

Review

Optimizing Phage Translation Initiation

Xuhua Xia^{1, 2, *}

1. Department of Biology, Faculty of Science, University of Ottawa, Ottawa, Canada; E-Mail: xxia@uottawa.ca
2. Ottawa Institute of Systems Biology, Ottawa, Ontario, Canada

* **Correspondence:** Xuhua Xia; E-Mail: xxia@uottawa.ca

Academic Editor: Joep Geraedts

OBM Genetics

2019, volume 3, issue 4

doi:10.21926/obm.genet.1904097

Received: June 09, 2019

Accepted: October 14, 2019

Published: October 17, 2019

Abstract

Phage as an anti-bacterial agent must be efficient in killing bacteria, and consequently needs to replicate efficiently. Protein production is a limiting step in replication in almost all forms of life, including phages. Efficient protein production depends on the efficiency of translation initiation, elongation and termination, with translation initiation often being rate limiting. Initiation signals such as Shine-Dalgarno (SD) sequences and start codon are decoded by anti-SD sequences and initiation tRNA, respectively. While the decoding machinery cannot be readily modified, the signals can be engineered to increase the efficiency of their decoding. Here I review our understanding of the translation machinery to facilitate the engineering of optimal translation initiation signals for facilitating the design of phage protein-coding genes, including 1) accurate characterization of the 3' end of 16S rRNA by using RNA-Seq data, 2) identification of the optimal SD/aSD interaction, and 3) reduction of secondary structure in sequences flanking the start codon.

Keywords

Translation initiation; coevolution; bacteriophage; anti-bacterial agents



© 2019 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

1. Introduction

Protein production is typically the rate-limiting step in organism growth and reproduction. Among the three sub-processes of translation, i.e., translation initiation, elongation and termination, initiation is often the limiting step [1-6]. Phages generally do not have their own translation machinery, and rely on the host translation machinery for protein production [7]. Although certain phages encode tRNA genes in their genome, e.g., some coliphage genomes encode as many as 20 tRNA genes [8], these tRNA genes may only primarily affect the tRNA pool and translation elongation, but not translation initiation. With this dependence of phage translation initiation on host translation machinery, phage protein-coding genes are expected to adapt to their host translation machinery and evolve to acquire sequence features characteristic of host protein-coding genes, especially highly expressed ones. Thus, understanding sequence features that enhance translation initiation in host proteins will facilitate the design of phage protein-coding genes with high translation initiation efficiency.

Efficient initiation in bacterial species depend mainly on three factors [4]: 1) the base-pairing interaction between Shine-Dalgarno (SD) sequence upstream of the start codon and anti-SD (aSD) sequence at the 3' tail of the small subunit rRNA [9-13], 2) the nature of start codon, and 3) the secondary structure that may embed SD and start codon to obscure these essential translation initiation signals. These factors contribute to the efficiency of ribosomes being properly positioned at start codon to transit from translation initiation to elongation. Much progress has been made towards a precise understanding of bacterial translation initiation, including the optimal pairing strength and positioning between SD and aSD, and the effect of secondary structure on translation initiation. This review covers recent advances in these two areas and their relevance in optimizing phage mRNA for efficient translation initiation.

2. A Model of SD/aSD Interaction

One key task of the translation machinery is to identify start codon based on the signals at, and flanking, the start codon. In mammalian species, the Kozak consensus on mRNA represents an important translation initiation signal [14-16]. In bacteria, the interaction between SD sequence located upstream of the start codon and aSD located at the free 3' end of the small subunit ribosomal rRNA [9-13, 17] serves to localize the start codon, most likely by physically bringing together the start codon and the initiation tRNA. The model of base-pairing between SD and aSD [4, 18, 19] illustrates how the start codon is positioned against the anticodon of tRNA^{fMet} (Figure 1A,B). The graphic model is partly motivated by the observation that highly expressed genes in *Escherichia coli* use different SDs with different distances between SD and start codon (e.g., L₁ and L₂ for SD₁ and SD₂ in Figure 1A). However, they share the same distance designated by D_{toStart}, which is defined as the number of bases from the 3' end of the ssu (small subunit) rRNA to the nucleotide just before the start codon (Figure 1C,D). D_{toStart} can be calculated by software DAMBE [20]. As an increase or decrease of D_{toStart} will change the optimal juxtaposition between the start codon and the initiation tRNA (Figure 1 A, B), D_{toStart} is constrained within a narrow range [4, 18, 21] (Figure 1E).

