Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome containing the hut and wapA loci

Ken-ichi Yoshida,1 Hidetoshi Sano,1 Shin Seki,1 Masanao Oda,2 Miyuki Fujimura1 and Yasutaro Fujita1

Author for correspondence: Yasutaro Fujita. Tel: +81 849 36 2111. Fax: +81 849 36 2023.

Within the framework of an international project for the sequencing of the entire Bacillus subtilis genome, a 29 kb chromosome segment, which contains the hut operon (335') and the wapA gene, has been cloned and sequenced. This region (28954 bp) contains 21 complete ORFs and one partial one. The 5th, 6th and 17th genes correspond to hutH encoding histidase, hutP encoding the positive regulator for the hut operon and wapA encoding a precursor of three major wall-associated proteins, respectively. A homology search for their products deduced from the 21 complete ORFs revealed that nine of them exhibit significant homology to known proteins such as urocanase (Pseudomonas putida), a protein involved in clavulanic acid biosynthesis (Streptomyces griseus), amino acid permeases (lysine, Escherichia coli; histidine, Saccharomyces cerevisiae; and others), β-glucoside-specific phosphotransferases (E. coli and Erwinia chrysanthemi) and 6-phospho-β-glucosidases (E. coli and Erw. chrysanthemi). Based on the features of the determined sequence and the results of the homology search, as well as on genetic data and sequence of the hut genes reported by other groups, it is predicted that the B. subtilis hut operon may consist of the following six genes (6th-1st), the last of which is followed by a typical ρ-independent transcription terminator: hutP, hutH, EE57A (hutU) encoding urocanase, EE57B (hutI) encoding imidazolone-5-propionate hydrolase, EE57C (hutG) encoding formiminoglutamate hydrolase and EE57D (tentatively designated as hutM) possibly encoding histidine permease. Interestingly, the direction of transcription of these hut genes is opposite to that of the movement of the replication fork.

Keywords: Bacillus subtilis, genome sequencing, wapA locus, hut operon

INTRODUCTION

Within the framework of an international project for the sequencing of the entire Bacillus subtilis genome, we are responsible for the sequencing of an approximately 180 kb chromosomal region between sacS (333') and gnt (344').

In our first communication on this project (Yoshida et al., 1994) we described our sequencing strategy and reported the sequencing and gene features of the 15 kb region containing the iol operon involved in myo-inositol utilization.

In this second communication we describe the sequencing and gene features of the 29 kb region containing the hut operon involved in histidine utilization (Chasin & Magasanik, 1968; Kimhi & Magasanik, 1970; Oda et al., 1988) and the wapA gene encoding a precursor of three major wall-associated proteins (Foster, 1993). We also predict a plausible primary structure for the hut operon.

Abbreviations: GP, translated protein sequence database from the NCBI-GenBank Nucleotide Sequence Database; SP, Swiss-Prot Protein Sequence Database.

The GSDB, DDBJ, EMBL and NCBI accession number for the nucleotide sequence (28954 bp) reported in this paper is D31856.
METHODS

Bacterial strains, phages, plasmids and growth media. *B. subtilis* strain 1A1 (Marburg 168; *tpcC2*) was chosen as our target strain. *Escherichia coli* strains P2392, a P2 lysogen of strain LE392 [hisD514(r5 m5) supE44 supF58 lacY1 or (lacZYE) galK2 galT2 metB1 trpR35], and XLI-Blue {endA1 hisD717(r5 m5) supE44 thi-1 recA1 gyrA96 relA1 (larO lacF1 trpAB' lacIq lacZAM15 Tn10Tet') were used as cloning hosts for *B. subtilis* phage cloning and phage plasmid cloning, respectively. *E. coli* strain JM109 {recA endA1 hisD717 thi-1 relA1 (larO lacF1 trpAB' lacIq lacZAM151) was the host for the plasmids described below. Strain 1A1 was obtained from the *Bacillus* Genetic Stock Center (Ohio) via M. Itaya (Mitsubishi Kasei Institute of Life Sciences, Tokyo), strain P2392, strain XLI-Blue and phage λ DASH II were obtained from Stratagene and phage M13mp19 was obtained from Takara Shuzo.

