Sequence analysis of a 50 kb region between *spo0H* and *rrnH* on the *Bacillus subtilis* chromosome

Kenji Yasumoto, Hongtu Liu, Sang Min Jeong, Yukio Ohashi, Shinji Kakinuma,[†] Kan Tanaka, Fujio Kawamura,[‡] Hirofumi Yoshikawa and Hideo Takahashi

Author for correspondence: Hideo Takahashi. Tel: +81 3 3812 2111 ext. 7825. Fax: +81 3 3813 0539.

Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

The 49630 bp *spo0H*–*rrnH* region of the *Bacillus subtilis* genome has been fully sequenced. The sequence contains one partial and 62 complete ORFs, one partial and three complete rRNA genes and a cluster of six tRNA genes. The direction of the transcription and translation of 61 ORFs is the same as that of the movement of the replication fork. A homology search of 40 ORFs in newly determined sequence revealed that 27 of them had significant similarity to known proteins such as elongation factor G, elongation factor Tu, pseudouridine synthase I and ribosomal proteins. Two adjacent genes, *ybaD* and *ybaE*, appeared to encode proteins belonging to the ATP-binding cassette (ABC) family.

Keywords: *Bacillus subtilis*, genome sequencing, ribosomal protein gene cluster

INTRODUCTION

Within the framework of the international project for the sequencing of the entire *Bacillus subtilis* genome, we are responsible for sequencing a chromosomal region of about 150 kb between *spo0H* (10°) and pNEXT9 (24°). In this study, we determined the sequence of about 30 kb in the region between 10° and 15°. We combined the results with known sequences and describe here a continuous 49 630 bp sequence from *spo0H* (10°) to *rrnH* (15°).

METHODS

Bacterial strains, phage and plasmids. *B. subtilis* 168 (*trpC2*), from our laboratory stock, was used as a template for PCR. *Escherichia coli* DH5 α (*supE44 lacU169* [ϕ 80 *lacZ*ΔM15] *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) was used as the host for cloning with plasmid pTZ18R (Toyobo). The λ phage K4F811

derived from EMBL 3a was supplied by F. Morohoshi (National Cancer Center, Tokyo, Japan).

Sequence determination. All PCR products to be sequenced were self-ligated and fragmented by sonic oscillation, then 400–700 bp fragments were purified by means of agarose gel electrophoresis. These fragments were cloned into a vector (pTZ18R) and randomly overlapping libraries for shotgun sequencing were prepared. Inserts of pTZ18R were amplified by PCR using a forward (5' CGTCAAGGCGATTAAGTTGGTAA 3') and a reverse (5' TGCTTCCGGCTCGTATGT-TGTGTG 3') sequencing primer based on the vector sequence. The amplified PCR products were purified by precipitation with polyethylene glycol and used as templates for sequencing. Sequencing reactions were performed on both strands using the CATALYST Molecular Biology LabStation – Model 800 and Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems), and were then analysed with a Model 373A DNA Sequencer (Applied Biosystems).

Data handling and computer analysis. DNA sequences were compiled using the GENETYX-MAC ATSQ (Software Development) and GENEWORKS programs (IntelliGenetics). We used sequence outputs only from the regions where every base was well separated (about 400 bp in one run). The compiled sequence was further searched for possible ORFs using GENETYX-MAC and GENEWORKS. A homology search of the identified ORFs was performed with the FASTA program using GENETYX-CD (Software Development) and a GenomeNet FASTA server in the Human Genome Center, Institute of Medical Science, University of Tokyo, Japan.

[†]Present address: Patent Department, Research and Technology Headquarters, Showa Denko K.K., Minato-ku, Tokyo 105, Japan.

[‡]Present address: Laboratory of Molecular Biology, College of Science, Rikkyo (St Paul's) University, Toshima-ku, Tokyo 171, Japan.

The GenBank accession numbers for the nucleotide sequences reported in this paper are D50303, D64127, D50302, D64125, D64124 and D64126.

Fig. 1. Gene organization of the *spo0H-rrnH* region of the *B. subtilis* genome. ORFs are indicated by boxes. ORFs above the lines are those of which the direction of translation is the same as that of the movement of replication fork, and those below the lines are those translated in the opposite direction. ORFs were categorized into five groups: A, function assigned; B, high similarity to known protein; C, low similarity to known protein; D, high similarity to protein of unknown function; E, no significant similarity to other protein. The shaded bars indicate the regions that we sequenced in this study (accession numbers D50303, D64127, D50302, D64125, D64124, D64126). Previously known sequences are indicated by arrows. Accession numbers for each of the regions are also shown. The numbers with asterisks are late entry sequences and were not available when we started sequencing. The sequence for the region ref** is in the report of Wawrousek & Hansen (1983). The tRNA gene cluster (#) contains six tRNA genes for Asn, Thr, Gly, Arg, Pro and Ala.