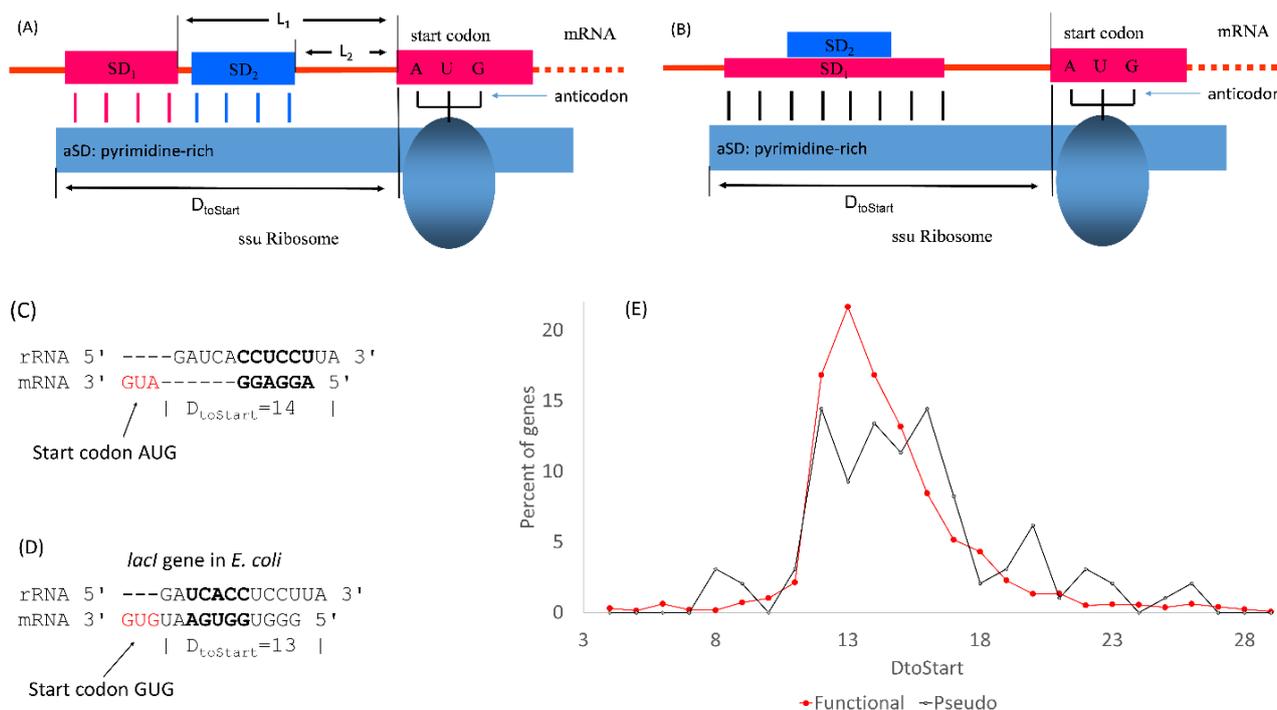


Figure 1 A graphic model of SD/aSD interaction for defining a new variable $D_{toStart}$. (A,B) schematic illustration of SD/aSD interaction, with start codon juxtaposed against the anticodon of initiation tRNA, with the same $D_{toStart}$ but different leash lengths (L_1 and L_2). (C,D) $D_{toStart}$ as the number of nucleotides between the 3' end of ssu rRNA and the nucleotide just before the start codon. (E) $D_{toStart}$ is strongly constrained within *E. coli* functional genes relative to pseudogenes, obtained with the last 13 nucleotide at the 3' end of the ssu rRNA of *E. coli*. (F) Weak secondary structure around translation initiation region to avoid embedding SD sequence and start codon in a second structure. Redrawn from Wei et al. [21] and Xia [19].

Is the narrow range of $D_{toStart}$ really constrained by selection favoring optimal SD/aSD pairing? We may address this question by reasoning that, if the selection is relaxed, then $D_{toStart}$ distribution will deviate from the narrow range. *Escherichia coli* K12 has 306 annotated pseudogenes in its genome (NC_000913). These pseudogenes are not subject to the selection for increased translation initiation efficiency. The distribution of $D_{toStart}$ for these pseudogenes misses the peak, suggesting that the distribution of $D_{toStart}$ is indeed maintained by selection. If one separates *E. coli* genes into highly expressed and lowly expressed genes (designated by HEGs and LEGs, respectively), by ranking them either by protein abundance or codon adaptation indices such as I_{TE} (Index of translation elongation) [6, 22], then HEGs also exhibit a more pronounced peak than LEGs. The calculation of $D_{toStart}$, as well as a variety of other analysis related to SD/aSD interaction, has been implemented in software DAMBE [20].

The conceptual model in Figure 1A and Figure 1B eliminates confusions in textbooks in molecular biology that take a simple-minded definition of SD as AGGAGG. One can find AGGAGG upstream of the start codon with $D_{toStart}$ far away from the peak region of $D_{toStart}$ (Figure 1E). Such an AGGAGG is unlikely to be an SD. One can also find non-AGGAGG that can pair with the 3' end of 16S rRNA to have a $D_{toStart}$ right at the peak region. For example, *lacI* does not have AGGAGG, but

has GGUGA forming SD/aSD base-pairing which has a $D_{\text{toStart}} = 13$. Thus, defining SD simply as AGGAGG (or AGGAGGU) is misleading.

One frequently encounters statements that six nucleotides between AGGAGG and the start codon is an optimal distance. Such statements may be true for *E. coli* because such an SD would have a D_{toStart} of 14 that is near the peak of D_{toStart} (Figure 1E). That is, an AGGAGG six nucleotide upstream of the start codon indeed has an optimal D_{toStart} . However, this definition of SD (i.e., AGGAGG spaced about six nucleotide upstream of the start codon) would not be applicable to the SD in *lacI* which is GGUGA spaced two nucleotides upstream of the start codon (Figure 1D). With D_{toStart} , we now have a consistent definition for SD sequences in *E. coli*, i.e., any sequence that can base-pair against aSD to achieve a D_{toStart} close to 14. Different bacterial species have different 3' end of ssu rRNA and consequently have different optimal D_{toStart} . For example, the optimal D_{toStart} is longer in *Bacillus subtilis* than in *E. coli* [18]. In short, without referring to the 3' end of 16S rRNA, it is often meaningless to say what the optimal distance is for a SD/aSD interaction.

While SD/aSD interaction is important for translation initiation in a majority of bacterial mRNAs, there are cases where SD/aSD interaction is not required for translation initiation, based on two lines of evidence. First, the small ribosomal subunit assembled with a 16S rRNA without the 30 nt at the 3' end can still initiate translation at the start site [23]. Second, leaderless mRNA can be efficiently translated in *Halobacterium salinarum* [24], which contributes to the formulation of a new hypothesis that the minimum requirement for translation initiation in bacteria is an accessible initiation codon [25]. In Gram-negative bacteria, both a downstream box [26], i.e., a motif downstream of the start codon contributing to translation initiation, and ribosomal protein S1 (RPS1) [27, 28] both contribute to translation initiation in mRNAs with a weak SD or no SD at all. There are ~20% of functional genes in *E. coli* K12 genome (NC_000913) that do not have an identifiable SD within 20 bases upstream of the start codon [19].

In short, most genes in bacteria use SD/aSD interaction for localizing the start codon, especially in highly expressed genes. The graphic model of SD/aSD interaction in Figure 1 may serve as a conceptual framework for understanding optimal positioning of SD/aSD base-pairing. However, the minimum requirement for translation initiation appears to be an accessible initiation codon that is not embedded in secondary structure.

3. Accurate Characterization of 3' End of 16S Rrna by Rna-Seq Data

The main difficulty with characterizing SD/aSD interaction is that, while almost all aspects of 16S rRNA maturation have been elucidated, the maturation step involving the 3' end of 16S rRNA remains unknown for most bacterial species [29, 30]. Consequently, the 3' end of 16S rRNA is known for only a few bacterial species since 1980 [31], but unknown for most bacterial species. Computation prediction [32, 33] is often unreliable [32, 34-36], with many notable errors [21]. For example, 16S rRNA entries for *Streptococcus pyogenes* (NC_002737), *Bacillus anthracis* (NC_005945), and *Legionella pneumophila* (NC_005823) were suspected to have been incorrectly annotated in their GenBank files because their 3' ends do not include the core CCUCC motif. The wrong annotation was confirmed by analyzing RNA-Seq data [21]. In addition, the 16S rRNA pool in the cell may be heterogeneous due to differential RNA degradation, and therefore should not be represented by a single sequence.

