J Mol Evol (1995) 40:585–593

JOURNAL OF MOLECULAR EVOLUTION

Identification and Chromosomal Distribution of DNA Sequence Segments Conserved Since Divergence of *Escherichia coli* and *Bacillus subtilis*

Takashi Kunisawa

Department of Applied Biological Sciences, Science University of Tokyo, Noda, 278, Japan

Received: 21 December 1993 / Revised: 11 July 1994 / Accepted: 16 July 1994

Abstract. DNA sequence segments conserved since divergence of *Escherichia coli* and *Bacillus subtilis* were identified, using the GenBank sequence database. Chromosomal locations of the conserved segments were compared between the two bacteria, and the following three features were observed. (1) Although the two genomes are nearly identical in size, chromosomal arrangements of the conserved segments are considerably different from each other. (2) In many cases, chromosomal locations of a conserved segment in the two species have deviated from each other by a multiple of 60°. (3) There are many instances in which a contiguous segment in one genome is split into two or more segments located at distinct positions in the other genome, and these split segments were found to tend to lie on the *E. coli* or *B. subtilis* genome separated by distances of multiples of 60°. On the basis of these observations, genome organizations of the two bacteria were discussed in terms of genome doublings as well as random chromosomal rearrangements.

Key words: Genome evolution — Gene locations — Sequence similarity — Genome doublings

Introduction

Recent progress in genetic and biochemical techniques has permitted the chromosomal assignment of numerous biochemically or genetically defined genetic loci and the development of genetic maps for a variety of species. (For compilation, see O'Brien 1990.) Analysis of these maps provides a useful means with which to study genome organization and evolution. For example, the organizations of *Escherichia coli* and *Salmonella typhimurium* genomes are very similar with a few exceptions, and gene order has essentially been preserved over a long period of evolutionary time (Riley and Krawiec 1987). A similar gene order is also found in the linkage maps of rice and maize (Ahn and Tanksley 1993). Although the maize genome is six times that of rice, the two genomes have extensive conservation of gene order over large segments and the difference in genome size is accounted for by the difference in copy number of duplicated segments. Furthermore, comparative linkage maps that show the relative order of homologous genes along the genomes provide a basis for interpreting genetic information between divergent species (Tanksley et al. 1992).

Here genetic maps of *E. coli* and *Bacillus subtilis* are compared. Such a comparison may provide useful information regarding the extent to which the two genomes are shuffled since the divergence and the pattern of recombination. Several reports have suggested that contemporary bacterial genomes may have evolved from small to big by several genome doublings (Hopwood 1967; Wallace and Morowitz 1973; Riley et al. 1978; Herdman 1985; Kunisawa and Otsuka 1988). If this hypothesis is true, some regularity is to be expected in the distribution of conserved segments within the genomes. This study addresses the question of whether chromosomal rearrangements are randomly distributed within the genomes or whether evolutionary relics of ancient genome doublings are observed in the contemporary genome structures.

E. coli is the most extensively studied gram-negative bacterium. According to the genetic map compiled by

Bachmann (1990), nearly 1,400 genes are mapped. DNA sequence libraries currently contain more than 2,400 kb of nonredundant sequences (Kröger et al. 1993). These 2,400 kb represent approximately 50% of the entire *E. coli* genome. The other bacterium, *B. subtilis,* has also been extensively studied. The genome size (4,200 kb) of this gram-positive bacterium is nearly the same as the size of *E. coli* (4,700 kb). About 800 genetic loci have been localized on its genetic map (Anagnostopoulos et al. 1993) and DNA sequence libraries contain 650 kb of nonredundant sequences, which correspond to about 15% of the entire *B. subtilis* genome (Glaser et al. 1993). A phylogenetic study of 5S rRNA sequences has given an estimate of the divergence time of ca. 1.2 billion years between the two bacteria (Hori and Osawa 1987).

Data and Computer Method

Genetic map location in *E. coli* is usually shown in units of minutes as measured by time of entry in interrupted-conjugation experiments, and the entire *E. coli* genome is represented by a circular 100-min map (Bachmann 1990). On the other hand, the *B. subtilis* genome is represented by a circular 360° map (Anagnostopoulos et al. 1993). Data of gene location were obtained from these two genetic maps. To facilitate comparison between the two maps, the *E. coli* genetic map was converted to a 360° map by simply multiplying gene locations shown in units of minutes by a factor of 3.6.

