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.

¹ Department of Biotechnology, Faculty of Engineering, Fukuyama University, Fukuyama 729-02, Japan

² National Institute of Bioscience and Human Technology, Agency of Industrial Science and Technology, MITI, Tsukuba, Tsukuba City, Ibaraki 305, Japan

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 *p*-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°).

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.

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.

METHODS

Bacterial strains, phages, plasmids and growth media. *B. subtilis* strain 1A1 (Marburg 168; *trpC2*) was chosen as our target strain. *Escherichia coli* strains P2392, a P2 lysogen of strain LE392 [*hsdR514*(r_k^- m_k^+) *supE44* *supF58* *lacY1* or Δ (*lacIZY*)6 *galK2* *galT22* *metB1* *trpR55*], and XL1-Blue [*endA1* *hsdR17*(r_k^- m_k^+) *supE44* *thi-1* *recA1* *gyrA96* *relA1* Δ (*lac*) F' [*proAB*⁺ *lacI*^q *lacZAM15* Tn10Tet^r]] were used as cloning hosts for λ DASH II and M13mp19 phage clones, respectively. *E. coli* strain JM109 [*recA1* *supE44* *endA1* *hsdR17* *gyrA96* *relA1* *thi* Δ (*lac-proAB*) F' [*traD36* *proAB*⁺ *lacI*^q *lacZAM15*]] 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 XL1-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 *hutH* 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 pUC1.13HE, pUC1.7ES, pUC1.2EA and pUC0.55AS were subclones obtained using plasmid pUC19, the inserts of which were derived from the unsequenced region (2.8 kb) upstream of the *hut* operon of the 8 kb insert of *B. subtilis* temperate phage ϕ 105*hut11* (Oda *et al.*, 1988) (Fig. 1). Phage lysates and DNA were prepared as described previously (Oda *et al.*, 1988).

NZY broth and L-broth were used for growth of the *E. coli* strains to propagate phages λ DASH II and M13mp19, respectively (Yoshida *et al.*, 1994); the latter medium was also used for cultivation of plasmid-bearing strain JM109.

Cloning and sequencing of the 29 kb *hut-wpaA* region. A *B. subtilis* genomic library was prepared using λ DASH II and screened by means of plaque hybridization as described previously (Yoshida *et al.*, 1994) to obtain λ DASH II recombinant clones (EH2.75-1 and N1.8-3) containing the *hut* and *wapA* genes, respectively.

To sequence the 5.7 kb region comprising two *EcoRI* fragments (1.3 and 4.4 kb) within the 6.4 kb insert of one λ DASH II recombinant clone, EH2.75-1, the *EcoRI* fragments were mixed and ligated, randomly fragmented with DNase I, and cloned into phage M13mp19 as described previously (Yoshida *et al.*, 1994). The PCR products of inserts of phage M13mp19 recombinant clones (400–1000 bp in length) were sequenced with a DNA sequencer (373A; Applied Biosystems) using a Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems). The compiled sequences were connected to form separate sequences of the two *EcoRI* fragments, which were aligned with reference to a physical map of the 6.4 kb insert. To extend the sequence of the 5.7 kb region downstream of the *hut* operon, we employed inverse PCR (Triglia *et al.*, 1988) as follows. As the 4.0 kb *HindIII* insert of plasmid pSOFT11 was found to overlap the 5.7 kb region by approximately 3 kb (Fig. 1), 100 ng of chromosomal DNA was digested with *HindIII* and circularized with T4 DNA ligase at a DNA concentration of less than 2 μ g ml⁻¹. The inverse PCR product was sequenced as described above. The gap was filled by sequencing PCR products prepared using synthetic primers, as described previously (Yoshida *et al.*, 1994).

The insert (17 kb) of the other λ DASH II recombinant clone (N1.8-3) was obtained as two *NotI* fragments (3 and 14 kb) (Fig. 1). Each fragment was isolated, ligated separately, randomly fragmented and cloned into phage M13mp19 as described

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 *hut* (8963 bp) and *wapA* (17060 bp) regions was filled as follows. The inserts of plasmids pUC1.13HE, 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 (Intelligenetics). 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 *hut-wapA* region

To clone the *B. subtilis* *hut* 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 *hutH* 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 *hut* 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 *hut*, 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 ϕ 105*hut11* (Oda *et al.*, 1988), as well as from the results of the above Southern analysis, a physical map of the *hut-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*

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 *Aval* 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.

operon was inferred from the fact that the 2.75 kb insert and the *Sfi*I 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 *Sfi*I and *Not*I 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 *Eco*RI 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 *Eco*RI 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 *Sfi*I 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 *Sfi*I 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.

