

한국에서의 고초균 유전체 연구: *Bacillus subtilis* 염색체상 180°-185° 부위 53 kb DNA 단편의 염기서열 분석

김사열 · 최수근 · 정영미 · 신병식 · 박승환*

한국과학기술연구원 생명공학연구소 미생물분자유전학 RU

The *Bacillus subtilis* Genome Sequencing Project in Korea: Sequence Analysis of the 53 kb DNA Fragment at 180°-185° of *B. subtilis* 168 Chromosome. Ghim, Sa-Youl, Soo-Keun Choi, Young-Mee Jeong, Byung-Sik Shin, and Seung-Hwan Park*. *Bacterial Molecular Genetics RU, Korea Research Institute of Bioscience and Biotechnology (KRIBB), P.O. Box 115, Yusong, Taejeon 305-600, Korea* - The entire sequence of a 4,214,810 bp genome of the *Bacillus subtilis* 168 has been determined by an international project, and the completion has been announced on July 19, 1997. For the sequencing project an international consortium was established and 25 European, 7 Japanese laboratories, 2 biotechnology companies, and our laboratory participated in the project. Within this framework we determined the complete nucleotide sequence of a 53,289 bp fragment upstream of the *odhA* gene (181°) of the *B. subtilis* 168 chromosome. On the basis of the published DNA sequences of the *B. subtilis* *sspC* and *odhA* genes, we obtained genomic fragments by plasmid rescue and long-range PCR. The sequenced fragment contains 56 putative open reading frames (designated *yojA-yoJl*) and 9 known genes (*sspC*, *cge* cluster, *orfE5*, *orfRM1* and *odhA*), in which we found many interesting features. In addition, the entire nucleotide sequence of a 53,289 bp region enabled us to revise the current genetic map of this region.

Key words: *Bacillus subtilis*, genome sequencing, *yojP-odhA* region, genetic map

상당한 기간 동안 과학계에서 논란을 빚어왔던 생명체의 유전체 연구에 대한 회의론적 시각은 1995년 J. Craig Venter와 그 동료들에 의하여 *Haemophilus influenzae*[23]의 전체 염기서열 해독이 완료되면서부터 긍정적인 시각으로 변하였다. 그 후 불과 수년 사이에 다음과 같은 다양한 생명체들에 대한 유전체 염기서열 해독이 완료된 사실로부터 현대 생물학에서 유전체 연구의 중요성이 얼마나 큰 것인지 짐작할 수 있게 되었다: *Mycoplasma genitalium*[25], *Synechocystis* sp.[40, 41], *Mycoplasma pneumoniae*[36], *Saccharomyces cerevisiae*[29], *Methanococcus jannaschii*[13], *Helicobacter pylori*[63] 및 *Escherichia coli*[10]. 생태학적으로, 산업적으로 혹은 의학적으로 중요한 생명체들에 대한 전체 염기서열의 해독은 단순히 흥미롭고 유용한 많은 유전자원의 확보 뿐 아니라 생명정보학적 접근을 통하여 생명현상에 대한 통합적 이해의 새로운 세계를 열어가고 있다[19].

*Bacillus subtilis*균(고초균)이 유전체 연구의 대상이 된 데는 몇가지 이유가 있다. 유해하지 않은 세균으로서 그동안 많은 기초 연구 결과가 축적되어 왔고, 산업적으로 중요한 위치를 차지하고 있으며, 포자형성과 같은 세포 분화과정을 가진다는 특징 외에 유전체의 크기가 비교적

작고 G+C 함량이 낮은 점 등이 그것이다. 고초균의 유전체 연구는 1989년 유럽의 5개 연구 팀에 의해 시작되어 35개 그룹이 참여하는 국제협력과제로 발전하였으며 [48], 이들의 유기적 협력에 의하여 1997년 5월에 전체 염기서열이 밝혀지게 되었다. 대부분의 다른 유전체 연구가 단순히 염기서열을 결정하는 차원에 머물고 있는데 반하여, 고초균의 염기서열 결정에 참여했던 많은 연구 팀들은 국제적 협력과 조정의 경험을 살려 유전체의 기능 연구를 계속 수행하고 있다.

우리 연구팀은 국제협력과제의 일원으로 1995년 5월부터 참여하였고, *B. subtilis* 168의 유전자 지도[3, 9] 상에서 *attSPβ*와 *odhA* 사이의 부위(181° 근처)를 할당받아 총 53 kb의 염기서열을 결정하였다. 이 염기서열을 분석하여 65개의 open reading frame(ORF)을 발견하였는데, 그 중 *sspC*[20], *cge* cluster[52], *orfE5*[24], *orfRM1*[45] 및 *odhA* [51] 등 9개 유전자는 이미 연구되어 보고된 것이었으나 나머지 56개 ORF는 새로운 유전자인 것으로 분석되었다.

본 논문에서는 53 kb DNA 단편의 염기서열 및 그 일차적인 분석을 통하여 얻은 결과를 보고하며, 아울러 해당 지역의 새로운 유전자 지도를 제시하고자 한다.

재료 및 방법

사용 균주, 플라스미드 및 배지

*Corresponding author
Tel. 82-42-860-4412, Fax. 82-42-860-4594
E-mail: shpark@kribb4680.kribb.re.kr

본 연구에서 사용한 모든 세균과 플라스미드 벡터는 Table 1에 정리하였다. 세균 배양에는 주로 LB 배지[8] 또는 Spizizen salt 최소 배지[61]에 0.5% glucose와 0.1% glutamate를 첨가한 배지를 사용하였다. 필요시 ampicillin(100 µg/ml)과 chloramphenicol(대장균, 20 µg/ml; 고초균, 5 µg/ml)을 첨가하였다. β-galactosidase chromogenic 기질인 X-gal(5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)은 대장균의 경우 40 µg/ml, 고초균의 경우 80 µg/ml을 각각 사용하였다. 세균은 37°C에서 배양하였다.

크로닝 및 DNA 조작

고초균의 염색체 DNA 단편은 PCR 증폭 혹은 plasmid rescue 방법[28]에 의하여 확보하였다. Long range PCR[16]은 염색체 DNA를 주형으로 하여 oligonucleotide primer(Table 2)와 GeneAmp® XL PCR Kit(Perkin-Elmer)를 써서 수행하였다. 2 kb 미만의 비교적 짧은 유전체의 단편을 증폭하기 위한 PCR은 PreMix™-Top Kit(Bioneer)를 사용하였다. Long range PCR에 의하여 얻은 염색체 DNA 단편은 적절한 primer를 사용하여 직접 그 염기서열을 결정하였다. 벡터 플라스미드에 크로닝된 유전체 DNA의 단편은 Erase-a-Base® System(Promega)을 이용하여 순차적으로 크기를 줄여 대장균을 숙주로 형질전환을 하였다. 그 결과로 얻은 0.5-5 kb 크기의 DNA 단편은 세균 배양액을 직접 전기영동하는 방법[17]에 의하여 검색하였다.

염색체 DNA는 Sambrook 등[54]이 서술한 바대로 분

리하였고, 대장균과 고초균으로부터의 플라스미드 DNA 분리는 QIAprep Plasmid Kit(QIAGEN)를 사용하였다. Competence cell 준비와 형질전환은 고초균의 경우 Albano 등의 방법[1]을 따랐고 대장균의 경우 Cohen 등의 방법[18]을 따랐다. 다른 여러 가지 DNA 조작 기술은 Sambrook 등이 서술한 일반적인 실험방법[54]을 따라 실행하였다.