DNA sequence data were obtained from GenBank, release 81.0. These sequences were translated to identify conserved DNA segments at the amino acid sequence level. The computer program FASTA developed by Lipman and Pearson (1985) was used to compare individual *B. subtilis* amino acid sequences with each *E. coli* amino acid sequence. According to them, distantly related protein sequences have initial scores greater than 50 and optimized scores greater than 100 in the FASTA search. Therefore, sequence pairs fulfilling the above criteria were possible candidates for evolutionary common origins. For pairs thus selected by the FASTA search, statistical significance of the sequence similarity was evaluated (Dayhoff et al. 1983; Kunisawa and Otsuka 1988). In accordance with Lipman and Pearson (1985), sequence pairs found by the FASTA search exhibit high degrees of statistical significance; the chance probability is, in most cases, less than 1.28×10^{-12}. In addition to the sequence similarity, most of the pairs thus identified appear to serve a similar function. Furthermore, we take account of only large segments containing two or more genes that show high sequence similarity and are arranged in the same order on the two genomes. Therefore, conserved sequence segments identified here are most likely to have originated from common ancestors and to be homologous.

Results and Discussion

Identification of Conserved DNA Segments

The present computer search has identified 47 sequence segments showing the same or a very similar gene order between *B. subtilis* and *E. coli.* These conserved segments are summarized in Table 1, where gene orders, map positions, and percent values of amino acid identity for each pair of corresponding genes are listed. In this table, each segment is ordered according to map location with respect to the *B. subtilis* genome and corresponding genes are displayed side by side for easy comparison. A contiguous DNA sequence segment is shown by a solid line and the direction of transcription is indicated with arrows. For several conserved DNA segments which are listed in Table 1, simple explanations are given below.

A large conserved DNA segment consisting of ca. 13 kb is found in the region of replication origin. The nucleotide sequence of the *B. subtilis* replication origin located at map position 0° reveals a gene order *gidBA*-orf50K-orf208-orf261-*rnpA-rpmH-dnaAN-orf71-recF-gyrB.* The identical gene order including two open reading frames, orf50K and orf261 (orf60K), is also found in *E. coli* around its replication origin, which is located at map position 301° (83.5 min). In *E. coli,* however, *gidA* and *gidB* genes are located at 303° (84.3 min) and therefore 40 kb away from the segment comprised of orf60K to *gyrB.*

According to the genetic map of *B. subtilis,* 30 genes coding for ribosomal proteins are present in a large cluster at map position 10°. Although the nucleotide sequence of this cluster is not completely determined, the cluster may be assumed to form a continuous DNA segment. This is due to the following observation. In *E. coli,* an identical gene order can be found in two DNA sequence segments located at 263° (73 min) and 324° (90 min); if these two segments are fused, one obtains the gene arrangement found in *B. subtilis,* as can be seen in Table 1. These *B. subtilis* genes with known sequences show high degrees of sequence similarity to *E. coli* counterparts. Thus, the *B. subtilis* ribosomal gene cluster of 32 kb length may be regarded as a large conserved DNA segment.

In *B. subtilis,* a cluster of genes involved in cell-wall synthesis and cell division lies in the 133°–135° region. The order of genes in this large cluster of ca. 15 kb is very similar to a cluster of *E. coli* genes with similar functions, which is located in the 2-min region. The products of these genes are closely related in amino acid sequence (23–54% identity) with one another, as shown in Table 1.

Homologous operons are also found between the two bacteria. The *B. subtilis cta* operon, which encodes cytochrome-c oxidase and is located at map position 127°, is clearly homologous with its *E. coli* counterpart *cyo* operon, which is located at 36° (10 min). The gene order in the two operons is identical except that the first gene *ctaB* of the *cta* operon corresponds to the last gene *cyoE* of the *cyo* operon. This *cyo* operon is also homologous with *B. subtilis* quinol oxidase operon *qox,* which is mapped at 331°. Although a homologue of the *E. coli cyoE* gene is not found in the *B. subtilis qox* operon, the rest of the *cyo* genes are very similar in amino acid sequence to corresponding *qox* genes. Another example

Table 1. Homologous DNA segments in which gene order is conserved since divergence of *B. subtilis* and *E. coli*