As a prerequisite for deducing the functions of the newly found genes, we searched for homology of their deduced products with known protein sequences present in a non-redundant protein sequence database (Institute for Chemical Research, Kyoto University) made up of the PIR, SP and GP Protein Databases, using the FASTA program to obtain the FASTA optimized score (Pearson & Lipman, 1988). (The last homology search was performed in May, 1994.) This search revealed that nine of them exhibit significant homology to known proteins, as shown in Table 2 (FASTA optimized score > 165), the other 12 proteins showing low homology (FASTA optimized score < 110), implying that they might be unique proteins.

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 *p*-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-

Table 1. General features of ORFs found in the *hut-wapA* region

ORF	Endpoints (nucleotides)		Size of product (amino acids)	Translation start*
EE57D	1788	364	475	AAGGAGGgcgagacag tg
EE57C	2823	1867	319	AgGGAGccattc ttg
EE57B	4081	2819	421	AAGGAGtgatgac atg
EE57A	5752	4097	552	AAGGgGGatgaacat atg
<i>hutH</i>	7275	5752	508	AAGGAGtTGggctt atg
<i>hutP</i>	7843	7391	151	tAGGAatTGag tg
J3A	8451	9857	469	AgaGAGGcGaattgtttag atg
J3B	10167	10529	121	cAcGAGGTGttgcggc atg
J3C	10544	10810	89	AgtGAGGTGagcaca atg
J3D	10833	12539	569	cgGGAGGacaaggga ttg
N17A	12539	12979	147	AAGGAGGcaaaaaa atg
N17B	13028	13333	102	AAGGAGcaagatgac atg
N17C	13719	15545	609	AAGGAGGTatacac atg
N17D	15535	16977	481	AAGaAGcgctgc ttg
N17E	17083	17526	148	AAGGcGGgaagagc atg
N17F (<i>orf1</i>)	17681	18613	311	ttGGAGGTagaatacgtttg atg
N17G (<i>wapA</i>)	18778	25779	2334	AAGGAGagaggaa atg
N17H (<i>orf3</i>)	25844	26269	142	gAGGAGGaaat atg
N17I	26598	27065	156	gcGGAGGgaatgcggctc atg
N17J	27481	27894	138	AAaGAGGcagggg atg
N17K	28020	28346	109	tAGGAGcataacca atg
N17L†	28889			AgGGAtGaGatt atg

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

† Partial ORF.

workers (Chasin & Magasanik, 1968; Kimhi & Magasanik, 1970). As regards the product of *hutP* there was no homologous protein exhibiting significant similarity in the database (FASTA optimized score < 80). The *hutH* 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* (Conseville & 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* *hutU* gene encoding urocanase is genetically located next to the *hutH* gene in the order *hutHUIG* (Kimhi & Magasanik, 1970). Therefore, EE57A most probably corresponds to *hutU*. The EE57B protein showed high homology to a protein with unknown function of *Streptomyces gordonii* (GP accession number strscaa 6). Although no evidence was obtained in this homology search, this gene could be considered to be *hutI* encoding imidazolone-5-propionate hydrolase for the following reasons. The *B. subtilis* *hutI* gene is genetically located between *hutU* and *hutG* (Kimhi & Magasanik, 1970). The EE57C product exhibited significant homology to the HutG protein, formimino-glutamate hydrolase, of *Klebsiella pneumoniae* (Schwacha &

Bender, 1990), implying that EE57C might be the *hutG* 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 & Cole, 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 *hutM*. At present, we do not understand why Magasanik and coworkers could not isolate any mutants of *hutM*.

