# Preferential use of A- and U-rich codons for Mycoplasma capricolum ribosomal proteins S8 and L6

Akira Muto, Yasushi Kawauchi\*, Fumiaki Yamao and Syozo Osawa

Laboratory of Molecular Genetics, Department of Biology, Faculty of Science, Nagoya University, Furo-Cho, Chikusa-Ku, Nagoya 464, Japan

Received 17 September 1984; Accepted 15 October 1984

# ABSTRACT

The nucleotide sequence of the 1.3 kilobase-pair DNA segment, which contains the genes for ribosomal proteins S8 and L6, and a part of L18 of Mycoplasma capricolum, has been determined and compared with the corresponding sequence in Escherichia coli (Cerretti et al., Nucl. Acids Res. 11, 2599, 1983). Identities of the predicted amino acid sequences of S8 and  $\overline{L6}$  between the two organisms are 54% and 42%, respectively. The A + T content of the M. capricolum genes is 71%, which is much higher than that of E. coli (49%). Comparisons of codon usage between the two organisms have revealed that M. capricolum preferentially uses A- and U-rich codons. More than 90% of the codon third positions and 57% of the first positions in M. capricolum is either A or U, whereas E. coli uses A or U for the third and the first positions at a frequency of  $51\overline{2}$  and 36%, respectively. The biased choice of the A- and U-rich codons in this organism has been also observed in the codon replacements for conservative amino acid substitutions between M. capricolum and E. coli. These facts suggest that the codon usage of M. capricolum is strongly influenced by the high A + T content of the genome.

#### INTRODUCTION

The DNA of mycoplasmas is extremely rich in A and T (1,2). It is thus interesting to see how this characteristic feature effects on the structure of the mycoplasma genes. Recently, we have cloned a DNA segment containing a cluster of ribosomal protein genes of <u>Mycoplasma capricolum</u> (3). This paper reports the nucleotide sequence of the DNA segment coding for ribosomal proteins S8 and L6, demonstrating that the codon usage in <u>M. capricolum</u> genes is greatly deviated from that in E. coli.

# MATERIALS AND METHODS

Plasmid DNA was prepared according to Oka <u>et al</u>. (4). DNA sequence determinations were performed by the chain-termination method of Sanger <u>et</u> al. (5) as described by Messing and Vieira (6).

Restriction-endonucleases, T4 DNA-ligase and DNA polymerase I (large fragment) were purchased from Takara-Shuzo Co. Ltd.(Kyoto). The enzyme reaction conditions were those recommended by the comercial supplier.



Fig. 1. (a) Restriction map of pMCB1088. The thick line shows the inserted <u>M. capricolum</u> DNA segment and the thin line represents the vector pBR322 DNA. (b) The strategy for DNA sequence determination of the 1.3 Kb <u>HindIII-fragment from pMCB1088</u>. The arrows indicate the direction of the sequence determinations. (c) The locations of open reading frames.

 $(\alpha - {}^{32}P)$  dATP was purchased from Amersham Japan. The construction of plasmid pMCB1088 containing a DNA segment from <u>M</u>. <u>capricolum</u> ATCC27343(KID) was described elsewhere (3).

# RESULTS

Plasmid pMCB1088 contained a 9 kilobase-pair (Kb) <u>Bgl</u>II-fragment of <u>M</u>. <u>capricolum</u> DNA that had been inserted into the <u>Bam</u>HI-site of pBR322 (3). This DNA segment coded for at least eight ribosomal proteins (3). The restriction map of the insert is shown in Fig. 1(a). Deletion mapping experiments revealed that the 1.3 Kb <u>Hind</u>III-fragment derived from the central part of the insert contained the gene for at least a ribosomal small subunit protein of <u>M</u>. <u>capricolum</u> (MS5; see ref. 3).

The complete nucleotide sequence of the 1.3 Kb (1,251 base-pairs) fragment was determined according to the strategy shown in Fig. 1(b). More than 90% of the sequences was confirmed by sequencing both strands. Computer analyses for potential protein coding sequences revealed the presence of three open reading frames (ORF-1, ORF-2 and ORF-3) on the same DNA strand (Fig. 1(c); see also Fig. 2). The amino acid sequences of the ORF-1 and ORF-2 deduced from the nucleotide sequences were found to have 54% and 42% identities with those of  $\underline{E}$ . <u>coli</u> ribosomal protein S8 (7) and L6 (8),

respectively. The reported N-terminal sequences of proteins S8 and L6 from Bacillus subtilis also showed high homologies with the corresponding sequences of the ORFs (9, 10). Thus, we considered the ORF-1 and ORF-2 as the genes for M. capricolum S8 or rpsH and L6 or rplF. respectively. The ORF-3 could encode 81 amino acid residues, the sequence of which resembled that from the N-terminus of the E. coli ribosomal protein L18 (31% identity). We tentatively identified the ORF-3 as a part of the L18 gene or rplR. The ORFs were interrupted by relatively short (19 and 28 base-pairs) spacers, that did not contain promoter- or terminater-like sequences, suggesting that these genes are a part of the operon for ribosomal protein genes. It is interesting that the order of the genes (rpsH-rplF-rplR) on this fragment was the same as that in the spc-operon of E. coli, where the genes for ten ribosomal proteins and two membrane proteins are clustered as a transcriptional unit (11). This indicates that the order of at least the three genes mentioned above has been conserved between M. capricolum and E. coli.