Plasmids pBC275HE, pSOFT2 and pSOFT11 were used for the preparation of DNA probes for hybridization. pBC275HE carries the 3' part of the *butH* gene (Oda et al., 1988), and pSOFT2 and pSOFT11 are two of the *SfiI* linking clones located at 4017 and 4032 kb on the *B. subtilis* chromosome, respectively (Itaya & Tanaka, 1991), and were supplied by M. Itaya. Plasmids pUC1.13HEY, pUC1.7ES, pUC1.2EA and pUC0.55AS were derived from the unsequenced region (2.8 kb upstream of *hut* (Oda et al., 1988)) (Fig. 1). Phage lysates and DNA were prepared as described previously (Oda et al., 1988). Strain 1Al (Marburg 168; *trpC2*) was chosen as our cloning host for propagation of phages and plasmids to propagate phages and plasmids, respectively.

Cloning and physical mapping of the *hut–wapA* region

To clone the *B. subtilis* *but* region, a *B. subtilis* genomic library constructed using λ DASH II was screened with the DNA of plasmid pBC275HE (Oda et al., 1988) as probe (Fig. 1), the 2.75 kb insert of which contained the 3' half of the *butH* gene (Oda et al., 1988). A recombinant clone, EH2.75-1, carrying a 6.4 kb insert with the above 2.75 kb segment and a *SfiI* site, was isolated (Fig. 1).

When the library was screened with the DNA of a *SfiI* linking plasmid clone, pSOFT2, whose *SfiI* site is located at 4017 bp on the *B. subtilis* chromosome (Itaya & Tanaka, 1991), as probe, a total of 15 positive recombinant clones were obtained. Of these, 10 carried a *NotI* site within their insert, which seemed to correspond to the *NotI* site at 4029 kb on the *B. subtilis* chromosome map (Itaya & Tanaka, 1991). Since the direction from this site to *but* is the same as that from *SfiI* (4017 kb) to *NotI* (4029 kb), the insert (17 kb) of recombinant clone N1.8-3, being the closest to *but*, was used for sequencing.

The insert of plasmid pBC275HE hybridizes with the chromosomal 15 kb *SfiI* *XS* fragment (4017–4032 kb; Itaya & Tanaka, 1991; M. Itaya, personal communication), indicating that the 2.75 kb insert is located between 4017 and 4032 kb. Moreover, the 4.0 kb insert of a *SfiI* linking clone, pSOFT11, the *SfiI* site of which is located at 4032 kb (Itaya & Tanaka, 1991), hybridized with the 6.4 kb insert of λ DASH II clone EH2.75-1 (data not shown), indicating that the *SfiI* site of the 6.4 kb insert is the one at 4032 kb. From the physical maps of the 6.4 and 4.0 kb inserts, and that of the 8 kb *HindIII* insert of *B. subtilis* temperate phage ϕ105but11 (Oda et al., 1988), as well as from the results of the above Southern analysis, a physical map of the *but–wapA* region, with respect to the *EcoRI*, BamHI, HindIII, *SalI*, *NotI* and *SfiI* sites, was constructed (the upper part of Fig. 1). The location of the three inserts is shown on this map. The direction of the *gnt* above. The PCR products of inserts of phage M13mp19 clones were sequenced, and the sequences were compiled and connected. Islands of connected sequences were ordered and the gaps were filled, as described previously (Yoshida et al., 1994).

The 3 kb gap between the *but* (8963 bp) and *wapA* (17060 bp) regions was filled as follows. The inserts of plasmids pUC1.13HEY, pUC1.7ES, pUC1.2EA and pUC0.55AS (Fig. 1) were first sequenced using a Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems) with each plasmid as the template. Several gaps between the resulting islands as well as that between the *wapA* region and the nearest island to it were filled as described above.

Computer analysis. The DNA sequences determined were analysed with the GENETYX-MAC 6.2.0 program (Software Development, Tokyo) and the GeneWorks 2.3 program (Intelli-genetics). Homology searches for putative gene products were performed with the FASTA program (Pearson & Lipman, 1988) using a GenoMeNet FASTA server (Supercomputer Laboratory, Institute for Chemical Research, Kyoto University; e-mail: fasta@genome.ad.jp).