The rationale of characterizing 3' end of 16S rRNA by RNA-Seq data is straightforward [21, 37], and the method is implemented in software ARSDA [38] which also include a variety of other functions for RNA-Seq data analysis. One maps the transcriptomic reads onto 16S rRNA gene and its downstream region on the genome, as illustrated in Figure 2. Some bacterial species such as *Lactococcus lactis* have relatively homogeneous 16S rRNA pool with most 16S rRNA species in the cell ending at the same location (Figure 2). In such cases, the 3' end is unambiguous. However, some species, such as *Deinococcus deserti*, have mapped reads ending at predominantly two locations [21]. For example, in addition to a large number of mapped reads ending at site 50 in Figure 2, there are also a large number of mapped reads ending at site 53, but few mapped reads ending at sites 51 and 52, creating a bimodal distribution of mapped reads over sites. Some other species have even more heterogeneous 16S rRNA pools [21].

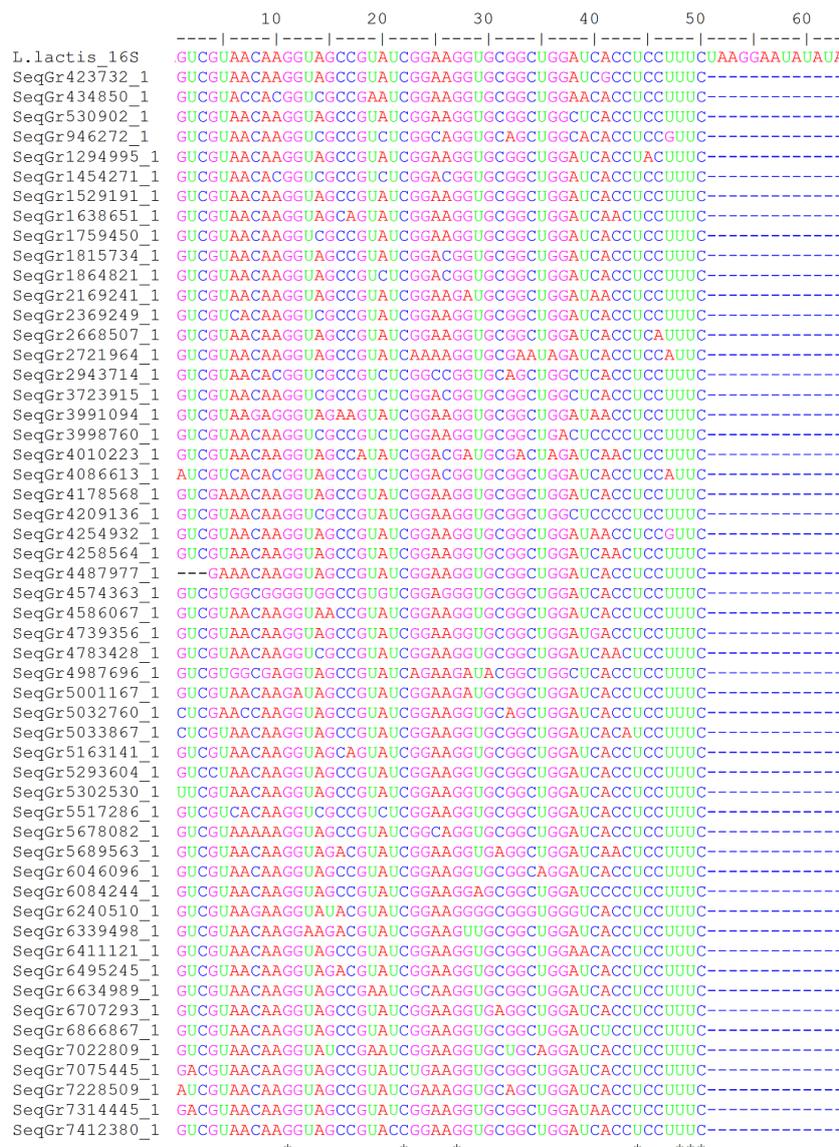


Figure 2 Characterizing the 3' end of 16S rRNA in *Lactococcus lactis* by RNA-Seq data, which is 5'...GAU GAUCACCUCCUUUC 3'. The top sequence is *L. lactis* 16S rRNA gene and its downstream sequence, and the other sequences are transcriptomic reads mapped to the top sequence. Extracted from Figure 1 in Silke et al. [37].

Given different 16S rRNA pools in different species, with potentially different aSD availability, one expects SD usage to evolve and adapt to available aSD motifs [18]. The characterization of 3' end of 16S rRNA not only paves the way for studying the coevolution between SD and aSD in bacterial species, but also facilitates the design of phage protein-coding genes with efficient translation initiation. Such research advances would contribute to the design of phage species against bacterial pathogens when antibiotics fail. For example, mRNA from gene *gp6* in *B. subtilis* phage $\phi 29$ cannot form SD/aSD base-pairing interactions with the 3' tail of *E. coli* 16S rRNA [18], and it cannot be translated in *E. coli* [28]. Thus, if we are to engineer phage $\phi 29$ against *E. coli*-like bacteria such as pathogenic *Shigella* species, we would have to modify the *gp6* gene in phage $\phi 29$ so that its mRNA can form proper SD/aSD interaction for translation initiation in *Shigella* species.

4. SD/aSD Pairing in Phage Genes Mirrors that of Their Host

As I have mentioned before, phages do not have their own translation machinery for protein production. Their dependence on the host translation machinery implies that phage genes should evolve sequence features mirroring the host genes, especially host HEGs. I illustrate the SD/aSD adaptation between bacterial pathogen *Listeria monocytogenes* (genome sequence NC_003210) and its two sequenced phage genomes: phage PSA (NC_003291) and phage A511 (NC_009811). *L. monocytogenes* encodes 2867 protein-coding genes in its genome, whereas phages PSA and A511 encode 59 and 199 protein-coding genes, respectively.

