

FUNCTIONAL ROLE OF BACTERIOPHAGE TRANSFER RNAs: CODON USAGE ANALYSIS OF GENOMIC SEQUENCES STORED IN THE GENBANK/EMBL/DDBJ DATABASES

T Kunisawa

Dept of Applied Biological Sciences, Science University of Tokyo, Noda 278-8510, Japan

Email: kunisawa@rs.noda.tus.ac.jp

ABSTRACT

Complete genomic sequence data are stored in the public GenBank/EMBL/DDBJ databases so that any investigator can make use of the data. This report describes a comparative analysis of codon usage that is impossible without such a public and open data system. A limited number of bacteriophages harbor their own transfer RNAs. Based on a comparison between T4 phage-encoded tRNA species and the relative cellular amounts of host Escherichia coli tRNAs, it is hypothesized that T4 tRNAs could serve to supplement host isoacceptor tRNA species that are present in minor amounts and thus enhance the translational efficiency of phage proteins. When compared to their respective host bacteria, the codon usage data of bacteriophages D3, ϕ C31, HP1, D29 and 933W all show an increased frequency of synonymous codons or amino acids that correspond to phage tRNA species, suggesting their supplemental role in the efficient production of phage proteins. The data-analysis presents an example in which the availability of an open and fully accessible database system would allow one to obtain comprehensive insights into a fundamental problem in molecular biology.

Keywords: Codon usage, Bacteriophage, Transfer RNA, Translation, Comparative genomics

1 INTRODUCTION

Recognizing the importance of establishing an international database system that can be freely accessible by any researcher, molecular biologists have created a public database system (GenBank/EMBL/DDBJ) through the collaborative efforts of many workers. In this system, a researcher who has experimentally determined a nucleic acid sequence of DNA sends the data to the international database so that other researchers can make use of his data. At the same time the contributing researcher can compare his sequence data with those already stored in the database, making further development of his research possible. Effective use of such an open data-sharing system could lead the researcher to novel insights into fundamental problems that might not otherwise be gained from the researcher's own original data alone. In this report the author would like to introduce such an attempt.

Bacteriophages utilize their host's cellular translational machinery in the synthesis of their own proteins. So far there are no phage genomes known to encode ribosomal proteins, translation factors, or aminoacyl-tRNA synthetases. The presence of transfer RNA-specifying genes in the phage T4 genome, however, was reported about thirty years ago (McClain, Guthrie, & Barrell, 1972; Broida & Abelson, 1985). Since then, the functional role of the phage-encoded tRNAs has attracted some interest. One suggestion is that the presence of tRNA genes in phage genomes may provide a suitable phage integration sequence by homologous recombination (reviewed in Cheethman & Katz, 1995). Another is that phage tRNAs may boost cellular tRNA populations to enhance translational

efficiency (Kan, Kano-Sueoka, & Sueoka, 1968; Daniel, Sarid, & Littauer, 1970; Scherberg & Weiss, 1972). The latter idea was put forward following the discovery of phage-encoded tRNAs but this has not been verified. In 1992, the advantage of a phage having its own tRNAs was examined in the light of the synonymous codon-usage data of phage T4 and its host *Escherichia coli*; the phage tRNA species could serve to supplement the host's isoacceptor tRNA species that are present in minor cellular amounts and recognize those synonymous codons that are rare in the host genes but are frequent in the phage genes (Kunisawa, 1992). At that time, T4 presented a unique example in which complete data on phage-encoded tRNAs were available.

Currently, however, more than 90 phage genomic sequences are stored in the GenBank/EMBL/DDBJ databases (GenBank, n.d.; EMBL, n.d.; DDBJ, n.d.). In this report phage genomes encoding tRNA genes are searched for in the databases, and their significance in translation is examined using codon usage analysis. This is to obtain more comprehensive insights into the host-phage translation system. Through this project, the author would like to emphasize the importance of open and fully accessible database system.

2 DATA AND METHODS

All the genomic sequence information used in this study was retrieved with the use of WWW from the major public GenBank/EMBL/DDBJ databases. Genomic sequence data are available from any of these databases. *Escherichia coli* K12, database accession No. U00096; phage T4, X51522; *Mycobacterium tuberculosis*, AF123456, phage D29, AF022214; *Pseudomonas aeruginosa*, AE004091; phage D3, AF077308; *Streptomyces coelicolor* AL645882; phage ϕ C31, AJ006589; *Haemophilus influenzae*, L42023; phage HP1, U24159; *E. coli* O157:H7, AE005174; phage 933W, AF125520; phage VT2-Sa, AP000363.

Besides the major DNA sequence databases, T4 genomic databases are available on the Web (Evergreen Bacteriophage Group, n.d.; Suzuki, Arisaka, Kutter, Thomas, Tsugita, & Kunisawa, 1998). In the latter database, codon usage data of each T4 gene are stored. A complete list of entirely sequenced phage genomes is available from the homepage of NCBI (NCBI, n.d.).

Computer programs were written to extract protein coding regions from complete genomic sequences in the database to analyze codon usage patterns.