RESULTS AND DISCUSSION

Sequencing of the *spo0H-rrnH* region of the *B. subtilis* genome

Several parts of the *spo0H-rrnH* region have already been sequenced (Fig. 1). In addition, we sequenced a clone containing a fragment that proved to be derived from *ybaC* and had received the λ clone K4F811 that was thought to contain the region upstream of *rrnI* (F. Morohoshi, personal communication). Starting with these seven 'islands' (known sequences) from the *spo0H-rrnH* region, we amplified additional segments by means of PCR to try to bridge the gaps using synthesized primers based on known sequences.

Assignment of ORFs in the *spo0H-rrnH* region

ORFs were searched for in all six possible translation frames. We considered the longest possible ORF starting with ATG, GTG or TTG and preceded by a Shine-Dalgarno (SD) sequence to be the most likely coding sequence in any region. By these criteria, 40 new ORFs were identified in the sequenced regions, giving a total of 63 putative ORFs, 4 rRNAs and 6 tRNAs in the *spo0H-rrnH* interval (Table 1, Fig. 1). Those genes to which functions could not be assigned were named according to the guidelines of the SwissProt protein sequence database as *ybaA-ybaO*. The initiation codon for 53 ORFs was ATG, GTG for 5 and TTG for 4. The direction of transcription and translation of all but two of

Table 1. ORFs in the 49630 bp sequence between *spo0H* and *rrnH* of the *B. subtilis* genome

ORF	Endpoints (nt)	Size of product (aa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>spo0H</i>	1 > 319	79	
<i>rpmG</i>	412 > 561	49	AgAAAGGtGtaatgacatg
<i>secE</i>	595 > 774	59	AttgtGGAGGTcttttacatg
<i>nusG</i>	953 > 1486	177	GgAAGGActGGAcAGTctctgaatagaatg
<i>rplK</i>	1654 > 2079	141	AAGGAGGTgtgtctcgtg
<i>rplA</i>	2173 > 2871	232	AAGGAGGaaattttaaatg
<i>rplJ</i>	3123 > 3623	166	AcAGGAGGTgtaaccatg
<i>rplL</i>	3670 > 4041	123	AaAAtGGAGGaatcaaatg
<i>ybaA</i>	4131 > 4736	201	AaAAAGGAGGTtctaataatg
<i>rpoB</i>	4982 > 8563	1193	GAttGAGGggtgaatcagtgtg
<i>rpoC</i>	8625 > 12224	1199	AaAAgGGAGGTaggcccttg
<i>ybaB</i>	12405 > 12653	82	ActttGGAGGatattgttatg
<i>rpsL</i>	12767 > 13183	138	GAAgGGAGGaaaaattcatg
<i>rpsG</i>	13225 > 13695	156	GAAgGGAGGatattcagatg
<i>fus</i>	13749 > 15827	692	AGgAAGGAGaaattaccaatg
<i>tufA</i>	15947 > 17137	396	AAGGAGGattttagaatg
<i>ybaC</i>	17236 > 18192	318	AtAttGAGGggttaatttatg
<i>rpsJ</i>	18428 > 18736	102	AAAGGAGGgaaaataatg
<i>rplC</i>	18776 > 19405	209	AtAGGAGGTgtgacgaatg
<i>rplD</i>	19433 > 20056	207	GAAAGGAGGaaattgattatg
<i>rplW</i>	20056 > 20343	95	AgAGGAGGTgcttgcataatg
<i>rplB</i>	20375 > 21208	277	AagAAGGAGGgaaattcaaaatg
<i>rpsS</i>	21266 > 21544	92	GAAgGGAGGTtccaaaatg
<i>rplV</i>	21561 > 21902	113	AGAgAGGAGGcttttaaatg
<i>rpsC</i>	21906 > 22562	218	AAGGAGGgataagtagtg
<i>rplP</i>	22564 > 22998	144	AGgAAGGAGGaaaataaatg
<i>rpmC</i>	22988 > 23188	66	AttGGtGGTgaatcaatg
<i>rpsQ</i>	23211 > 23474	87	GAggGGAGGTcgtgagcatg
<i>rplN</i>	23515 > 23883	122	GAAgGGAGGTaacaataatg
<i>rplX</i>	23921 > 24232	103	AAGGtGGTgcgatcagatg
<i>rplE</i>	24259 > 24798	179	GAAgGGAGGTctatctcatg
<i>rpsN</i>	24821 > 25006	61	AAgGGAGGcgaaatcgtg
<i>rpsH</i>	25038 > 25436	132	GgAAGGAGGTaattatataatg
<i>rplF</i>	25466 > 26005	179	GAAAtGGAGGTgtttggtatg
<i>rplR</i>	26038 > 26400	120	AGAAAGGAGtgactagatg
<i>rpsE</i>	26425 > 26925	166	AAGGAGGgacacatgattatg
<i>rpmD</i>	26939 > 27118	59	AAGGAGGgaacattaatg
<i>rplO</i>	27149 > 27589	146	AAGGAGGTgtcctaataatg
<i>secY</i>	27591 > 28886	431	AcAgctGAGGTgatctaacttg
<i>adk</i>	28941 > 29594	217	AAgaGAGGAcGGgGGaaatccgaatg
<i>map</i>	29591 > 30337	248	cttGGAGGatataaaaaatg
<i>infA</i>	30649 > 30867	72	AGAAgGGAGaacttgaaatg
<i>rpmJ</i>	30901 > 31014	37	AAGGAGGTtggaacacatg
<i>rpsM</i>	31037 > 31402	121	AAGGAGGTgccagagaatg
<i>rpsK</i>	31423 > 31818	131	AAgGGAGGTtcttaacaatg
<i>rpoA</i>	31995 > 32939	314	AAGGAGGgttttaagtaatg
<i>rplQ</i>	33017 > 33379	120	AtAAAGGAGGggacatcacatg
<i>ybaD</i>	33507 > 34352	281	GAgAGGAGGccggttttccatg
<i>ybaE</i>	34367 > 35197	276	tAAaGAtGTagagcATCgctatcagatg
<i>ybaF</i>	35194 > 35463	89	cAgGGAGagaagactttatg
<i>ybaG</i>	35729 > 35992	87	gAtGGAGGagacagataagatcatg
<i>ybaH</i>	36002 > 36745	247	gcttaGAGGTgctgtatg

