Aeromonas phages encode tRNAs for their overused codons

Ramanandan Prabhakaran
and Shivapriya Chithambaram

Department of Biology,
University of Ottawa,
Ottawa, Ontario, K1N6N5, Canada
E-mail: rprab028@uottawa.ca
E-mail: schit057@uottawa.ca

Xuhua Xia*

Department of Biology,
Center for Advanced Research in Environmental Genomics,
Ottawa Institute of Systems Biology,
University of Ottawa,
Ottawa, Ontario, K1N6N5, Canada
E-mail: Xuhua.Xia@uottawa.ca
*Corresponding author

Abstract: The GC-rich bacterial species, *Aeromonas salmonicida,* is parasitised by both GC-rich phages (*Aeromonas* phages- phiAS7 and vB_AsaM-56) and GC-poor phages (*Aeromonas* phages – 25, 31, 44RR2.8t, 65, Aes508, phiAS4 and phiAS5). Both the GC-rich *Aeromonas* phage phiAS7 and *Aeromonas* phage vB_AsaM-56 have nearly identical codon usage bias as their host. While all the remaining seven GC-poor *Aeromonas* phages differ dramatically in codon usage from their GC-rich host. Here, we investigated whether tRNA encoded in the genome of *Aeromonas* phages facilitate the translation of phage proteins. We found that tRNAs encoded in the phage genome correspond to synonymous codons overused in the phage genes but not in the host genes.

Keywords: *Aeromonas* phages; dsDNA phage; codon usage; tRNA; tRNA gain or loss events; selection; adaptation.

Reference to this paper should be made as follows: Prabhakaran, R., Chithambaram, S. and Xia, X. (2014) '*Aeromonas* phages encode tRNAs for their overused codons', *Int. J. Computational Biology and Drug Design*, Vol. 7, Nos. 2/3, pp.168–182.

Biographical notes: Ramanandan Prabhakaran received his Bachelors degree in Bioinformatics from VIT University, School of Bio Sciences and Technology, Vellore, Tamil Nadu, India in 2008. Currently, he is an MSc student in the Department of Biology at University of Ottawa. His research interests include studying codon adaptation of phages to their host, bacterial isoelectric point profiles, functional genomics and bioinformatics.

Shivapriya Chithambaram received her Bachelors degree in Bioinformatics from VIT University, School of Bio Sciences and Technology, Vellore, Tamil Nadu, India in 2008. Currently, she is an MSc student in the Department of Biology at University of Ottawa. Her research interests include studying codon adaptation of phages to their host, comparative genomics and bioinformatics.

Xuhua Xia received his PhD from the University of Western Ontario, Canada. Currently, he is a Full Professor in Bioinformatics and Molecular Evolution in the Department of Biology at University of Ottawa. His research interests include bioinformatics, molecular evolution and phylogenetics, comparative genomics, host-parasite interactions at the molecular level, microbial genomic evolution and development of powerful computational tools.

This paper is a revised and expanded version of a paper entitled 'Aeromonas phages encode tRNAs for their overused codons' presented at *International Conference on Intelligent Biology and Medicine (ICIBM 2013)*, Nashville, TN, USA, August 11–13, 2013.

1 Introduction

Differential preference and usage of synonymous codons has been reported in a wide range of species (Grantham et al., 1980). Numerous studies have been carried out to understand the factors shaping codon usage in different organisms. Selection and mutation are proposed to be the universal factors driving codon usage choices. Selection for increased translational efficiency (Ikemura, 1981; Robinson et al., 1984; Sorensen et al., 1989), translational accuracy (Akashi, 1994; Eyre-Walker, 1996) and energetically optimal codon-anticodon pairing (Grosjean and Fiers, 1982; Grosjean et al., 1978) appears to be shaping codon usage bias in organisms. A positive correlation between tRNA abundance and codon usage was first demonstrated in *Escherichia coli* (Ikemura, 1981). Even in *Saccharomyces cerevisiae* a similar correlation between tRNA abundance and usage of corresponding codons was shown (Bennetzen and Hall, 1982; Ikemura, 1982). Experimental evidences suggest that highly expressed genes tend to experience a greater degree of codon usage bias than poorly expressed genes which further established the influence of selection on codon usage bias (Bennetzen and Hall, 1982; Ikemura, 1985; Sharp and Devine, 1989). Theoretical models have been developed to explain the role of selection at the levels of transcription (Xia, 1996) and translation (Bulmer, 1987; Xia, 1998, 2008) affects the codon usage preferences. A number of codon usage indices, including RSCU (Sharp et al., 1986), Nc (Sun et al., 2013; Wright, 1990), and CAI (Sharp and Li, 1987; Xia, 2007) have been proposed and improved to facilitate the study of factors affecting codon usage.

The effect of mutation on codon usage bias has been demonstrated through GC-biased mutations (Bernardi, 1986; Muto and Osawa, 1987; Sueoka, 1988), methylation mediated mutations (Beletskii and Bhagwat, 1996; Xia, 2005) and strand asymmetry in organisms (Lobry, 1996; Lobry and Sueoka, 2002; Marin and Xia, 2008). Mutation mediated by DNA methylation has been proposed to be an important factor in shaping codon usage bias among coronaviruses (Woo et al., 2007). Mutation bias was invoked to explain the deviation in codon usage among some of the prokaryotes with

extreme high GC or AT content (Ohama et al., 1990). In general, it is known that mutational and translational selection pressures are the factors accountable for codon usage bias.

Viruses that parasitise bacterium are termed as bacteriophages (hereafter we will refer to them as phages in short).When a phage infects a bacterial cell, it is advantageous to have a codon usage bias concordant with the bacteria, as this would expedite the time consuming and expensive translation process. Accordingly, phage species have been shown to exhibit codon adaptation to their host translation machinery (Carbone, 2008). A comparative codon usage study among different viruses infecting a broad range of hosts spanning from bacteria to humans illustrated that phages displayed the highest degree of concordance both in terms of codon usage and GC content with respect to their hosts (Bahir et al., 2009).