3 PHAGE-ENCODED TRANSFER RNAS

Table 1 summarizes results of a database search for bacteriophage genomes that encode tRNA genes. All the phage genomes and their host genomes listed here have been completely sequenced. Gene orders of phage tRNAs are shown in the direction of 5' to 3' in Table 1. Two general points can be made. Phage tRNA genes almost always exist in a cluster on the genome and they are not sufficient for the translation of the 61 sense codons. Coliphage T4 harbors eight tRNA species, the anticodons of which are also shown in the table within the square brackets (Broida & Abelson, 1985). The *Streptomyces* phage ϕ C31 (Smith, Burns, Wilson & Gregory, 1999) harbors a Thr-tRNA with the anticodon ACG. The genome of the *Pseudomonas* phage D3 encodes four tRNA genes for Met, Gly, Asn, and Thr (Kropinski & Sibbald, 1999). From sequence similarity, the *Haemophilus influenzae* phage HP1's genome (Esposito, Fitzmaurice, Benjamin, Goodman, Waldman & Scocca, 1996) is reported to code for Lys- and Leu-tRNA genes, although the latter Leu-tRNA is a pseudogene (Kropinski & Sibbald, 1999). The virulent mycobacteriophage D29 genome encodes five tRNA

genes for Asn, Trp, Gln, Glu, and Tyr in a cluster (Ford, Sarkis, Belanger, Hendrix & Hatfull, 1998), while its close relative L5 lacks the last two tRNAs in the D29 cluster (Hatfull & Sarkis, 1993). Phage 933W and its close relative VT2-Sa, both of which infect enterohemorrhagic *Escherichia coli* O157:H7 (Perna, Plunkett III, Burland, Mau, Glasner, Rose, et al., 2001), harbor one tRNA for Ile and two isoacceptor tRNAs for Arg (Plunkett III, Rose, Durfee & Blattner, 1999; Miyamoto, Nakai, Yajima, Fujibayashi, Higuchi, Sato et al., 1999).

Table 1. Bacteriophage tRNA genes found in completely sequenced genomes.

phage	host	tRNA gene cluster [anticodon]
T4	<i>Escherichia coli</i>	Gln - Leu - Gly - Pro - Ser - Thr - Ile - segB - Arg [UUG] [UAA] [UCC] [UGG] [UGA] [UGU] [CAU] [UCU]
D29	<i>Mycobacterium tuberculosis</i>	Asn - Trp - Gln - Glu - Tyr [GUU] [CCA] [CUG] [CUC] [GUA]
L5	<i>Mycobacterium tuberculosis</i>	Asn - Trp - Gln [GUU] [CCA] [CUG]
D3	<i>Pseudomonas aeruginosa</i>	Met - ORF - ORF - Gly - Asn - Thr [CAU] [UCC] [GUU] [UGU]
phiC31	<i>Streptomyces coelicolor</i>	Thr [ACG]
HP1	<i>Haemophilus influenzae</i>	Lys - Leu (pseudogene) [UUU] [UAA]
933W/ VT2-Sa	enterohemorrhagic <i>Escherichia coli</i>	Ile - Arg - Arg [CAU] [UCG] [UCU]

In addition to these, the coliphage T5, for which the complete genomic sequence is not known, is reported to harbor at least 24 tRNAs (McCordquodale & Warner, 1988). Furthermore, the *Vibrio eltol* phage e4 genome is known to encode 5 tRNAs in a cluster (Chattopadhyay & Ghosh, 1988). However, their sequence data are not available.

4 TRNA USAGE AND SYNONYMOUS CODON USAGE OF T4

In Table 2, the anticodon species of tRNAs from *E. coli* and from phage T4 are listed, together with codon usage data. The cellular amounts of host *E. coli* tRNAs relative to that of Leu-tRNA with anticodon CAG are indicated within the square brackets (Ikemura, 1985; 1992). The gene copy number of respective tRNAs is also tabulated. As a first approximation, the relative cellular amount of tRNA is proportional to the gene copy number (see Fig. 3, Ikemura 1992). It is interesting to note that in Table 2 all the eight anticodon species of T4 tRNAs can be found in the host *E. coli*'s tRNAs, and the phage tRNAs do not carry any novel anticodon species. In *E. coli* and in most other bacteria, there are multiple isoacceptor tRNAs for the eight types of amino acids, Gln, Leu, Gly, Pro, Ser, Thr, Ile and Arg that can be charged by T4 tRNAs. It is remarkable to observe that phage tRNAs do not correspond in every case to the major *E. coli* tRNA isoacceptors present in larger cellular amounts. Rather, they correspond to the minor isoacceptors. An extreme case is seen with Ile; the cellular amount of the minor isoacceptor having the CAU anticodon (the base C at the first anticodon position may be modified to cytidine; for details of anticodon modification, see Osawa (1995)) is only 0.05 compared to that of the major isoacceptors with the GAU anticodon, and the phage's Ile-tRNA corresponds to the minor one. Thus, phage tRNAs seem to supplement host tRNAs that are present in small amounts.