염기서열의 결정

DNA 염기서열의 결정은 dideoxy chain termination 방법[55]을 기본으로 하였으며 다음의 두 가지 방법을 병행하여 그 결과를 보완하였다: Taq DyeDeoxy Termination Cycle Sequencing Kit(Applied Biosystems)와 ABI 373 A automated sequencer를 쓰거나 DIG system(Boehringer Mannheim)과 GATC 1500 DNA Sequencer(GATC-Gesellschaft für Analyse-Technik und Consulting)를 사용하였다. DNA 단편 사이의 떨어진 지역은 해당 부위를 포함하는 subclone을 새로 준비하거나 염색체 DNA를 주형으로 하여 직접 염기서열을 결정함으로써 확인하였다. 할당받은 염색체 DNA의 전 지역에서 양쪽 방향으로 그 염기서열을 결정하였다.

데이터 분석

DNA 염기서열 데이터를 종합하여 contig를 형성하는 과정은 EditSeq와 Seqman 프로그램(DNASTAR Inc.)을 써서 수행하였다. 결정된 DNA 염기서열의 분석은 MacDNASIS Pro(Hitachi)와 MACMOLLY TETRA

Table 1. Bacterial strains and plasmids

Strain or plasmid	Description	Source or reference
<i>B. subtilis</i> 168	<i>trpC2</i>	[4]
<i>E. coli</i> XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac[F' proABlac⁺ZAM15 Tn10 (Tet^r)]</i>	[12]
Plasmids		
pUC18/19	High-copy number <i>E. coli</i> vector, Ap ^r	[65]
pDIA5304	<i>E. coli</i> cloning vector, Ap ^r Cm ^r	P. Glaser, Institut Pasteur
pHPS9	A shuttle vector for <i>E. coli</i> and <i>B. subtilis</i> Cm ^r Er ^r <i>cat-86::lacZ_α</i> T1060-pUC9 derivative	[34]

Table 2. Sequences of the primers used for PCR

Primer	Sequence
ND-F	5'-GAC GCA GCA AAC TGG TCA AAG CAT G-3'
ND-R	5'-CGT AAA GTT CCT GCA GTA AGT-3'
ssp-F	5'-CAT TCC AAG GAA TCC AAT GA-3'
ssp-R	5'-GAA TTC TTT GAG CTG AAA GAA G-3'
odh-F	5'-CGA TCA GTA TAC CCA GGA TCC GTA C-3'
odh-R	5'-GCT GAT TCC AAT GGA TCC TC-3'
LR-F	5'-GGC AAT TTC ATA TGG ACA TTG ATG TAT GGG-3'
LR-R	5'-AAG CAT CGC GCC AAG TCC AAA CAC TAA AGC-3'

(Soft gene GmbH) 소프트웨어 패키지를 사용하여 실시하였다. RNA 2차 구조는 FOLD 프로그램[66]으로 분석하였다. 발견된 ORF들의 상동성 조사를 단백질 데이터베이스에서 실행하는 작업은 BLAST 연산법[2]을 사용하였다.

기존의 유전자 지도[3]를 바탕으로 본 연구팀에게 할당된 부위인 *odhA*와 *attSPβ* 사이의 염색체 DNA 단편의 확보는 *sspC*[20]와 *odhA*[51] 유전자의 알려진 염기서열을 이용하여 실시하였다. 먼저 *sspC*(182°) 유전자의 양쪽 방향으로 인접한 DNA 단편을 얻기 위하여 두 primer, *ssp-F*와 *ssp-R*(Fig. 1, Table 2)을 사용하여 PCR을 실시하였고 그 결과 700 bp DNA 산물을 얻었다. 이를 pDIA5304 벡터에 크로닝하여 pSQ5를 제작한 후 homo-

결과 및 고찰

할당된 고초균 유전체 부위의 확보

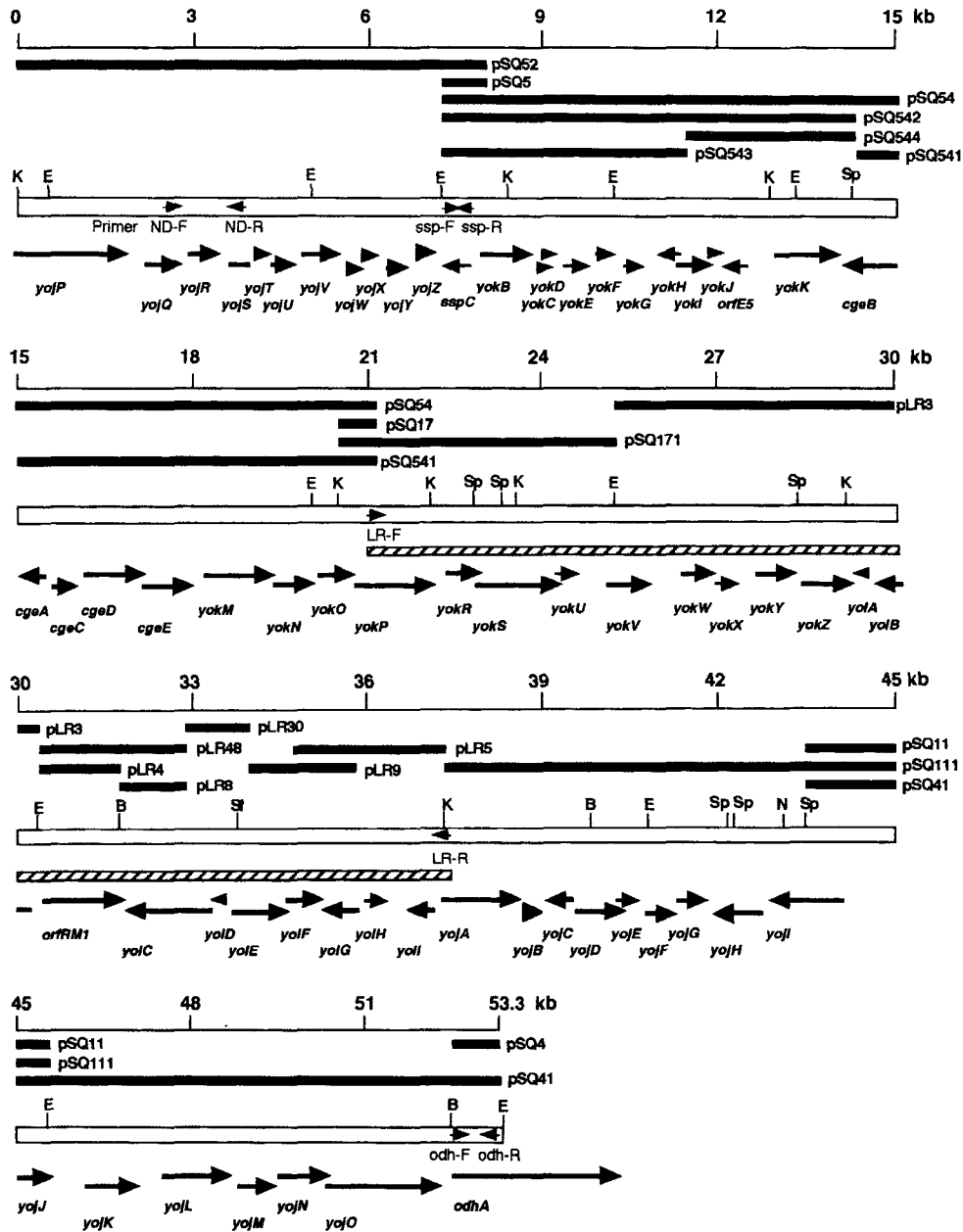


Fig. 1. Physical map and gene organization of the *B. subtilis* genomic region upstream of the *odhA* gene 181° [51]. The vector used was pDIA5304 and the location of the cloned DNA in pDIA5304 is indicated with filled bars. Restriction sites are abbreviated as K, *KpnI*; E, *EcoRI*; Sp, *SphI*; B, *BamHI*; Sf, *SfiI*; and N, *NoI*. Primers for PCR are indicated by arrows and letters in and right below the solid bar, respectively. A DNA fragment produced by long range PCR is indicated by a hatched bar. The letters below each thick arrows and a solid bar represent the name of the ORF and the arrowheads show the direction of their transcription and translation.