RESULTS AND DISCUSSION

Cloning and physical mapping of the *but–wapA* region

To clone the *B. subtilis but* region, a *B. subtilis* genomic library constructed using λ DASH II was screened with the DNA of plasmid pBC275HE (Oda et al., 1988) as probe (Fig. 1), the 2.75 kb insert of which contained the 3' half of the *butH* gene (Oda et al., 1988). A recombinant clone, EH2.75-1, carrying a 6.4 kb insert with the above 2.75 kb segment and a *SfiI* site, was isolated (Fig. 1).
The **hut-wapA** region of the *B. subtilis* genome

**Fig. 1.** Physical map and gene organization of the **hut-wapA** region deduced from maps of the inserts (thick solid lines) of λ DASH II recombinant clones EH2.75-1 and N1.8-3 obtained by plaque hybridization using plasmids pBC275HE and pSOFT2 as probes (indicated by dashed boxes), respectively, and φ105hut11 and plasmid pSOFT11, as well as the results of genomic Southern analysis described in the text. The map was confirmed by sequencing. The direction to gnt and sacS is indicated at the top. The locations of the two SfiI sites and one NotI site on the physical map of the *B. subtilis* chromosome (Itaya & Tanaka, 1991) are indicated, but their distance requires minor correction. EcoRI (E), BamHI (B), HindIII (H) and SalI (S) sites are also shown. Regions previously sequenced (Oda et al., 1988; Foster, 1993) are shown as meshed boxes. The two sequences of the **hut** and **wapA** regions (1–8963 and 11895–28954) were connected by filling the gap region (8964–11894), as described in Methods. The locations of inserts of plasmid subclones used are indicated: pUC1.7ES (ES), pUC1.13HE (HE), pUC0.55AS (AS) and pUC1.2EA (EA). These were derived from the SalI–HindIII fragment (2.8 kb) of the insert of clone φ105hut11. The AvaI site between the inserts of plasmids pUC0.55AS and pUC1.2EA is also shown. Beneath the physical map, the deduced 21 complete ORFs and one partial one (N17L) found in the sequence of the region are indicated by thick arrows. The first six genes (**hutP**, **hutH**, **EE57A**, **EE57B**, **EE57C** and **EE57D**) are predicted to constitute the **hut** operon. The promoter of the **hut** operon (P**hut**) and a putative **hut** transcription terminator (T**hut**) are also indicated. The N17F, N17G and N17H genes correspond to the orf1, wapA and orf3 genes described by Foster (1993), respectively. The direction of movement of the replication fork is indicated at the bottom of the figure.
the SfiI site of the 6.4 kb insert are located between 4017 and 4032 kb, and at 4032 kb on the B. subtilis chromosome, respectively. The lower part of Fig. 1 shows the physical map of the 17 kb insert of the other λ DASH II recombinant clone (N1.8-3) constructed by the method of Kohara et al. (1987). The distance between the SfiI and NotI sites (4017 and 4029 kb) on the B. subtilis chromosome was calculated to be 12 kb (Itaya & Tanaka, 1991), but the actual distance is only 3.8 kb, indicating that the relative map locations of these sites in the hut–wapA region should be corrected.

**Sequencing of the hut–wapA region**

The sequence (3932 bp) containing the hutP and hutH genes has been reported (indicated as M20659 in Fig. 1; Oda et al., 1988). To extend the sequence of the hut operon downstream, the two EcoRI fragments (4.4 and 1.3 kb) in the 6.4 kb insert of λ DASH II clone EH2.75-1 (Fig. 1) were sequenced systematically, resulting in a 5580 bp sequence. Further sequence extension downstream of hut (929 bp) was performed by means of inverse PCR. These three sequences (3932, 5580 and 928 bp) were connected to give a 8963 bp sequence containing the hut genes.

