, and E. Ferrari, unpublished data) also describe an antisense \textit{B. subtilis} oligonucleotide array that detects $>70\%$ of the known coding regions. In addition to antisense arrays, Affymetrix has commercialized a sense \textit{E. coli} array. This array is reported to detect transcripts from only ca. $50\%$ of the known coding regions (7). It is not clear why the sensitivity level of the sense microarray is lower than the antisense microarray. An underlying cause might be the fact that these microarray systems used two different targets during hybridization (cDNA for antisense arrays and total RNA for sense arrays). Differences in nucleic acid quantity, labeling efficiency of the targets, and hybridization conditions could also contribute to the performance differences of these microarrays.

Analysis of the transcript levels of three vitamin biosynthetic operons for riboflavin, biotin, and thiamine showed that the expression pattern of these operons was dramatically different. Whereas derepression of thiamine- and biotin-regulated biosynthetic genes in wild-type cells varied between 30- and 100-fold, only a 3-fold change was detected for riboflavin-regulated genes. However, a higher level of riboflavin derepression (10- to 30-fold) was observed in the \textit{ribC} mutant strains. These differences could be related to the dissimilar regulatory mechanisms that these pathways exhibit. Biotin-regulated genes are reported to be regulated by a repressor-operator mechanism which controls the ability of the RNA polymerase to bind to the promoter. On the other hand, riboflavin biosynthetic genes are regulated by a termination-antitermination mechanism, which modulates the transcriptional flow from the promoter to the structural genes by means of an active or inactive transcription terminator. Thiamine-regulated genes have regulatory elements of both a highly conserved 39-bp sequence referred to as the \textit{thi} box and a transcriptional terminator within the predicted 5’ leader region (27).

Transcription patterns that are common among a set of genes can be used to identify additional genes that are similarly regulated. Accordingly, we were able to use this fact to identify several new biotin-regulated genes by using a computer-based algorithm. As expected of biotin-regulated genes, the predicted 5’ leader regions of \textit{yuiG} and \textit{yhfU} contained sequences that were homologous to known \textit{bio} operator sequences. However, it is unclear why the level of derepression was significantly lower than that observed for the biosynthetic genes. Further work is necessary to determine the function of these genes and whether they are involved in biotin biosynthesis.

One advantage of short single-stranded oligonucleotide arrays is the ability to monitor expression over a short DNA sequence or a specific region of a transcript without undue cross-hybridization (8, 9, 19, 30). We were able to make use of this property to detect transcripts from intergenic regions of the chromosome that were not previously annotated to contain coding regions. Although most of these represent either 3’ readthrough transcripts or untranslated 5’ leader transcripts from existing genes, analysis of a subclass of transcripts oriented in the opposite direction relative to transcripts from known flanking genes indicated that at least 20 transcripts were synthesized from specific coding regions. Many of these coding regions showed similarity to known genes from recently sequenced \textit{Bacillus} species and from other microorganisms. However, several others did not. As the genomes of more bacilli and other microorganisms are sequenced, determination of the identity of these genes should be possible. In many cases, the expression level of these transcripts differed depend-

![FIG. 6. Detection of antisense RNA transcripts. False-colored images of probe sets to the coding and noncoding (reverse complement) strand of \textit{B. subtilis} \textit{yocS} (top) and \textit{dgkA} (bottom) are shown. The images were derived from hybridization experiments by using cDNA samples prepared from total RNA of \textit{B. subtilis} PY79 grown in minimal medium to either the stationary (\textit{yocS}/\textit{yocS-rc}) or the exponential (\textit{dgkA}/\textit{dgkA-rc}) growth phase. The 20 probe pairs in these probe sets are outlined by a white grid, with the PM (perfect match) features on the top row and the MM (mismatch) features on the bottom row. Black indicates no hybridization.](http://jb.asm.org/Downloadedfrom)
ing on the medium in which the cells were grown (Table 4). A systematic comparison of transcript levels from minimal and rich medium-grown samples should provide leads about the function and regulation of these transcripts. We have also begun to construct null mutations in several of these putative genes to ascertain their function in B. subtilis physiology. To date mutations within GAPI23, GAPI172, GAP206, and GAP240 have not resulted in a discernible phenotype when grown on complex or minimal medium. For several of these transcripts, a coding region could not be detected. It is possible that these regions encode small RNAs (sRNA) that have been shown to have a regulatory function (36). Alternatively, failure to identify an ORF might be caused by sequencing mistakes within the intergenic region that result in an incorrectly predicted ORF start or stop codon (30). In any event, analysis of the remaining intergenic regions is expected to yield more new transcripts. Additional work is necessary to determine their function and role in B. subtilis physiology. Recently, Wassarman et al. (37) reported on the detection of new sRNA in E. coli by a similar method, which also resulted in the detection of putative new short ORFs.

A smaller number of new transcripts detected by the microarray appeared to originate from the noncoding strand of the genome. Approximately 18% of 565 known genes generated an antisense transcript. This suggests that a significant portion of the noncoding strand of the genome is transcribed at a low level. This observation is consistent with a similar analysis of the E. coli genome with a nucleotide-based microarray (30). It remains to be determined, however, whether these “antisense” transcripts simply represent readthrough transcription from known coding regions or cryptic genetic elements or whether they are independent transcripts. Construction of gene fusions, as well as promoter and transcript mapping, should resolve the source of these transcripts. It is also not clear whether they have a regulatory function or indicate the presence of overlapping genes. However, analysis of several of these antisense transcripts did reveal the presence of possible ORFs. For example, a 156-bp ORF was located on the reverse complement of dgkA (Fig. 7). The TTG start of this putative ORF was preceded by a sequence (5'-GCAAGAGGTTG-3') that strongly resembled a Bacillus-like RBS, with a calculated ΔG of −11 kcal/mol (33). The predicted amino acid sequence did not show significant homology to any known protein. In addition, the predicted amino acid sequences from several other ORFs also did not show significant homology to any known protein. Thus, it remains to be determined whether these putative ORFs encode protein products or are just artifacts. However, it is important to point out that the B. subtilis genome, like bacteriophage genomes, does contain genes encoded on the opposite strand of known genes. Integrated Genomics, Inc. has recently shown that ca. 50 overlapping genes have been identified in the B. subtilis genome (M. D’Souza, unpublished results). Further analysis of these transcripts is necessary to resolve these issues.

ACKNOWLEDGMENTS

We gratefully acknowledge Detlef Wolf and Clemens Broger for their help with bioinformatics. We thank Antoine de Saizieu and Ulrich Certa for helpful discussions and critical reading of the manuscript. We also thank Integrated Genomics for providing unpublished genomics data.

REFERENCES