Table 2. Codon usage and tRNA availability

	codon usage			tRNA			codon usage		tRNA	
	codon	cumulative frequency (per thousand)		anti- codon	gene copy number [content]		<i>M. tuber- culosis</i>	D29	<i>M. tuber- culosis</i>	D29
Arg	<u>CGU</u>	20.9	19.0	ACG	4 [0.9]		8.4	6.5	1	
	<u>CGC</u>	22.0	5.8				28.4	20.6		
	CGA	3.6	5.5				7.2	9.6	1	
	CGG	5.4	1.2	CCG	1 [minor]		24.6	21.2		
	AGA	2.1	10.2	UCU	2 [minor]	1	1.3	0.9	1	
	AGG	1.2	1.9	CCU	1 [minor]		3.2	5.3	1	
		55.2	43.6				73.0	64.2		
Leu	UUA	13.9	27.6	UAA	1 [0.25]	1	1.6	0.0	1	
	UUG	13.7	10.8	CAA	1 [0.2]		17.9	5.6	1	
	CUU	11.0	18.3				5.4	4.6		
	CUC	11.1	4.1	GAG	1 [0.3]		17.3	26.4	1	
	CUA	3.9	6.9	UAG	1 [minor]		4.8	1.6	1	
	<u>CUG</u>	52.6	5.7	CAG	4 [1.0]		50.4	44.1	1	
		106.2	73.5				97.4	82.3		
Ser	UCU	8.5	24.6				2.2	3.4		
	UCC	8.6	3.6	GGA	2 [0.25]		11.5	11.6	1	
	UCA	7.2	18.1	UGA	1 [0.25]	1	3.5	3.0	1	
	UCG	8.9	3.7	CGA	1 [0.05]		19.3	17.9	1	
	AGU	8.8	10.7				3.5	2.4		
	AGC	16.1	5.6	GCU	1 [0.25]		14.5	12.1	1	
		58.1	66.3				54.5	50.4		
Ala	<u>GCU</u>	15.3	31.2				10.9	17.5		
	GCC	25.5	5.3	GGC	2 [0.3]		59.9	36.9	1	
	<u>GCA</u>	20.1	19.6	UGC	3 [1.0]		12.8	12.1	1	
	<u>GCG</u>	33.6	6.3				48.5	33.9	1	
		94.6	62.4				132.1	100.4		
Gly	<u>GGU</u>	24.7	27.5				18.9	15.4		
	<u>GGC</u>	29.6	8.1	GCC	4 [1.1]		51.4	42.6	1	
	GGA	8.0	19.5	UCC	1 [0.15]	1	9.9	8.5	1	
	GGG	11.1	3.9	CCC	1 [0.1]		19.3	15.8	1	
		73.5	59.1				99.5	82.3		
Pro	CCU	7.0	14.4				3.4	5.5		
	CCC	5.5	1.2	GGG	1 [minor]		17.0	15.0	1	
	CCA	8.4	13.7	UGG	1 [0.3]	1	6.1	4.5	1	
	<u>CCG</u>	23.2	5.0	CGG	1 [0.3]		31.3	28.9	1	
	44.1	34.2				57.8	53.9			
Thr	<u>ACU</u>	9.0	27.8				3.7	4.3		
	<u>ACC</u>	23.4	6.3	GGU	2 [0.8]		35.3	31.0	1	
	ACA	7.1	17.1	UGU	1 [0.1]	1	4.5	3.9	1	
	ACG	14.4	5.3	CGU	1 [0.1]		15.6	19.3	1	
	53.9	56.4				59.1	58.5			
Val	<u>GUU</u>	18.3	31.5				8.0	6.5		
	GUC	15.3	5.5	GAC	2 [0.4]		32.7	39.6	1	
	<u>GUA</u>	10.9	20.0	UAC	5 [1.05]		4.7	3.7	1	
	<u>GUG</u>	26.4	6.1				40.0	23.7	1	
	70.8	63.2				85.4	73.5			
Ile	AUU	30.3	50.9				6.4	2.8		
	<u>AUC</u>	25.1	11.5	GAU	3 [1.0]		33.9	46.7	1	
	AUA	4.4	12.2	CAU	2 [0.05]	1	2.2	0.6		
	59.9	74.6				42.5	50.1			

Table 2. (Continued)

	codon	codon usage		tRNA		codon usage		tRNA		
		cumulative frequency (per thousand)		anti- codon	gene copy number [content]	cumulative frequency		gene copy number		
		<i>E. coli</i>	T4		<i>E. coli</i>	T4	<i>M. tuber- culosis</i>	D29	<i>M. tuber- culosis</i>	D29
Asn	AAU	17.7	42.8				5.3	1.9		
	<u>AAC</u>	21.7	15.0	GUU	4 [0.6]		20.0	30.6	1	1
		39.4	57.8				25.3	32.5		
Asp	GAU	32.1	47.2				15.7	10.7		
	GAC	19.1	14.4	GUC	3 [0.8]		42.1	53.4	1	
		51.2	61.6				57.8	64.1		
Cys	UGU	5.2	7.3				2.2	1.3		
	UGC	6.5	3.7	GCA	1 [minor]		6.6	7.4	1	
		11.7	11.0				8.8	8.7		
Gln	CAA	15.3	21.7	UUG	2 [0.3]	1	8.1	3.8	1	
	<u>CAG</u>	28.8	11.1	CUG	2 [0.4]		22.7	31.4	1	1
		44.2	32.8				30.8	35.2		
Glu	<u>GAA</u>	39.4	60.0	UUC	4 [0.9]		16.1	14.7	1	
	GAG	17.8	10.8				30.5	51.8	1	1
		57.3	70.8				46.6	66.5		
His	CAU	12.9	13.5				6.4	2.6		
	CAC	9.7	3.7	GUG	1 [0.4]		15.8	19.0	1	
		22.7	17.3				22.2	21.6		
Lys	<u>AAA</u>	33.6	64.1	UUU	6 [1.0]		5.3	3.4	1	
	AAG	10.3	17.5				15.0	41.9	1	
		43.9	81.6				20.3	45.3		
Phe	UUU	22.3	33.3	GAA	2 [0.35]		6.2	1.2		
	<u>UUC</u>	16.6	11.1				23.3	33.9	1	
		38.9	44.4				29.5	35.1		
Tyr	UAU	16.2	33.8				6.1	1.5		
	<u>UAC</u>	12.2	9.7	GUA	3 [0.5]		14.7	27.5	1	1
		28.4	43.5				20.8	29.0		
Met	AUG	27.9	26.7	CAU	5 [0.8]		18.4	21.2	3	
Trp	UGG	15.2	14.1	CCA	1 [0.3]		14.7	20.3	1	1