B. subtilis		*E. coli*			
Map position (°)	Genes	Genes	Map position (°) [min]	Amino acid identity (%)	Map difference (°)
0	*gidB*	*gidB*	303 [84.3]	34	–57
	gidA	*gidA*		52	
	orf50K	*orf50K*	301 [83.5]	32	–59
	orf208				
	orf261	*orf60K*		39	
		orf9K			
	rnpA	*rnpA*		29	
	rpmH	*rpmH*		67	
	dnaA	*dnaA*		49	
	dnaN	*dnaN*		27	
	orf71				
	recF	*recF*		23	
	gyrB	*gyrB*		59	
0	*dnaX*	*dnaZX*	40 [11]	31	+40
	orf	*orf*		40	
	recR	*recR*		43	
10	*secE*	*secE*	324 [90]	24	–46
	nusG	*nusG*		44	
	rplK	*rplK*		67	
	rplA	*rplA*		50	
	rplJ	*rplJ*			
	rplL	*rplL*		51	
	rpoB	*rpoB*			
	rpoC	*rpoC*			
	rpsL	*rpsL*	263 [73]		–107
	rpsG	*rpsG*			
	fusA	*fusA*			
	tufA	*tufA*			
	rpsJ	*rpsJ*			
	rplC	*rplC*			
	rplD	*rplD*			
	rplW	*rplW*			
	rplB	*rplB*			
	rpsS	*rpsS*			
	rplV	*rplV*			
	rpsC	*rpsC*			
	rplP	*rplP*		64	
	rpmC	*rpmC*		40	
	rpsQ	*rpsQ*		54	
	rplN	*rplN*		63	
	rplX	*rplX*		46	
	rplE	*rplE*		57	
	rpsN	*rpsN*		57	
	*rpsH**	*rpsH*		81	
	rplF	*rplF*			
	rplR	*rplR*			
	rpsE	*rpsE*		52	
	rpmD	*rpmD*		43	
	rplO	*rplO*		47	
	secY	*secY*		41	
	adk				
	map				
	infA				
	rpmJ	*rpmJ*		58	
	rpsM	*rpsM*		55	
	rpsK	*rpsK*		63	
		rpsD			
	rpoA	*rpoA*		46	
	rplQ	*rplQ*		45	
55	*purE*	*purE*	43 [12]	59	–12
	purK	*purK*		32	

Table 1. Continued

B. subtilis		*E. coli*			
Map position (°)	Genes	Genes	Map position (°) [min]	Amino acid identity (%)	Map difference (°)
	purB				
	purC				
	purQ				
	purL				
	purF				
	purM	*purM*	194 [54]	50	+139
	purN	*purN*		31	
	purH	*purH*	324 [90]	52	–91
	purD	*purD*		53	
75	*glpF*	*glpF*	317 [88]	33	–118
	glpK	*glpK*		63	
104	*oppA*	*oppA*	101 [28]	34	–3
	oppB	*oppB*		50[b]	
	oppC	*oppC*		44[b]	
	oppD	*oppD*		53[b]	
	oppF	*oppF*		57[b]	
118	*ptsX*	*(crr)*	187 [52]	40	+69
	ptsH	*ptsH*		36	
	*ptsI**	*ptsI*		35	
		crr			
120	*motA*	*motA*	151 [42]	27	+31
	motB	*motB*		27	
126	*pdhC*	*aceF*	11 [3]	35	–115
	pdhD	*lpd*		47	
127	*ctaB*	*(cyoE)*	36 [10]	35	–91
	ctaC	*cyoA*		26	
	ctaD	*cyoB*		45	
	ctaE	*cyoC*		44	
	ctaF	*cyoD*		30	
		cyoE			
134	*orf*[a]	*orf*	7 [2]	44	–127
		ftsL			
	orf				
	pbp2B	*(ftsI)*		25	
	spoVD[a]	*ftsI*		23	
	murE	*murE*		38	
	(murE)	*murF*		30	
	mraY	*mraY*		54	
	murD	*murD*		32	
	spoVE	*ftsW*		42	
	murG	*murG*		30	
	orf				
	divIB				
	orf				
	orf				
	sbp				
		murC			
		ddl			
		ftsQ			
	ftsA	*ftsA*		34	
	ftsZ	*ftsZ*		48	
139	*pyrAA*	*carA*	4 [1]	46	–135
	pyrAB	*carB*			
140	*flgB*	*flgB*	86 [24]	25[b]	–54
	flgC	*flgC*		37[b]	
	fliE	*fliE*[a]	153 [42.5]	30	+13
	fliF	*fliF*		23	
	fliG	*fliG*		36	
	.	.			
	.	.			

