Non-random arrangement of synonymous codons in archaea coding sequences

Yan-Mei Zhang,1, Zhu-Qing Shao,1, Le-Tian Yang, Xiao-Qin Sun, Ya-Fei Mao, J.-Q. Chen, Bin Wang

ARTICLE INFO

Article history:
Received 28 August 2012
Accepted 9 April 2013
Available online 17 April 2013

Keywords:
mRNA translation
Genetic selection
Codon
Regulation of gene expression
Archaeal genes

ABSTRACT

Non-random arrangement of synonymous codons in coding sequences has been recently reported in eukaryotic and bacterial genomes, but the case in archaeal genomes is largely undetermined. Here, we systematically investigated 122 archaeal genomes for their synonymous codon co-occurrence patterns. We found that in most archaeal coding sequences, the order of synonymous codons is not arranged randomly, but rather some successive codon pairs appear significantly more often than expected. Importantly, such codon pairing bias (CPB) pattern in archaea does not seem to completely follow the co-tRNA codon pairing (CCP) rule previously reported for eukaryotes, but largely obeys an identical codon pairing (ICP) rule. Further, synonymous codon permutation test demonstrated that in many archaeal genomes, random mutation alone is unable to cause the observed high level of ICP bias, which strongly indicates that selection force has been involved to shape synonymous codon orders, potentially meeting a global requirement to optimize translation rate.

1. Introduction

Translation is a fundamental cellular activity shared by all living organisms on earth and different organisms appear to exercise similar strategies so as to optimize their translation. One well-known strategy, codon usage bias (CUB), has been reported for many organisms, including bacteria, eukaryotes and also some archaea [1–9]. The key point of CUB strategy is that for a given genome, depending on its tRNA composition and wobble base modification, some codons are translated more efficiently and/or accurately by the ribosome than their synonyms. These ‘optimal’ codons were often found to be preferred in high expression genes, reflecting translational selection on codon choices of these genes [1,2,4,10–13].

However, in a genome, a majority of genes do not use optimal codons at high frequencies [14]. Are codon choices of these genes also subject to selection? Shedding light on this, two recent studies reported interesting roles for synonymous codon arrangement order in translational dynamics [15,16]. By investigating 27 organisms covering all three life domains, Tuller et al. [16] reported a universal ‘Ramp strategy’: 30–50 non-optimal codons are preferably used at the start region of a gene, presumably to slow down the elongation rate at an early stage, thereby reducing ribosomal traffic and minimizing the overall cost of translation. Also, Cannarozzi et al. [15] revealed that in eukaryotic coding sequences, two successively appearing synonymous codons (need not to be neighbors) are significantly more often to share the same tRNA than expected. This interesting pattern of co-tRNA codon pairing (CCP) bias was thought to be beneficial in improving translational efficiency. Indeed, a version of the GnP gene with its synonymous codon order obeying the CCP rule showed a 29% improvement in translational elongation rate comparing to a version fully against CCP rule [15]. It should be noted here that co-tRNA codon pairs include two types: 1) pairs formed by two identical codons (certainly read by same tRNAs) and 2) pairs formed by two non-identical but co-tRNA codons. In the original data reported by Cannarozzi et al. [15], identical codon pairing (ICP) is often found more biased than the other type. Overall, these recent findings revealed that translational selection can exert influences on their synonymous codon orders so as to optimize translation.

Both the CUB and the ‘Ramp’ strategies have been universally observed among all three life domains on earth. However, the codon pairing bias (CPB) strategy was only revealed in eukaryotes. We had posit the question: Is it also the case for other life domains? In an earlier study, we investigated 773 bacterial genomes for their synonymous codon orders and found that the CPB pattern is also prevalent in prokaryotes [17]. However, mainly ICP was found to be preferentially biased, whereas non-identical CCP was often not biased significantly in many prokaryotes. Herein, we aimed to uncover whether CPB is true in archaea and whether its pattern is consistent with the CCP rule previously reported in eukaryotes.