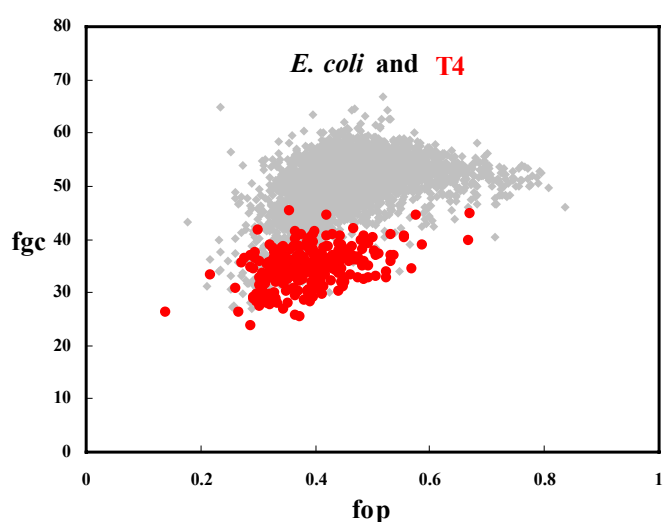
The total number of codons is 1363498 (4290 protein-coding genes) for *E. coli*, 53183 (271 genes) for T4, 1335687 (3924 genes) for *M. tuberculosis*, and 14789 (77 genes) for D29. Sums over synonymous codons are shown in italic. Optimal codons of *E. coli* are underlined..

In Table 2 the codon usage of T4's protein-coding sequences are compared with that of its host *E. coli*'s genes (Blattner, Plunkett III, Bloch, Perna, Burland, Riley et al., 1997). The occurrences per thousand codons are shown in the table. The amino acid frequencies are shown in italic for an easy comparison of amino acid usage. It is worth noting that the frequency of synonymous codons read by phage tRNAs is always higher in phage genes than in host genes. For example, the frequency of two Arg codons, AGA and AGG, both of which are read by the phage tRNA with anticodon UCU (wobble rules, see Osawa (1995)), is 13.1 per thousand codons for phage genes, while only 3.3 for host genes. Similarly, the frequency of AUA codon for Ile in phage genes is about three times as high as that in host genes. From these observations, we can hypothesize that phage tRNAs could serve to supplement host tRNAs present in minor amounts and thus enhance the efficiency of translation of phage genes (Kunisawa, 1992).

Based on the synonymous codon usage patterns of *E. coli* genes, Ikemura (1985; 1992) has indicated

a correlation between heavily used synonymous codons and isoacceptor tRNA species that are abundant in the cell. Thus, the synonymous codon choice is constrained by the tRNA availability and seems to be designed to optimize the translational efficiency of *E. coli* genes, in particular its highly expressed genes. Therefore, foreign genes such as phage genes are not translated efficiently unless their synonymous codon choice is similar to that of the highly expressed genes of *E. coli*. The synonymous codon usage pattern of T4 genes, however, is clearly different from that of *E. coli*, as can be seen in Figure 1 (Kunisawa, Kanaya & Kutter, 1998). In this figure, for each of the 4290 *E.*

Figure 1. Codon usage patterns of *E. coli* and phage T4



coli (shown in gray) or 271 T4 (in red) protein-coding genes, the average G+C content, fgc, is plotted against the fraction, fop, of optimal codons, which are thought to be translationally optimal (Ikemura, 1985; 1992). The fraction of optimal codons in each gene is known to serve as a surrogate measure of the gene expression level. The optimal codon species defined by Ikemura (1985; 1992) are underlined in Table 2. As can be seen in Figure 1, the G+C contents of T4 genes are much lower than those of *E. coli* genes. Table 2 shows a tendency for T4 genes to frequently use codons ending in U or A owing to the low G+C content, while *E. coli* genes tend to use codons that are recognized by the abundant tRNA species and thus often use codons ending in G or C rather than codons ending in U or A. Figure 1 also indicates that the majority of T4 genes exhibit low fop values centered around 0.4, suggesting their lower translational efficiency. Thus, it is reasonable to speculate that T4 supplies its own tRNAs to the host tRNA populations that are present in minor amounts for more efficient production of phage proteins.