Since the complete sequence of the <u>E</u>. <u>coli spc</u>-operon was reported (11), we can directly compare the sequences of the ribosomal protein genes between M. <u>capricolum</u> and <u>E</u>. <u>coli</u>. Fig. 2 shows the nucleotide sequences of the 1.3 Kb DNA (mRNA-like strand) and the predicted amino acid sequences aligned with those of <u>E</u>. <u>coli</u>. The <u>M</u>. <u>caprilolum</u> S8 gene was 384 base-pairs long (128 codons including initiation codon), 6 base-pairs shorter than the <u>E</u>. <u>coli</u> gene. The L6 gene of <u>M</u>. <u>capricolum</u> was 540 base-pairs long (180 codons), 9 base-pairs longer than that of <u>E</u>. <u>coli</u>. The spacers between the genes of <u>M</u>. <u>capricolum</u> were longer than those of <u>E</u>. <u>coli</u>.

#### DISCUSSION

The A + T content of <u>M</u>. <u>capricolum</u> genomic DNA is 75% (12), one of the highest in prokaryotes so far known, whereas that of <u>E</u>. <u>coli</u> DNA is about 50%. Thus, the high A + T content of the genome may directly dictate the codon usage of the <u>M</u>. <u>capricolum</u> genes. In fact, the A + T content of the <u>M</u>. <u>capricolum</u> ribosomal protein genes is about 71%, which is much higher than that of <u>E</u>. <u>coli</u> (49%). In Table 1 is compared the codon usage in S8 and L6 genes between <u>M</u>. <u>capricolum</u> and <u>E</u>. <u>coli</u>. The choice of synonymous codons in <u>M</u>. <u>capricolum</u> is clearly different from that in <u>E</u>. <u>coli</u> and strongly biased. For example, UUA(Leu), AUU(IIe) and AGA(Arg) are the most frequently used codons among synonyms in <u>M</u>. <u>capricolum</u>, while they are rarely used in <u>E</u>. <u>coli</u> [CUG(Leu), AUC(IIe) and CGU(Arg) are predominant]. The most outstanding feature of the codon usage in <u>M</u>. <u>capricolum</u> can be seen in a very high