While codon adaptation may lead to the benefit of more efficient and accurate translation, there are several factors associated with phages that may act against codon adaptation. First, phages typically have high mutation rate which would prevent them from reaching an optimal state of codon adaptation. Second, phages and their hosts may experience different mutation spectrum leading to different compositional bias and consequently discordant codon usage bias.

Several recent studies have suggested a role of phage-encoded tRNA in codon adaptation, i.e., phage-encoded tRNA may enhance the translation of phage proteins (Kunisawa, 1992; Sahu et al., 2004) and expand the host range of phage species (Bailly-Bechet et al., 2007; Limor-Waisberg et al., 2011). A conceptually similar codon adaptation has been recently documented in HIV-1, which has many A-ending codons with few cognate tRNA in the human cell. Empirical evidence strongly suggests that tRNA species decoding A-ending codons are enriched when late HIV-1 genes are translated (van Weringh et al., 2011). This tRNA enrichment has also been reported in vaccinia and influenza A viruses (Pavon-Eternod et al., 2013).

Aeromonas phages present a unique case for studying codon adaptation mediated by tRNA encoded by the phage genome. *Aeromonas* phages are double-stranded DNA (dsDNA) tailed phages parasitising the bacterial host *Aeromonas salmonicida*. They belong to two families, Podoviridae (Kim et al., 2012) and Myoviridae (Ackermann et al., 1985), classified by their tail morphology. Myoviridae have a long and contractile tail while Podoviridae have a very short tails. Most of the analysed myoviruses have longer genome than podoviruses. The GC-rich bacterial species, *A. salmonicida*, is parasitised by both GC-rich and GC-poor *Aeromonas* phages (Table 1). While GC-rich *Aeromonas* phages phiAS7 and vB_AsaM-56 have nearly identical codon usage bias as their GC-rich host, the remaining seven GC-poor *Aeromonas* phages differ dramatically in codon usage from their GC-rich host. Since the seven GC-poor *Aeromonas* phages encode a set of their own tRNAs, it is natural for one to ask if these phage-encoded tRNAs decode codons that are overused in the phage, especially those with few cognate host tRNAs.

A GC-rich host and a GC-poor phage typically would have dramatically different codon usage bias. For a given R-ending codon family (where R stands for either A or G), the GC-rich host may overuse G-ending codons and may have many tRNAs decoding G-ending codons, whereas the GC-poor phage may overuse A-ending codons which would be difficult to find their decoding tRNAs in the host cytoplasm. It would therefore be beneficial if the phage encodes its own tRNAs decoding these A-ending codons by the phage genes, i.e., the phage tRNA should have a wobble U instead of a wobble C if A-ending codons are overused in the phage genes but not in the host genes.

This argument can also be applied to the R-ending codons within a four-fold codon family. For example, GGA and GGG codons for glycine are translated by $tRNA^{Gly/UCC}$ and $tRNA^{Gly/CCC}$. If GGA is overused in the phage but not in the host, then we expect the phage tRNA for glycine should have a wobble U instead of a wobble C.

We cannot generate the same prediction concerning Y-ending codons. While R-ending codons are frequently translated by tRNAs with a wobble U or wobble C, Y-ending codons are always translated by tRNA with a wobble G but not by tRNA with a wobble nucleotide A. A wobble A in a tRNA is always modified to inosine. This avoidance of a wobble A in tRNA has been explained by structural modelling, i.e., a tRNA with a wobble A, once moved to the P-site, will interfere with the entry/pairing of a tRNA at the A-site (Lim, 1994). Thus, for a Y-ending codon family, even if the host overuses the C-ending codon and the phage overuses the U-ending codon, we would not predict that the phage should carry a tRNA with a wobble A to facilitate the decoding of the overused U-ending codons, simply because a tRNA with a wobble A is so deleterious that it is not observed unmodified in nature. Here, we investigate the codon usage of the GC-rich bacterial host and its phages and evaluate the prediction concerning the wobble nucleotide of tRNAs encoded in the *Aeromonas* phage genomes.

Among the nine phages parasitising *A. salmonicida*, two are GC-rich (*Aeromonas* phage phiAS7 and *Aeromonas* phage vB_AsaM-56) and the remaining seven are GC-poor phages (*Aeromonas* phage 25, 31, 44RR2.8t, 65, Aes508, phiAS4 and phiAS5). We report that the seven *Aeromonas* phages, being GC-poor (AT-rich), have codon usage dramatically different from their GC-rich host. These GC-poor phage genomes encode 11 to 24 tRNAs, with a maximum of 11 for R-ending codons, and these tRNAs for R-ending codons correspond to A-ending codons that are overused in the phage but not in the host.

2 Materials and methods

2.1 Sequence selection

The complete genome sequences of *A. salmonicida* host and nine *Aeromonas* phages belonging to two main phage families of dsDNA (see Table 1) were downloaded from GenBank (http://ncbi.nlm.nih.gov) on 2nd October 2013. For all phages, their host names were extracted from the '/HOST' tag, under 'FEATURES' header from their respective GenBank files using their unique accession numbers. All the protein-coding sequences (CDS) of phages and their host were examined for being full length and possessing proper start and stop codons. In order to avoid bias in the samples, we excluded the CDS shorter than 300 base pairs in length from codon usage analysis.

2.2 Relative synonymous codon usage

To study the overall codon usage heterogeneity in the host and phage genomes, we used a normalised index relative synonymous codon usage (RSCU) (Sharp et al., 1986). RSCU is defined as the ratio of the observed frequency of codons to the expected frequency if all the synonymous codons for those amino acids are used equally. RSCU values greater than 1.0 indicate that the corresponding codons are used more frequently than expected,

while RSCU values less than 1.0 would mean that such codons are underused. In this study we computed RSCU values using the codon usage functionality implemented in the software DAMBE (Xia, 2013).