logous recombination을 이용하여 고초균의 염색체 내로 삽입하였다. 형질전환된 고초균(Cm^r)으로부터 분리한 염색체 DNA를 제한효소 *KpnI* 및 *EagI*으로 절단하여 각각 자체 ligation시킨 후 대장균을 형질전환시켜 ampicillin 저항성을 갖는 형질전환주를 선별하였다. 형질전환주로부터 분리한 플라스미드는 *KpnI*의 경우 시초의 700 bp DNA 절편을 포함하는 7.8 kb의 insert를 함유하고 있었으며 *EagI*의 경우 13.8 kb의 insert를 함유하고 있었다. 이들을 각각 pSQ52와 pSQ54(Fig. 1)로 명명하였다. 그 다음 *odh-F*와 *odh-R* primer를 사용한 PCR 생성물로부터 출발하여 같은 방법으로 pSQ41 플라스미드를 확보하였다. pSQ54 및 pSQ41의 말단 부위를 sub-cloning하여 pSQ17 및 pSQ11을 제작한 후 앞에서와 같은 plasmid rescue 방법[28]을 사용하여 pSQ171과 pSQ111을 얻었다. pSQ171과 pSQ111 사이의 부위(Fig. 1, hatched bar)는 LR-F와 LR-R의 primer를 사용하여 long range PCR[16]에 의하여 확보하였다.

염기서열의 결정과 분석

고초균의 전체 염색체중 본 연구팀에 할당된 부위는 염기서열을 해독한 결과 53,289 bp의 염기서열로 이루어져 있음이 밝혀졌으며 그 physical organization은 Fig. 1에 표시하였다. 이 부위에 존재하는 ORF를 찾기 위하여 6가지의 가능한 translation frame을 조사하였고, 예상되는 번역 개시 코돈의 3-9 bp 상류에 위치하는 전형적인 리보솜 결합 부위[46]의 존재가 전제 조건으로 고려되었다. 그 결과 53,289 bp의 염기서열 내에 모두 65개의 ORF가 존재하는 것으로 분석되었다(Table 3). 번역의 개시는 52개의 ORF가 AUG 코돈, 13개의 ORF가 UUG

코돈을 이용하고 있었다. 종결 코돈의 빈도는 UAA (58%)>UGA(25%)>UAG(17%)의 순서이었다. 이 부위의 G+C 함량은 41.0%이었는데, 이는 전체 유전체의 G+C 함량인 43.5%[42]에 비교하면 다소 낮은 값이다. 그 이유는 비록 길지는 않지만 *yojP-orfE5*에 걸쳐있는 12,060 bp 부위가 *SPβ* prophage에 해당하며 이 부위는 매우 낮은 34.3%의 G+C 함량을 가지기 때문이다.

이 지역에서 발견된 65개의 ORF중 이미 밝혀진 9개의 ORF를 제외한 나머지 56개의 ORF는 그 염기서열이 결정되는 대로 *yojA*에서 *yolI*까지 알파벳 순서로 명명하였다(Fig. 1, Table 3). 분석된 데이터와 염기서열은 모두 GenBank에 등록하였다(AF012906, U80600, AF006665, AF026147). 본 연구팀이 제출한 GenBank 데이터의 유전자 명칭이 *SubtiList* ([48]; URL <http://www.pasteur.fr/Bio/SubtiList.html>) 데이터베이스의 명칭과 부분적으로 일치하지 않는데, 이는 명명과정에서 다른 그룹과 사전 조정을 할 수 없었던 불가피한 사정으로 발생한 것이며, 따라서 Table 3에는 두가지 명칭을 모두 수록하였다. 또한 다른 그룹에 의하여 중복 수행된 연구 결과의 일부가 본 연구팀의 결과와 다소 차이를 보였다. 예를 들자면, 서로 다른 번역 개시 코돈을 채택함으로써 N-말단 아미노산에 차이를 보이는 것으로 *yokP*, *yokX* 및 *yokZ* 등이 있으며, *yokB*의 경우 *SubtiList* 데이터베이스에서 2개의 ORF로 표시되었고, *yosU*의 경우 그 방향이 서로 반대로 나타났다. *SubtiList* 데이터베이스에서 *yokU*와 *yolD*는 나타나지 않았고, *yokF-yokG* 부근에서 3개의 매우 작은 ORF가 더 발견되어 표시되었다. 이러한 몇 군데 분석결과의 불일치는 대부분 연관된 ORF가 기존의 단백질들의 데이터와 지극히 낮은 유사성을 보여줌으로써, 그 진위

Table 3. General features of the 65 ORFs in the 53,289 bp sequence at 3' flanking region of *odhA* 181° of the *B. subtilis* 168 chromosome

Gene name*		Endpoints (nucleotides)	Molecular size		Putative ribosome-binding site (upper case) and initiation codon (underlined)
GenBank	<i>SubtiList</i>		(kDa)	(aa)	
<i>yoyP</i>	<i>yosO</i>	28>1986	74.8	652	GtAGaTGcaattcca <u>ATG</u>
<i>yoyQ</i>	<i>yosP</i>	2316>2896	(38.2) [†]	193	GAAAGGAcatgattaa <u>TTG</u>
<i>yoyR</i>	<i>yosQ</i>	3008>3529	20.1	173	AAAGGAGGtgaaa <u>ATG</u>
<i>yoyS</i>	<i>yosR</i>	3705>4113		136	
<i>yoyT</i>	<i>yosS</i>	4110>4352	9.1	80	AGAAAGGAAtgtattca <u>ATG</u>
<i>yoyU</i>	<i>yosT</i>	4398>4826	16.2	142	AAAGGAGctaacata <u>ATG</u>
<i>yoyV</i>		4921>5370	17.0	149	GAAAGGgGattaa <u>ATG</u>
<i>yoyW</i>	<i>yosV</i>	5807>5974	6.6	55	GAAAtatgacgac <u>ATG</u>
<i>yoyX</i>	<i>yosW</i>	5975>6265	10.9	96	GGtGGgGgattaa <u>TTG</u>
<i>yoyY</i>	<i>yosX</i>	6418>6753	12.7	111	GAGGTcAggtgatt <u>ATG</u>
<i>yoyZ</i>	<i>yosZ</i>	6984>7337	14.0	117	AAGGAGGTGAttca <u>TTG</u>
<i>sspC</i>	<i>sspC</i>	7855<7637	7.8	72	AGGAGaTGAataag <u>ATG</u>
<i>yokB</i>		7974>8801	32.2	275	AGAAAGGtGGagacaaa <u>ATG</u>
<i>yokC</i>	<i>yotD</i>	8845>9036	7.5	63	AGAAgGGAaGgggagac <u>ATG</u>
<i>yokD</i>	<i>yotE</i>	9076>9207	5.0	43	AagAAaGAGGTgtagtata <u>ATG</u>