The 3' end of 16S rRNA in annotated *L. monocytogenes* genome turns out to be the same as that characterized by RNA-Seq, as shown in the shaded region in Figure 3A. With this 3' end, the distribution of D_{toStart} for the 2867 protein-coding genes in *L. monocytogenes* is constrained within the range of 15-21, with the peak at 17 (Figure 3B). Similarly, the distribution of D_{toStart} from the 258 phage genes (59 genes from Phage PSA and 199 genes from phage A511) is confined within effectively the same range and the same peak at 17 (Figure 3B). The distribution curve for the phage (blue line in Figure 3B) is less smooth than that for the host genes (red line in Figure 3B), which is expected given the large number of host genes and the relatively small number of phage genes.

Figure 3B suggests that, if we are to develop a phage against *L. monocytogenes*, then we should design the SD in such a way that its D_{toStart} should be 17. However, such theoretical prediction has become available only lately with the accurate characterization of the 16S rRNA pool in the host cell. There has been little experimental validation of the prediction. In addition, for bacterial hosts with heterogeneous 16S rRNA pool, the prediction will be less obvious, and we will need to characterize HEGs and contrast their SDs with those of LEGs to derive the optimal D_{toStart} .

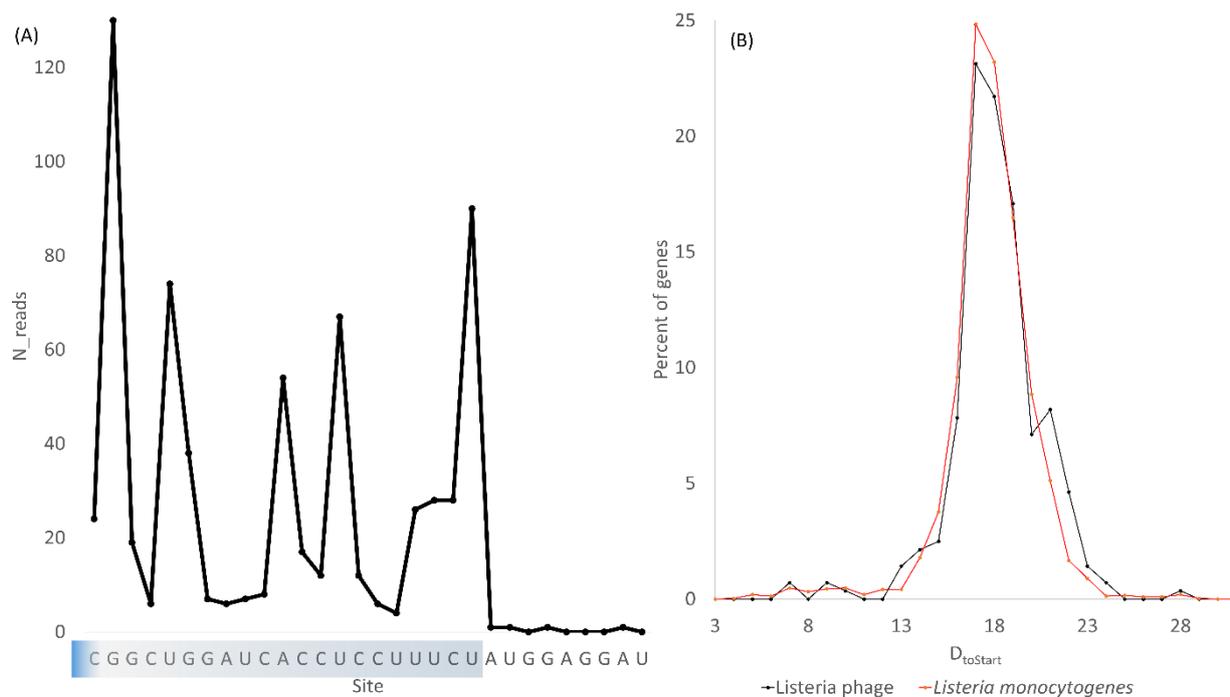


Figure 3 SD/aSD pairing in protein coding genes of two *Listeria* phages mirrors that of their host, *Listeria monocytogenes*. (A) Identification of the 3' end of 16S rRNA (shaded sequence) in *L. monocytogenes*, with Y-axis being the number of transcriptomic reads mapped to the genomic sequence flanking the core CCUCC motif. The dots represent number of mapped transcriptomic reads that end at the shown sequence at the X-axis. Effectively no reads map beyond the shaded sequence, marking the 3' end of 16S rRNA. (B) Distribution of $D_{toStart}$ for host genes (red line) and for phage genes (black line). The latter is less smooth because of smaller number of genes in the two phages (a combined total of 258 genes in the two phage genomes) than in the host (2867 genes in the host genome NC_003210).

5. Secondary Structure of mRNA and Translation Initiation Efficiency

Secondary structure can either increase or decrease translation initiation efficiency [4, 39, 40], depending on where the secondary structure is located. If a start codon is embedded in a stable secondary structure, then it will not be accessible and its signal is obscured. Stable secondary structure in sequences flanking the start codon has been experimentally shown to inhibit translation initiation [40], and mRNAs in bacterial species and unicellular eukaryotes tend to have much weaker secondary structure near the start codon than elsewhere, especially those from highly expressed genes as shown in Figure 8-6 in [19].

Messenger RNAs in both phages and their bacterial hosts have translation initiation signals in the form of SD and start codon. These translation initiation signals in mRNA, when embedded in a secondary structure, may become obscured and unavailable to its signal decoder. If an SD or a start codon is embedded in a stable secondary structure, then it will not be available for decoding by aSD and initiation tRNA, respectively. The classical example is the start codon AUG in the replicase gene of MS2 phage, which is embedded in a stem formed by part of the replicase gene and the upstream coat gene. When the upstream coat genes is translated, separating the