The 17 kb insert of λ DASH II recombinant clone N1.8-3 was randomly sequenced. When several islands of connected sequences were formed, we found that a 8307 bp sequence containing the wapA gene (indicated as L05634 in Fig. 1; Foster, 1993) was located in this insert. Therefore, sequencing of the flanking regions of the wapA sequence was completed, and resulted in a 17060 bp sequence. The wapA sequence (8307 bp) was corrected by elimination of the first 7 bp which was derived from an EcoRI linker attached in the course of subcloning (S. G. Foster, personal communication) and by inserting another G in front of the G at nt 558 in his sequence (Foster, 1993) to create a SfiI site (4017 kb on the chromosome map of Itaya & Tanaka, 1991).

According to the locations of the regions sequenced above (Fig. 1), the size of the unsequenced region between the two SfiI sites (4017 and 4032 kb) was estimated to be approximately 3 kb. However, we failed to obtain any λ DASH II clone carrying this region. Thus, we filled this gap as described in Methods and obtained a 2931 bp sequence. Final connection of the three sequences (8963, 2931 and 17060 bp) resulted in a 28954 bp sequence for the hut–wapA region.

**Genes found in the hut–wapA region and their features**

Computer analysis of the 28954 bp sequence revealed 21 complete ORFs (EE57D, EE57C, EE57B, EE57A, hutH, hutP, J3A, J3B, J3C, J3D, N17A, N17B, N17C, N17D, N17E, N17F, N17G, N17H, N17I, N17J and N17K) and one partial ORF (N17L) (Fig. 1). Sequencing of hutP (ORF1), hutH (ORF2), and part of the EE57A gene (ORF3) was reported previously by Oda et al. (1988). N17F, N17G and N17H correspond to orf1 (partial ORF), wapA and orf3, respectively, the sequencing of which was reported by Foster (1993). The endpoints of these genes in the sequence and their features, such as molecular size, putative Shine–Dalgarno sequence and initiation codon, are shown in Table 1. Seventeen of the 22 genes were predicted to use ATG as initiation codon, though EE57D and hutP, and EE57C, J3D and N17D might use GTG and TTG, respectively. This suggests that the genes in the hut–wapA region might utilize these initiation codons with a relative frequency similar to that reported recently by Glaser et al. (1993) and Ogasawara et al. (1994). The direction of transcription of these 22 genes, except for the first six, is the same as that of the movement of the replication fork. This is often observed in regions located relatively close to the chromosome origin, like the hut–wapA region (Zeigler & Dean, 1990; Yoshida et al., 1994). Interestingly, the first six genes were predicted to constitute the hut operon, as described below.

**Functions of the first six genes identified in the hut–wapA region, supposedly belonging to the B. subtilis hut operon**

*B. subtilis* histidine degradation requires the four enzymes, histidase, urocanase, imidazolone-5-propionate hydrolase and formiminoglutamate hydrolase (Magasanik et al., 1971), which are encoded by the hutH, hutU, hutI and hutG genes, respectively (Chasin & Magasanik, 1968; Kimhi & Magasanik, 1970). These genes are likely to be organized as a single transcriptional unit in the order hutH, hutU, hutI and hutG, that is, the hut operon (Chasin & Magasanik, 1968; Kimhi & Magasanik, 1970). Oda et al. (1988) cloned the 5' region of the hut operon and determined its nucleotide sequence, demonstrating that the first gene is hutP encoding a positive regulator for this operon and the second is hutH encoding histidase. Transcription starts 32 bp upstream of hutP, to produce a polycistronic mRNA for these hut genes (Oda et al., 1988).