Table 3. Continued

Gene name*		Endpoints (nucleotides)	Molecular size		Putative ribosome-binding site (upper case) and initiation codon (underlined)
GenBank	<i>SubtilList</i>		(kDa)	(aa)	
<i>yokE</i>	<i>yotF</i>	9525>9791	10.5	88	AtgAAGGAaagtctc <u>ATG</u>
<i>yokF</i>	<i>yotH</i>	10012>10233	8.5	73	GGAGGattattca <u>ATG</u>
<i>yokG</i>	<i>yotJ</i>	10439>10651	8.4	70	AgtGGAGGTatt <u>ATG</u>
<i>yokH</i>	<i>yotL</i>	10922<11203	10.8	93	AttGaAGGTaaatt <u>ATG</u>
<i>yokI</i>	<i>yotM</i>	11239>11826	22.8	195	AGGAGagtgtga <u>TTG</u>
<i>yokJ</i>	<i>yotN</i>	11829>12005	7.1	58	AaAaCGGGGGTgtagcc <u>TTG</u>
<i>ortE5</i>	<i>yodU</i>	12040<12465	15.9	141	GGAGGataagcc <u>ATG</u>
<i>yokK</i>	<i>yodV</i>	12835>13983	42.0	382	AaAgAGGAGGgtaccaa <u>ATG</u>
<i>cgeB</i>	<i>cgeB</i>	14054<15007	36.9	317	AtAACatccg <u>ATG</u>
<i>cgeA</i>	<i>cgeA</i>	15014<15415	14.2	133	AtAGGAGGTgtgtga <u>ATG</u>
<i>cgeC</i>	<i>cgeC</i>	15620>15925	11.3	101	AtAAGGGAGgGgatgat <u>TTG</u>
<i>cgeD</i>	<i>cgeD</i>	15990<17270	50.1	426	GcggGGAGGcGgtatag <u>ATG</u>
<i>cgeE</i>	<i>cgeE</i>	17299>18078	29.6	259	AtAAAGaAatTGAata <u>TTG</u>
<i>yokM</i>	<i>yodT</i>	18401>19735	48.6	444	GtttGAGGTGggagc <u>ATG</u>
<i>yokN</i>	<i>yodS</i>	19742>20431	24.4	229	AaAAAGGAttgacattaa <u>ATG</u>
<i>yokO</i>	<i>yodR</i>	20416>21069	23.3	217	AGcgAGGgaGTGAagtggaa <u>ATG</u>
<i>yokP</i>	<i>argE</i>	21084>22388	48.3	434	AaGAGGgGgaatgagg <u>ATG</u>
<i>yokR</i>	<i>yodP</i>	22366>23193	31.2	275	GGAGGatgceg <u>ATG</u>
<i>yokS</i>	<i>yodO</i>	23222>24637	54.2	471	GcgAGGAGGagttogt <u>TTG</u>
<i>yokU</i>		24616>24912	11.5	98	AGAAAagactga <u>ATG</u>
<i>yokV</i>	<i>yodN</i>	25374>26054	26.9	226	GcgAGGgGGaGAcacc <u>ATG</u>
<i>yokW</i>	<i>yodM</i>	26567>27178	22.8	203	AaAAttGAGGTGgttaaa <u>TTG</u>
<i>yokX</i>	<i>yodL</i>	27242>27553	12.0	103	AaAaGAGGgGAtcttgat <u>ATG</u>
<i>yokY</i>	<i>deoD</i>	27919>28620	25.4	233	AcAGGAGGatagag <u>ATG</u>
<i>yokZ</i>	<i>yodJ</i>	28720>29523	30.2	267	AaAAAGtccgtaa <u>ATG</u>
<i>yolA</i>	<i>yodI</i>	29594<29845	9.2	83	GAtAGGAGGGAaacaata <u>TTG</u>
<i>yolB</i>	<i>yodH</i>	29932<30633	26.1	233	AaAAAGGAGGatcgata <u>TTG</u>
<i>orfRM1</i>	<i>ctpA</i>	30786>32186	43.0	466	AaAAAGGTGGTgttaa <u>TTG</u>
<i>yolC</i>	<i>yodF</i>	32220<33710	53.8	496	AagAtGGAGGactgaa <u>ATG</u>
<i>yolD</i>		33714<33929	8.4	71	GTGAgcattatc <u>ATG</u>
<i>yolE</i>	<i>yodE</i>	34093>35004	33.8	303	AaAtAGGAGtgagaaa <u>ATG</u>
<i>yolF</i>	<i>yodD</i>	35020>35622	22.4	200	AAAGGgGGaattata <u>ATG</u>
<i>yolG</i>	<i>yodC</i>	35665<36273	23.3	202	AAGGAGGaaacact <u>ATG</u>
<i>yolH</i>	<i>yodB</i>	36403>36741	12.8	112	GAAgGGAtGAagact <u>ATG</u>
<i>yolI</i>	<i>yodA</i>	37147<37536	14.6	129	AgAGGAGGaGtacac <u>ATG</u>
<i>yojA</i>	<i>yojA</i>	37907>39241	46.8	444	GGAGGgGtaccgtacc <u>ATG</u>
<i>yojB</i>	<i>yojB</i>	39326>39562	8.1	78	AtAAAGGAGGaaAaatc <u>ATG</u>
<i>yojC</i>	<i>yojC</i>	39670<40071	15.7	133	AGAAAGGAtGTGAagaca <u>ATG</u>
<i>yojD</i>	<i>yojD</i>	40170>40961	29.2	263	AGAAAGGcaGgGAtatca <u>ATG</u>
<i>yojE</i>	<i>yojE</i>	41143>41421	10.8	92	AtAAAGGAGcaagacg <u>ATG</u>
<i>yojF</i>	<i>yojF</i>	41418>41768	13.0	116	AGAAAagGgGagcaatc <u>ATG</u>
<i>yojG</i>	<i>yojG</i>	42022>42450	16.7	142	AAAtGGAtaTcAatgactcg <u>ATG</u>
<i>yojH</i>	<i>yojH</i>	42476<43324	32.1	282	AAAaGAGGTGAtcaatc <u>ATG</u>
<i>yljI</i>	<i>yojI</i>	43340<44698	49.3	452	AACGGAGaTcAatgt <u>ATG</u>
<i>yojJ</i>	<i>yojJ</i>	44972>45568	22.7	198	AcAGGAGaTGAaggct <u>ATG</u>
<i>yojK</i>	<i>yojK</i>	46174>47037	31.9	287	GcgGGTaAagtccatgccaac <u>ATG</u>
<i>yojL</i>	<i>yojL</i>	47419>48663	44.2	414	AGGAGGTtgaatga <u>ATG</u>
<i>yojM</i>	<i>yojM</i>	48756>49346	20.9	196	AAGGgGGaGAcagt <u>ATG</u>
<i>yojN</i>	<i>yojN</i>	49410>50324	33.5	304	GAAAGGgaaTctgtct <u>ATG</u>
<i>yojO</i>	<i>yojO</i>	50266>52251	76.2	661	AGcggGGArtGTgagaaaagc <u>TTG</u>
<i>odhA</i>	<i>odhA</i>	52480>53289	N-terminal region of <i>odhA</i>		GGgGGTaAtattcaa <u>ATG</u>

*Our data in GenBank; *SubtilList* of Institut Pasteur.†The molecular mass in parenthesis is for the putative product of combined *yojQ/S* with an excision of *yojR* region.

Putative terminator	Sequence
<i>TerVIII</i>	3186 <u>TTTGGCGACTAACCTATGCTACTAAATAGTTC</u> 3157
<i>TerIX</i>	10681 <u>ATATGCAAAAAGTAATGAACTAAATATTGC</u> 10710
<i>TerV</i>	12629 <u>AAACTAAATAAATAATGCTACTAAATATTCA</u> 12658
<i>TerX</i>	23197 <u>TGACTTACAGTTGCTTGCCACCAAATATTCA</u> 23168
Consensus for DNA terminator	$\begin{matrix} A & A & TAAATTA & T & T & A & \\ N & C & ACTG & AN & AGCTCAG & TG & A & C & C & AAAT & C & TTCA \end{matrix}$

Fig. 2. DNA sequence alignment of 3 potential DNA terminators with the known consensus of *B. subtilis* DNA terminator [33].