										<u>1</u> S	8 st	art			_						
										Met	Thr	Thr		Asp	Val	Ile	Ala	Asp	Met	Leu	
M.c.	AAGC	TTCA	TGAT	AGAA	AGAG	AGAA	GATT	CAAA	AGT	ATG	ACA	ACA		GAT	GTT	ATT	GCA	GAT	ATG	CTA	0065
E.c.	ATTG	TCAC	CAAT	TGAA	TCAC	GGGA	GGTA	AAGA	CAG	ATG	AGC	ATG	CAA	GAT	CCG	ATC	GCG	GAT	ATG	CTG	
										Met	Serj	Met	Gln	Asp	Pro	Ile	Ala	Asp	Met	Leu	
					<u> </u>	77.77		01-1		1	1	7	·		<del>77.1</del> 7	7.7.21	w-1 f	D	C	<u></u>	
м.	Thr	Arg	lle	Arg	Asn	ALAI	ASD	GIN	Arg	TYT	Leu	Lys	ACT		CTA	Ser	CTT	PTO CCC	Jer	ACC	01 22
п.с. Р.	ACC	AGA	ATT	AGA	AAL	CCTI	AA 1	CAC	CCC	CCC		***	CCT	CCC	CTC	ACC	ATC	CCT	TCC	TCC	0122
E.C.	The	0G1	TIA	Arg	AAC			Cin	419	419	Aen	Ive	419	41 a	Val	Thr	Mot	Pro	Ser	Ser	
	1111	ALS	116	ALS	ASI	250	1	<u>ur</u>			noul	293	ara	ma (	144			110	Jer	561	
	lvs	งล์เ	Lvs	Leu	Glu	Île	Ala	Arg	Ile	Leu	Lvs	Glu	Glu	Glv	Phe	Ile	Ser	Asp	Phe	Thr	
M.c.	AAA	GTA	AAA	TTA	GAA	ATA	GCA	AGA	ATT	TTA	AAA	GAA	GAA	GGA	TTT	ATC	TCA	GAC	ттс	ACT	0182
E.c.	AAG	CTG	AAA	GTG	GCA	ATC	GCC	AAC	GTG	CTG	AAG	GAA	GAA	GGT	TTT	ATT	GAA	GAT	TTT	AAA	
	Lys	Leu	Lys	Val	Ala	Ile	Ala	Asn	Val	Leu	Lys	Glu	Glu	Gly	Phe	Ile	Glu	Asp	Phe	Lys	
								-													
	Val	Glu	Gly	Asp	Val	Lys	Lys	Thr	Ile	Asn	Ile	Glu	Leu	Lys	Tyr		Gln	Gly	Lys	Thr	
M.c.	GTT	GAA	GGT	GAT	GTT	AAA	AAA	ACT	ATT	AAT	ATT	GAA	TTA	AAA	TAC		CAA	GGA	AAA	ACT	0239
E.c.	GTT	GAA	GGC	GAC	ACC	AAG	CCT	GAA	CTG	GAA	CTT	ACT	CTG	AAG	TAT	TTC	CAG	GGC	AAA	GCT	
	Val	GLu	GLY	Asp	Thr	Lys	rro	GIU	Leu	GLU	Leu	Inr	Leu	Lys	Tyr	rne	GIn	GIY	Lys	AIA	
	Aral	Val	T1.	61.	C1 v <sup>1</sup>	Tau	Ivel	1 ve	112	Sar	1551	Pro	<u>C1v</u>	I AU	Arg	Val I	Twr	41.0	Gla	41 a	
M.c.	AGA	GTA	ATT	CAA	GGA	TTA	AAG	AAA	ATT	TCT	AAA	CCA	GGT	TTA	AGA	GTT	TAT	GCA	CAA	GCT	0299
E.c.	GTT	GTT		GAA	AGCI	ATT	CAG	CGT	GTC	AGC	CGC	CCA	GGC	TTG	CGC	TAC	TAT	AAA	CGT	AAA	
	Val	Val		Glu	Ser	Ile	Gln	Arg	Val	Ser	Arg	Pro	Gly	Leu	Arg	Tyr	Tyr	Lys	Arg	Lys	
						'													-		
	Asn	Glu	Ile	Pro	Gln	Val	Leu	Asn	Gly	Leu	Gly	Ile	Ser	Ile	Val	Ser	Thr	Ser	Gln	Gly	
M.c.	AAT	GAA	ATT	CCA	CAA	GTA	TTA	AAC	GGA	TTA	GGT	ATC	TCA	ATT	GTT	TCA	ACA	TCA	CAA	GGA	0359
E.c.	GAT	CAG	CTG	CCC	AAA	GTT	ATG	GCG	GGT	CTG	GGT	ATC	GCA	GTT	GTT	TCT	ACC	TCT	AAA	GGT	
	Asp	Gln	Leu	Pro	Lys	Val	Met	Ala	Gly	Leu	Gly	Ile	Ala	Val	Val	Ser	Thŕ	Ser	Lys	G1y	
	TIC	Mat	The	ic1	5	Twe	A1 2	Arg	17	111	Lan	41 -	<u>C1</u> .	C1	C1	17-1	Tan	41-	Phai	<u> </u>	
Мс	ATA	ATC	ACT	CCT	LyS	LyS	CCT	CGA	CTA	GCT	ASI	CCT	CCT	CCA	CAA	GTT	CTA	GCA	TTC	ATTI	0419