As shown in Fig. 1, hutP and hutH are followed by four unidentified genes in the order hutP, hutH, EE57A, EE57B, EE57C and EE57D. EE57D was followed by a palindromic sequence (complementary to that of nt 320–358), considered to be part of a typical ρ-independent transcription terminator (Fig. 1). We assume that the hut operon might comprise these six genes. We discuss below the functions of the products of these six genes, based on their homology with known proteins (Table 2) and the genetic data of Magasanik and co-
The **but−wapA** region of the *B. subtilis* genome

### Table 1. General features of ORFs found in the **but−wapA** region

<table>
<thead>
<tr>
<th>ORF</th>
<th>Endpoints (nucleotides)</th>
<th>Size of product (amino acids)</th>
<th>Translation start*</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE57D</td>
<td>1788 364</td>
<td>475</td>
<td>AAGGAGGgcgagacagt</td>
</tr>
<tr>
<td>EE57C</td>
<td>2823 1867</td>
<td>319</td>
<td>AgcGAGGcctttcttg</td>
</tr>
<tr>
<td>EE57B</td>
<td>4081 2819</td>
<td>421</td>
<td>AAGGAGTagtgcgagat</td>
</tr>
<tr>
<td>EE57A</td>
<td>5752 4097</td>
<td>552</td>
<td>AAGGGGatgaacatatt</td>
</tr>
<tr>
<td>butH</td>
<td>7275 5752</td>
<td>508</td>
<td>AAGAGGTGgctttcag</td>
</tr>
<tr>
<td>butP</td>
<td>7843 7391</td>
<td>151</td>
<td>tAGGAatTgagagtg</td>
</tr>
<tr>
<td>J3A</td>
<td>8451 9857</td>
<td>469</td>
<td>AgaGAGGcGaatgtttag</td>
</tr>
<tr>
<td>J3B</td>
<td>10167 10529</td>
<td>121</td>
<td>cAGcGAGGTTgacgcg</td>
</tr>
<tr>
<td>J3C</td>
<td>10544 10810</td>
<td>89</td>
<td>AGgGAGGcGaatgtcct</td>
</tr>
<tr>
<td>J3D</td>
<td>10833 12539</td>
<td>569</td>
<td>cGGGAGGcCagagagat</td>
</tr>
<tr>
<td>N17A</td>
<td>12539 12979</td>
<td>147</td>
<td>AAGGAGGGcaaaaaaat</td>
</tr>
<tr>
<td>N17B</td>
<td>13028 13333</td>
<td>102</td>
<td>AAGGAGGcagagat</td>
</tr>
<tr>
<td>N17C</td>
<td>13719 15545</td>
<td>609</td>
<td>AAGGAGGcCagagat</td>
</tr>
<tr>
<td>N17D</td>
<td>15555 16977</td>
<td>481</td>
<td>AAGGAgGcGcGccttt</td>
</tr>
<tr>
<td>N17E</td>
<td>17083 17526</td>
<td>148</td>
<td>AAGGAGGcCagagagat</td>
</tr>
<tr>
<td>N17F</td>
<td>17681 18613</td>
<td>311</td>
<td>tGGGAGG7agaatcGtttag</td>
</tr>
<tr>
<td>N17G</td>
<td>18778 25779</td>
<td>2334</td>
<td>AAGGAGGcGcG</td>
</tr>
<tr>
<td>N17H</td>
<td>25844 26269</td>
<td>142</td>
<td>gAGGAGGaaatag</td>
</tr>
<tr>
<td>N17I</td>
<td>26598 27065</td>
<td>156</td>
<td>gcGAGGAgatgCgGcct</td>
</tr>
<tr>
<td>N17J</td>
<td>27481 27894</td>
<td>138</td>
<td>AAGGAGGcGgGcG</td>
</tr>
<tr>
<td>N17K</td>
<td>28020 28346</td>
<td>109</td>
<td>tAGGAGcataaaccaat</td>
</tr>
<tr>
<td>N17L</td>
<td>28889</td>
<td></td>
<td>AAGGcGagttttag</td>
</tr>
</tbody>
</table>

*Putative initiation codons are shown in bold. Nucleotides identical to the consensus Shine–Dalgarno sequence of AAGGAGGGTG are shown in upper case letters.