Table 1 (cont.)

ORF	Endpoints (nt)	Size of product (aa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>rplM</i>	36908 > 37345	145	ttAGGAGGggaattcaatg
<i>rpsI</i>	37366 > 37758	130	AaAAAGGAGGTtaccaattg
<i>ybaI</i>	37733 > 37957	74	AAAGGcGcTcgtcgtg
<i>ybaJ</i>	38222 > 38989	255	AGtAAGGAGaagattcatg
<i>ybaK</i>	39175 > 39618	147	AAAGGAGagtgaatttatg
<i>cwlD</i>	39678 > 40391	237	AagcgGGAGGggaagggatg
<i>ybaL</i>	40487 > 41545	352	AagggtGAGGTtcgatg
<i>gerD</i>	41581 < 42138	185	AAAGGAGcTtaagcatg
<i>ybaM</i>	42248 > 42844	198	GAggGGAtGaataaagaaatg
<i>ybaN</i>	42845 < 43609	254	tAAGGAGGcatcatcagtg
<i>ybaO</i>	43612 > 43773	53	AtaGAaGTgattcactgatg
<i>rrnI</i>			
16S	43959 > 45510		
23S	45676 > 48601		
5S	48657 > 48773		
<i>tRNA</i>			
Asn	48820 > 48891		
Thr	48896 > 48968		
Gly	49025 > 49099		
Arg	49130 > 49206		
Pro	49234 > 49310		
Ala	49319 > 49394		
<i>rrnH</i>			
16S	49577 >		

the ORFs in the *spo0H-rrnH* region was the same as that of the movement of the replication fork. Such a distribution is often observed in chromosomal regions relatively close to the origin of replication (Zeigler & Dean, 1990; Glaser *et al.*, 1993; Yoshida *et al.*, 1994; Ogawa *et al.*, 1995).

Similarity analysis of the putative gene products to known proteins

The results of a homology search for each of the new, complete ORFs are shown in Table 2.

Concluding remarks

Among the 63 ORFs in the *spo0H-rrnH* region, 51 (81 %) have now been assigned functions by experimental analysis or by similarity to known gene products. This high ratio of known ORFs can be ascribed to the many

ribosomal protein genes in the region. Our results coincide well with the most recent genetic map (Anagnostopoulos *et al.*, 1993) of this region, with two exceptions. The first is the *mecB* gene which has been mapped at 8°, upstream of *sigH* (Msadek *et al.*, 1994). The other is the *gerD* gene (Yon *et al.*, 1989) which was mapped at 16°, downstream of *rrnH* gene, but is, in fact, upstream of *rrnI*. In addition, there should be a phage SPO2 attachment site between *rplQ* and *rrnI*. Since no sequence information for this site is available, we do not know if it really lies within the region sequenced.