Table 1 Basic genome features of the *Aeromonas* host and their phages

Name	*Accession No.*	*Length*	*No. CDS*	*GC%*
Aeromonas salmonicida (host)	NC_009348	4702,402	3373	58.51
Aeromonas phage 25 (M)	NC_008208	161,475	242	41.04
Aeromonas phage 31 (M)	NC_007022	172,963	247	43.91
Aeromonas phage 44RR2.8t (M)	NC_005135	173,591	252	43.88
Aeromonas phage 65 (M)	NC_015251	235,229	437	37.20
Aeromonas phage Aes508 (M)	NC_019543	160,646	230	41.23
Aeromonas phage phiAS4 (M)	NC_014635	163,875	271	41.30
Aeromonas phage phiAS5 (M)	NC_014636	225,268	343	43.00
Aeromonas phage vB_AsaM-56 (M)	NC_019527	43,551	83	55.42
Aeromonas phage phiAS7 (P)	NC_019528	41,572	51	56.92

*M stands for Myoviridae phage family and P stands for Podoviridae phage family.

2.3 tRNA dataset

As tRNA abundance data is not available for the species examined in this analysis, we used tRNA gene copy number as a proxy (Duret, 2000; Kanaya et al., 1999; Percudani et al., 1997). We obtained the tRNA gene copy number information for *Aeromonas* phages and their hosts from tRNAscan-SE (www.genetics.wustl.edu/eddy/tRNAscan-SE) (Lowe and Eddy, 1997) and Genomic tRNA database (http://gtrnadb.ucsc.edu/) (Chan and Lowe, 2009) respectively.

2.4 Phylogenetic analysis

We extracted the CDS of DNA ligase genes from 22 phage genomes (including the outgroup species *Pseudomonas* phage vB_Pae-TbilisiM32) and translated them into amino acid sequences. We first aligned the amino acid sequences and then aligned the CDS sequences against the aligned amino acid sequences by using DAMBE (Grantham et al., 1980; Xia, 2013).

We used both distance-based and maximum likelihood (ML) methods for phylogenetic reconstruction. For the distance-based reconstruction, we used the simultaneously estimated distance based on the TN93 model (Tamura and Nei, 1993), i.e., MLCompositeTN93 in DAMBE, and the neighbour-joining method (Saitou and Nei, 1987). We have chosen the *Pseudomonas* phage vB_Pae-TbilisiM32 as outgroup in this analysis. We used 1000 bootstrap iterations to measure the confidence of the tree topology. All the above-mentioned analyses were carried out using the comprehensive bioinformatics tool DAMBE (Xia, 2013).

We have also performed phylogenetic reconstruction with the maximum likelihood method based on K80, F84 and TN93 substitution models. The resulting topologies are identical to that from the distance-based method. We have also checked the sufficiency of

the substitution models by likelihood ratio test and information-theoretic indices by using DAMBE.

3 Results and discussion

3.1 The GC-rich and GC-poor phages differ dramatically in their codon usage relative to their host

The bacterial host *A. salmonicida* and *Aeromonas* phage phiAS7 are both GC-rich and exhibit similar codon usage, with their RSCU values showing a positive and highly significant correlation (Figure 1(A)). In contrast, the GC-poor *Aeromonas* phage 65 has dramatically different codon usage from that of the host, with its RSCU values showing a highly significant negative correlation (Figure 1(B)). Similarly, all other GC-poor *Aeromonas* phages also maintain an extremely poor correlation with the host. A conventional mutationist explanation for the pattern in Figure 1(B) is that the host and the phage differ in codon usage because of differential mutation spectra experienced by the host and the phage, partially reflected in their differences in nucleotide composition. This difference in mutation bias between the host and the phage counteracts against codon adaptation of the phage to the host. However, the fact that the GC-poor *Aeromonas* phages encode a set of tRNA genes in their genome suggests an additional dimension of the system, i.e., the phage genome may encode tRNA that decodes specifically the synonymous codons that are overused in the phage but not in the host.

What phage would benefit from carrying its own tRNAs? Two factors may be major contributors. The first is the difference in nucleotide composition between the phage and the host and the second is when the host tRNA pools is uncertain i.e., when a phage has a broad range of host such as KVP40 (Sau et al., 2007). This hypothesis leads to a testable prediction that the number of tRNAs carried by the phage should increase with the differences between the host and phage GC% (the greater the difference in GC%, the greater the difference in codon usage bias). The result for this prediction is strongly substantiated by data from the nine *Aeromonas* phages (Figure 2). The general trend is a positive relationship, i.e., more phage tRNA is associated with greater differences in GC% between the host and the phages.

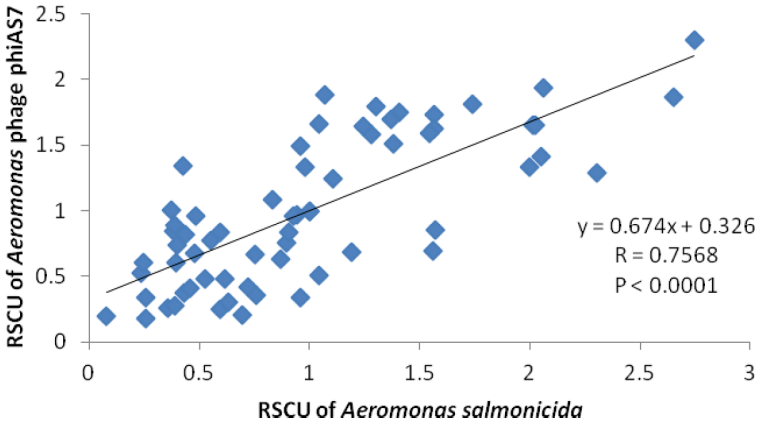
3.2 Aeromonas phages encode tRNA for their overused codons

The GC-rich *Aeromonas* phage phiAS7 and phage vB_AsaM-56 does not encode any tRNA genes. Given their similarity in codon usage to their host, one would expect little necessity or selection pressure for the phage to maintain its own tRNA genes. In contrast, the GC-poor *Aeromonas* phages, with their codon usage dramatically different from that of their host (Figure 1(B)), could benefit from having their own tRNA genes decoding synonymous codons that are overused in the phages but not in the host (and consequently with few host cognate tRNAs). In particular, for R-ending codons, the GC-rich host invariably overuses G-ending codons whereas GC-poor *Aeromonas* phages tend to overuse A-ending codons (Table 2 and Figure 3).