Table 1. Continued

B. subtilis		*E. coli*			
Map position (°)	Genes	Genes	Map position (°) [min]	Amino acid identity (%)	Map difference (°)
	fliL	*flaAI*		21	
	fliM	*flaAII*		29	
	fliY	*motD*		41	
	.				
	.				
	cheA	*cheA*	149 [41.5]	34	+9
	cheW	*cheW*		27	
145	*ffh*	*ffh*	205 [57]	54	+60
	rpsP	*rpsP*		51	
147	*P15A*	*P15A*	248 [69]	37	+101
	nusA	*nusA*		40	
	orf				
	orf				
	infB	*infB*		48	
	orf				
	P15B	*P15B*		42	
	P35	*P35*		35	
181	*odhA*	*sucA*	58 [16]	37	−123
	odhB	*sucB*		47	
203	*trpE*	*trpE*	101 [28]	38	−102
	trpD	*trp(G)D*		35	
	trpC	*trpC*		35	
	trpF	*(trpC)*		30	
	trpB	*trpB*		56	
	trpA	*trpA*		35	
208	*orf17*	*phoB*	32 [9]	41	+176
	orf18	*phoR*		30	
209	*ribG*	*orf2*	36 [10]	41	−173
	ribB				
	ribA				
	ribH	*orf3*		53	
		nusB			
224	*dnaK*	*dnaK*	1 [0.2]	54	+137
	dnaJ	*dnaJ*		50	
	rpoD	*rpoD*	241 [67]	31	+17
	dnaG	*dnaG*		63	
233	*levD*	*(ptsL)*	144 [40]	37	−89
	levE	*ptsL*		48	
	levF	*ptsP*		59	
	levG	*ptsM*		61	
242	*mreB*	*mreB*	256 [71]	58	+10
	mreC	*mreC*		23	
	mreD	*mreD*		21	
	minC	*minC*	94 [26]	18	−148
	minD	*minD*		44	
	spoIVFA				
	spoIVFB				
	rplU	*rplU*	248 [69]	46	+6
	orf				
	rpmA	*rpmA*		46	
244	*hemC*	*hemC*	306 [85]	46	+62
	hemD	*hemD*		25	
252	*sdhA*	*sdhA*	58 [16]	31	+166
	sdhB	*sdhB*		30	
	sdhA	*frdA*	338 [94]	33	+86
	sdhB	*frdB*		25	
258	*phoP*	*phoB*	32 [9]	40	+134
	phoR	*phoR*		37	
263	*ccpA*		151 [42]		−112
	orfA	*motA*		19	

Table 1. Continued

B. subtilis		*E. coli*			
Map position (°)	Genes	Genes	Map position (°) [min]	Amino acid identity (%)	Map difference (°)
	orfB	*motB*		23	
273	*menD*	*menD*	176 [49]	28	–97
	menB	*menB*		70	
275	*glgB*	*glgB*	273 [76]	43	–2
		glgX			
	glgC	*glgC*		38	
	glgD	*(glgC)*		22	
	glgA	*glgA*		35	
	glgP	*glgP*		44	
284	*thrC*	*thrC*	0 [0]	26	+76
	thrB	*thrB*		34	
317	*alsR*	*leuO*	7 [2]	18	+50
	alsS	*ilvI*		24	
	alsS	*ilvG*	306 [85]	26	–11
		ilvE			
		ilvD			
		ilvA			
	alsR	*ilvY*		28	
327	*orf71*	*(orf292)*	306 [85]	42	–26
	orf72	*rffE*		48	
		orf292			
330	*sacT*	*bglC*	302 [84]	35	–28
	sacP	*bglS*		30	
331	*qoxA*	*cyoA*	36 [10]	36	+64
	qoxB	*cyoB*		52	
	qoxC	*cyoC*		47	
	qoxD	*cyoD*		31	
333	*sacX*	*bglC*	302 [84]	29	–31
	sacY	*bglS*		35	
344	*groES*	*groES*	338 [94]	45	–6
	groEL	*groEL*		61	
357	*rpsF*	*rpsF*	342 [95]	23	–15
	ssb				
		orf			
	rpsR	*rpsR*		49	
	(30 kb)				
	rplI	*rplI*		33	