ACKNOWLEDGEMENTS

The authors thank Dr F. Morohoshi (National Cancer Center, Tokyo) for providing the λ phage K4F811 and Ms N. Foster for critical reading of the manuscript. This research was supported by a Grant-in-Aid for Creative Research (Human Genome Project) from the Ministry of Education, Science and Culture of Japan.

Table 2. Similarity of the putative gene products of newly sequenced genes to other proteins

ORF	Similar protein in database	Identity observed	FASTA score	Reference*
<i>rplA</i>	50S Ribosomal protein L1 <i>B. stearothermophilus</i>	80.4% in 230 residues	941	(1)
	<i>Streptomyces griseus</i>	56.9% in 225 residues	679	(2)
<i>rplJ</i>	50S Ribosomal protein L10 <i>Streptomyces griseus</i>	42.0% in 138 residues	310	(2)
	<i>Salmonella typhimurium</i>	35.2% in 159 residues	274	(3)
<i>rplL</i>	50S Ribosomal protein L7/L12 <i>B. stearothermophilus</i>	79.5% in 122 residues	422	(4)
	<i>Micrococcus luteus</i>	78.5% in 121 residues	403	(5)
<i>rpoC</i>	RNA polymerase β' subunit <i>E. coli</i>	47.3% in 1219 residues	2836	(6)
	<i>Mycobacterium leprae</i>	54.2% in 1060 residues	2695	(7)
<i>ybaB†</i>	30S Ribosomal protein HS6 <i>Haloarcula marismortui</i>	34.6% in the middle region 78 residues	115	(9)
	50S Ribosomal protein L30ER <i>Sulfolobus acidocaldarius</i>	27.3% in the middle region 66 residues	101	(10)
<i>rpsL</i>	60S Ribosomal protein L30 <i>Leishmania major</i>	27.3% in the middle region 66 residues	69	(11)
	30S Ribosomal protein S12 <i>B. stearothermophilus</i>	88.2% in 136 residues	580	(12)
<i>rpsG</i>	<i>Streptococcus pneumoniae</i>	85.4% in 137 residues	545	(13)
	30S Ribosomal protein S7 <i>B. stearothermophilus</i>	87.7% in 155 residues	702	(12)
<i>fus</i>	<i>Mycobacterium smegmatis</i>	61.9% in 155 residues	537	(15)
	Elongation factor G (EF-G) <i>Thermus aquaticus</i>	66.7% in 685 residues	2319	(16)
<i>tufA</i>	<i>Anacystis nidulans</i>	63.2% in 695 residues	2257	(17)
	Elongation factor Tu (EF-Tu) <i>B. subtilis</i>	100% in 396 residues	–	(18)
<i>rpsJ</i>	30S Ribosomal protein S10 <i>Mycobacterium leprae</i>	72.3% in 101 residues	384	(12)
	<i>Mycoplasma capricolum</i>	71.6% in 102 residues	376	(19)
<i>rplC</i>	50S Ribosomal protein L3 <i>B. stearothermophilus</i>	79.9% in 209 residues	849	(20)
	<i>Cyanophora paradoxa</i>	46.2% in 208 residues	506	(21)
<i>rplD</i>	50S Ribosomal protein L4 <i>B. stearothermophilus</i>	78.3% in 207 residues	814	(22)
	<i>Yersinia pseudotuberculosis</i>	42.1% in 195 residues	431	(23)
<i>rplW</i>	50S Ribosomal protein L23 <i>B. stearothermophilus</i>	75.5% in 94 residues	368	(12)
	<i>Thermotoga maritima</i>	50.0% in 82 residues	214	(1)
<i>rplB</i>	50S Ribosomal protein L2 <i>B. stearothermophilus</i>	85.5% in 275 residues	1199	(24)
	<i>Cyanophora paradoxa</i>	61.1% in 275 residues	877	(25)
<i>rpsS</i>	30S Ribosomal protein S19 <i>B. stearothermophilus</i>	82.4% in 91 residues	409	(22)
	<i>Cyanophora paradoxa</i>	67.4% in 92 residues	355	(26)
<i>rplV</i>	50S Ribosomal protein L22 <i>B. stearothermophilus</i>	88.5% in 113 residues	450	(12)
	<i>Acholeplasma laidlawii</i>	57.3% in 111 residues	309	(25)
<i>rpsC</i>	30S Ribosomal protein S3 <i>B. stearothermophilus</i>	85.3% in 217 residues	928	(26)
	<i>E. coli</i>	56.4% in 218 residues	691	(27)

Table 2 (cont.)