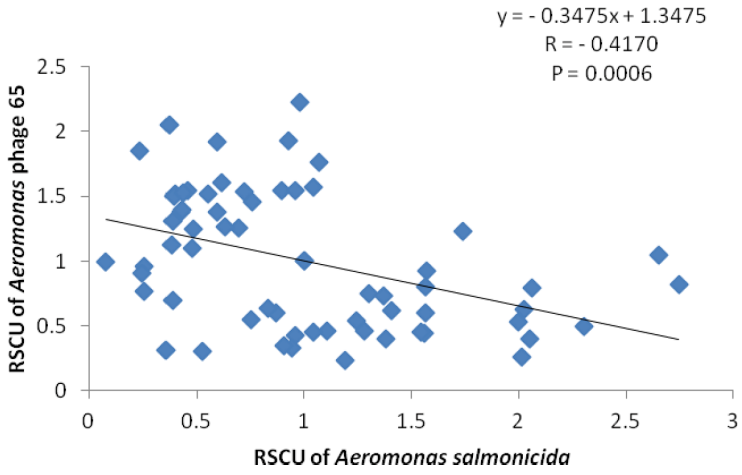
As per our prediction mentioned in the introduction that tRNA genes encoded in the *Aeromonas* phage genomes, if present for R-ending codons, should have a wobble U to decode these A-ending codons overused in the phages but not in the host.

Of the 24 tRNA genes encoded in the genome of *Aeromonas* phage phiAS5, 10 tRNAsare for R-ending codons (Table 2). In all these 10 sets of R-ending codons, the phage uses more A-ending codons than the host, and the phage-encoded tRNAs invariably have a wobble U to enhance the translation of the A-ending codons that are overused in the phage but not in the host. Take the two-fold arginine codons (AGA and AGG) for example. The host overuses the AGG codon whereas the phage overuses the AGA codon. One would predict that the phage-encoded tRNA should have a wobble U to decode the A-ending codon overused in the phage but not in the host. This prediction is consistent with the empirical data (Table 2). Similarly, we tested this prediction in the other six GC-poor *Aeromonas* phage 25, phage 31, phage 44RR2.8t, phage 65, phage Aes508 and phage phiAS4 (Figure 3). Again the results from these GC-poor *Aeromonas* phages are also in agreement with our prediction.

Figure 1 Comparison of codon usage of GC-rich and GC-poor *Aeromonas* phages with their host: (A) RSCU plot of GC-richdsDNAAeromonas phage phiAS7 and its host and (B) RSCU plot of GC-poordsDNAAeromonas phage 65 and its host (see online version for colours)



(A)



(B)

Figure 2 The number of phage encoded tRNA genes plotted against the difference in GC content between the phage and its host (see online version for colours)

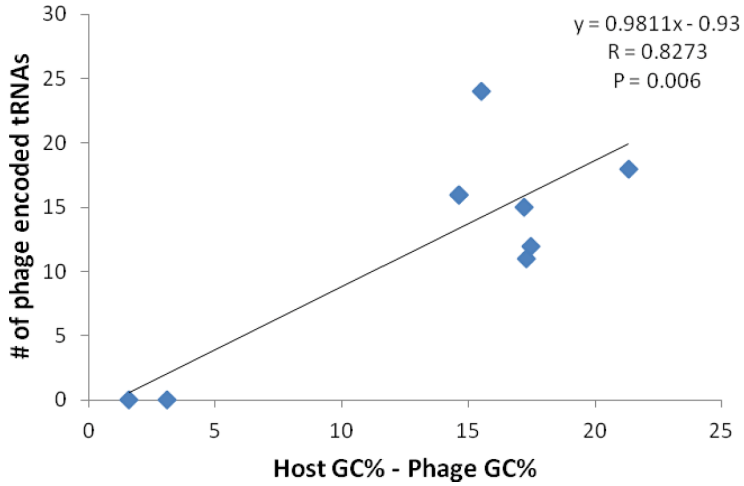
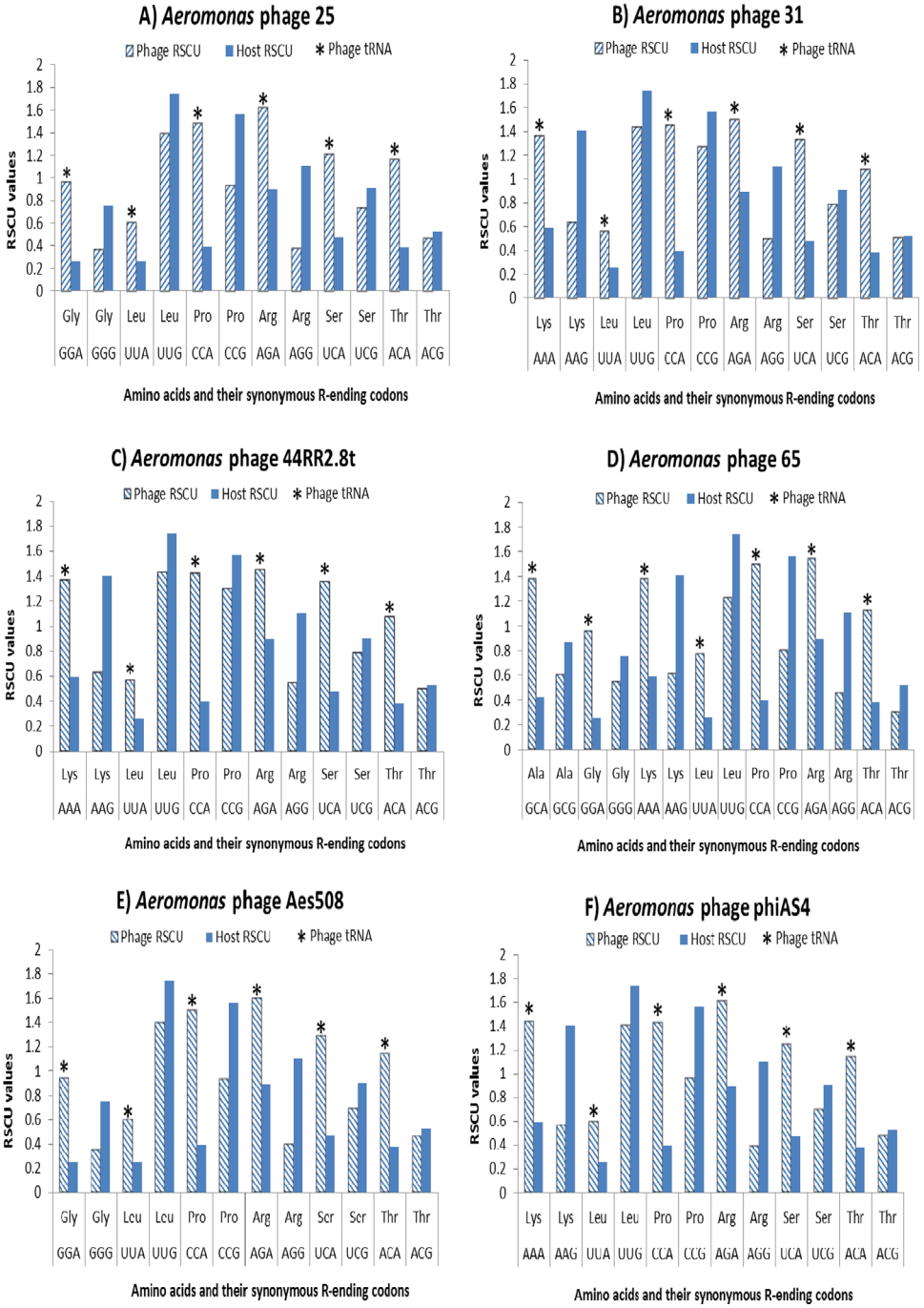