E.c.	GTT	ATG	ACT	GAT	CGT	GCA	000	CGC	CAA	GCT	GGT	CTT	GGT	GGC	GAA	ATT	ATC	TGC	TAC	GTA	0415
2	Val	Met	Thr	ASD	Arg	Ala	Ala	Arg	Gln	Ala	Glv	Leu	Glv	Glv	Glu	Ile	Ile	Cvs	Tvr	Val	
		<u> </u>			Y		11	.6 st	tart												
						1	Met	Ser	Arg	Île	Gly	Asn	Arg	Leu	Leu	Gln	Ile	Pro	Asn	Gly	
M.c.	TGA	TAA	TAG	GAGT	TAA/	T	ATG	TCT	CGT	ATA	GGT	AAT	AGA	TTA	TTA	CAA	ATT	CCA	AAT	GGT	0480
E.c.	GCC	TAA	TCG	GACG	AAAA/	۹-	ATG	TCT	CGT	GTT	GCT	AAA	GCA	CCG	GTC	GTT	GTT	CCT	GCC	GGC	
	Ala						Met	Ser	Arg	Val	Ala	Lys	Ala	Pro	Val	Val	Val	Pro	Ala	Gly	
	130	127 -			79.1		1	a		- <del>-</del> -				1	<u>(1)</u>	1	-	<u> </u>		<u> </u>	
м.	CTT	GIU	Val	Lys	11e	ALA	GIU			Leu	(Val	Int	ATT	Inr	GIY	Ser	Lys	GIY	INT	Leu	05/0
n.c.	CTT	GAA	CTA	AAA	ATC	GUA	COT	CAC	I ARI	ATT	I GIA	ACA	ATC	ALA	COT		AAA	GGA	ACI	CTC	0540
5.C.	Val	Aen	Val	Lve	TIe	Aen	Glv	Gln	Val	111	1	Thr	TIO	Lve		Ive	Aen	C1 v	Clu	TAN	
		1		2,0			•_)		1.02		1	<u> </u>		12,0	(42)	, 2, 0			010	200	
	Ser	Lys	Gln	Phe	Ser	Pro	Leu	Ile	Lys	Ile	]G1u	Val	Glu	Glū	Asn	Lys	Leu	Ile	Thr	Lys	
M.c.	TCA	AAA	I CAA	TTT	TCA	CCT	TTA	ATC	AAA	ATT	GAA	GTT	GAA	i gaa	AAC	AAA	TTA	ATC	ACT	AAA	0600
E.c.	ACT	CGT	ACT	CTC	AAC	GAT	GCT	GTT	GAG	GTT	<b>! AAA</b>	CAT	GCA	GAT	AAT	ACC	CTG	ACC	TTT	GGT	
	Thr	Arg	, Thr	Leu	Asn	Asp	Ala	Val	Glu	Val	Lys	His	Ala	Asp	Asn	Thr	Leu	Thr	Phe	Gly	
	•			~ 1	~	• .		-		~ 1	-		<u>(a)</u>					·			
× .	Arg	Leu	Asn	Glu	Gin	Lys	His	Thr	Lys	Gin	Leu	His	Gly	Thr	Thr	Asn	Ser	Leu	Leu	Gln	
M.C.	AGA	CCT	CAT	GAA	TAC	AAA	CAT	ACA	TCC		CAC	CAC	GGA	ACT	ACT	AAT	TUA	TTA	CTA	CAA	0660
E.C.	Pro	470	Acn	C1 v	Tur	41 a	Acn	C1v	Tro	41.9	Cln	A1 a	1001	The	110	470	41 9	Tau	Lau	Ano	
	110		voh	GLY	1,1	n1a	vah	519	119	714	orn	лца			ייען	TL R	A14	Leu	Leu	<u>nau</u>	
	Gly	Met	Leu	Thr	Gly	Val	Ser	Glu	]G1 y	Phe	Lys	Lys	]G1u	Leu	Gln	lle	Thr	Glv	Val	Glv	
M.c.	GGT	ATG	TTA	ACT	GGA	GTT	AGT	GAA	GGA	TTT	AAA	AAA	GAA	TTA	CAA	ATT	ACT	GGG	GTT	GGG	0720
E.c.	TCA	ATG	GTT	ATC	GGT	GTT	ACC	GAG	GAC	TTC	ACT	AAG	AAG	CTG	CAG	CTG	GTT	GGT	GTA	GGT	
	Ser	Met	Val	Ile	Gly	Val	Thr	Glu	Asp	Phe	<b>j</b> Thr	Lys	Lys	Leu	Gln	Leu	Val	G1y	Val	Gly	
	Tyr	Lys	Ala	Ala	Val	Asn	Gly	Ser	Lys	Leu	Asn	Leu	Ser	Leu	Gly	Tyr	Ser	His	Pro	Val	1
M.c.	TAT	AAA	GCT	GCA	GTT	AAT	GGT	TCT	AAA	TTA	AAT	TTA	AGT	TTA	GGT	TAT	TCA	CAT	ССТ	GTT	0780
E.c.	TAC	CGT	GCA	GCG	GTT	AAA	GGG	AAT	GTG	ATT	AAC	CTG	TCT	CTG	GGT	TTC	TCT	CAT	CCT	GTT	1
	Tyr	Arg	Ala	Ala	Val	Lys	Gly	Asn	Val	Ile	Asn	Leu	Ser	Leu	Gly	Phe	Ser	His	Pro	Val	J