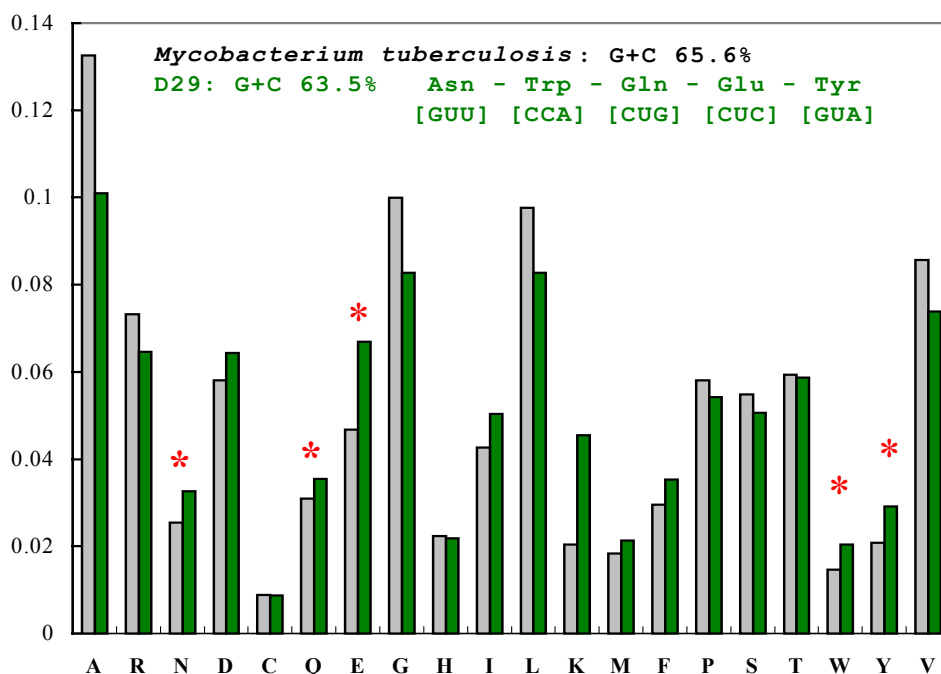
Deletions of all eight T4 encoded tRNAs have shown a decreased burst size and a decreased rate of synthesis for several proteins when grown on laboratory host strains, suggesting the existence of a selective pressure to maintain the T4 tRNA genes (Wilson, 1973). In accordance with this reasoning,

a T4 mutant that grows normally on most laboratory strains but is unable to grow on a specific strain of *E. coli* (CT439) has been isolated (Guthrie & McClain, 1973). On the other hand, the tRNA gene repertoire appears to be variable among phages T2, T6 and RB69, which are closely related to T4 (Moen, Seidman, & McClain, 1978). The genome of RB69 has recently been completely sequenced (Karam & Krisch, 2002), and comparisons of codon usage among these T4 related phages would be very interesting future studies.

5 MYCOBACTERIOPHAGE TRNAS

As listed in Table 1, mycobacteriophage D29 harbors 5 tRNAs for Asn, Trp, Gln, Glu, Tyr (Ford et al., 1998). The presence of Asn-, Trp- and Tyr-tRNAs is somewhat surprising, as there is a single anticodon species for each of these D29 tRNAs, in contrast to the T4 tRNAs case. Therefore, the supply of isoacceptor tRNAs seems improbable in the case of D29 tRNAs. The host bacterium, *M. tuberculosis*, encodes tRNA genes in a single copy on the genome except for the Met-tRNA genes (Cole, Brosch, Parkhill, Garnier, Churcher, Harris et al., 1998). This is in sharp contrast to tRNA genes in *E. coli* where the tRNA-gene copy number varies from one to six for Lys-tRNA (see Table 2), and the cellular content of tRNA molecules is approximately proportional to its gene copy number (Ikemura, 1992). Therefore, constraints on synonymous codon usage from the tRNA availability may not be significant in *M. tuberculosis*, and the genomic G+C content seems a principal factor in determining synonymous codon preference. However, the host and phage show almost identically high genomic G+C contents, 65.6% for the host and 63.5% for phage D29. Thus, synonymous codon usage patterns are also similar, as can be seen in Table 2.

Figure 2. Comparison of amino acid usage between *M. tuberculosis* (shown by gray bars) and its phage D29 (green bars).



Despite the similarity in synonymous codon choice, Figure 2 shows that the frequency of amino acid occurrences slightly increases in D29 for all five amino acid species, Asn, Gln, Glu, Tyr and Trp, which correspond to D29 tRNAs. In this figure, amino acid usage is compared between the host (shown in gray) and phage (in green), where a red asterisk indicates if phage tRNA exists. The fraction of Asn, for instance, increases to 32.5 per thousand amino acids in D29 from 25.3 in *M. tuberculosis*. In the *E. coli* - T4 system, corresponding increases in amino acid usage cannot be recognized; conversely, the fraction of amino acid usage for Arg, Lys, Gly, Pro, and Gln decreases in T4 (see Table 2). These observations suggest that D29 tRNAs may serve to tune host tRNA populations to the amino acid usage of phage proteins (Kunisawa, 2000). However, this tuning seems imperfect, since the frequency of Lys increases from 20.3 in the host to 45.3 in D29 but there is no phage tRNAs for this kind of amino acid. Furthermore, the phage tRNAs may not be required for growth, since another mycobacteriophage L5, which is closely related to D29, lacks the last two tRNAs in the D29 gene cluster.

A comparison of codon usage between the mycobacteriophage D29 and *Mycobacterium leprae* was also carried out, since D29 may also infect *M. leprae*, which has a lower genomic G+C content of 57.8%. As with *M. tuberculosis*, our analysis showed an almost identical codon (amino acid) usage in *M. leprae*. The host range of mycobacteriophage D29 is broad and this phage is known to grow on both fast- and slow-growing mycobacteria. Although available data are currently confined to *M. tuberculosis* and *M. leprae*, future comparisons of codon usage data among a broad range of host bacteria would provide a new horizon for phage evolution.