Table 2 *Aeromonas* phage phiAS5 encodes tRNAs for its overused codons

		Host			*Anti-*	*Phage*		
Codon	*AA*	*#Codon*	*RSCU*	*tRNA*	*codon*	*#Codon*	*RSCU*	*tRNA*
GCA	Ala	13759	0.427	3	UGC	1293	1.243	1
GCG	Ala	28030	0.870		CGC	772	0.742	
GAA	Asp	26196	0.722	6	UUC	2754	1.406	1
GAG	Asp	46416	1.278		CUC	1163	0.594	
GGA	Gly	6322	0.257	2	UCC	870	0.874	1
GGG	Gly	18504	0.753	1	CCC	189	0.190	
AAA	Lys	14791	0.593	3	UUU	2552	1.173	1
AAG	Lys	35059	1.407	1	CUU	1798	0.827	
CUA	Leu	2538	0.077	1	UAG	411	0.566	1
CUG	Leu	90545	2.747	6	CAG	1386	1.909	
UUA	Leu	2559	0.259	1	UAA	333	0.538	1
UUG	Leu	17201	1.741	2	CAA	906	1.462	1
CCA	Pro	5833	0.395	2	UGG	624	1.249	1
CCG	Pro	23154	1.566	1	CGG	722	1.445	
CAA	Gln	13243	0.433	5	UUG	1294	1.304	1
CAG	Gln	47870	1.567		CUG	690	0.696	
AGA	Arg	2773	0.896	1	UCU	562	1.816	1
AGG	Arg	3420	1.104	1	CCU	57	0.184	
ACA	Thr	5944	0.384	1	UGU	955	1.050	1
ACG	Thr	8132	0.525	1	CGU	348	0.383	

*See the text for reasons of including only R-ending codons.

Figure 3 Relationship between Myoviridae *Aeromonas* phage encoded tRNAs and their overused codons (see online version for colours)



Only NNR codons are considered for this analysis. * represents presence of tRNAs in phages for their respective codons. Filled bars represent host codon usage and striped bars represent phage codon usage.

More and more studies have reported the relevance of phage-encoded tRNA in enhancing the translation of phage proteins. Such phages include T4 (Kunisawa, 1992), BXZ1 (Sahu et al., 2004), Phikz (Sau et al., 2005), Aeh1 (Sau, 2007). However, it is unknown whether the phage-encoded tRNA genes constitute an ancestral state or a derived one, i.e., acquired in response to colonising a host with a dramatically different codon usage and a tRNA pool unfavourable for translating phage genes. It is conceivable that the ancestor of Myoviridae GC-poor *Aeromonas* phages may also be GC-poor (AT-rich) and may encode even more tRNA genes than its descendent lineages. If a descendent lineage parasitises an AT-rich host with a tRNA pool favourable to translate the phage genes, then there would be little selection pressure to maintain such phage-encoded tRNA genes, which may then degrade and disappear from the genome. If a descendent lineage parasitises a GC-rich host such as *A. salmonicida*, then it is beneficial to maintain at least a subset of tRNAs for translating codons overused in the phage genes but not in the host genes. On the other hand, it is also possible for the phage genome to acquire tRNA genes after parasitising a new host. To address such questions, one needs to build a phylogenetic tree so that ancestral states can be identified.

There are at least two possible evolutionary hypotheses concerning the presence of tRNA in phages:

- the ancestral state hypothesis (ASH) according to which tRNA gene was present in the ancestor of *Aeromonas* phages
- the derived state hypothesis (DSH) which states that tRNA gene was absent in the ancestor, but gained along subsequent descendent lineages, likely in response to compositional differences with their hosts.

We assessed which of the above two hypotheses received more empirical support by mapping the presence/absence of tRNA genes onto the phylogenetic tree built using shared genes of phages (Figures 4 and 5).