ORF	Similar protein in database	Identity observed	FASTA score	Reference*
<i>rplP</i>	50S Ribosomal protein L16			(28)
	<i>Thermotoga maritima</i>	68.1 % in 141 residues	549	(24)
	<i>Mycoplasma capricolum</i>	66.9 % in 136 residues	523	(20)
<i>rpsH</i>	30S Ribosomal protein S8			(12)
	<i>Micrococcus luteus</i>	57.9 % in 133 residues	383	(29)
	<i>Marchantia polymorpha</i>	48.1 % in 129 residues	352	(30)
<i>rplF</i>	50S Ribosomal protein L6			(12)
	<i>B. stearothermophilus</i>	80.3 % in 178 residues	693	(31)
	<i>Cyanophora paradoxa</i>	53.1 % in 179 residues	501	(32)
<i>rplR</i>	50S Ribosomal protein L18			
	<i>B. stearothermophilus</i>	73.3 % in 120 residues	427	(31)
	<i>Mycoplasma capricolum</i>	59.8 % in 117 residues	329	(20)
<i>ybaD</i>	Cobalt transport ATP-binding protein CBIO			
	<i>Salmonella typhimurium</i>	33.1 % in 251 residues	367	(33)
	Probable ABC transporter in <i>ntrA/rpoN</i> 5' region			
<i>ybaE</i>	<i>Azorhizobium caulinodans</i>	28.4 % in 229 residues	289	(34)
	<i>Thiobacillus ferrooxidans</i>	28.1 % in 235 residues	285	(35)
	Sulfate transport ATP-binding protein CysA			
<i>ybaH</i>	<i>Synechococcus</i> sp.	38.1 % in 223 residues	364	(36)
	<i>E. coli</i>	36.7 % in 226 residues	326	(37)
	Dipeptide transport ATP-binding protein DPPF			
<i>ybaI</i>	<i>E. coli</i>	31.9 % in 216 residues	299	(38)
	Pseudouridylyl synthase I			
	<i>E. coli</i>	37.6 % in 242 residues	434	(39)
<i>rplM</i>	50S Ribosomal protein L13			
	<i>Staphylococcus carnosus</i>	69.0 % in 145 residues	484	(40)
	<i>E. coli</i>	58.4 % in 137 residues	424	(41)
<i>rpsI</i>	30S Ribosomal protein S9			
	<i>B. stearothermophilus</i>	83.7 % in 129 residues	545	(42)
	<i>E. coli</i>	55.0 % in 129 residues	358	(41)
<i>ybaJ</i>	3-Demethylubiquinone-9 3-methyltransferase			
	<i>Salmonella typhimurium</i>	20.4 % in the middle region 142 residues	100	(43)
	<i>E. coli</i>	20.4 % in the middle region 142 residues	98	(44)
<i>cwlD</i>	Cell wall hydrolase CwlD			
	<i>B. subtilis</i>	100 % in 237 residues	—	(45)
<i>ybaL</i>	MetG-related protein			
	<i>E. coli</i>	32.5 % in 342 residues	528	(46)
<i>ybaN</i>	Nodulation protein B			
	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i>	23.9 % in 218 residues	234	(47)
	<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i>	27.8 % in the middle region 162 residues	209	(48)

*References: 1, Kimura *et al.* (1985); 2, Kuberski *et al.* (1994); 3, Paton *et al.* (1990); 4, Garland *et al.* (1987); 5, Itoh (1981); 6, Boor *et al.* (1995); 7, Ovchinnikov *et al.* (1982); 8, Honoré *et al.* (1993); 9, Kimura *et al.* (1987); 10, Pühler *et al.* (1989); 11, D. C. Robyr, unpublished (SwissProt accession no. P39095); 12, Higo *et al.* (1980); 13, Kimura (1991); 14, Salles *et al.* (1992); 15, Kenny & Churchward (1994); 16, Yakhnin *et al.* (1989); 17, Meng *et al.* (1989); 18, Ludwig *et al.* (1990); 19, Honoré *et al.* (1993); 20, Ohkubo *et al.* (1987); 21, Herwig *et al.* (1992); 22, Evrard *et al.* (1990); 23, Gross *et al.* (1989); 24, Sanangelantoni *et al.* (1994); 25, Krömer *et al.* (1990); 26, Lim & Sears (1992); 27, Zurawski & Zurawski (1985); 28, Henkin *et al.* (1989); 29, Ohama *et al.* (1989); 30, Fukuzawa *et al.* (1988); 31, Ramakrishnan & Gerchman (1991); 32, Michalowski *et al.* (1990); 33, Roth *et al.* (1993); 34, Stigter *et al.* (1993); 35, Berger *et al.* (1990); 36, Green *et al.* (1989); 37, Sirko *et al.* (1990); 38, Sofia *et al.* (1994); 39, Kammen *et al.* (1988); 40, Overhoff-Freundlieb & Freudl (1991); 41, Isono *et al.* (1985); 42, Kimura & Chow (1984); 43, Jordan *et al.* (1994); 44, Wu *et al.* (1992); 45, Sekiguchi *et al.* (1995); 46, Dardel *et al.* (1990); 47, Rossen *et al.* (1984); 48, Schofield & Watson (1986).