The numbering indicates position of nucleotide residue. Underlined indicates identical nucleotide residues to the known consensus sequence.

를 가리기가 쉽지 않았으며, 현재 진행중인 국제협력 프로젝트에 의한 유전자 기능분석 연구를 통하여 이러한 문제점이 해소되리라 전망한다.

우리가 염기서열을 밝힌 부위가 염색체 상에서 180° 부근이어서 DNA 복제 종결 단백질이 결합하는 부위[24, 33]가 몇 군데서 발견되었다. 그 동안 모두 7개의 복제 종결자 sequence(*TerI-TerVII*)가 알려져 왔고, 그 중 *TerV*가 이 부위의 *orfE5*에 인접한 non-coding region에서 발견되었다. 본 연구팀은 이외에 3개의 새로운 복제 종결자 후보 sequence를 발견하였는데, 이들은 *yojR*의 중간(*TerVIII*), *yokG*와 *yokH* 사이(*TerIX*) 및 *yokR*의 C-말단 부위(*TerX*)에 각각 위치하는 것으로 밝혀졌다 (Fig. 2). 이들이 진정한 복제 종결자 sequence인지는 복제 종결 단백질과의 결합 여부를 통해 향후 실험적으로 밝혀질 수 있을 것이다.

새로운 ORF 산물과 알려진 단백질과의 상동성 조사

65개 ORF의 염기서열로부터 유도된 단백질의 아미노산 서열과 데이터베이스에 수록된 기존 단백질의 아미노산 서열의 상동성을 조사하여 비교적 높은 BLAST score를 보여주는 것들을 Table 4에 요약 정리하였다. 9개의 알려진 유전자 *sspC*[20], *cge cluster*[52], *orfE* 5[24], *orfRM1*[45] 및 *odhA*[51] 등은 이미 그 염기서열과 기능이 밝혀진 것들이다. 다음의 27개 ORF는 데이터베이스 속의 어느 단백질과도 유의성 있는 상동성을 보여주지 않았다: *yojV*, *yojW*, *yojX*, *yojY*, *yojZ*, *yokB*, *yokC*, *yokD*, *yokE*, *yokF*, *yokG*, *yokH*, *yokI*, *yokJ*, *yokR*, *yokU*, *yokV*, *yokX*, *yolA*, *yolD*, *yolF*, *yojB*, *yojC*, *yojE*, *yojF*, *yojG* 및 *yojO*. 앞으로 이들 27개의 새로운 유전자들에 대한 기능의 해명을 위하여 많은 연구가 뒤따라야 할 것이다.

인트론 구조의 발견

*yojQ*와 *yojS*로부터 유도된 아미노산 서열을 분석한 결과, *B. subtilis* ribonucleotide reductase F의 N-말단, C-말단 지역과 각각 97%와 77%의 매우 높은 상동성을

보여 주었다. 두 가지 이상한 점은 *yojQ*와 *yojS* 두 ORF가 *yojR*을 사이에 두고 떨어져 있으며 *yojS*는 자신의 고유한 리보솜 결합 부위를 가지지 않는다는 것이다. 만약 *yojR* 부위를 제거하여 *yojQ*와 *yojS* 부위를 바로 연결할 경우, 아미노산의 과부족 없이 *B. subtilis*의 NrdF와 86%의 상동성을 보여 주었다. *yojR* 부위가 실제로 *B. subtilis* 168의 염색체상에 이와 같은 양상으로 존재하는지 확인하기 위하여 염색체 DNA를 주형으로 하고 ND-F와 ND-R primer(Fig. 1)를 사용하여 PCR을 수행하였다. 그 결과 522 bp 크기의 예상되는 부위가 *yojQ*와 *yojS* 사이에 존재함을 확인하였다(data not shown).

또한 흥미롭게도 *yojR*의 양 끝 부분 가까이에 group I intron에 특이적으로 존재한다고 알려져 온 P, Q, R 및 S라 불리는 잘 보존된 consensus element가 발견되었다 (Fig. 3; [60]). 더욱이 *yojR*을 포함하는 부위는 고등생물의 group I intron의 2차구조와 매우 흡사한 구조를 가지는 것으로 분석되었다(data not shown). 비록 *yojR*에서 유도된 단백질의 아미노산서열이 여러 인트론 경우의 endodeoxyribonuclease와 23%의 약한 상동성을 보여 주었지만, His-X-Asn-X-His의 특이적 아미노산 서열을 포함하고 있음이 밝혀졌다. 그 특이적 서열은 몇몇 group I intron-encoded endonuclease와 여러 group II intron ORF의 Zn-finger-like domain 내에서 항상 발견되는 것들이다[32, 59]. 그 동안 병원성 박테리오파지의 DNA 합성에 관여하는 몇 개의 유전자에서 인트론의 존재가 보고되었는데[31] 3종의 T4-파지 유전자(*nrdB*, *td*, 및 *sunY*)와 SPO1의 DNA polymerase를 코드하는 유전자에서 발견된 인트론이 그 예이다. 세균 염색체에 상주하는 template 박테리오파지에서 인트론으로 여겨지는 구조가 발견된 것은 이번이 처음이며 앞으로 실험적 증명을 해야 하는 과제가 남아있다.

유전자 지도의 재조정

그 동안 고초균의 유전자 지도 상에서 *odhAB*와 *sspC*의 위치는 181°와 182°로 각기 정해져 있었다[3, 9]. 그래서 두 유전자 사이의 거리는 12 kb 정도일 것으로 예상되어 왔지만, 염기서열을 해독한 결과 실제로는 40 kb 이상임이 판명되었다. 따라서 지금까지 사용되어온 이 부위의 유전자 지도는 정정되어야 한다.

이 부위에 위치하는 유전자들로서 이미 염기서열이 밝혀진 9개의 유전자중 *orfRM1*([45]; *ctpA*)은 그간 유전자 지도상의 위치가 알려지지 않았었는데, 염기서열이 완성되면서 그 위치가 밝혀졌다(Fig. 4). 반면에 그 동안 이 부근에 있을 것으로 예상되던 branched-chain α -keto acids를 위한 수송 단백질을 코드하는 *kauA*[30]는 이 지역에 포함되지 않는 것으로 밝혀졌다.