secondary structure as a consequence, the downstream replicase gene is then translated [41]. However, one could take advantage of this and design phage protein genes in such a way that proteins are produced only when the host has a fever, by embedding the SD or start codon in a secondary structure that would melt when host body temperature increases. However, fever can be triggered by a variety of factors not involving a bacterial pathogen, so this temperature-mediated translation regulation can only be used as a fail-safe mechanism. Another pathogen-specific genetic switch is needed to achieve specificity. Nature has furnished many genetic switches that are best known in phage lambda [42], bacteria such as the *Lac operon* [43] and yeasts such as the *IRE1-HAC1* switch controlling the unfolded protein response [44]. Optimal design of anti-bacterial phages ultimately depends on our understanding of these genetic switches, especially those in response to bacterial infection. Minimum free energy (MFE) is frequently used to measure secondary structure stability, with zero indicating no secondary structure and more negative values indicating stronger secondary structure. Calculation of MFE is implemented in Vienna RNA Folding Library [45], which is also used in software DAMBE [20, 46] for MFE computation. This function in DAMBE has been used not only to study secondary structure and translation of phage genes [4], but also in mitochondrial genes [47] as well as internal ribosomal entry in the yeast, *Saccharomyces cerevisiae*, and fruit fly, *Drosophila melanogaster* [48]. A consistent reduction in secondary structure stability near the translation start site, measured by MFE, has been observed in all these cases, suggesting the importance of not embedding the translation initiation signal within stable secondary structure. Secondary structure is also weak in sequences flanking stop codons [20]. Note that more negative MFE corresponds to stronger secondary structure.

A dramatic reduction of secondary structure stability in sequences flanking the start codon in the Gram-negative *E. coli* was previously shown [49]. Here I provide more details by separating genes into highly and lowly expressed genes (HEG and LEG as well as pseudogenes in *E. coli* (Figure 4A), and add the Gram-positive *Bacillus subtilis* (Figure 4B). The two species are phylogenetically highly divergent, but they both exhibit a consistent and dramatic reduction in secondary structure in sequences flanking the start codon. In particular, the pattern is more pronounced in HEGs than in LEGs (Figure 4), where HEGs are 200 genes with the highest I_{TE} (Index of Translation Elongation) values [6] and LEGs are 200 genes with the lowest I_{TE} values. Ranking genes with protein abundance exhibits the same pattern. Pseudogenes annotated in the *E. coli* genome exhibits the weakest pattern (Figure 4A). These results suggest that the weak secondary structure in sequences flanking the start codon is maintained by selection.

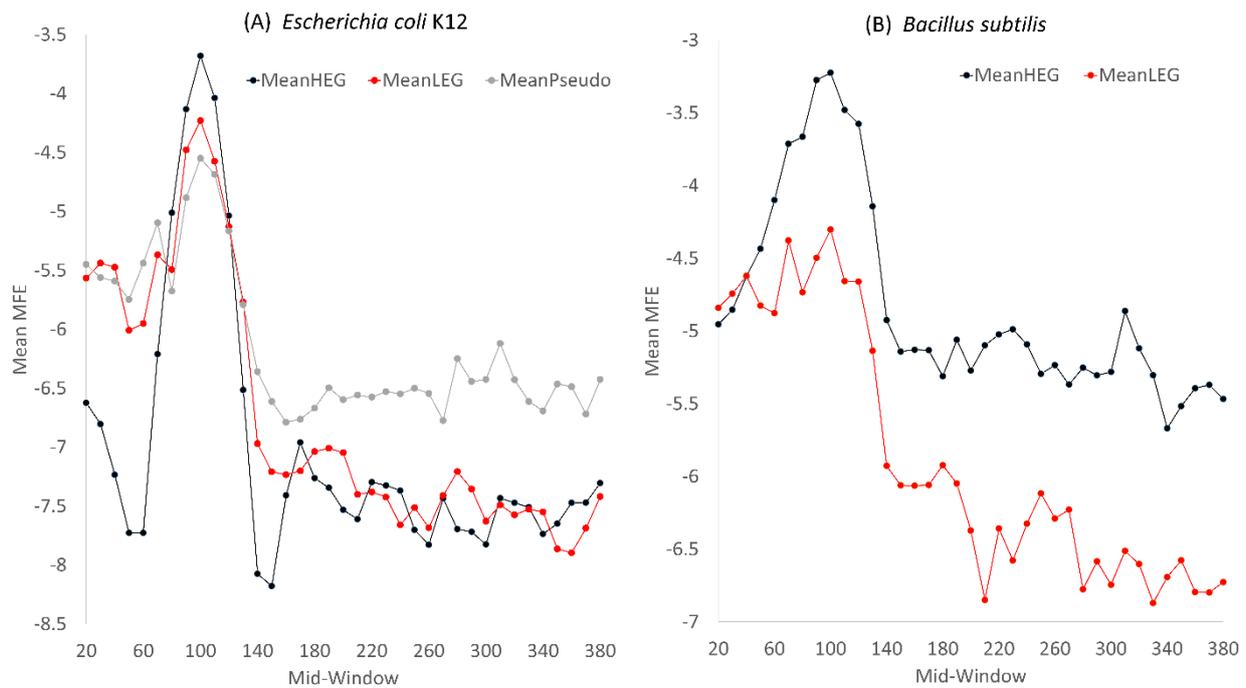


Figure 4 Mean MFE (minimum folding energy) of a sliding window of 40 nt, with a step size of 10 nt, along mRNA sequences plus 100 nt upstream (i.e., Window 1 spans sites 1 to 40, Window 2 spans sites 11 to 50, and so on). The start codon is located at 101-103 in the X-axis. (A) *Escherichia coli* K12 MG1665 (NC_000913), (B) *Bacillus subtilis* (NC_000964). Computed with DAMBE [46]. Secondary structure stability decreases as MFE becomes less negative. HEG: 200 highly expressed genes; LEG: 200 lowly expressed genes.

Do phage genes exhibit the same pattern of reduced secondary structure stability in sequences flanking the start codon? The answer is generally yes, but I will only illustrate it here with the bacterial pathogen *Listeria monocytogenes* and its two phages (their GenBank accession numbers and number of genes per genome have been mentioned previously). HEGs in *Listeria monocytogenes* exhibit a more dramatic reduction in secondary structure in sequences flanking the start codon than LEGs (Figure 5A), which is consistent with the pattern in Figure 4. What is noteworthy is that *Listeria* phage genes exhibit exactly the same pattern, with a dramatic decrease in secondary structure stability in sequences flanking the start codon (Figure 5B). In particular, the pattern is much stronger in HEGs than in LEGs (Figure 5B). Thus, phage genes mimic the host gene not only in SD/aSD interaction, but also in secondary structure in sequences flanking the start codon. In other words, both the phage genes and host HEGs have evolved and adapted to the same host translation machinery and acquired similar SD/aSD interaction and secondary structure stability. This highlights the relevance of studying the host translation machinery to understand the optimization of phage mRNA translation. The results in Figure 5 suggest that we should reduce secondary structure in designing phage genes to facilitate its translation initiation.