The presence of tRNA genes in all phages along lineage B except *Bacillus* phage SP10 (lineage E) and *Enterobacteria* phage Bp7 (lineage F) and the absence of tRNAs in the descendent phage lineages C and D could be the outcome of two possible evolutionary scenarios. Consider scenario one as ancestor A encoding tRNA genes (ASH), under this assumption the possible events are listed below

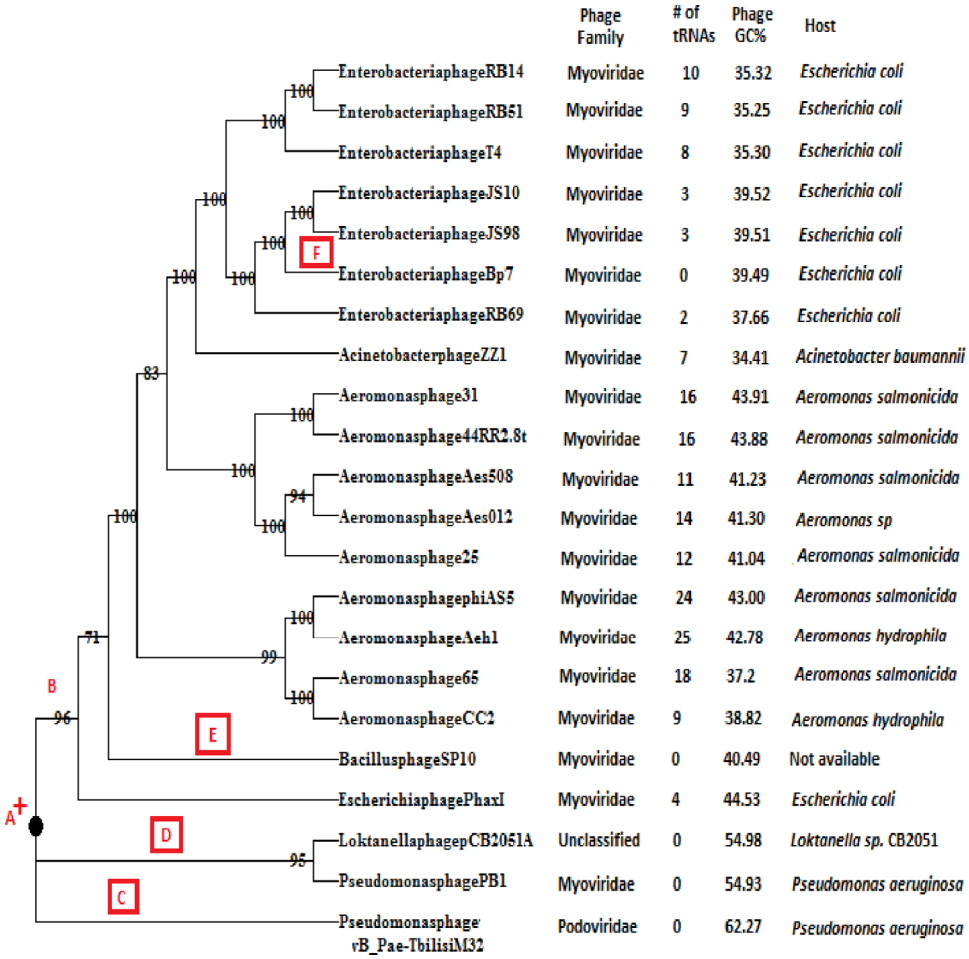
- tRNA genes encoded by ancestor A must have been passed on to all subsequent descendent lineages through lineage B but *Bacillus* phage SP10 (lineage E) and *Enterobacteria* phage Bp7 (lineage F) have experienced tRNA gene loss events in lineage B (Figure 4).
- The descendent lineages C and D have experienced tRNA gene loss events (Figure 4).

Consider another scenario where ancestor A does not encode tRNA genes (DSH). Again, under this assumption the two possible events are listed below, lineage B experienced a tRNA gene gain event and passed on this trait to all its descendents but *Bacillus* phage SP10 (lineage E) and *Enterobacteria* phage Bp7 (lineage F) have experienced tRNA gene loss events (Figure 5).

ASH predicts four tRNA gene loss events while DSH predicts 1 tRNA gene gain and 2 tRNA gene loss events. During the process of evolution loss/gain of a trait is always gradual; however ASH predicts that there are four complete tRNA gene loss

events which are very unlikely to occur. Hence as per the principle of parsimony, DSH gains more support. Therefore phylogenetic analyses indicate that tRNA genes may be a derived trait for *Aeromonas* phages.

Figure 4 Phylogeny based comparison of tRNA gene loss and gain events in phages based on ASH (see online version for colours)