6 CODON USAGE OF D3, ϕ C31, HP1 AND 933W/VT2-Sa

As listed in Table 1, the *Pseudomonas* phage D3 harbors four tRNAs for Met, Gly, Asn, and Thr (Kropinski & Sibbald, 1999). The genomic G+C content of the host bacterium *Pseudomonas aeruginosa* (Stover, Pahm, Erwin, Mizoguchi, Warren, Hickey et al., 2000) is 66.6%, while the phage genome is 57.8%. Although the difference in G+C content is not large, codon usage shows that the occurrences of either synonymous codons or amino acids corresponding to the phage's tRNAs are greater in the phage genes compared to the host genes; e.g. for Met it is 20.2 (per thousand codons) in the host and 24.2 in the phage, for Gly it is 14.1 and 28.6, for Asn it is 24.8 and 33.8, and for Thr it is 7.2 and 17.9 (see Table 3), being consistent with the hypothesis of the translational role of phage tRNAs (Kropinski & Sibbald, 1999). Another example is seen with *Streptomyces* phage ϕ C31 (Smith, Burns, Wilson, & Gregory, 1999) harboring a Thr-tRNA with the CGU anticodon, which reads the ACG codon. In the host *S. coelicolor* (Bentley, Chater, Cerdeno-Tarraga, Challis, Thomson, James et al., 2002) genes, the Thr codon ACC is most heavily used (32.8 occurrences per thousand codons, see Table 3) and ACG codon is less frequently used (6.3). In the phage ϕ C31 genes, the frequency of the ACG codon increases to 11.8. Similarly, in the case of the Lys-tRNA found in the *Haemophilus influenzae* phage HP1 (Esposito, Fitzmaurice, Benjamin, Goodman, Waldman, & Scocca, 1995), the occurrence of Lys per thousand amino acids slightly increases to 77.4 in HP1 from 63.1 in its host genes (Fleischmann, Adams, White, Clayton, Kirkness, Kerlavage et al., 1995). Furthermore, an increased frequency of codons that are read by phage 933W (Plunkett III, Rose, Durfee, & Blattner, 1999) Ile-tRNA or Arg-tRNA with anticodon UCU can also be seen in Table 3. In the Shiga toxin II converting phage VT2-Sa, which is closely related to phage 933W and encodes an identical tRNA gene cluster, it has been reported that phage encoded Arg-tRNAs recognize rare codons in the host but these rare codons occur in the toxin genes at a high frequency (Kanjo & Inokuchi, 1999). All these examples suggest a strategy in which phage tRNAs may adjust tRNA populations in infected cells to produce the efficient production of phage proteins.

Table 3. Codon usage of tRNA-harboring phages and their host bacteria.

	<i>P. aeruginosa</i>	D3	<i>S. coelicolor</i>	phiC31	<i>H. influenzae</i>	HP1	O157	933W	
G+C (%)	66.6	57.8	72.1	63.6	38.2	40	50.4	49.4	
genes	5565	93	7759	55	1717	41	5283	80	
codons	1865096	16839	2566997	12489	521683	9923	1596900	18379	
Arg	CGU	7.92	8.43	5.38	13.21	23.82	17.23	20.24	19.1
	CGC	49.2	28.62	38.98	27.3	10.11	15.02	20.86	15.07
	CGA	2.37	8.31	2.44	4.72	5.28	6.35	3.82	6.64
	CGG	14.07	13.24	32.18	21.06	1.21	1.51	6.11	8.87
	AGA	0.48	4.81	0.75	0.64	3.71	5.04	2.85	9.9
	AGG	2.01	8.85	3.61	1.2	0.51	1.21	1.77	6.04
Leu	UUA	0.28	1.72	0.06	0.08	49.93	38.09	13.89	9.74
	UUG	8.74	9.26	2.37	6.57	18.26	17.64	13.03	6.04
	CUU	3.09	14.37	1.48	25.3	20.11	13.81	11.4	14.53
	CUC	27.77	15.97	36.59	14.65	5.27	4.23	10.51	8.43
	CUA	1.4	6.12	0.31	2.32	6.74	9.17	3.89	3.1
	CUG	82.93	38.72	61.37	27.22	4.46	5.54	51.1	33.79
Ser	UCU	0.83	7.78	0.6	4.16	16.68	9.98	8.7	8.54
	UCC	12.06	11.58	20.26	12.97	4.39	5.74	8.93	7.89
	UCA	0.58	5.58	1.02	2.96	12.58	11.69	8.09	13.77
	UCG	13.03	12.95	13.87	20.66	4.12	5.34	8.83	6.75
	AGU	2.62	5.17	1.49	3.2	11.97	16.43	9.36	11.21
	AGC	25.93	19.72	12.33	11.85	8.53	12.29	16.05	12.13
Ala	GCU	4.81	22.21	2.83	19.78	21.19	14.01	15.43	20.46
	GCC	67.69	38.6	78.58	46.92	10.9	17.33	25.16	20.68
	GCA	4.85	21.79	5.33	9.77	32.86	32.55	20.64	28.08
	GCG	38.84	29.04	50.19	47.16	16.92	16.63	32.35	24.97
Gly	GGU	8.27	12.53	9.25	18.18	28.69	18.34	24.21	22.58
	GGC	61.91	32.96	61.55	44.68	19.54	24.39	28.14	18.5
	GGA	4.16	13.96	7.12	6.97	10.9	8.87	8.81	16.32
	GGG	9.9	14.67	18.46	17.3	7.17	10.48	11.7	15.18
Pro	CCU	2.12	8.73	1.44	5.69	12.37	8.36	7.22	9.03
	CCC	13.04	8.43	25.47	13.13	2.9	4.33	5.55	5.17
	CCA	2.17	7.78	1.25	1.12	16.83	11.79	8.39	8.98
	CCG	33.29	19.66	33.68	33.79	5.02	7.16	22.43	20.19
Thr	ACU	1.66	8.25	1.11	8.01	16.08	12.5	9.09	10.45
	ACC	32.77	20.73	39.69	17.86	11.33	15.42	22.84	15.23
	ACA	0.81	6.12	1.53	2.08	15.72	17.23	8.04	16.54
	ACG	6.34	11.76	19.07	39.71	8.74	10.68	14.94	14.75
Val	GUU	2.72	16.57	1.36	17.22	20.67	14.31	18.15	20.73
	GUC	28.8	20.73	47.18	29.87	6.56	7.26	14.84	9.85
	GUA	3.99	8.25	2.63	3.92	18.79	15.92	10.91	11.26
	GUG	33.42	18.88	35.39	29.55	20.94	21.97	26.15	23.07
Ile	AUU	2.86	13.54	0.59	14.25	50.54	43.13	29.75	22.85
	AUC	37.8	33.32	27.31	19.7	14.27	14.71	23.94	19.37
	AUA	0.95	4.45	0.65	0.96	5.99	6.15	5.44	10.17
Asn	AAU	3.73	10.39	0.68	3.84	36.52	36.68	19.12	22.96
	AAC	22.59	23.4	16.22	24.9	12.22	21.47	21.69	21.38
Asp	GAU	10.46	24.05	2.88	7.05	42.18	41.32	32.82	34.99
	GAC	42.6	33.14	58.22	61.01	7.51	13.5	19.2	23.02
Cys	UGU	1	3.09	0.69	2.4	6.79	4.94	5.24	5.55
	UGC	9	11.58	7.01	7.53	3.52	4.03	6.38	6.58
Gln	CAA	6.25	10.69	1.26	3.04	38.72	37.69	14.77	10.01
	CAG	36.21	29.16	24.97	20.74	7.54	8.97	29.31	36.73
Glu	GAA	23.39	27.44	8.41	26.18	54.14	56.64	39.31	40.7
	GAG	37.31	41.87	48.38	33.55	10.46	11.39	18.63	27.1
His	CAU	6.27	6.47	1.61	2.96	13.39	9.88	12.81	10.01
	CAC	15.38	12.71	21.72	14.17	7.08	9.27	9.39	6.91
Lys	AAA	3.58	10.63	0.97	1.28	56.27	65.61	33.98	38.74
	AAG	25.01	33.02	19.44	42.76	6.81	11.79	10.95	21.49
Phe	UUU	1.72	6.89	0.42	3.04	32.24	31.14	22.27	19.48
	UUC	33.72	25.18	26.04	27.94	12.27	12.5	15.91	12.51
Tyr	UAU	5.26	9.32	0.95	6.01	24.33	24.09	16.57	16.87
	UAC	20.04	18.53	19.54	21.62	6.86	10.18	12.27	11.86
Met	AUG	20.2	24.23	15.74	17.78	24.04	21.97	27.24	27.75
Trp	UGG	14.82	18.53	15.09	19.86	11.24	13	15.29	15.07
Tern	UAA	0.28	0.77	0.13	0.88	2.46	2.82	1.96	1.85
	UAG	0.34	0.65	0.53	0.8	0.47	0.4	0.27	0.27
	UGA	2.36	4.1	2.35	2.96	0.36	0.91	1.07	2.23