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

ORF*	Homologous protein	Database: entry†	FASTA optimized score
<i>hutH</i>	Histidase		
	(Man)	GP: D16626_1	980
	(Mouse)	PIR: a46128	975
	(<i>P. putida</i>)	SP: huth_psepu	933
	(<i>Strep. griseus</i>)	SP: huth_strgr	807
<i>EE57A</i>	Urocanase (<i>P. putida</i>)	PIR: s17184	1898
<i>EE57B</i>	Unidentified gene product (<i>Strep. griseus</i>)	GP: strscas_6	552
<i>EE57C</i>	A protein involved in clavulanic acid biosynthesis (<i>Strep. clavuligerus</i>)	GP: stmclvacd_1	311
	Agmatinase (<i>E. coli</i>)	SP: speb_ecoli	186
	HutG protein (fragment) (<i>K. pneumoniae</i>)	PIR: a36730	167
<i>EE57D</i>	<i>ipa-78d</i> gene product (<i>B. subtilis</i>)	GP: bsgenr_79	1607
	Lysine-specific permease (<i>E. coli</i>)	PIR: s24560	1055
	Aromatic amino acid transport protein (<i>E. coli</i>)	PIR: grecaa	912
	Histidine permease (<i>Sacch. cerevisiae</i>)	PIR: a24519	603
<i>J3A</i>	Endo-1,5- α -L-arabinase (<i>A. niger</i>)	GP: asnabna_1	187
<i>N17C</i>	Phosphotransferase enzyme II, β -glucoside-specific		
	(<i>Erw. chrysanthemi</i>)	PIR: b42603	1164
	(<i>E. coli</i>)	GP: ecouw82_86	1152
<i>N17D</i>	6-Phospho- β -glucosidase		
	BglB protein (<i>E. coli</i>)	PIR: a25977	1609
	AscB protein (<i>E. coli</i>)	PIR: s27553	1544
	ArbB protein (<i>Erw. chrysanthemi</i>)	SP: arbb_erwch	1580
<i>N17G (wapA)</i>	RhsD protein precursor (<i>E. coli</i>)	PIR: js0625	228

* 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 hut operon might consist of the six genes in the following order from the 5' end; *hutP* encoding the positive regulator for this operon, *hutH* encoding histidase, *EE57A* (*hutU*) encoding urocanase, *EE57B* (*hutI*) encoding imidazolone-5-propionate hydrolase, *EE57C* (*hutG*) encoding formiminoglutamate hydrolase and *EE57D* (*hutM*) possibly encoding histidine permease. To verify this prediction, analysis of the *hut* transcript and gene disruption of the last four unidentified genes are in progress.

Homology of the other gene products encoded in the *hut-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 -L-arabinase of *Aspergillus niger* (Flippin *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 4.3 kb region containing the *N17C* (*yltA*; 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