† Partial ORF.

workers (Chasin & Magasanik, 1968; Kimhi & Magasanik, 1970). As regards the product of *butP* there was no homologous protein exhibiting significant similarity in the database (FASTA optimized score < 80). The *butH* product, histidase, exhibited very high levels of homology with histidases of several organisms, including man (Suchi et al., 1993), rat (Taylor et al., 1990), mouse (Taylor et al., 1993), *Pseudomonas putida* (Consevage & Phillips, 1990) and *Streptomyces griseus* (Wu et al., 1992), suggesting that histidase is highly conserved among evolutionarily diverse organisms. The EE57A protein showed very high homology to urocanase of *P. putida* (Fessenmaier et al., 1991). The *B. subtilis* *butU* gene encoding urocanase is genetically located next to the *butH* gene in the order *butHUIG* (Kimhi & Magasanik, 1970). Therefore, EE57A most probably corresponds to *butU*. The EE57B protein showed high homology to a protein with unknown function of *Streptomyces gordonii* (GP accession number strsgc_6). Although no evidence was obtained in this homology search, this gene could be considered to be *butI* encoding imidazolone-5-propionate hydrolase for the following reasons. The *B. subtilis* *butI* gene is genetically located between *butU* and *butG* (Kimhi & Magasanik, 1970). The EE57C product exhibited significant homology to the HutG protein, formiminoglutamate hydrolase, of *Klebsiella pneumoniae* (Schwacha & Bender, 1990), implying that EE57C might be the *butG* gene encoding this enzyme. This EE57C product also showed high levels of homology to a protein possibly involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus* (GP accession number stmclvacd_1) and the agmatinase of *E. coli* (Szumanski & Boyle, 1990), implying that these enzymes might have evolved from a common ancestor and/or catalyse similar reactions.

The product of the last gene, EE57D, showed very high levels of homology to many amino acid permeases, such as lysine permease (PIR accession number s24560) and aromatic amino acid permease (Honore et al., 1990) of *E. coli*, histidine permease of *Sacch. cerevisiae* (Tanaka & Fink, 1985) and the *ipa-78d* gene product, which was reported to be an amino acid permease of *B. subtilis* (Glaser et al., 1993). These similarities suggest that this product is a member of the large family of amino acid permeases and is most probably the histidine permease of *B. subtilis*. We therefore tentatively designate the EE57D gene as *butM*. At present, we do not understand why Magasanik and coworkers could not isolate any mutants of *butM*.

From the results of our homology search and the previous findings (Oda et al., 1988; Chasin & Magasanik, 1968; Kimhi & Magasanik, 1970), it is predicted that the B.
Table 2. Results of comparison of nine ORF products with proteins in databases

<table>
<thead>
<tr>
<th>ORF*</th>
<th>Homologous protein</th>
<th>Database: entry†</th>
<th>FASTA optimized score</th>
</tr>
</thead>
<tbody>
<tr>
<td>butH</td>
<td>Histidase</td>
<td>GP: D16626 1</td>
<td>980</td>
</tr>
<tr>
<td></td>
<td>(Man)</td>
<td>PIR: a46128</td>
<td>975</td>
</tr>
<tr>
<td></td>
<td>(Mouse)</td>
<td>SP: hutn,psedu</td>
<td>933</td>
</tr>
<tr>
<td></td>
<td>(P. putida)</td>
<td>SP: hutn_strgr</td>
<td>807</td>
</tr>
<tr>
<td>EE57A</td>
<td>Urocanase (P. putida)</td>
<td>PIR: s17184</td>
<td>1898</td>
</tr>
<tr>
<td>EE57B</td>
<td>Unidentified gene product (Strep. griseus)</td>
<td>GP: strsca_6</td>
<td>552</td>
</tr>
<tr>
<td>EE57C</td>
<td>A protein involved in clavulanic acid biosynthesis (Strep. clavuligerus)</td>
<td>GP: stmelvacd_1</td>
<td>311</td>
</tr>
<tr>
<td></td>
<td>Agmatinase (E. coli)</td>
<td>SP: speb,ecoli</td>
<td>186</td>
</tr>
<tr>
<td></td>
<td>HutG protein (fragment) (K. pneumonia)</td>
<td>PIR: a36730</td>
<td>167</td>
</tr>
<tr>
<td>EE57D</td>
<td>ipa-78d gene product (B. subtilis)</td>
<td>GP: bsgenr.79</td>
<td>1607</td>
</tr>
<tr>
<td></td>
<td>Lysine-specific permease (E. coli)</td>
<td>PIR: s24560</td>
<td>1055</td>
</tr>
<tr>
<td></td>
<td>Aromatic amino acid transport protein (E. coli)</td>
<td>PIR: grecaa</td>
<td>912</td>
</tr>
<tr>
<td></td>
<td>Histidine permease (Sacch. cerevisiae)</td>
<td>PIR: a24519</td>
<td>603</td>
</tr>
<tr>
<td>J3A</td>
<td>Endo-1,5-α-1-arabinase (A. niger)</td>
<td>GP: asnabna_1</td>
<td>187</td>
</tr>
<tr>
<td>N17C</td>
<td>Phosphotransferase enzyme II, β-glucoside-specific (Erw. chrysanthemi)</td>
<td>PIR: b42603</td>
<td>1164</td>
</tr>
<tr>
<td></td>
<td>(E. coli)</td>
<td>GP: ecouw82,86</td>
<td>1152</td>
</tr>
<tr>
<td>N17D</td>
<td>6-Phospho-β-glucosidase</td>
<td>PIR: a25977</td>
<td>1609</td>
</tr>
<tr>
<td></td>
<td>BglB protein (E. coli)</td>
<td>PIR: s27553</td>
<td>1544</td>
</tr>
<tr>
<td></td>
<td>AscB protein (E. coli)</td>
<td>SP: arbb,erwch</td>
<td>1580</td>
</tr>
<tr>
<td>N17G (wapA)</td>
<td>RhsD protein precursor (E. coli)</td>
<td>PIR: j0625</td>
<td>228</td>
</tr>
</tbody>
</table>