	้ตีมี	Phe	Glu	lle	Pro	Aspi	Gly	Val	Val	Ile	Gln	Ala	Val	Lys	Pro	Thr	Glu	Leu!	Ala	Ile	
M.c.	GAA	TTT	GAA	ATT	CCT	GAT	GGT	GTA	GTG	ATT	CAA	GCA	GTT	AAA	CCA	ACT	GAA	TTA	GCT	ATT	0840
E.c.	GAC	CAT	CAG	CTG	CCT	GCG	GGT	ATC	ACT	GCT	GAA	TGT	CCG	ACT	CAG	ACT	GAA	ATC	GTG	CTG	
	Asp	His	Gln	Leu	Pro	Ala	Gly	Ile	Thr	Ala	Glu	Cys	Pro	Thr	Gln	Thr	Glu	Ile	Val	Leu	
		_																			
	Thr	Gly	Ile	Asp	Lys	Gln	Leu	Val	Gly	Gln	Val	Ala	Ala	Asn	Ile	Arg	Ala	Tyr	Arg	Lys	
M.c.	ACT	GGA	ATC	GAT	AAA	CAA	TTA	GTT	GGT	CAA	GTA	GCA	GCA	AAT	ATT	AGA	GCA	TAT	AGA	AAA	0900
E.c.	AAA	GGC	GCT	GAT	AAG	CAG	GTG	ATC	GGC	CAG	GTT	GCA	GCG	GAT	CTG	CGC	GCC	TAC	CGT	CGT	
	Lys	GIY	ALA	ASP	Lys	GIR	var	TTel	GIY	GIN	var	ALA	ALA	wab	Leu	Arg	ATa	IVE	Arg	Arg	
	Pro	<u>61 11</u>	Pro	Tvr	Lvs	Glv	Lvs	G1v	Tie	Lvs	Tvr	I.vs	Asn	Glu	Thr	TIE	TIE	Arol	Lvs	Glu	l
M.c.	CCT	GAA	CCA	TAT	AAA	GGT	AAA	GGA	ATT	AAA	TAC	AAA	AAT	GAA	ACT	ATT	ATT	AGA	AAA	GAA	0960
E.c.	CCT	GAG	CCT	TAT	AAA	GGC	AAG	GGT	GTT	CGT	TAC	GCC	GAC	GAA	GTC	GTC	CGC	ACC	AAA	GAG	
	Pro	Glu	Pro	Tyr	Lys	Gly	Lys	Gly	Val	Arg	Tyr	Ala	Asp	Glu	Val	Val	Arg	Thr	Lys	Glu	
						180							-		11	18 8	start	:		_	_
	GIy	Lys	Ala	Ala	Gly	Lys									Met	Lys	Phe	Thr	Lys	Thr	
M.c.	GGG	AAA	GCA	GCT	GGT	AAA	TAG	TAC	CAGA	GCTT/	\GGG[	ATTAC	CTAAC	ST	ATG	AAA	TTT	ACT	AAA	ACT	1025
E.c.	GCT	AAG	AAG	AAG	TAA			GGT	ACAG	.T					ATG	GAT		AAG	AAA	TCT	
	VISI	Lys	Lys	LYS											Met	Asp		Lys	Lys	Ser	i
	61i	A1 9	470	Lvei	Aro	Are	His	Phe	Are	Val	Aro	Hiel	LVG	ival.		Glv	Thr	<b>6</b> 72	61.1	470	
M.c.	GAA	GCT	AGA	AAA	CGT	AGA	CAT	TTC	AGA	GTA	AGA	CAT	AAA	GTT	GTT	GGT	ACT	GCT	GAA	AGA	1085
E.c.		GCT	CGT	ATC	CGT	CGT	GCG	ACC	CGC	GCA	CGC	CGC	AAG	CTC	CAG	GAG	CTG	GGC	GCA	ACT	
		Ala	Arg	Ile	Arg	Arg	Ala	Thr	Arg	Ala	Arg	Arg	Lys	Leu	Gln	Glu	Leu	Gly	Ala	Thr	
										•											
	Pro	Arg	Leu	Asn	Val	Phe	Lys	Ser	Asn	Thr	Asn	Phe	Tyr	Ala	Gln		Ile	Ile	Asp	Asp	
M.c.	CCT	AGA	TTA	AAT	GTA	TTT	AAA	TCA	AAT	ACT	AAT	TTC	TAT	GCT	CAA		ATT	ATT	GAT	GAT	1142
E.C.		CGC	CIG	GIG	GIA	CAT	CGT	ACC	Dee	CGT	CAC	ATT	TAC	GCA	CAG	GTA	ATT	GCA	CCG	AAC	
		Arg	Leu	Ivar	Val	Iurs	LAIS	Inr	Pro	ALA	118	TTe	lyr	ALa	GIU	Ivar	LITE .	ALA	rro	ASU	
	Thr	Lvs	Glv	Val	Thr	Leu	Val	Ser	Ala	Ser	Thr	Len	I.vs	Met	Asn	Ten	Tve	Ser	Lve	Ser	
M.c.	ACT	AAA	GGA	GTT	ACA	TTA	GTA	TCT	GCT	TCT	ACA	TTA	I AAA	ATG	GAT	TTA	AAG	AGT	AAA	TCT	1202
E.c.	GGT	TCT	GAA	GTT		CTG	GTA	GCT	GCT	TCT	ACT	GTA	GAA	AAA	GCT	ATC	GCT	GAA	CAA	CTG	
	Gly	Ser	Glu	Val		Leu	Val	Ala	Ala	Ser	Thr	Val	Glu	Lys	Ala	Ule_	Ala	Glu	Gln	Leu	
															-	81					
	Asn	Ile	Gln	Ala	Ala	Glu	Lys	Val	Ala	Glu	Glu	Leu	Thr	Lys	Lys	Ala	I_				
M.c.	AAT	ATT	CAA	GCT	GCT	GAA	AAA	GTT	GCT	GAA	GAA	I TTA	ACT	AAA		GCT	T				1251
E.C.	AAG	Tar	ACC	GGT C1		AAA	GAC	41 -	ALA	ALA	GCT	GIG Val		GGT		GCT					
	Lys	ràt	THE	Lota	nsu	гуз	лэр	AT 9	הגמ	Iura	AL B		, <u> </u>	GLY	Lys	77	i				