전체 염기서열이 밝혀짐에 따라, Fig. 5에 나타난 바와

Table 4. Similarity of the predicted ORF products to other known proteins

Product	Description or similar protein in database	BLAST score	Percentage identity*	References or database access [†]
YojP	Ribonucleotide reductase E (NrdE) of <i>Bacillus subtilis</i>	1325	46(651)	[56]
YojQ	N-terminal region of ribonucleotide reductase F (NrdF) of <i>B. subtilis</i>	955	97(193)	[56]
YojR	Orf36.1 of bacteriophage SPP1	124	23(173)	E: X67865
	A hypothetical protein in <i>Paramecium bursaria</i> Chlorella virus 1	114	23(173)	[43]
	Endodeoxyribonuclease of bacteriophage SP82	102	23(173)	[32]
YojS	C-terminal region of ribonucleotide reductase F	473	77(136)	[56]
YojT	Glutaredoxin-like protein (NrdH) of <i>Lactococcus lactis</i>	74	25(80)	[39]
	Thioredoxin (TrxA) of <i>B. subtilis</i>	64	39(80)	[15]
YojU	dUTPase of the temperate lactococcal bacteriophage r1t	117	32(142)	[64]
SspC	Small, acid-soluble spore protein C of <i>B. subtilis</i>	301	100(72)	[20]
OrfE5	A potential ORF adjacent to the replication terminator (TerV) of <i>B. subtilis</i>	444	100(88)	[24]
YokK	Orf181 at 3' region of <i>B. subtilis</i> <i>cgeB</i> corresponds to C-terminal region of YokK	687	100(129)	[52]
CgeB	A gene product in the <i>cge</i> cluster	1632	100(316)	[52]
CgeA	A gene product in the <i>cge</i> cluster	657	100(133)	[52]
CgeC	A gene product in the <i>cge</i> cluster	451	100(101)	[52]
CgeD	A gene product in the <i>cge</i> cluster	2288	100(426)	[52]
CgeE	A gene product in the <i>cge</i> cluster	546	100(118)	[52]
YokM	Biotin synthesis protein BioA homologue of <i>Rhizobium</i> sp.	181	32(444)	[26]
YokN	Acetyl-CoA: acetoacetyl-CoA transferase A subunit of <i>Escherichia coli</i>	220	47(229)	[37]
YokO	Probable 3-oxoadipate CoA-transferase subunit B of <i>B. subtilis</i>	521	48(217)	SP: P42316
	3-oxoacid CoA-transferase of <i>Sus scrofa domestica</i>	404	51(217)	[44]
yokP	Succinyl-diaminopimelate desuccinylase (DapE) of <i>E. coli</i>	133	24(179)	[11]
	Acetylornithine deacetylase (ArgE) of <i>E. coli</i>	82	23(434)	[47]
YokS	Hypothetical protein H10329 of <i>Haemophilus influenzae</i>	177	32(358)	[23]
YokW	Hypothetical protein 2 of <i>B. megaterium</i>	119	26(203)	PIR: S32217
YokY	Purine nucleoside phosphorylase II (PunB) of <i>B. stearothermophilus</i>	932	75(233)	[35]
YokZ	D,D-carboxypeptidase (VanY) of <i>Enterococcus faecalis</i>	159	31(267)	[22]
YolB	Hypothetical 21.6 kDa protein CY251.07 in <i>Mycobacterium tuberculosis</i>	101	20(233)	SP: Q10886
	Biotin synthesis protein BioC of <i>Serratia marcescens</i>	75	22(233)	SP: P36571
	BioC of <i>E. coli</i>	66	21(233)	[50]
OrfRM1	<i>B. subtilis</i> OrfRM1 similar to <i>E. coli</i> Prc	2289	98(466)	[45]
	Carboxyl-terminal processing protease (CtpA) of <i>Synechocystis</i> sp.	214	28(466)	[58]
YolC	Orfc0207 of <i>Sulfolobus solfataricus</i>	283	29(496)	[57]
	Proline permease homologue of <i>B. subtilis</i>	88	20(496)	GB: D50453
YolE	A pentachlorophenol-induced periplasmic protein homologue of <i>B. subtilis</i>	540	48(303)	DDBJ: AB001488
YolG	A protein similar to nitroreductase of <i>B. subtilis</i>	157	40(202)	DDBJ: AB001488
YolH	Hypothetical 14.7 kDa protein of <i>B. subtilis</i>	186	36(112)	[49]
YolI	Orf129 of <i>B. subtilis</i>	470	67(129)	[6]
YojA	Gluconate permease GntP of <i>B. subtilis</i>	806	65(444)	[27]
YojD	Chloramphenicol-sensitive protein RarD of <i>E. coli</i>	259	41(256)	[21]
YojH	Hypothetical 31.8 kDa protein YqhA of <i>B. subtilis</i>	412	48(147)	SP: P54504
YojI	Hypothetical protein YdhE of <i>E. coli</i>	371	34(452)	GB: U68703
YojJ	Hypothetical protein YbbP of <i>B. subtilis</i>	120	30(198)	DDBJ: AB002150
YojK	Hypothetical protein YdhE of <i>B. subtilis</i>	463	38(287)	[53]
	Glycosyl transferase of <i>Streptomyces lividans</i>	212	30(287)	[38]
YojL	Phosphatase-associated protein PapQ precursor of <i>B. subtilis</i>	437	46(414)	SP: P54421
	Autolysin of <i>Streptococcus faecalis</i>	171	26(414)	[7]
YojM	Cu/Zn superoxide dismutase of <i>Caulobacter crescentus</i>	107	31(196)	[62]
YojN	Nitric oxide reductase (NorQ) of <i>Rhodobacter sphaeroides</i>	128	26(304)	[5]
OdhA	2-Oxoglutarate dehydrogenase of <i>B. subtilis</i>	1359	99(270)	[51]

*The number of amino acids over which the percentage match was determined is shown in parentheses.

[†]E, EMBL; SP, SwissProt; PIR, NBRF-PIR; DDBJ, DNA Data Bank of Japan; GB, GenBank.

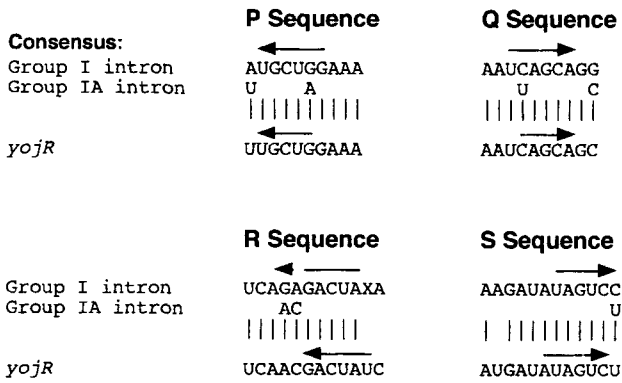


Fig. 3. DNA sequence alignment of the *yojR* region with highly conserved P, Q, R and S sequences of eukaryotic group I and group IA introns [60].

Arrows above sequences indicate regions of dyad symmetries. Vertical lines indicate identical nucleotide residues.

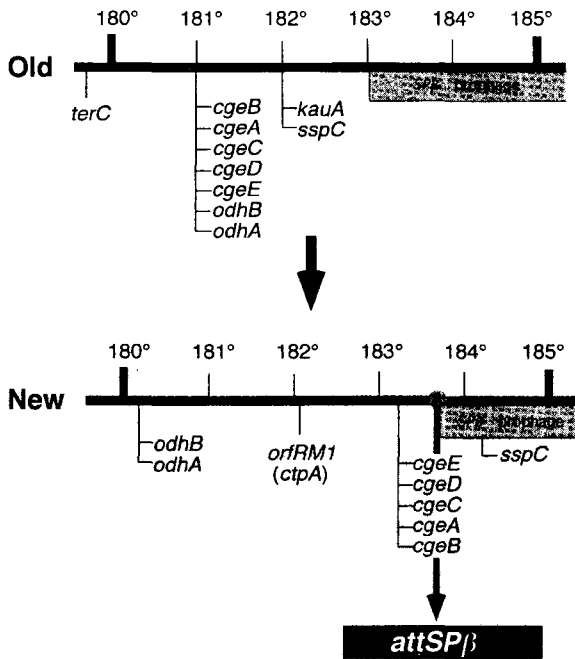


Fig. 4. Rearrangement of the current 'computerized genetic map of *B. subtilis* 168' [9].

'Old' and 'New' indicates the current map and the revised map around 180° on the chromosome, respectively. The letters under a filled bar represent the name of the known genetic loci. The *attSPβ* indicated by a filled circle and an arrow means the right attachment site for *SPβ* prophage.