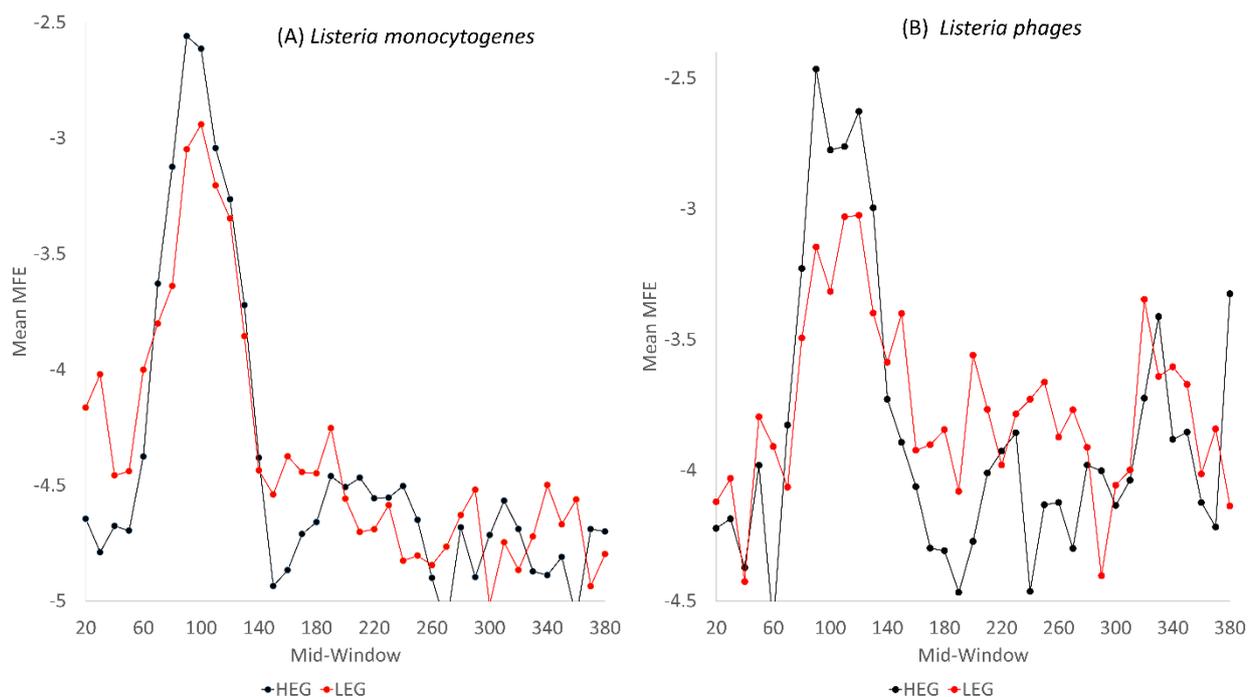


Figure 5 Mean MFE (minimum folding energy) of sliding windows along mRNA sequences, plus 100 nt upstream. Window size is 40 and step size is 10. The start codon corresponds to positions 101-103 on the X-axis. (A) Distribution of MFE for 200 HEGs (highly expressed genes) and 200 LEGs (lowly expressed genes) in *Listeria monocytogenes* (B) Distribution of MFE in 100 HEGs and 100 LEGs in *Listeria phages* PSA and A511 (with a total of 258 genes). Computed with DAMBE [46]. Secondary structure stability decreases as MFE becomes less negative. HEG: high expressed genes; LEG: lowly expressed genes. Relationship generated from DAMBE.

Many bacterial and phage genomes have start codon overlapping stop codon, e.g., in configurations of UAAUG (where AUG is the start codon of the downstream gene and UAA the stop codon of the upstream gene) and AUGA (where AUG is the start codon of the downstream gene and UGA the stop codon of the upstream gene). For such genes, it is not clear whether the reduction in secondary structure is to facilitate the decoding of the start codon or the stop codon. To avoid this complication, Figure 4 and Figure 5 are generated from non-overlapping genes with an intergenic sequence of at least 50 nt. It is important to take such overlapping genes into consideration when studying start or stop signals. For example, a surplus of nucleotide U following stop codons (+4U) has been documented in a number of studies [50, 51], and is interpreted to contribute to the strength of the stop signal. However, such a surplus of +4U could just be due to a large number of genes with a UAAUG configuration and may have nothing to do with contributing to the strength of the stop signal.

6. Translation Initiation and Phage Host Specificity

It is likely that specific requirement of mRNA (e.g., optimal position and binding affinity in SD/aSD interaction and weak secondary structure in sequences flanking the start codon) may limit phage host specificity. For example, given the difference in 3' end of 16S rRNA among different

bacterial species, a phage mRNA that has optimal SD/aSD interaction in one bacterial species may fail to have good SD/aSD interaction in another bacterial species, i.e., the phage is limited to the former as a host.

Some bacterial species can translate a variety of messages, whereas others are more restrictive. For example, *E. coli* possesses a more permissible translation machinery than *B. subtilis*, likely because of the ribosomal protein S1 in *E. coli* that allows translation of mRNA with weak or poorly positioned SD [27, 28]. However, there is one rare exception [28]. Gene 6 (*gp6*) of the *B. subtilis* phage $\phi 29$ can be translated efficiently in *B. subtilis* but not in *E. coli*. It turned out *gp6* is the only one out of 16 predicted genes in phage $\phi 29$ that cannot form a well-positioned SD/aSD in *E. coli* [18]. The failure of *gp6*, which is essential for phage survival, to form a proper SD/aSD in *E. coli* would explain why it cannot expand its host range to *E. coli*. That is, even if it gains entry into an *E. coli*-like host, it will not be able to produce the essential *gp6* protein and consequently will not survive and reproduce successfully.

7. Effect of Translation Initiation on Elongation

Just as SD is recognized by aSD, sense codons are decoded by tRNAs. Many studies have shown that codon-anticodon adaptation increased translation elongation efficiency, by correlating codon usage with tRNA abundance [52-59], by following evolutionary changes in tRNA genes and the associated change in codon usage [60, 61], by experimentally altering codon usage and monitoring protein production [62-65], and by theoretical justification with models [1, 56, 66-68]. Codon adaptation has also been demonstrated in phage [4, 8, 69, 70] and in HIV-1 virus [71] which previously were thought to exhibit little codon adaptation due to its high mutation rate. In this context it is surprising to encounter a seemingly well-substantiated claim that translation elongation efficiency has little relevance to the rate of protein production [2].