[a] Partial sequence data

[b] Compared with *Salmonella* sequence

is the *opp* operon responsible for oligopeptide transport. This operon is found at position 104° in *B. subtilis* and at 101° (28 min) in *E. coli.* Since *E. coli oppBCDE* genes are not sequenced, the percent values of amino acid identity listed in Table 1 were obtained using *Salmonella* sequences. However, since the *E. coli oppA,* for which sequence data are available, is homologous to both *B. subtilis* and *Salmonella oppA* genes, the *E. coli* and *B. subtilis* oligopeptide transport operons are very likely homologous. Another homologous operon can be seen in the tryptophan biosynthesis operon. In *E. coli* the *trp* operon is mapped at 101° (28 min), while in *B. subtilis* this operon is split into two distinct positions; a major cluster consisting of *trpEDCFBA* genes is located at position 203° but the *trpG* gene is separated from this cluster and is located at 9°.

Thus, the present computer-assisted analysis reveals that conserved DNA segments are found in various regions of the two distantly related genomes. However, their relative chromosomal positions on one genome differ considerably from those on the other genome, reflecting chromosomal rearrangements during a long evolutionary time.

Comparison of Chromosomal Arrangements in E. coli *and* B. subtilis

To make a quantitative comparison between the chromosomal arrangements of conserved DNA segments in *B. subtilis* and *E. coli,* the number of segments is plotted in Fig. 1 against the difference *D* between map positions of the two genomes. The value of *D* for a pair of conserved segments was evaluated with respect to the map position

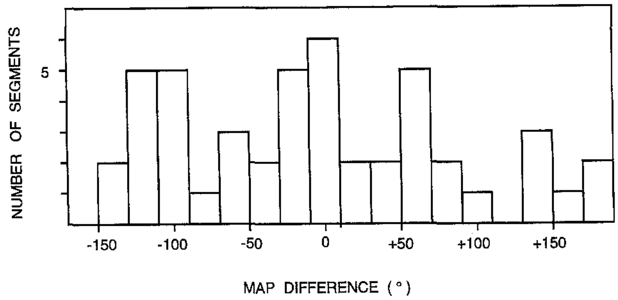

Fig. 1. Frequency distribution of map differences between *E. coli* and *B. subtilis* at the resolution of 20°.

of *B. subtilis*. Thus, +*D* indicates that the position for *E. coli* is larger by *D*° in the clockwise direction and –*D* indicates that *E. coli* position deviates by *D*° from its counterpart of *B. subtilis* in the counterclockwise direction. These map differences for each pair of conserved DNA segments are shown at the last column in Table 1. If, on one extreme, the chromosomal arrangement is identical with a rotation of one genome by *X*° with respect to the other, one gets a single peak at *X* in this plot. If, on the other extreme, the chromosomal arrangement is completely shuffled, then one gets a uniform distribution. The frequency distribution shown in Fig. 1 reveals three major peaks located around –110°, 0°, and +60°, and frequencies at map differences other than these are close to the average 2.6 expected from a random distribution of a total of 47 segments into 18 cells. The three major peaks are located with an interval of multiples of 60°, suggesting some regularity in the chromosomal arrangement.

As mentioned, in *B. subtilis* the replication origin is at 0°, but in *E. coli* the origin is at 301°. It is to be noted that the misalignment of the conserved segments is not corrected by this 60° shift; if we rotate the *E. coli* map by +60° so that the locations of the two replication origins become identical, this gives rise only to the shift of the abscissa by +60° and the histogram itself is not altered.