- Bramley, H. F. & Kornberg, H. L. (1987).** Nucleotide sequence of *bglC*, the gene specifying enzyme II^{bgl} of the PEP:sugar phosphotransferase system in *Escherichia coli* K12, and overexpression of the gene product. *J Gen Microbiol* **133**, 563–573.
- Chasin, L. A. & Magasanik, B. (1968).** Induction and repression of the histidine-degrading enzymes of *Bacillus subtilis*. *J Biol Chem* **243**, 5165–5178.
- Conseville, M. W. & Phillips, A. T. (1990).** Sequence analysis of the *hutH* gene encoding histidine ammonia-lyase in *Pseudomonas putida*. *J Bacteriol* **172**, 2224–2229.
- El Hassouni, M., Henrissat, B., Chippaux, M. & Barras, F. (1992).** Nucleotide sequences of the *arb* genes, which control β -glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli bgl* operon and evidence for a new β -glycosidase family including enzymes from eubacteria, archaeobacteria, and humans. *J Bacteriol* **174**, 765–777.
- Fessenmaier, M., Frank, R., Retey, J. & Schubert, C. (1991).** Cloning and sequencing the urocanase gene (*butU*) from *Pseudomonas putida*. *FEBS Lett* **286**, 55–57.
- Flipphi, M. J. A., Panneman, H., van der Veen, P., Visser, J. & Graaff, L. H. (1993).** Molecular cloning, expression and structure of the endo-1,5- α -L-arabinase gene of *Aspergillus niger*. *Appl Microbiol Biotechnol* **40**, 318–326.
- Foster, S. J. (1993).** Molecular analysis of three major wall-associated proteins of *Bacillus subtilis* 168: evidence for processing of the product of a gene encoding a 258 kDa precursor two-domain ligand-binding protein. *Mol Microbiol* **8**, 299–310.
- Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweizer, J., Vertès, A., Rapoport, G. & Danchin, A. (1993).** *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.
- Hall, B. G. & Xu, L. (1992).** Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. *Mol Biol Evol* **9**, 688–706.
- Honore, N. & Cole, S. T. (1990).** Nucleotide sequence of the *aroP* gene encoding the general aromatic amino acid transport protein of *Escherichia coli* K-12: homology with yeast transport proteins. *Nucleic Acids Res* **18**, 653.
- Itaya, M. & Tanaka, T. (1991).** Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J Mol Biol* **220**, 631–648.
- Kimhi, Y. & Magasanik, B. (1970).** Genetic basis of histidine degradation in *Bacillus subtilis*. *J Biol Chem* **245**, 3545–3548.
- Kohara, Y., Akiyama, K. & Isono, K. (1987).** The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495–508.
- Magasanik, B., Kaminskas, E. & Kimhi, Y. (1971).** Histidine degradation (*Bacillus subtilis*). *Methods Enzymol* **17B**, 45–62.
- Oda, M., Sugishita, A. & Furukawa, K. (1988).** Cloning and nucleotide sequences of histidase and regulatory genes in the *Bacillus subtilis hut* operon and positive regulation of the operon. *J Bacteriol* **170**, 3199–3205.
- Ogasawara, N., Nakai, S. & Yoshikawa, H. (1994).** Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res* **1**, 1–14.
- Pearson, W. R. & Lipman, D. J. (1988).** Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Sadosky, A. B., Gray, J. A. & Hill, C. W. (1991).** The RhsD–E subfamily of *Escherichia coli* K-12. *Nucleic Acids Res* **19**, 7177–7183.
- Schnetz, K., Toloczky, C. & Rak, B. (1987).** β -Glucoside (*bgl*) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. *J Bacteriol* **169**, 2579–2590.
- Schwacha, A. & Bender, R. A. (1990).** Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Klebsiella aerogenes*. *J Bacteriol* **172**, 5477–5481.
- Steinmetz, M. & Richter, R. (1994).** Easy cloning of mini-Tn10 insertions from the *Bacillus subtilis* chromosome. *J Bacteriol* **176**, 1761–1763.
- Suchi, M., Harada, N., Wada, Y. & Takagi, Y. (1993).** Molecular cloning of a cDNA encoding human histidase. *Biochim Biophys Acta* **1216**, 293–295.
- Szumanski, M. B. W. & Boyle, S. M. (1990).** Analysis and sequence of the *speB* gene encoding agmatine ureohydrolase, a putrescine biosynthetic enzyme in *Escherichia coli*. *J Bacteriol* **172**, 538–547.
- Tanaka, J. & Fink, G. R. (1985).** The histidine permease gene (HIP1) of *Saccharomyces cerevisiae*. *Gene* **38**, 205–214.
- Taylor, R. G., Lambert, M. A., Sexsmith, E., Sadler, S. J., Ray, P. N., Mahuran, D. J. & McInnes, R. R. (1990).** Cloning and expression of rat histidase, homology to two bacterial histidases and four phenylalanine ammonia-lyases. *J Biol Chem* **265**, 18192–18199.
- Taylor, R. G., Grieco, D., Clarke, G. A., McInnes, R. R. & Taylor, B. A. (1993).** Identification of the mutation in murine histidinemia (*his*) and genetic mapping of the murine histidase locus (*Hal*) on chromosome 10. *Genomics* **16**, 231–240.
- Triglia, T., Peterson, M. G. & Kemp, D. J. (1988).** A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res* **16**, 8186.
- Wu, P.-C., Kroening, T. A., White, P. J. & Kendrick, K. E. (1992).** Purification of histidase from *Streptomyces griseus* and nucleotide sequence of the *hutH* structural gene. *J Bacteriol* **174**, 1647–1655.
- Yoshida, K., Sano, H., Miwa, Y., Ogasawara, N. & Fujita, Y. (1994).** Cloning and nucleotide sequencing of a 15 kb region of the *Bacillus subtilis* genome containing the *iol* operon. *Microbiology* **140**, 2289–2298.
- Zeigler, D. R. & Dean, D. H. (1990).** Orientation of genes in the *Bacillus subtilis* chromosome. *Genetics* **125**, 703–708.

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