†YbaB also showed 23.9 % similarity with the N-terminal region of the endonuclease of *Mycobacterium leprae*. Since the flanking regions are highly conserved between the two organisms, the evolution of the nuclease gene in *M. leprae* is notable.

REFERENCES

- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Berger, D. K., Woods, D. R. & Rawlings, D. E. (1990). Complementation of *Escherichia coli* σ^{54} (NtrA)-dependent formate hydrogen lyase activity by a cloned *Thiobacillus ferrooxidans* ntrA gene. *J Bacteriol* **172**, 4399–4406.
- Boor, K. J., Duncan, M. L. & Price, C. W. (1995). Genetic and transcriptional organization of the region encoding the β subunit of *Bacillus subtilis* RNA polymerase. *J Biol Chem* **270**, 20329–20336.
- Dardel, F., Panvert, M. & Farat, G. (1990). Transcription and regulation expression of the *Escherichia coli* methionil-tRNA synthetase gene. *Mol Gen Genet* **223**, 121–133.
- Evrard, J. L., Kuntz, M. & Weil, J. H. (1990). The nucleotide sequence of five ribosomal protein genes from the cyanelles of *Cyanophora paradoxa*: implications concerning the phylogenetic relationship between cyanelles and chloroplasts. *J Mol Evol* **30**, 16–25.
- Fukuzawa, H., Kohchi, T., Sano, T., Shirai, H., Umesono, K., Inokuchi, H., Ozeki, H. & Ohyama, K. (1988). Structure and organization of *Marchantia polymorpha* chloroplast genome III. Gene organization of the large single copy region from *rbcL* to *trnI* (CAU). *J Mol Biol* **203**, 333–351.
- Garland, W. G., Louie, K. A., Mathson, A. T. & Liljas, A. (1987). The complete amino acid sequence of the ribosomal 'A' protein (L12) from *Bacillus stearothermophilus*. *FEBS Lett* **220**, 43–46.
- 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., Schweitzer, J., Vertes, 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.
- Green, L. S., Laudénbach, D. E. & Grossman, A. R. (1989). A region of a cyanobacterial genome required for sulfate transport. *Proc Natl Acad Sci USA* **86**, 1949–1953.
- Gross, U., Chen, J.-H., Kono, D. H., Lobo, J. G. & Yu, D. T. Y. (1989). High degree of conservation between ribosomal proteins of *Yersinia pseudotuberculosis* and *Escherichia coli*. *Nucleic Acid Res* **17**, 3601–3602.
- Henkin, T. K., Moon, S. H., Mattheakis, L. C. & Nomura, M. (1989). Cloning and analysis of the *spc* ribosomal protein operon of *Bacillus subtilis*: comparison with the *spc* operon of *Escherichia coli*. *Nucleic Acid Res* **17**, 7469–7486.
- Herwig, S., Kruff, V. & Wittmann-Liebold, B. (1992). Primary structures of ribosomal proteins L3 and L4 from *Bacillus stearothermophilus*. *Eur J Biochem* **207**, 877–885.
- Higo, K., Itoh, T., Kumazaki, T. & Osawa, S. (1980). Ribosomal proteins from mesophilic and thermophilic bacteria. In *Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes*, pp. 655–666. Edited by S. Osawa, H. Ozeki, H. Uchida & T. Yura. University of Tokyo: University of Tokyo Press.
- Honoré, N., Berg, S., Chanteau, S., Doucet-Populaire, F., Eiglmeier, K., Garnier, T., Georges, C., Launois, P., Limpaboon, T., Newton, S., Niang, K., del Portillo, P., Ramesh, G. R., Reddi, P., Ridel, P. R., Sittisombut, N., Wu-Hunter, S. & Cole, S. T. (1993). Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rlf-Str region. *Mol Microbiol* **7**, 207–214.
- Isono, S., Thamm, S., Kitakawa, M. & Isono, K. (1985). Cloning and nucleotide sequencing of the genes for ribosomal proteins S9 (*rpsI*) and L13 (*rplM*) of *Escherichia coli*. *Mol Gen Genet* **198**, 279–282.