Fig. 2. The total nucleotide sequence of the 1.3 Kb <u>HindIII-fragment</u> and deduced amino acid sequences of the ORFs. The sequences are aligned with the corresponding <u>E</u>. <u>coli</u> sequences reported by Cerretti <u>et al</u> (11). M.c.: <u>M. capricolum</u>; E.c.: <u>E</u>. <u>coli</u>.

frequency of the codons ending in either A or U; 91% (280/308) of the <u>M</u>. <u>capricolum</u> codons has A or U at the third position, in contrast to only 51% (155/307) in <u>E</u>. <u>coli</u> (Table 2). The A + U content of the first position of the codons in <u>M</u>. <u>capricolum</u> is also significantly higher (57%;174/308) than that in <u>E</u>. <u>coli</u> (36%;110/307). It is thus evident that the choice of synonymous codon in <u>M</u>. <u>capricolum</u> is biased to the codons much richer in A and U as compared with E. coli.

To see further the preferential use of A- and U-rich condons in  $\underline{M}$ . <u>capricolum</u>, individual codons for the S8 and L6 genes of the two organisms are compared. There exists 143 identical amino acids at the homologous positions between M. capricolum and <u>E. coli</u> S8 and L6 (Fig. 2). Among them,

						Secor	ıd							
		τ	J		C			1	A			G		
		<u>M.c.</u>	<u>E.c.</u> *		<u>M.c.</u>	<u>E.c.</u>		<u>M.c.</u>	<u>E.c.</u>		<u>M.c.</u>	<u>E.c.</u>		
	Phe	4	3	Ser	5	5	Tyr	5	3	Cys	0	1	U	
Пп	Phe	2	3	Ser	0	2	Tyr	3	5	Cys	0	1	C	
ľ	Leu	24	0	Ser	8	1		0	2		1	0	A	
	Leu	0	1	Ser	0	0		1	0	Trp	0	1	G	
	Leu	0	2	Pro	4	7	His	2	3	Arg	1	13	U	
	Leu	0	1	Rro	0	1	His	1	0	Arg	0	5	c	
1	Leu	4	0	Pro	5	1	Gln	15	1	Arg	1	0	A	
	Leu	0	18	Pro	1	4	Gln	0	11	Arg	0	0	G	
	Ile	21	5	Thr	15~	8	Asn	13	2	Ser	.3	0	U	
	Ile	5	12	Thr	0	9	Asn	4	9	Ser	ī	2	C I	
A	Ile	4	0	Thr	6	0	Lys	34	17	Arg	11	0	A	
	Met	5	8	Thr	0	1	Lys	1	11	Arg	0	0	G	
Γ	Val	16	21	Ala	8	9	Asp	5	10	G1v	14	19	π	
	Val	0	5	Ala	Õ	7	Asp	1	6	GIV	_0	-9	l c	l
G	Val	7	3	Ala	11	9	Glu	21	12	Glv	13	ó	Ă	l
1	Val	1	5	Ala	0	9	Glu	0	5	G1v	3	1	G	
1			-		-	-		•	2	1 1		-	1	

Table 1. Codon usage in S8 and L6 genes of M. capricolum and E. coli

Taken from Cerretti et al. (11) M.c. : M. capricolum ; E.c. : E. coli

the corresponding codons for 46 amino acids are identical, while the rest 97 codons are different (synonymous codon replacement). All of these silent nucleotide substitutions in the 97 synonymous codons are listed in Table 3. Here, the preferential use of A- and U-rich codons in M. capricolum can clearly be seen. Sixty-six codons having G or C at the third position in E. coli are replaced by synonymous codons ending in A or U with M. capricolum.

Table 2.	Nucleotide composition at three different positions of codons in
	S8 and L6 genes of M. capricolum and E. coli

	F	irst	S	econd	T	hird	Total	
	<u>M.c.</u>	<u>E.c.</u>	<u>M.c.</u>	<u>E.c.</u>	<u>M.c.</u>	<u>E.c.</u>	<u>M.c.</u>	<u>E.c.</u>
U	51	26	93	87	116	111	260	224
С	34	67	63	73	17	77	114	217
A	123	84	105	95	164	44	392	223
G	100	130	47	52	11	75	158	257
A + U	174	110	198	182	280	155	625	447
(%)	(56.5)	(35.8)	(64.3)	(59.3)	(90.9)	(50.5)	(70.6)	(48.5)