Map Distance Between Split Segments in E. coli *or* B. subtilis

In Table 1 one sees that a large gene cluster in *B. subtilis* tends to be split into two separate clusters in *E. coli.* The cluster of ribosomal protein genes in *B. subtilis* is such an example. Conversely, *sdhDAB* and *sucABC* operons form a contiguous segment in *E. coli,* while their homologues of *B. subtilis, sdhAB* and *odhAB,* are located at distinct positions, 252° and 181°, respectively. Figure 2 illustrates such segments that are present in a single cluster in one genome but are split into two or more segments in the other genome. In this figure, the gene order of a contiguous segment is specified, but split segments are indicated using solid lines. Chromosomal positions are also listed. Here we analyze map distances, designated as *d* in Fig. 2, between the split segments found in the genome of either *E. coli* or *B. subtilis.*

The two major clusters of ribosomal protein genes in *E. coli* are present in two regions 61° apart (Fig. 2a). Operons *sdhAB* and *odhAB* are located 71° apart in the *B. subtilis* genome (Fig. 2b). Genes affecting flagellar, chemotaxis, and motility functions in *B. subtilis* are clustered in a region 140°. Among them, *flgBC* and *flgG* genes are separated from the rest in *E. coli.* The separation distance is 67° (Fig. 2c). The *dnaK-dnaJ* and *ropD-dnaG* clusters are located 120° apart (Fig. 2d). The 0–2.4-min region of *E. coli* is completely sequenced and *thrBC, dnaKJ, carAB, araDAB, leuDCBA,* and *ftsI* to *ftsZ* genes are arranged in this order. Their counterparts in *B. subtilis* are found in four regions (Fig. 2e), though sequence data for *araDAB* and *leuDCBA* genes of *B. subtilis* are not available. The map distances between these *B. subtilis* counterparts are 60°, 85°, 117°, and 113°. In *B. subtilis* genes involved in purine biosynthesis form a large cluster in the order *purEKBC*-orf-*purQLFMNHD.* In *E. coli,* gene orders *purEK, purMN, purHD* are found in three clusters with map distances 151°, 130°, and 79° (Fig. 2f). Similarly, map distances 162° and 154° are obtained in the comparison between *mreBCD* and *minCD* clusters and between *minCD* and *rplU-rpmA* clusters, respectively (Fig. 2g). Finally, *cyo* operon and *bglCS* cluster are separated by 93° in *E. coli* (Fig. 2h).

Thus, it can be noted that most of the split segments are separated by 60° or 120°. Among a total of 15 distances, 13 distances fall into three major regions between 60° and 80°, between 110° and 130°, and between 150° and 170°, and only two cases are exceptional. The three major regions are located with an approximately regular interval and, in particular, the regions 60°–80° and 110°–130° are close to those expected from multiples of 60°.

Regularity in Chromosomal Arrangement

In this way, it can be seen that numerous DNA segments conserved between *E. coli* and *B. subtilis* are distributed on each genome with map differences of multiples of 60° (Fig. 1) and that split segments also tend to be separated by multiples of 60° (Fig. 2). Although many factors are conceivable that may cause these regularities, e.g., distribution of recombination hot spots and uneven distribution of gene density, ancient genome doublings seem more straightforward. If the whole genome has been duplicated several times, homologous genes are distributed with a regular distance of the original genome length in the duplicated genome. If redundant copies are differentially eliminated in different lineages, those duplicated genes remaining are separated by a distance of multiples of the original genome length in the contemporary genomes. Furthermore, if differential silencing of duplicated genes occurs at distinct chromosomal positions in one lineage, split segments are observed with a regular