* Only ORFs whose products exhibited significant homology to known proteins (FASTA optimized score > 165) are listed.
† GP database entries are followed by an underlined space and an ordinal number for each coding sequence.

"subtilis but" operon might consist of the six genes in the following order from the 5' end; butP encoding the positive regulator for this operon, butH encoding histidase, EE57A (butU) encoding urocanase, EE57B (butI) encoding imidazolone-5-propionate hydrolase, EE57C (butG) encoding formiminoglutamate hydrolase and EE57D (butM) possibly encoding histidine permease. To verify this prediction, analysis of the but transcript and gene disruption of the last four unidentified genes are in progress.

Homology of the other gene products encoded in the "but-wapA" region

Although 11 putative gene products out of the remaining 15 did not show any significant homology to known proteins, four products ([J3A, N17C, N17D and N17G (wapA)] had homologous proteins (Table 2). The J3A protein showed significant homology to endo-1,5-α-1-arabinase of Aspergillus niger (Filiphi et al., 1993). The N17C and N17D proteins exhibited very high levels of homology to E. coli and Erwinia chrysanthemi β-glucoside-specific phosphotransferase enzymes II (Bramley & Kornberg, 1987; El Hassouni et al., 1992) and 6-phospho-β-glucosidases of the same organisms (Schnetz et al., 1987; Hall & Xu, 1992), suggesting that the two proteins might have the same functions as those respective enzymes. D. Le Coq (INAPG, France) communicated to us that he and C. Lindner recently sequenced the 43 kb region containing the N17C (yatA; Steinmetz & Richter, 1994) and N17D genes independently. Their sequence was found to be identical to our sequence, except for six positions. The N17G protein (WapA) encoding a precursor of three major wall-associated proteins showed significant homology to the RhsD protein of E. coli (Sadosky et al., 1991), as reported by Foster (1993).

ACKNOWLEDGEMENTS

We wish to thank the following undergraduate students, T. Tokuyama, T. Watanabe, Y. Yonezawa and T. Yamada, for their help with the experiments. We are also grateful to M. Itaya...
for providing plasmids pSOFT2 and pSOFT11, and communicating his unpublished results. We also thank D. Le Coq and S. J. Foster for their personal communications. This work was supported by a Grant-in-Aid for Creative Basic Research on 'Human Genome Analysis' from the Ministry of Education, Science and Culture of Japan.

REFERENCES


Received 13 July 1994; revised 13 August 1994; accepted 28 September 1994.