M.c. : M. capricolum ; E.c. : E. coli

	<u>M.c.</u>	<u>E.c.</u>	(AU) <sup>a</sup>	No.b	a x b		<u>M.c.</u>	<u>E.c.</u>	(AU) <sup>a</sup>	No. <sup>b</sup>	a x b
Ala	GCA	GCG	(+1)	3	3	Lys	AAA	AAG	(+1)	8	8
	GCA	GCC	(+1)	2	2	-			-		
	GCU	GCG	(+1)	1	1	Phe	ບບບ	UUC	(+1)	1	1
	GCU	GCA	( 0)	1	0		UUC	UUU	(-1)	1	1
Asn	AUU	AAC	(+1)	2	2	Pro	CCA	CCC	(+1)	1	1
	AAC	AAU	(-1)	1	-1		CCA	CCU	$\dot{(0)}$	2	Ō
			· -/	-			CCG	CCU	(-1)	ī	-1
Asp	GATI	GAC	(+1)	1	1				· -/	-	-
	GAC	GAU	(-1)	ĵ	-1	Ser	UCU	UCC	(+1)	1	1
			· -/	-	-		UCU	AGC	(+1)	1	1
Aro	AG4	CCC	(+2)	2	4			UCU		4	ň
<sup> 8</sup>	ACA	CCT	(+1)	2	2		ACTI		(0)	1	0
1	CCA	000	(+1)	1	1		AGO	1100	(0)	1	ñ
	UGA	000	(11)	Ŧ	T		AGO	000	( 0)	Ŧ	U
Gln	CAA	CAG	(+1)	5	5	Thr	ACA	ACG	(+1)	1	1
1			•				ACA	ACC	(+1)	1	1
Glu	GAA	GAG	(+1)	3	3		ACU	ACC	(+1)	2	2
			、-/	-	-			•	、 <i>_/</i>	-	-
G1y	GGA	GGC	(+1)	4	4	Tyr	UAU	UAC	(+1)	2	2
Í	GGA	GGU	( 0)	7	0		UAC	UAU	(-1)	ĩ	-1
	GGU	GGC	(+1)	5	5				,		-
	GGU	GGG	(+1)	ĩ	1	Val	GUA	GUC	(+1)	1	1
	GGG	GGU	(-1)	2	-2		GUA	GUII	(0)	3	ō
1			· -/	-	-		GUU	GUA	ì	2	õ
Ile	AUU	AAC	(+1)	3	3				/	-	-
	AUA	AUC	(+1)	2	2						
	AUC	AUU	(-1)	1	-1						_
Leu	UUA	CUG	(+2)	9	18		То	tal		97	74 <sup>C</sup>
	IIIIA	IIIC	(+1)	í	1						
I	CUA	CUG	(+1)	2	2						
	JUN	000	(,1)	-	-						

Table 3. Synonymous nucleotide substitutions in S8 and L6 genes between M. capricolum and E. coli

M.c. : M. capricolum ; E.c. : E. coli

a : A + U gains in M. capricolum as compared with E. coli

b : Number of occurrence

c : Total A + U gains in 97 synonymously substituted codons (291 nucleotide residues) in M. capricolum as compared with E. coli

A striking example is that eight AAG condons for Lys in <u>E</u>. <u>coli</u> are substituted by AAA in M. capricolum.

Fifty conservative amino acid replacements (i.e., Lys/Arg, Ser/Thr, Leu/Ile/Val, etc.) can be seen at the homologous positions of the S8 and L6 proteins between the two organisms (Fig. 2). In Table 4 are collected all of these amino acid substitutions with their codon replacements. Most of the conservative amino acid substitutions take place so that the <u>M. capricolum</u> maintains much higher A + U content in the codons than <u>E. coli</u>. For example, seven Arg with codon CGU or CGC in <u>E. coli</u> are replaced at the corresponding