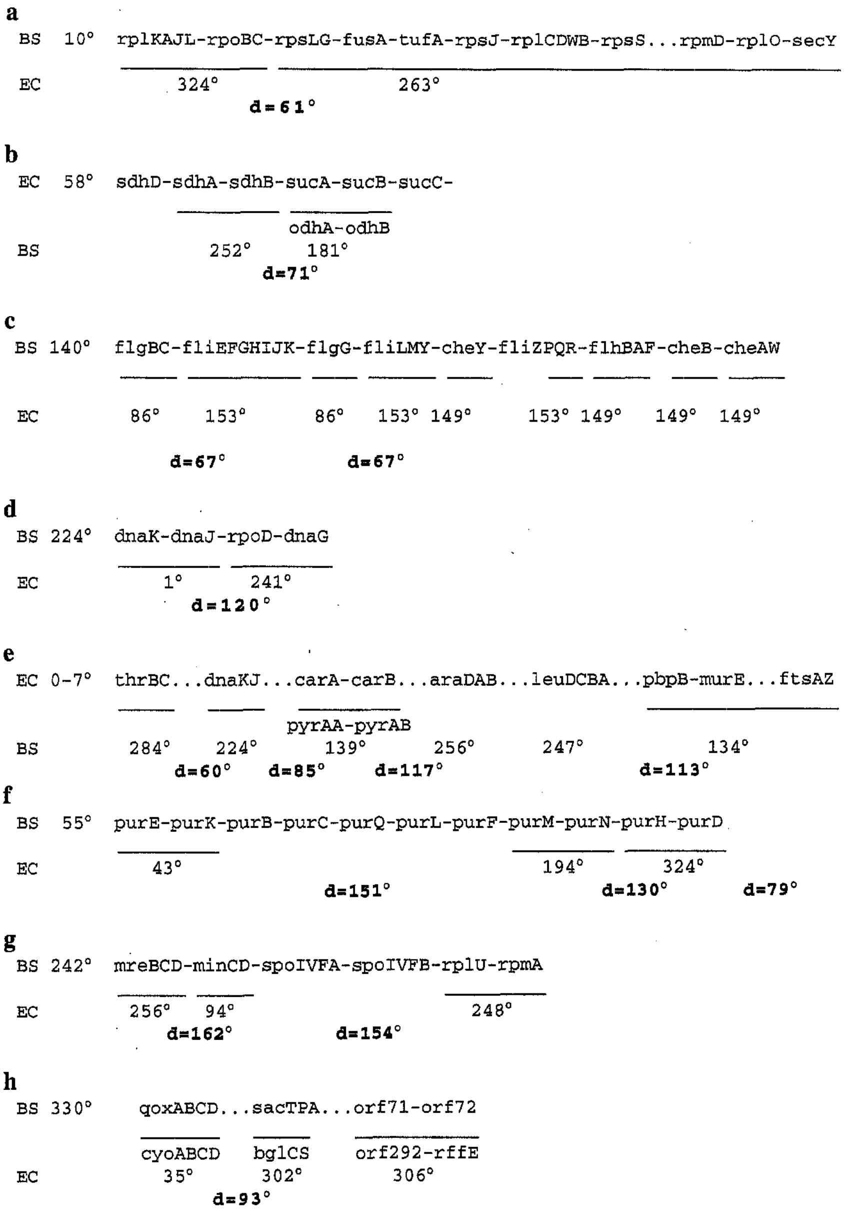
a
BS 10° rplKAJL-rpoBC-rpsLG-fusA-tufA-rpsJ-rplCDWB-rpsS...rpmD-rplO-secY
EC 324° 263°
d=61°
b
EC 58° sdhD-sdhA-sdhB-sucA-sucB-sucC-
odhA-odhB
BS 252° 181°
d=71°
c
BS 140° flgBC-fliEFGHIJK-flgG-fliLMY-cheY-fliZPQR-flhBAF-cheB-cheAW
EC 86° 153° 86° 153° 149° 153° 149° 149° 149°
d=67° d=67°
d
BS 224° dnaK-dnaJ-rpoD-dnaG
EC 1° 241°
d=120°
e
EC 0-7° thrBC...dnaKJ...carA-carB...araDAB...leuDCBA...pbpB-murE...ftsAZ
pyrAA-pyrAB
BS 284° 224° 139° 256° 247° 134°
d=60° d=85° d=117° d=113°
f
BS 55° purE-purK-purB-purC-purQ-purL-purF-purM-purN-purH-purD
EC 43° 194° 324°
d=151° d=130° d=79°
g
BS 242° mreBCD-minCD-spoIVFA-spoIVFB-rplU-rpmA
EC 256° 94° 248°
d=162° d=154°
h
BS 330° qoxABCD...sacTPA...orf71-orf72
cyoABCD bglCS orf292-rffE
EC 35° 302° 306°
d=93°

Fig. 2. Conserved segments that are split in *E. coli* or *B. subtilis.* Gene order of a contiguous segment in one genome is illustrated and its homologous but split segments in the other genome are indicated with *solid lines.* Map distances between the split segments in *E. coli* or *B. subtilis* are designated as *d. BS, B. subtilis; EC, E. coli.*

interval. Genome doubling or polyploidization has been suggested in plants (Stebbins 1950), fish and amphibians (Ohno 1970; Ferris and Whitt 1979), as well as in bacteria (Hopwood 1967; Wallace and Morowitz 1973; Riley et al. 1978; Herdman 1985; Kunisawa and Otsuka 1988).