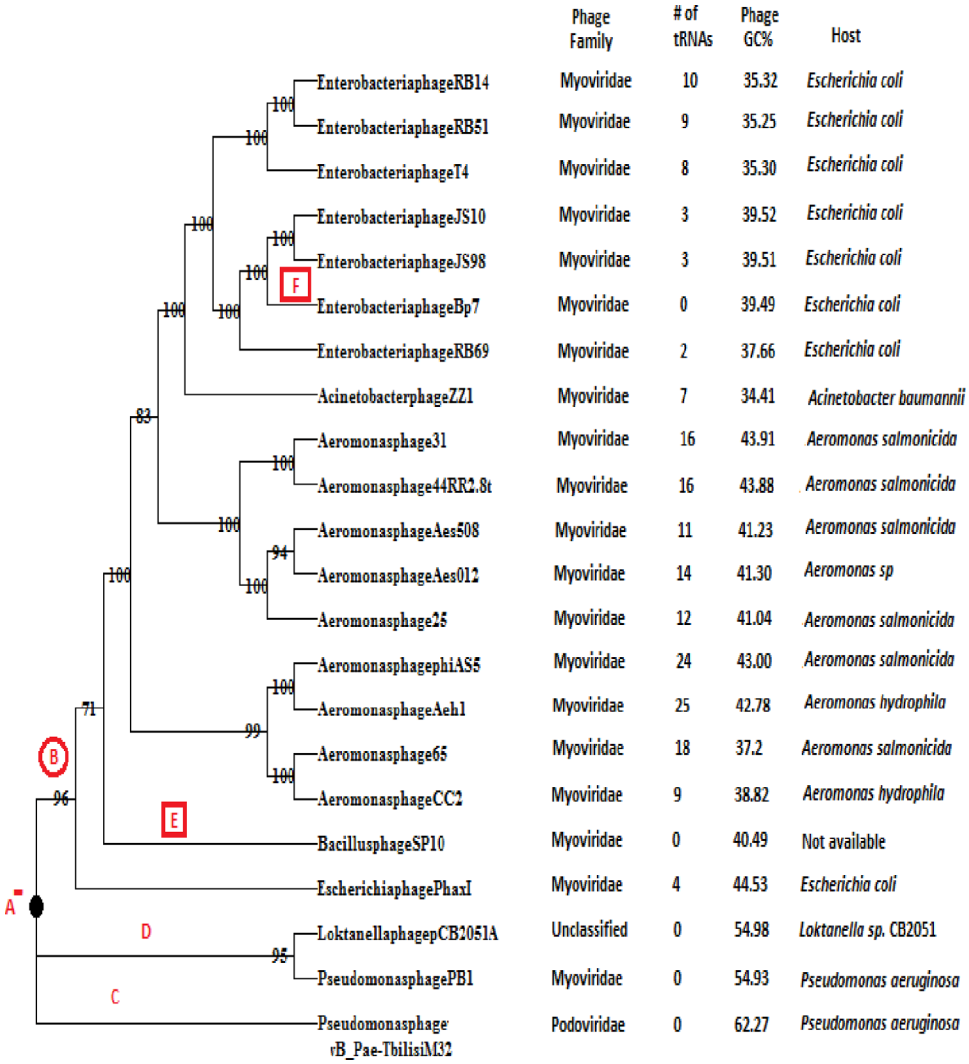


A^+ represents ancestor with tRNA genes, lineage names marked with square represents tRNA gene loss events.

4 Conclusion

To summarise, our study revealed considerable differences between synonymous codon preferences of GC-rich and GC-poor *Aeromonas* phages with respect to their host. Compositional differences between phages and host appear to be the causative factor for the presence of phage-encoded tRNA genes in the GC-poor *Aeromonas* phages. Finally, our phylogenetic analyses suggest that the presence of tRNAs in *Aeromonas* phages is a derived trait.

Figure 5 Phylogeny based comparison of tRNA gene loss and gain events in phages based on DSH (see online version for colours)



A⁻ represents ancestor without tRNA genes, lineage names marked with circle represents tRNA gene gain events, lineage names marked with square represents tRNA gene loss events.

Acknowledgements

This study is supported by NSERC's Discovery and Strategic Grants to XX.

References

Ackermann, H.W., Dauguet, C., Paterson, W.D., Popoff, M., Rouf, M.A. and Vieu, J.F. (1985) 'Aeromonas bacteriophages: reexamination and classification', *Annales de l'Institut Pasteur/Virologie*, Vol. 136, pp.175–199.

Akashi, H. (1994) 'Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy', *Genetics*, Vol. 136, pp.927–935.

Bahir, I., Fromer, M., Prat, Y. and Linial, M. (2009) 'Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences', *Mol. Syst. Biol*., Vol. 5, p.311.

Bailly-Bechet, M., Vergassola, M. and Rocha, E. (2007) 'Causes for the intriguing presence of tRNAs in phages', *Genome Res*., Vol. 17, pp.1486–1495.

Beletskii, A. and Bhagwat, A.S. (1996) 'Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli', *Proc. Natl. Acad. Sci. USA*, Vol. 93, pp.13919–13924.

Bennetzen, J.L. and Hall, B.D. (1982) 'Codon selection in yeast', *J. Biol. Chem*., Vol. 257, pp.3026–3031.

Bernardi, G. (1986) 'Compositional constraints and genome evolution', *J. Mol. Evol*., Vol. 24, pp.1–11.

Bulmer, M. (1987) 'Coevolution of codon usage and transfer RNA abundance', *Nature*, Vol. 325, pp.728–730.

Carbone, A. (2008) 'Codon bias is a major factor explaining phage evolution in translationally biased hosts', *J. Mol. Evol*., Vol. 66, pp.210–223.

Chan, P.P. and Lowe, T.M. (2009) 'GtRNAdb: a database of transfer RNA genes detected in genomic sequence', *Nucleic Acids Res*., Vol. 37, pp.D93–97.

Duret, L. (2000) 'tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes', *Trends Genet*., Vol. 16, pp.287–289.

Eyre-Walker, A. (1996) 'Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy?', *Mol. Biol. Evol*., Vol. 13, pp.864–872.

Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980) 'Codon catalog usage and the genome hypothesis', *Nucleic Acids Res.*, Vol. 8, pp.r49–r62.

Grosjean, H. and Fiers, W. (1982) 'Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes', *Gene*, Vol. 18, pp.199–209.

Grosjean, H., Sankoff, D., Jou, W.M., Fiers, W. and Cedergren, R.J. (1978) 'Bacteriophage MS2 RNA: a correlation between the stability of the codon: anticodon interaction and the choice of code words', *J. Mol. Evol*., Vol. 12, pp.113–119.

Ikemura, T. (1981) 'Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system', *J. Mol. Biol*., Vol. 151, pp.389–409.

Ikemura, T. (1982) 'Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs', *J. Mol. Biol*., Vol. 158, pp.573–597.

Ikemura, T. (1985) 'Codon usage and tRNA content in unicellular and multicellular organisms', *Mol. Biol. Evol*., Vol. 2, pp.13–34.

Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T. (1999) 'Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis', *Gene*, Vol. 238, pp.143–155.

Kim, J.H., Son, J.S., Choresca, C.H., Shin, S.P., Han, J.E., Jun, J.W., Kang, D.H., Oh, C., Heo, S.J. and Park, S.C. (2012) 'Complete genome sequence of bacteriophage phiAS7, a T7-like virus that infects Aeromonas salmonicida subsp. Salmonicida', *J. Virol*., Vol. 86, pp.2894–2895.

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*., Vol. 159, pp.287–298.

Lim, V.I. (1994) 'Analysis of action of wobble nucleoside modifications on codon-anticodon pairing within the ribosome', *J. Mol. Biol*., Vol. 240, pp.8–19.

Limor-Waisberg, K., Carmi, A., Scherz, A., Pilpel, Y. and Furman, I. (2011) 'Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies', *Nucleic Acids Res*., Vol. 39, pp.6016–6028.