같이 *odhA*의 위치는 180°, *cge* 유전자군은 183°, *sspC*는 184° 부근인 것으로 밝혀졌다. 또한 195° 부근의 *pipL* (Phage-Interrupted Protein near the Left attachment site; [14])외에 *SPβ* prophage의 다른 쪽 attachment site가 *orfE5*의 C-말단 지역에서 발견되었다(D. Karamata, personal communication; GenBank 등록 번호 M81760). 그 동안 182°에 자리한 *sspC*의 12 kb 하류 지

역인 183° 부근에 있을 것으로 예상되던 *attSPβ* 부위가 오히려 *sspC*보다 약 4 kb 상류 지역에 위치하는 것으로 판명된 것이다. 결과적으로 *sspC*는 *SPβ* prophage 내에 위치하는 것으로 밝혀졌다.

요 약

고초균 유전체 전체 염기서열을 밝히는 연구가 1997년 5월에 종료되어 전체 4,214,810 bp의 염기서열이 *SubtiList* 데이터베이스에 공식적으로 입력되었다. 과제의 진행은 약 8년 동안 국제적인 협력에 의하여 이루어져 왔으며, 유럽의 25개 연구팀, 일본의 7개 연구팀, 두 개의 회사 연구팀 그리고 한국의 본 연구팀이 참여했다. 고초균 유전체 염기서열 해독을 위한 위한 국제협력과제의 일환으로 본 연구팀은 *odhA* 유전자(181°) 상류지역 53,289 bp 부위의 염기서열을 해독하였다. 할당된 부위의 양 끝 부분에 위치한 *sspC*와 *odhA* 유전자의 알려진 염기서열을 시점으로하여, plasmid rescue와 long-range PCR 방법을 써서 염색체 DNA 단편을 획득하였다. 본 연구팀이 염기서열을 밝힌 염색체 DNA 부위에는 이미 보고된 9개 유전자(*sspC*, *cge* cluster, *orfE5*, *orfRM1* 및 *odhA*)를 포함하여 모두 65개의 ORF가 들어 있음이 밝혀졌다. 이 부위에서 얻은 흥미로운 결과중 하나는 인트론으로 여겨지는 한 ORF의 발견인데 세균의 염색체 상에서 인트론이 발견된 예는 흔치 않다. DNA 복제 종결 단백질의 결합이 예상되는 염기서열이 세곳에서 새로이 발견되었는데 이 역시 흥미로운 결과이다. 한편 이 부위 전체의 염기서열 해독을 통하여 기존의 유전자 지도상에 실제와는 매우 다르게 표시되어 온 여러 유전자들의 위치를 바로잡을 수 있었다.

감사의 말

이 연구는 과학기술처의 한·불 공동연구사업비 지원으로 수행되었으며 이에 감사드립니다.

참고문헌

1. Albano, M., J. Hahn, and D. Dubnau. 1987. Expression of competence genes in *Bacillus subtilis*. *J. Bacteriol.* **169**: 3110-3117.
2. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.
3. Anagnostopoulos, C., P. J. Piggot, and J. A. Hoch. 1993. The genetic map of *Bacillus subtilis*, pp. 425-461. In A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.), *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, American Society

- for Microbiology, Washington, DC.
4. Anagnostopoulos, C. and J. Spizizen. 1961. Requirements for transformation of *Bacillus subtilis*. *J. Bacteriol.* **81**: 741–746.
 5. Bartnikas, T. B., I. E. Tosques, W. P. Laratta, J. Shi, and J. P. Shapleigh. 1997. Characterization of the nitric oxide reductase-encoding region in *Rhodobacter sphaeroides* 2.4.3. *J. Bacteriol.* **179**: 3534–3540.
 6. Belitsky, B. and A. L. Sonenshein. 1997. Altered transcription activation specificity of a mutant form of *Bacillus subtilis* GltR, a *lysR* family member. *J. Bacteriol.* **179**: 1035–1043.
 7. Beliveau, C., C. Potvin, J. Trudel, A. Asselin, and G. Bellemare. 1991. Cloning, sequencing, and expression in *Escherichia coli* of a *Streptococcus faecalis* autolysin. *J. Bacteriol.* **173**: 5619–5623.
 8. Bertani, G. 1951. Studies of lysogeny. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**: 294–300.
 9. Biaudet, V., F. Samson, C. Anagnostopoulos, S. D. Ehrlich, and P. Bessieres. 1996. Computerized genetic map of *Bacillus subtilis*. *Microbiology* **142**: 2669–2729.
 10. Blattner, F. R. and 16 coauthors. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453–1474.
 11. Bouvier, J., C. Richard, W. Higgins, O. Bogler, and P. Stragier. 1992. Cloning, characterization, and expression of the *dapE* gene of *Escherichia coli*. *J. Bacteriol.* **174**: 5265–5271.
 12. Bullock, W. O., J. M. Fernandez, and J. M. Short. 1987. XL1-Blue, a high efficiency plasmid transforming *recA* *Escherichia coli* strain with β -galactosidase selection. *Bio-Techniques* **5**: 376–378.
 13. Bult, C. J. and 39 coauthors. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**: 1058–1073.
 14. Capuano, V., N. Galleron, P. Pujic, A. Sorokin, and S. D. Ehrlich. 1996. Organization of the *Bacillus subtilis* 168 chromosome between *kdg* and the attachment site of the SP β prophage: use of long accurate PCR and yeast artificial chromosomes for sequencing. *Microbiology* **142**: 3005–3015.
 15. Chen, N.-Y., J.-J. Zhang, and H. Paulus. 1989. Chromosomal location of the *Bacillus subtilis* aspartokinase II gene and nucleotide sequence of the adjacent genes homologous to *uvrC* and *trx* of *Escherichia coli*. *J. Gen. Microbiol.* **135**: 2931–2940.
 16. Cheng, S., C. Fockler, W. M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**: 5696–5699.
 17. Choi, S.-K., B.-T. Koo, B.-S. Shin, S.-H. Park, and J.-I. Kim. 1995. Screening of nested deletion mutants for DNA sequencing by direct electrophoresis of bacterial cultures. *Anal. Biochem.* **230**: 182–183.
 18. Cohen, S. N., A. C. Y. Chang, and L. Hsu. 1972. Non-chromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor. *Proc. Natl. Acad. Sci. USA* **69**: 2110–2114.
 19. Collado-Vides, J., B. Magasanik, and T. F. Smith. 1996. *Integrative Approaches to Molecular Biology*. The MIT Press, Massachusetts.
 20. Connors, M. J. and P. Setlow. 1985. Cloning of a small, acid-soluble spore protein gene from *Bacillus subtilis* and determination of its complete nucleotide sequence. *J. Bacteriol.* **161**: 333–339.
 21. Daniels, D. L., G. III. Plunkett, V. Burland, and F. R. Blattner. 1992. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* **257**: 771–778.
 22. Evers, S. and P. Courvalin. 1996. Regulation of VanB-type vancomycin resistance gene expression by the VanS_B-VanR_B two-component regulatory system in *Enterococcus faecalis* V583. *J. Bacteriol.* **178**: 1302–1309.
 23. Fleischmann, R. D. and 39 coauthors. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
 24. Franks, A. H., A. A. Griffiths, and R. G. Wake. 1995. Identification and characterization of new DNA replication terminators in *Bacillus subtilis*. *Mol. Microbiol.* **17**: 13–23.
 25. Fraser, C. M. and 28 coauthors. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397–403.
 26. Freiberg, C., X. Perret, W. J. Broughton, and A. Rosenthal. 1996. Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable 'sequenase': a beginning. *Genome Res.* **6**: 590–600.
 27. Fujita, Y., T. Fujita, Y. Miwa, J. Nishihashi, and Y. Aratani. 1986. Organization and transcription of the gluconate operon, *gnt*, of *Bacillus subtilis*. *J. Biol. Chem.* **261**: 13744–13753.
 28. Glaser, P., F. Kunst, M. Arnaud, M.-P. Coudart, W. Gonzales, M.-F. Hullo, M. Ionescu, B. Lubochinsky, L. Marcelino, I. Moszer, E. Presecan, M. Santana, E. Schneider, J. Schweizer, A. Vertes, G. Rapoport, and A. Danchin. 1993. *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol. Microbiol.* **10**: 371–384.
 29. Goffeau, A. and 632 coauthors. 1997. The yeast genome directory. *Nature* **387**: 5–105.
 30. Goldstein, B. J. and S. A. Zahler. 1976. Uptake of branched-chain α -keto acids in *Bacillus subtilis*. *J. Bacteriol.* **127**: 667–670.
 31. Goodrich-Blair, H., V. Scarlato, J. M. Gott, M.-Q. Xu, and D. A. Shub. 1990. A self-splicing group I intron in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. *Cell* **63**: 417–424.
 32. Goodrich-Blair, H. and D. A. Shub. 1994. The DNA po-

- lymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. *Nucleic Acids Res.* **22**: 3715–3721.
33. Griffiths, A. A. and R. G. Wake. 1997. Search for additional replication terminators in the *Bacillus subtilis* 168 chromosome. *J. Bacteriol.* **179**: 3358–3361.
 34. Haima, P., D. van Sinderen, H. Schotting, S. Bron, and G. Venema. 1990. Development of a β -galactosidase α -complementation for molecular cloning in *Bacillus subtilis*. *Gene* **86**: 63–69.
 35. Hamamoto, T., T. Noguchi, and Y. Midorikawa. 1997. Cloning of purine nucleoside phosphorylase II gene from *Bacillus stearothermophilus* TH 6-2 and characterization of its gene product. *Biosci. Biotech. Biochem.* **61**: 276–280.
 36. Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkl, B.-C. Li, and R. Herrmann. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**: 4420–4449.
 37. Itoh, T. and 31 coauthors. 1996. A 460-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 40.1-50.0 min region on the linkage map. *DNA Res.* **3**: 379–392.
 38. Jenkins, G. and E. Cundliffe. 1991. Cloning and characterization of two genes from *Streptomyces lividans* that confer inducible resistance to lincomycin and macrolide antibiotics. *Gene* **108**: 55–62.
 39. Jordan, A., E. Pontis, F. Aslund, U. Hellman, I. Gibert, and P. Reichard. 1996. The ribonucleotide reductase system of *Lactococcus lactis*: characterization of an NrdEF enzyme and a new electron transport protein. *J. Biol. Chem.* **271**: 8779–8785.
 40. Kaneko, T. and 23 coauthors. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**: 109–136.
 41. Kaneko, T., A. Tanaka, S. Sato, H. Kotani, T. Sazuka, N. Miyajima, M. Sugiura, and S. Tabata. 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res.* **2**: 153–166.
 42. Kunst, F. and 149 coauthors. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249–256.
 43. Kutish, G. F., Y. Li, Z. Lu, M. Furuta, D. L. Rock, and J. L. Van Etten. 1996. Analysis of 76 kb of the chloroella virus PBCV-1 330-kb genome: map positions 182 to 258. *Virology* **223**: 303–317.
 44. Lin, T. W. and W. A. Bridger. 1992. Sequence of a cDNA clone encoding pig heart mitochondrial CoA transferase. *J. Biol. Chem.* **267**: 975–978.
 45. Marasco, R., M. Varcamonti, E. Ricca, and M. Sacco. 1996. A new *Bacillus subtilis* gene with homology to *Escherichia coli* *prc*. *Gene* **183**: 149–152.
 46. McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz. 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β -lactamase gene. *J. Biol. Chem.* **256**: 11283–11291.
 47. Meinnel, T., E. Schmitt, Y. Mechulam, and S. Blanquet. 1992. Structural and biochemical characterization of *Escherichia coli* *argE* gene product. *J. Bacteriol.* **174**: 2323–2331.
 48. Moszer, I., F. Kunst, and A. Danchin. 1996. The European *Bacillus subtilis* genome sequencing project: current status and accessibility of the data from a new World Wide Web site. *Microbiology* **142**: 2987–2991.
 49. Ogasawara, N., S. Nakai, and H. Yoshikawa. 1994. Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res.* **1**: 1–14.
 50. Otsuka, A. J., M. R. Buoncristiani, P. K. Howard, J. Flamm, C. Johnson, R. Yamamoto, K. Uchida, C. Cook, J. Ruppert, and J. Matsuzaki. 1988. The *Escherichia coli* biotin biosynthetic enzyme sequences predicted from the nucleotide sequence of the *bio* operon. *J. Biol. Chem.* **263**: 19577–19585.
 51. Resnekov, O., L. Melin, P. Carlsson, M. Mannerlov, A. von Gabain, and L. Hederstedt. 1992. Organization and regulation of the *Bacillus subtilis* *odhAB* operon, which encodes two of the subenzymes of the 2-oxoglutarate dehydrogenase complex. *Mol. Gen. Genet.* **234**: 285–296.
 52. Roels, S. and R. Losick. 1995. Adjacent and divergently oriented operons under the control of the sporulation regulatory protein GerE in *Bacillus subtilis*. *J. Bacteriol.* **177**: 6263–6275.
 53. Sadaie, Y., K. Yata, M. Fujita, H. Sagai, M. Itaya, Y. Kasahara, and N. Ogasawara. Nucleotide sequence and analysis of the *phoB-rnE-groESL* region of the *Bacillus subtilis* chromosome. *Microbiology* **143**: 1861–1866.
 54. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: a Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, New York.
 55. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
 56. Scotti, C., A. Valbuzzi, M. Perego, A. Galizzi, and A. M. Albertini. 1996. The *Bacillus subtilis* genes for ribonucleotide reductase are similar to the genes for the second class I NrdE/NrdF enzymes of *Enterobacteriaceae*. *Microbiology* **142**: 2995–3004.
 57. Sensen, C. W., H.-P. Klenk, R. K. Singh, G. Allard, C. C.-Y. Chan, Q. Y. Liu, S. L. Penny, F. Young, M. E. Schenk, T. Gaasterlund, W. F. Doolittle, M. A. Ragan, and R. L. Charlebois. 1996. Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from *Sulfolobus solfataricus* P2. *Mol. Microbiol.* **22**: 175–191.
 58. Shestakov, S. V., P. R. Anbudurai, G. E. Stanbekova, A.

- Gadzhiev, L. K. Lind, and H. B. Pakrasi. 1994. Molecular cloning and characterization of the *ctpA* gene encoding a carboxyl-terminal processing protease. Analysis of a spontaneous photosystem II-deficient mutant strain of the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **269**: 19354–19359.
59. Shub, D. A., H. Goodrich-Blair, and S. R. Eddy. 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *TIBS* **19**: 402–404.
60. Shub, D. A., J. M. Gott, M.-Q. Xu, B. F. Lang, F. Michel, J. Tomaschewski, J. Pedersen-Lane, and M. Belfort. 1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. *Proc. Natl. Acad. Sci. USA* **85**: 1151–1155.
61. Spizizen, J. 1958. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *Proc. Natl. Acad. Sci. USA* **44**: 1072–1078.
62. Steinman, H. M. and B. Ely. 1990. Copper-zinc superoxide dismutase of *Caulobacter crescentus*: cloning, sequencing, and mapping of the gene and periplasmic location of the enzyme. *J. Bacteriol.* **172**: 2901–2910.
63. Tomb, J.-F. and 41 coauthors. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539–547.
64. van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Ruiters, G. Venema, and A. Nauta. 1996. Sequence analysis and molecular characterization of the temperature lactococcal bacteriophage r1t. *Mol. Microbiol.* **19**: 1343–1355.
65. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103–119.
66. Zuker, M. and P. Stiegler. 1981. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**: 133–148.

(Received October 22, 1997)