So far we have discussed the regular distribution. It is also noticeable, however, that there are exceptional gene regions that are not in discrete multiples of 60°. One simple explanation of these exceptional instances may be that they are translocated randomly from one chromosomal position to another by chromosomal recombination. Thus, the chromosomal arrangements of conserved DNA segments in the two bacteria seem to represent the superposition of a regular distribution and a random distribution. Much more must be done to establish the regular distribution of homologous segments within the two genomes. Bacterial genome-sequencing projects have recently been initiated, including *E. coli* (Yura et al. 1992; Daniels et al. 1992), *B. subtilis* (Kunst and Devine 1991; Glaser et al. 1993), *Mycoplasma pneumoniae* (Wenzel et al. 1992), and *Mycobacterium leprae* (Honoré et al. 1993). The outcome of these projects will allow us to study genome evolution in terms of entire genomic sequences.

Acknowledgments. The author would like to thank Professor Akira Tsugita for helpful discussions.

References

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980–7984

Anagnostopoulos C, Piggot PJ, Hoch JA (1993) The genetic map of *Bacillus subtilis.* In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington DC, pp 425–461

Bachmann BJ (1990) Linkage map of *Escherichia coli* K-12 edition 8. Microbiol Rev 54:130–197

Daniels DL, Plunkett GIII, Burland V, Blattner FR (1992) Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science 257:771–778

Dayhoff MO, Barker WC, Hunt LT (1983) Establishing homologies in protein sequences. Methods Enzymol 91:524–545

Ferris SD, Whitt GS (1979) Evolution of the differential regulation of duplicated genes after polyploidization. J Mol Evol 12:267–317

Glaser P, Kunst F, Arnaud M, Coudart MP, Gonzales W, Hullo MF, Ionescu M, Lubochinsky B, Marcelino L, Moszer I, Presecan E, Santana M, Schneider E, Schweizer 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

Herdman M (1985) The evolution of bacterial genomes. In: Cavalier-Smith T (ed) The evolution of genome size. John Wiley & Sons, Inc., New York, pp 37–68

Honore N, Bergh S, Chanteau S, Doucet-Populaire F, Eiglmeier K, Garnier T, Georges C, Launois P, Limpaiboon T, Newton S, Niang K, del Portill P, Ramesh GR, Reddi P, Ridel PR, Sittisombut N, Wu-Hunter S, Cole ST (1993) Nucleotide sequence of the first cosmid from *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. Mol Microbiol 7:207–214

Hopwood (1967) Genetic analysis and genome structure in *Streptomyces coelicolor.* Bacteriol Rev 31:373–403

Hori H, Osawa S (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol 4:445–472

Kröger M, Wahl R, Rice P (1993) Compilation of DNA sequences of *Escherichia coli* (update 1993), Nucleic Acids Res 21:2973–3000

Kunisawa T, Otsuka J (1988) Periodic distribution of homologous genes or gene segments on the *Escherichia coli* K-12 genome. Protein Seq Data Anal 1:263–267

Kunst F, Devine K (1991) The project of sequencing the entire *Bacillus subtilis* genome. Res Microbiol 142:905–912

Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. Science 227:1435–1441

O'Brien SJ, ed (1990) Genetic maps: locus maps of complex genomes. Cold Springer Harbor Laboratory Press, New York

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Berlin

Riley M, Solomon L, Zipkas D (1978) Relationship between gene function and gene location in *Escherichia coli.* J Mol Evol 11:47–56

Riley M, Krawiec S (1987) Genome organization. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella typhimurium:* cellular and molecular biology. American Society for Microbiology, Washington DC, pp 967–981

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandill S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular maps of the tomato and potato genomes. Genetics 132:1141–1160

Wallace DC, Morowitz HJ (1973) Genome size and evolution. Chromosoma 40:121–126

Wenzel R, Pirkl E, Herrmann R (1992) Construction of an *EcoRI* restriction map of *Mycoplasma pneumoniae* and localization of selected genes. J Bacteriol 174:7289–7296

Yura T, Mori H, Nagai H, Nagata T, Tshihama A, Fujita N, Isono K, Mizobuchi K, Nakata A (1992) Systematic sequencing of the *Escherichia coli* genome: analysis of the 0–2.4 min region. Nucleic Acids Res 20:3305–3308