- Itoh, T. (1981). Primary structure of an acidic ribosomal protein from *Micrococcus lysodeiketicus*. *FEBS Lett* **127**, 67–70.
- Jordan, A., Gibert, I. & Barbé, J. (1994). Cloning and sequencing of the genes from *S. typhimurium* encoding a new bacterial ribonucleotide reductase. *J Bacteriol* **176**, 3420–3427.
- Kammen, H. O., Marvel, C. C., Hardy, H. & Penhoet, E. E. (1988). Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. *J Biol Chem* **263**, 2255–2263.
- Kenney, T. J. & Churchward, G. (1994). Cloning and sequence analysis of the *rpsL* and *rpsG* genes of *Mycobacterium smegmatis* and characterization of mutations causing resistance to streptomycin. *J Bacteriol* **176**, 6153–6156.
- Kimura, J., Arndt, E. & Kimura, M. (1987). Primary structures of three highly acidic ribosomal proteins S6, S12 and S15 from the archaeobacterium *Halobacterium marismortui*. *FEBS Lett* **224**, 65–70.
- Kimura, M. (1991). The nucleotide sequences of *Bacillus stearothermophilus* ribosomal protein S12 and S7 genes: comparison with the *str* operon of *Escherichia coli*. *Agric Biol Chem* **55**, 207–213.
- Kimura, M. & Chow, C. K. (1984). The complete amino acid sequences of ribosomal proteins L17, L27 and S9 from *Bacillus stearothermophilus*. *Eur J Biochem* **139**, 225–234.
- Kimura, M., Kimura, J. & Ashman, K. (1985). The complete primary structure of ribosomal proteins L1, L14, L15, L23, L24 and L29 from *Bacillus stearothermophilus*. *Eur J Biochem* **150**, 491–497.
- Krömer, W. J., Hatakeyama, T. & Kimura, M. (1990). Nucleotide sequences of *Bacillus stearothermophilus* ribosomal protein genes: part of the ribosomal S10 operon. *Biol Chem Hoppe-Seyler* **371**, 631–636.
- Kuberski, S., Kasberg, T. & Distler, J. (1994). The *nusG* gene of *Streptomyces griseus*: cloning of the gene and analysis of A-factor binding properties of the gene product. *FEMS Microbiol Lett* **119**, 33–39.
- Lim, P. O. & Sears, B. B. (1992). Evolutionary relationships of a plant-pathogenic mycoplasma-like organism and *Acholeplasma laidlawii* deduced from two ribosomal protein gene sequences. *J Bacteriol* **174**, 2606–2611.
- Ludwig, W., Weizenegger, M., Betzl, D., Leidel, E., Lenz, T., Ludvigsen, A., Möllenhoff, D., Wenzig, P. & Schleifer, K. H. (1990). Complete nucleotide sequences of seven eubacterial genes coding for the elongation factor Tu: functional, structural and phylogenetic evaluations. *Arch Microbiol* **153**, 241–247.
- Meng, B. Y., Shinozaki, K. & Sugiura, M. (1989). Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-Tu of the cyanobacterium *Anacystis nidulans*: structural homology between 16S rRNA and S7 mRNA. *Mol Gen Genet* **216**, 25–30.
- Michalowski, C. B., Pfanzagl, B., Loeffelhardt, W. & Bohnert, H. J. (1990). The cyanelle S10 *spc* ribosomal protein gene operon from *Cyanophora paradoxa*. *Mol Gen Genet* **224**, 222–231.
- Msadek, T., Kunst, F. & Rapoport, G. (1994). MecB of *Bacillus subtilis*, a member of the ClpC ATPase, is a pleiotropic regulator controlling competence gene expression and growth at high temperature. *Proc Natl Acad Sci USA* **91**, 5788–5792.
- Ogawa, K., Akagawa, E., Nakamura, K. & Yamane, K. (1995). Determination of a 21 548 bp nucleotide sequence around the 24° region of the *Bacillus subtilis* chromosome. *Microbiology* **141**, 269–275.
- Ohama, T., Muto, A. & Osawa, S. (1989). Spectinomycin operon of *Micrococcus luteus*: evolutionary implications of organization and novel codon usage. *J Mol Evol* **29**, 381–395.

- Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F. & Ogawa, S. (1987). The ribosomal protein gene cluster of *Mycoplasma capricolum*. *Mol Gen Genet* **210**, 314–322.
- Ovchinnikov, Yu. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Salomatina, I. S., Shuvaeva, T. M., Lipkin, V. M. & Sverdlov, E. D. (1982). The primary structure of *E. coli* RNA polymerase. Nucleotide sequence of the *rpoC* gene and amino acid sequence of the β' -subunit. *Nucleic Acids Res* **10**, 4035–4044.
- Overhoff-Freundlieb, B. & Freudl, R. (1991). Suppression of an *Escherichia coli* *secA* ts mutant by a gene cloned from *Staphylococcus carnosus*. *FEMS Microbiol Lett* **84**, 143–150.
- Paton, E. B., Zolotukhin, S. B., Woodmaska, M. I., Kroupskaya, I. V. & Zhyvoloup, A. Z. (1990). The nucleotide sequence of gene *rplJ* encoding ribosomal protein L10 of *Salmonella typhimurium*. *Nucleic Acids Res* **18**, 2824.
- Pühler, G., Lottspeich, F. & Zillig, W. (1989). Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaeobacterium *Sulfolobus acidocaldarius*. *Nucleic Acids Res* **17**, 4517–4534.
- Ramakrishnan, V. & Gerchman, S. E. (1991). Cloning, sequencing and overexpression of genes for ribosomal proteins from *Bacillus stearothermophilus*. *J Biol Chem* **266**, 880–885.
- Rossen, L., Johnston, A. W. B. & Downie, J. A. (1984). DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAB* and *C* required for root hair curling. *Nucleic Acids Res* **12**, 9497–9508.
- Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S. & Church, G. M. (1993). Characterization of the cobalamin (vitamin B12) biosynthetic genes of *Salmonella typhimurium*. *J Bacteriol* **175**, 3303–3316.
- Salles, C., Créancier, L., Claverys, J.-P. & Méjean, V. (1992). The high level streptomycin resistance gene from *Streptococcus pneumoniae* is a homologue of the ribosomal protein S12 gene from *Escherichia coli*. *Nucleic Acids Res* **20**, 6103.
- Sanangelantoni, A. M., Bocchetta, M., Cammarano, P. & Tiboni, O. (1994). Phylogenetic depth of S10 and *spc* operons: cloning and sequencing of a ribosomal protein gene cluster from the extremely thermophilic bacterium *Thermotoga maritima*. *J Bacteriol* **176**, 7703–7710.
- Schofield, P. R. & Watson, J. M. (1986). DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. *Nucleic Acids Res* **14**, 2891–2903.
- Sekiguchi, J., Akeo, K., Yamamoto, H., Khasanov, F. K., Alonso, J. C. & Kuroda, A. (1995). Nucleotide sequence and regulation of a new putative cell wall hydrolase gene, *cwlD*, which affects germination in *Bacillus subtilis*. *J Bacteriol* **177**, 5582–5589.
- Sirko, A., Hryniewicz, M., Hulanicka, D. & Boeck, A. (1990). Sulfate and thiosulfate transport in *Escherichia coli* K-12 nucleotide sequence and expression of the *cysTWM* gene cluster. *J Bacteriol* **172**, 3351–3357.
- Sofia, H. J., Burland, V., Daniels, D. L., Plunkett, G., III & Blattner, F. R. (1994). Analysis of the *Escherichia coli* genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. *Nucleic Acids Res* **22**, 2576–2586.
- Stigter, J., Schneider, M. & de Bruijn, F. J. (1993). *Azorhizobium caulinodans* nitrogen fixation (*nif/fix*) gene regulation: mutagenesis of the *nifA* –24/–12 promoter element, characterization of a *ntrA* (*rpoN*) gene, and derivation of a model. *Mol Plant-Microbe Interact* **6**, 238–252.
- Wawrousek, E. F. & Hansen, J. N. (1989). Structure and organization of a cluster of six tRNA genes in the space between tandem ribosomal RNA gene sets in *Bacillus subtilis*. *J Biol Chem* **264**, 291–298.
- Wu, G., Williams, H. D., Zamanian, M., Gibson, F. & Poole, R. K. (1992). Isolation and characterization of *Escherichia coli* mutants affected in aerobic respiration: the cloning and nucleotide sequence of *ubiG*. Identification of an S-adenosylmethionine-binding motif in protein, RNA, and small-molecule methyltransferases. *J Gen Microbiol* **138**, 2101–2112.
- Yakhnin, A. V., Vorozheykina, D. P. & Matvienko, N. I. (1989). Nucleotide sequence of the *Thermus thermophilus* HB8 gene coding for elongation factor G. *Nucleic Acids Res* **17**, 8863.
- Yon, J. R., Sammons, R. L. & Smith, D. A. (1989). Cloning and sequencing of the *gerD* gene of *Bacillus subtilis*. *J Gen Microbiol* **135**, 3431–3445.
- 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 the genes in the *Bacillus subtilis* chromosome. *Genetics* **125**, 703–708.
- Zurawski, G. & Zurawski, S. M. (1985). Structure of the *Escherichia coli* S10 ribosomal protein operon. *Nucleic Acids Res* **13**, 4521–4526.

Received 18 April 1996; revised 22 July 1996; accepted 31 July 1996.