Kudla et al. [2] engineered a synthetic library of 154 genes, all encoding the same protein but differing in degrees of codon adaptation, to quantify the effect of differential codon usage on protein production in *E. coli*. They concluded that “codon bias did not correlate with gene expression” and that “translation initiation, not elongation, is rate-limiting for gene expression”. The main problem of the paper is that it did not consider differences in translation initiation among these engineered genes. Translation elongation efficiency becomes important and rate-limiting for protein production only in mRNAs with high translation initiation. When translation initiation and translation elongation are both taken into consideration, protein production increases highly significantly with codon adaptation [5, 6].

8. Conclusion

The translation machinery of bacterial hosts represents an environment that shape the evolution of protein-coding genes from both the host and the phage. Thus, studying host translation machinery will shed lights on how phage protein-coding genes should evolve. This review shows that phage genes resemble host highly expressed genes in both translation initiation signals and weakened secondary structure in sequences flanking the initiation signals.

Acknowledgements

This research was funded by Discovery Grant from Natural Science and Engineering Research Council (NSERC, RGPIN/2018-03878) of Canada. I thank J. Silke and Y. Wei for discussion and comments.

Author Contributions

Xuhua Xia did all the work.

Competing Interests

I declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics*. 1991; 129: 897-907.
2. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-Sequence determinants of gene expression in *Escherichia coli*. *Science*. 2009; 324: 255-258.
3. Liljenstrom H, von Heijne G. Translation rate modification by preferential codon usage: Intragenic position effects. *J Theor Biol*. 1987; 124: 43-55.
4. Prabhakaran R, Chithambaram S, Xia X. *Escherichia coli* and *Staphylococcus* phages: Effect of translation initiation efficiency on differential codon adaptation mediated by virulent and temperate lifestyles. *J Gen Virol*. 2015; 96: 1169-1179.
5. Tuller T, Waldman YY, Kupiec M, Ruppin E. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci USA*. 2010; 107: 3645-3650.
6. Xia X. A major controversy in codon-anticodon adaptation resolved by a new codon usage index. *Genetics*. 2015; 199: 573-579.
7. Bull JJ, Badgett MR, Springman R, Molineux IJ. Genome properties and the limits of adaptation in bacteriophages. *Evolution*. 2004; 58: 692-701.
8. Chithambaram S, Prabhakaran R, Xia X. Differential codon adaptation between dsDNA and ssDNA phages in *Escherichia coli*. *Mol Biol Evol*. 2014; 31: 1606-1617.
9. Shine J, Dalgarno L. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA*. 1974; 71: 1342-1346.
10. Shine J, Dalgarno L. Identical 3'-terminal octanucleotide sequence in 18S ribosomal ribonucleic acid from different eukaryotes. A proposed role for this sequence in the recognition of terminator codons. *Biochem J*. 1974; 141: 609-615.
11. Shine J, Dalgarno L. Determinant of cistron specificity in bacterial ribosomes. *Nature*. 1975; 254: 34-38.
12. Steitz JA, Jakes K. How ribosomes select initiator regions in mRNA: Base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc Natl Acad Sci USA*. 1975; 72: 4734-4738.

13. Taniguchi T, Weissmann C. Inhibition of Qbeta RNA 70S ribosome initiation complex formation by an oligonucleotide complementary to the 3' terminal region of *E. coli* 16S ribosomal RNA. *Nature*. 1978; 275: 770-772.
14. Kozak M. Mechanism of mRNA recognition by eukaryotic ribosomes during initiation of protein synthesis. *Curr Top Microbiol Immunol*. 1981; 93: 81-123.
15. Kozak M. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res*. 1981; 9: 5233-5252.
16. Xia X. The +4G site in Kozak consensus is not related to the efficiency of translation initiation. *PLoS One*. 2007; 2: e188.
17. Hui A, de Boer HA. Specialized ribosome system: Preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc Natl Acad Sci USA*. 1987; 84: 4762-4766.
18. Abolbaghaei A, Silke JR, Xia X. How changes in anti-SD sequences would affect SD sequences in *Escherichia coli* and *Bacillus subtilis*. *G3*. 2017; 7: 1607-1615.
19. Xia X. Bioinformatics and translation initiation. *Bioinformatics and the cell: Modern computational approaches in genomics, proteomics and transcriptomics*. Springer, Cham; 2018. p. 173-195.
20. Xia X. DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. *J Hered*. 2017; 108: 431-437.
21. Wei Y, Silke JR, Xia X. Elucidating the 16S rRNA 3' boundaries and defining optimal SD/aSD pairing in *Escherichia coli* and *Bacillus subtilis* using RNA-Seq data. *Sci Rep*. 2017; 7: 17639.
22. Xia X. Bioinformatics and translation elongation. *Bioinformatics and the cell: Modern computational approaches in genomics, proteomics and transcriptomics*. Springer, Cham; 2018. p. 197-238.
23. Melancon P, Leclerc D, Destroismaisons N, Brakier-Gingras L. The anti-Shine-Dalgarno region in *Escherichia coli* 16S ribosomal RNA is not essential for the correct selection of translational starts. *Biochemistry*. 1990; 29: 3402-3407.
24. Sartorius-Neef S, Pfeifer F. In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Mol Microbiol*. 2004; 51: 579-588.
25. Nakamoto T. A unified view of the initiation of protein synthesis. *Biochem Biophys Res Commun*. 2006; 341: 675-678.
26. Sprengart ML, Fuchs E, Porter AG. The downstream box: An efficient and independent translation initiation signal in *Escherichia coli*. *Embo J*. 1996; 15: 665-674.
27. Duval M, Korepanov A, Fuchsbauer O, Fechter P, Haller A, Fabbretti A, et al. *Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol*. 2013; 11: e1001731.
28. Vellanoweth RL, Rabinowitz JC. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol Microbiol*. 1992; 6: 1105-1114.
29. Sulthana S, Deutscher MP. Multiple exoribonucleases catalyze maturation of the 3' terminus of 16S ribosomal RNA (rRNA). *J Biol Chem*. 2013; 288: 12574-12579.
30. Deutscher MP. Twenty years of bacterial RNases and RNA processing: How we've matured. *RNA*. 2015; 21: 597-600.

31. Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, et al. Secondary structure model for bacterial 16S ribosomal RNA: Phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res.* 1980; 8: 2275-2293.
32. Lin YH, Chang BC, Chiang PW, Tang SL. Questionable 16S ribosomal RNA gene annotations are frequent in completed microbial genomes. *Gene.* 2008; 416: 44-47.
33. Nakagawa S, Niimura Y, Miura K, Gojobori T. Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci USA.* 2010; 107: 6382-6387.
34. Starmer J, Stomp A, Vouk M, Bitzer D. Predicting Shine-Dalgarno sequence locations exposes genome annotation errors. *PLoS Comput Biol.* 2006; 2: e57.
35. Jones CE, Brown AL, Baumann U. Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics.* 2007; 8: 170.
36. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007; 35: 3100-3108.
37. Silke JR, Wei Y, Xia X. RNA-Seq-Based analysis reveals heterogeneity in mature 16S rRNA 3' termini and extended anti-Shine-Dalgarno motifs in bacterial species. *G3.* 2018; 8: 3973-3979.
38. Xia X. ARSDA: A new approach for storing, transmitting and analyzing transcriptomic data. *G3.* 2017; 7: 3839-3848.
39. Kozak M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell.* 1980; 19: 79-90.
40. Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* 2013; 41: 474-486.
41. Jou WM, Haegeman G, Ysebaert M, Fiers W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature.* 1972; 237: 82-88.
42. Ptashne M. A genetic switch: Phage lambda revisited. 3rd ed. Cambridge, MA: Cell Press and Blackwell Scientific; 2004. 154 p.
43. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* 1961; 3: 318-356.
44. Xia X. Translation control of HAC1 by regulation of splicing in *Saccharomyces cerevisiae*. *International journal of molecular sciences.* 2019; 20: 2860.
45. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res.* 2003; 31: 3429-3431.
46. Xia X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol Biol Evol.* 2018; 35: 1550-1552.
47. Zid BM, Rogers AN, Katewa SD, Vargas MA, Kolipinski MC, Lu TA, et al. 4E-BP extends lifespan upon dietary restriction by enhancing mitochondrial activity in *Drosophila*. *Cell.* 2009; 139: 149-160.
48. Xia X, Holcik M. Strong eukaryotic IRESs have weak secondary structure. *PLoS One.* 2009; 4: e4136.
49. Xia X. Bioinformatics and translation termination in bacteria. *Bioinformatics and the cell: Modern computational approaches in genomics, proteomics and transcriptomics.* Springer, Cham; 2018. p. 239-254.
50. Wei Y, Xia X. The role of +4U as an extended translation termination signal in bacteria. *Genetics.* 2017; 205: 539-549.
51. Wei Y, Wang J, Xia X. Coevolution between stop codon usage and release factors in bacterial species. *Mol Biol Evol.* 2016; 33: 2357-2367.

52. Ikemura T. 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.* 1981; 151: 389-409.
53. Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol.* 1981; 146: 1-21.
54. Ikemura T. 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.* 1982; 158: 573-597.
55. Ikemura T. Correlation between codon usage and tRNA content in microorganisms. In: *Transfer RNA in protein synthesis.* Boca Raton: CRC Press; 1992. p. 87-111.
56. Xia X. How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics.* 1998; 149: 37-44.
57. Carullo M, Xia X. An extensive study of mutation and selection on the wobble nucleotide in tRNA anticodons in fungal mitochondrial genomes. *J Mol Evol.* 2008; 66: 484-493.
58. Xia X. Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene.* 2005; 345: 13-20.
59. Wei Y, Silke JR, Xia X. An improved estimation of tRNA expression to better elucidate the coevolution between tRNA abundance and codon usage in bacteria. *Sci Rep.* 2019; 9: 3184.
60. Xia X. Rapid evolution of animal mitochondria. In: *Evolution in the fast lane: Rapidly evolving genes and genetic systems.* Oxford: Oxford University Press; 2012. p. 73-82
61. Xia X, Huang H, Carullo M, Betran E, Moriyama EN. Conflict between translation initiation and elongation in vertebrate mitochondrial genomes. *PLoS One.* 2007; 2: e227.
62. Robinson M, Lilley R, Little S, Emtage JS, Yarranton G, Stephens P, et al. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res.* 1984; 12: 6663-6671.
63. Sorensen MA, Kurland CG, Pedersen S. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol.* 1989; 207: 365-377.
64. Ngumbela KC, Ryan KP, Sivamurthy R, Brockman MA, Gandhi RT, Bhardwaj N, et al. Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 *gag* in intact T cells. *PLoS One.* 2008; 3: e2356.
65. Haas J, Park E-C, Seed B. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr Biol.* 1996; 6: 315-324.
66. Bulmer M. The effect of context on synonymous codon usage in genes with low codon usage bias. *Nucleic Acids Res.* 1990; 18: 2869-2873.
67. Palidwor GA, Perkins TJ, Xia X. A general model of codon bias due to GC mutational bias. *PLoS One.* 2010; 5: e13431.
68. Xia X. The cost of wobble translation in fungal mitochondrial genomes: Integration of two traditional hypotheses. *BMC Evol Biol.* 2008; 8: 211.
69. Prabhakaran R, Chithambaram S, Xia X. Aeromonas phages encode tRNAs for their overused codons. *Int J Comput Biol Drug Des.* 2014; 7: 168-182.
70. Chithambaram S, Prabhakaran R, Xia X. The effect of mutation and selection on codon adaptation in *Escherichia coli* bacteriophage. *Genetics.* 2014; 197: 301-315.

71. van Weringh A, Ragonnet-Cronin M, Pranckeviciene E, Pavon-Eternod M, Kleiman L, Xia X. HIV-1 modulates the tRNA pool to improve translation efficiency. *Mol Biol Evol.* 2011; 28: 1827-1834.



Enjoy *OBM Genetics* by:

1. [Submitting a manuscript](#)
2. [Joining in volunteer reviewer bank](#)
3. [Joining Editorial Board](#)
4. [Guest editing a special issue](#)

For more details, please visit:

<http://www.lidsen.com/journals/genetics>