The phage 933W's Arg-tRNA with the UCG anticodon is very unusual; a search of the tRNA database (Sprinzl, Horn, Brown, Ioudovitch, & Steinberg, 1998) shows no other examples of the UCG anticodon in bacteria, although this anticodon can be found in animals. Therefore, it is likely that this unusual anticodon species was laterally transferred into the phage genome during evolution. As mentioned earlier, phage-encoded tRNA genes tend to appear in a cluster on the phage's genome. Was the gene cluster horizontally transferred? Or, was the cluster formed during evolution by disrupting unnecessary tRNA genes and retaining necessary ones? The evolutionary origins and variations of phage tRNAs are interesting issues.

7 CONCLUSION

The presently available data on the cellular concentrations of tRNAs in *E. coli*, T4 phage-encoded tRNA species, codon usage of other phages harboring their own tRNAs, suggest a strategy for phage survival. Phage tRNAs supplement host tRNA populations to enable efficient production of phage proteins. The author would like to emphasize the importance of open and fully accessible databases.

8 REFERENCES

- Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., & Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417 (6885), 141-147.
- Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., & Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331), 1453-1462.
- Broida, J., & Abelson, J. (1985) Sequence organization and control of transcription in the bacteriophage T4 tRNA region. *J. Mol. Biol.* 185(3), 545-563.
- Chattopadhyay, S., & Ghosh, R.K. (1988) Localization of transfer RNA genes on the physical map of *Vibrio eltor* phage e4 genome. *Virology* 162(2), 337-345.
- Cheetham, B.F., & Katz, M.E. (1995) A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.* 18(2), 201-208.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., & Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393(6685), 537-44.

Daniel, V., Sarid, S., & Littauer, U.Z. (1970) Bacteriophage induced transfer RNA in *Escherichia coli*. New transfer RNA molecules are synthesized on the bacteriophage genome. *Science* 167(926), 1682-1688.

DDBJ (n.d.) Homepage of the DNA Data Bank of Japan. Available from: <http://www.ddbj.nig.ac.jp>.

EMBL (n.d.) Homepage of the EMBL Nucleotide Sequence Database. Available from: <http://www.ebi.ac.uk/embl/index.html>

Esposito, D., Fitzmaurice, W.P., Benjamin, R.C., Goodman, S.D., Waldman, A.S., & Scocca, J.J. (1996) The complete nucleotide sequence of bacteriophage HP1 DNA. *Nucl. Acids Res.* 24(12), 2360-2368.