Lobry, J.R. (1996) 'Asymmetric substitution patterns in the two DNA strands of bacteria', *Mol. Biol. Evol*., Vol. 13, pp.660–665.

Lobry, J.R. and Sueoka, N. (2002) 'Asymmetric directional mutation pressures in bacteria', *Genome Biol*., Vol. 3, RESEARCH0058.

Lowe, T.M. and Eddy, S.R. (1997) 'tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence', *Nucleic Acids Res*., Vol. 25, pp.955–964.

Marin, A. and Xia, X. (2008) 'GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: new substitution models incorporating strand bias', *J. Theor. Biol*., Vol. 253, pp.508–513.

Muto, A. and Osawa, S. (1987) 'The guanine and cytosine content of genomic DNA and bacterial evolution', *Proc. Natl. Acad Sci. USA*, Vol. 84, pp.166–169.

Ohama, T., Muto, A. and Osawa, S. (1990) 'Role of GC-biased mutation pressure on synonymous codon choice in Micrococcus luteus, a bacterium with a high genomic GC-content', *Nucleic Acids Res*., Vol. 18, pp.1565–1569.

Pavon-Eternod, M., David, A., Dittmar, K., Berglund, P., Pan, T., Bennink, J.R. and Yewdell, J.W. (2013) 'Vaccinia and influenza A viruses select rather than adjust tRNAs to optimize translation', *Nucleic Acids Res*., Vol. 41, pp.1914–1921.

Percudani, R., Pavesi, A. and Ottonello, S. (1997) 'Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae', *J. Mol. Biol*., Vol. 268, pp.322–330.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) 'Codon usage can affect efficiency of translation of genes in Escherichia coli', *Nucleic Acids Res*., Vol. 12, pp.6663–6671.

Sahu, K., Gupta, S.K., Ghosh, T.C. and Sau, S. (2004) 'Synonymous codon usage analysis of the mycobacteriophage Bxz1 and its plating bacteria M. smegmatis: identification of highly and lowly expressed genes of Bxz1 and the possible function of its tRNA species', *J. Biochem. Mol. Biol*., Vol. 37, pp.487–492.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees', *Mol. Biol. Evol*., Vol. 4, pp.406–425.

Sau, K. (2007) 'Studies on synonymous codon and amino acid usages in Aeromonas hydrophila phage Aeh1: architecture of protein-coding genes and therapeutic implications', *J. Microbiol. Immunol. Infect*., Vol. 40, pp.24–33.

Sau, K., Gupta, S.K., Sau, S., Mandal, S.C. and Ghosh, T.C. (2007) 'Studies on synonymous codon and amino acid usage biases in the broad-host range bacteriophage KVP40', *J. Microbiol*., Vol. 45, pp.58–63.

Sau, K., Sau, S., Mandal, S.C. and Ghosh, T.C. (2005) 'Factors influencing the synonymous codon and amino acid usage bias in AT-rich Pseudomonas aeruginosa phage PhiKZ', *Acta Biochim. Biophys. Sin (Shanghai)*, Vol. 37, pp.625–633.

Sharp, P.M. and Devine, K.M. (1989) 'Codon usage and gene expression level in Dictyostelium discoideum: highly expressed genes do 'prefer' optimal codons', *Nucleic Acids Res*., Vol. 17, pp.5029–5039.

Sharp, P.M. and Li, W.H. (1987) 'The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications', *Nucleic Acids Res.*, Vol. 15, pp.1281–1295.

Sharp, P.M., Tuohy, T.M. and Mosurski, K.R. (1986) 'Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes', *Nucleic Acids Res*., Vol. 14, pp.5125–5143.

Sorensen, M.A., Kurland, C.G. and Pedersen, S. (1989) 'Codon usage determines translation rate in Escherichia coli', *J. Mol. Biol*., Vol. 207, pp.365–377.

Sueoka, N. (1988) 'Directional mutation pressure and neutral molecular evolution', *Proc. Natl. Acad. Sci. USA*, Vol. 85, pp.2653–2657.

Sun, X., Yang, Q. and Xia, X. (2013) 'An improved implementation of effective number of codons (nc)', *Mol. Biol. Evol*., Vol. 30, pp.191–196.

Tamura, K. and Nei, M. (1993) 'Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees', *Mol. Biol. Evol*., Vol. 10, pp.512–526.

van Weringh, A., Ragonnet-Cronin, M., Pranckeviciene, E., Pavon-Eternod, M., Kleiman, L. and Xia, X. (2011) 'HIV-1 modulates the tRNA pool to improve translation efficiency', *Mol. Biol. Evol*., Vol. 28, pp.1827–1834.

Woo, P.C., Wong, B.H., Huang, Y., Lau, S.K. and Yuen, K.Y. (2007) 'Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in coronaviruses', *Virology*, Vol. 369, pp.431–442.

Wright, F. (1990) 'The 'effective number of codons' used in a gene', *Gene*, Vol. 87, pp.23–29.

Xia, X. (1996) 'Maximizing transcription efficiency causes codon usage bias', *Genetics*, Vol. 144, pp.1309–1320.

Xia, X. (1998) 'How optimized is the translational machinery in Escherichia coli, Salmonella typhimurium and Saccharomyces cerevisiae?', *Genetics*, Vol. 149, pp.37–44.

Xia, X. (2005) 'Content sensors based on codon structure and dna methylation for gene finding in vertebrate genomes', in Kolchanov, N. and Hofestadt, R. (Eds.): *Bioinformatics of Genome Regulation and Structure II*, Springer Science+Business Media, Inc., pp.21–29.

Xia, X. (2007) 'An improved implementation of codon adaptation index', *Evol Bioinform Online*, Vol. 3, pp.53–58.

Xia, X. (2008) 'The cost of wobble translation in fungal mitochondrial genomes: integration of two traditional hypotheses', *BMC Evol. Biol*., Vol. 8, p.211.

Xia, X. (2013) 'DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution', *Mol. Biol. Evol*., Vol. 30, pp.1720–1728.