0.1	N			Number of <sup>b</sup>	h
Substitution	M. capricolum	<u>E. <u>coli</u></u>	(AU-gain)	occurrence	a x b
Lys/Arg	Lys(AAA)	Arg(CGU)	(+2)	6	12
	Lys(AAA)	Arg(CGC)	(+3)	1	3
Leu/Ile	Ile(AUU)	Leu(CUG)	(+2)	6	12
	Ile(AUU)	Leu(CUU)	(+1)	1	1
	Leu(UUA)	Ile(AUC)	(+1)	ĩ	1
	Leu(IIIIA)	Tle(AUU)	ĊŐ	3	0
	Leu(CUA)	Ile(AUC)	( 0)	1	Ō
Leu/Val	Leu (IIIIA)	Val (GUC)	(+2)	1	2
		Val (GUG)	(+2)	2	4
	Leu(UAA)		(+1)	1	1
	Val(GUA)	Leu(CUG)	(+1)	ī	1
Ile/Val	Tle(AIIII)	Val (GUU)	(+1)	4	4
	Ile(AUU)	Val(GUC)	(+2)	2	4
	Ile(AIIII)	Val (GUA)	(+1)	1	1
		Val (GIG)	(+2)	1	2
	Ile(AUA)	Val (GIII)	(+1)	2	2
	Tie (AUC)	Val (GUU)	(0)	1	ō
	Val (GUA)		( ů)	1	ŏ
1	Val (GIII)	Tle(AUC)	( Ó)	1	õ
	Val(GUU)	Ile(AUU)	(-1)	ī	-1
Ser/Thr	Ser(AGU)	Thr(ACC)	(+1)	2	2
	Ser(IICA)	Thr(ACU)	ĊŌ	1	0
	Thr (ACA)	Ser(AGC)	(+1)	1	1
Ala/Glv	Ala(GCU)	G1v(GGU)	(0)	1	0
	G1v(GGU)	Ala(GCU)	ĊŐ	1	0
	Gly(GGG)	Ala(GCU)	(-1)	ī	-1
Asn/Gln	Asn(AAC)	Gln(CAG)	(+1)	1	1
	Gln(CAA)	Asn (AAC)	( 0)	ī	Ō
Asp/Glu	Glu(GAA)	Asp(GAC)	(+1)	2	2
	Glu (GAA)	Asp(GAU)	( 0)	1	0
	Tota	al		50	54 <sup>C</sup>
1					

Table 4. Conservative amino acid substitutions in protein S8 and L6 between M. capricolum and E. coli with their codon replacements

a : A + U gains in M. capricolum codons as compared with E. coli

b : Number of occurrence

c : Total A + U gains in 50 conservatively substituted codons (150 nucleotide residues) in <u>M</u>. <u>capricolum</u> as compared with <u>E</u>. <u>coli</u>

positions by Lys with AAA in <u>M</u>. <u>capricolum</u>. Thus, the choice of codons in <u>M</u>. <u>capricolum</u> seems to occur to discriminate against G and C and to use A and U wherever possible.

The genomic DNAs of M. capricolum as well as all the known species

belonging to the class Mollicute (Mycoplasma, Acholeplasma, Ureaplasma and Spiroplasma) are rich in A and T (61-78%), suggesting that the Mollicute has evolved with a constraint keeping the high A + T contents in their genomes. The obvious preference of A and T in <u>M</u>. <u>capricolum</u> is clearly seen not only in the codons, but also in the spacers. In fact, the spacers between rRNA genes of <u>M</u>. <u>capricolum</u> are extremely rich in A and T (13). Also, the A + U content of rRNAs from mycoplasmas is highest in prokaryotes (14). Thus, in <u>M</u>. <u>capricolum</u>, the constraint for the preferential use of A and T has operated at the DNA level as a selection force upon the codon choice as well as the construction of other parts of the genome.

#### ACKNOWLEDGEMENTS

This work was supported by grants from the Ministry of Education, Science and Culture, Japan (Nos. 57121003, 57121009 and 58220012) and the Naito Science Foundation Research Grant (81-106).

\* On leave from Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University, Hiroshima 734, Japan.

#### REFERENCES

- 1. Maniloff, J. and Morowitz, H. J. (1972) Bacteriol. Rev. 36, 263-290.
- 2. Razin, S. (1978) Microbial. Rev. 42, 414-470.
- 3. Kawauchi, Y., Muto, A., Yamao, F. and Osawa, S. (1984) Mol. Gen. Genet. in press.
- 4. Oka, A., Sugisaki, H. and Takanami, M. (1981) J. Mol. Biol. 147, 217-226.
- Sanger, F., Wicken, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 6. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
- Allen, G. and Wittmann-Liebold, B. (1978) Hoppe-Seyler's Z. Physiol. Chem. 359, 1509-1525.
- Chen, R., Afrsten, U. and Chen-Schmeisser (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 531-535.
- 9. Higo, K., Otaka, E. and Osawa, S. (1982) Mol. Gen. Genet. 185, 239-244. 10. Higo, K., Itoh, T., Kumazaki, T. and Osawa, S. (1980) in Genetics and
- Evolution of RNA polymerase, tRNA and Ribosomes. Osawa, S., Ozeki, H., Uchida, H. and Yura, T. eds. pp.655-666, Tokyo Univ. Press, Tokyo.
- 11. Cerretti, D. P., Dean, D., Davis, G. R., Bedwell, D. M. and Nomura, M. (1983) Nucl. Acids Res. 11, 2599-2616.
- 12. Neimerk, H. C. (1970) J. Gen. Microbiol. 63, 249-263.
- 13. Sawada, M., Muto, A., Iwami, M., Yamao, F. and Osawa, S. (1984) Mol. Gen. Genet. in press.
- 14. Ryan, J. L. and Morowitz, H. J. (1969) Proc. Natl. Acad. Sci. USA 63, 1282-1289.