Evergreen Bacteriophage Group (n.d.) Homepage of the Evergreen Bacteriophage Group. Available from: <http://www.evergreen.edu/phage/home.html>.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., & Kerlavage, A.R. Bult, C.J., Tomb, J., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., & Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223), 496-512.

Ford, M.E., Sarkis, G.J., Belanger, A.E., Hendrix, R.W., & Hatfull, G.F. (1998) Genome structure of mycobacteriophage D29: implications for phage evolution. *J. Mol. Biol.* 279(1), 143-164.

GenBank (n.d.) Homepage of GenBank. Available from: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

Guthrie, C., & McClain, W.H. (1973) Conditionally lethal mutants of bacteriophage T4 defective in production of a transfer RNA. *J. Mol. Biol.* 81(2), 137-155.

Hatfull G.F., & Sarkis, G.J. (1993) DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol. Microbiol.* 7(3), 395-405.

Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2(1), 13-34.

Ikemura, T. (1992) Correlation between codon usage and tRNA content in microorganisms. In Hatfield, D.L., Lee, B.J. & Pirtle, R.M. (Eds.) *Transfer RNA in protein synthesis*, Boca Raton, Ann Arbor, London, Tokyo: CRC Press

Kan, J., Kano-Sueoka, T., & Sueoka, N. (1968) Characterization of leucine transfer ribonucleic acid in *Escherichia coli* following infection with bacteriophage T2. *J. Biol. Chem.* 243(21), 5584-5590.

Kanjo, N. & Inokuchi H. (1999) Genes for tRNA^{Arg} located in the upstream region of the Shiga toxin II operon in enterohemorrhagic *Escherichia coli* O157:H7. *DNA Res.* 681), 71-73.

- Karam, J.D., & Krisch, H.M. (2002) Homepage of the T4-like Phage Genome Project. Available from: <http://www.mcl.tulane.edu/sif/t4genome.html>.
- Kropinski, A.M., & Sibbald M.J. (1999) Transfer RNA genes and their significance to codon usage in the *Pseudomonas aeruginosa* lambdoid bacteriophage D3. *Can. J. Microbiol.* 45(9), 791-796.
- Kunisawa, T. (1992) Synonymous codon preferences in bacteriophage T4: a distinctive use of transfer RNAs from T4 and from its host *Escherichia coli*. *J. Theor. Biol.* 159(3), 287-298.
- Kunisawa, T., Kanaya, S., & Kutter, E. (1998) Comparison of synonymous codon distribution patterns of bacteriophage and hst genomes. *DNA Res.* 5(6), 319-326.
- Kunisawa, T. (2000) Functional role of mycobacteriophage transfer RNAs. *J. Theor. Biol.* 205(1), 167-170.
- McClain, W.H., Guthrie, C., & Barrell, B.G. (1972) Eight transfer RNAs induced by infection of *Escherichia coli* with bacteriophage T4. *Proc. Natl. Acad. Sci. USA* 69(12), 3703-3707.
- McCordquodale D.J., & Warner, H.R. (1988) Bacteriophage T5 and related phages. pp.439-475 In Calendar, R. (Eds) *The Bacteriophages*. New York: Plenum Press
- Miyamoto, H., Nakai, W., Yajima, N., Fujibayashi, A., Higuchi, T., Sato, K., & Matsushiro, A. (1999) Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions. *DNA Res.* 6(4), 234-240.
- Moen, T.L., Seidman, J.G., & McClain, W.H. (1978) A catalogue of transfer RNA-like molecules synthesized following infection of *Escherichia coli* by T-even bacteriophages. *J Biol Chem.* 253(21), 7910-7917.
- NCBI (n.d.) Homepage of the National Center for Biotechnology Information. Available from: <http://www.ncbi.nlm.nih.gov/>
- Osawa, S. (1995) *Evolution of the genetic code*, London: Oxford University Press
- Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamouisis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., & Blattner, F.R. (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409(6819), 529-533.
- Plunkett III, G., Rose, D.J., Durfee, T.J., & Blattner, F.R. (1999) Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* 181(6), 1767-1778.
- Scherberg, N.H., & Weiss, S.B. (1972) T4 transfer RNAs: codon recognition and translational properties. *Proc. Natl. Acad. Sci. USA.* 69(5), 1114-1148.
- Smith, M.C.M., Burns, R.N., Wilson, S.E., & Gregory, M.A. (1999) The complete genome sequence of the *Streptomyces* temperate phage ϕ C31: evolutionary relationships to other viruses. *Nucl. Acids*

Res. 27(10), 2145-2155.

Sprinzi M., Horn C., Brown M., Ioudovitch A., & Steinberg S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* 26(1) 148-153. The online-database is available from: <http://www.uni-bayreuth.de/departments/biochemie/sprinzi/trna/index.html>.

Suzuki, H., Arisaka, F., Kutter, E., Thomas, E., Tsugita, A., & Kunisawa, T. (1998) *Res. Commun. Biochem. Cell Mol Biol.* 2(3), 309-322. The online-database is available from: <http://www.rs.noda.tus.ac.jp/~kunisawa>.

Stover, C.K., Pahm, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., & Paulsen, I.T. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 406, 959-964.

Wilson, J.H. (1973) Function of the bacteriophage T4 transfer RNAs. *J. Mol. Biol.* 74(4), 753-757.