To answer this question, one first needs to understand the general pattern of tRNA composition in archaea as well as their decoding strategies. Although the correspondence between the 64 genetic codons and the 20 amino acids is common among three life domains, the type and
number of tRNAs assigned to degenerate codon families may vary among different life domains due to their distinct modification systems [18]. Different from bacterial genomes which often show varied tRNA copy numbers, archaeal genomes are known for their simple and stable tRNA composition. It can be divided into two major groups [19]. In non-Methanococcus-like archaea, usually three different tRNAs (G34-tRNA, C34-tRNA, and U34-tRNA) are available to decode four synonymous codons in a codon box (formed as four-fold-degenerate codon box). According to a ‘wobbling parsimony’ criterion, (i.e. avoiding third-position wobbling unless there is no canonically paired tRNA available [10]), the G34-tRNA would read both U3 and C3 codons, while C34-tRNA would read G3 codon and U34-tRNA would read A3 codons [18]. In Methanococcus-like archaea, generally only two different tRNAs (G34-tRNA and U34-tRNA) exist for a four-fold-degenerate codon box. In such cases, the G34-tRNA would still read both U3 and C3 codons, while U34-tRNA would read A3 and G3 codons [21]. Therefore, U- and C-ending codons and A- and G-ending codons can often be regarded as co-tRNA codons. It would be interesting to check whether these co-tRNA codons successively appeared more often than expected in archaeal genomes.

In this study, a thorough analysis of CPB pattern in 122 archaeal genomes was conducted. Methanosaeta acetivorans, an intensively studied methane-producing archaea, was chosen as a sample species. By analyzing its coding sequences, we demonstrated that its overall synonymous codon order is clearly biased toward ICPs, whereas enrichment of non-identical CCPs was rarely observed. Further analyses showed that the observed CPB pattern was neither influenced by CUB nor ‘Ramp’ strategies, nor caused by random mutation alone, suggesting that it is a new common strategy shared by all organisms on earth to optimize their translation.

2. Materials and methods

2.1. Data used

All coding sequences of 122 archaeal genomes were downloaded from the NCBI FTP server (ftp://ncbi.nlm.nih.gov) by May, 2012. A complete list of species as well as related information is provided in Supplementary Table S1.

2.2. Calculation of significantly biased codon pairs in all 122 archaeal genomes

A synonymous codon pair is formed by two successively appearing synonymous codons in a coding sequence, regardless of how many codons occur between them. Synonymous codon pairs can be divided into three categories: identical codon pairs (ICPs), non-identical but co-tRNA codon pairs (non-identical CCPs), and other codon pairs (OCPs). For each archaeal genome, the occurrence number and frequency of each synonymous codon pair were first calculated (Personal Perl script, available upon request). Then, the expected frequency of given codon pair was computed as the products of two individual codons which occurred in whole genome. Finally, the extent of deviations away from the expected value was assessed by assuming a binomial distribution same as in previous studies [15,17]. If the observed number of a synonymous codon pair deviated >3 standard deviations (SDs) away from the expected value, then the occurrence of this pair was regarded as significantly more often than expected (favored). For each of the 122 archaeal genomes, all significantly favored codon pairs were recorded.

Also, using Proteinortho software [20], we retrieved 15 conserved ribosomal protein gene sequences from all 122 archaeal genomes, then constructed a phylogenetic tree based on Maximum Likelihood method (PhyML v3.0.1, [21]) as previously described [22]. Information of significantly favored codon pairs for all the 122 archaea species was then projected onto the phylogenetic tree.

2.3. Testing the influences of CUB and Ramp strategies on observed CPB pattern inarchaeal species M. acetivorans

In order to test the influences of CUB on the detected CPB pattern in M. acetivorans, we need to exclude genes subject to CUB as more as we can. According to a recent study [14], about 12% genes in M. acetivorans genome showed characteristics of CUB, including those ribosomal protein genes and elongation factor genes. All these genes were then removed from the dataset. Using the remaining genes, the level of CPB was assessed by following the procedures described above.

To test the influences of Ramp strategies on the detected CPB pattern in M. acetivorans, we removed the beginning 50 codons from all coding genes. Using the remaining data, the level of CPB was assessed.

To exclude the influences of both strategies on CPB pattern, all genes subject to CUB and the beginning 50 codons of remaining genes were removed. Using the remaining data, the level of CPB was assessed.

2.4. Synonymous codon permutation test and gene resampling

To determine whether the observed pattern of CPB was due to differential influences of mutational force and drift on different parts of a gene, 1000 iterations of within-gene synonymous codon random shuffling were performed. For each shuffle, a permutation on the locations of all synonymous codons in every gene was done, but leaving amino acid sequence as well as codon usage frequencies unchanged. The level of CPB was then repeatedly assessed using 1000 permuted data. This test was basically done for all 122 archaeal genomes.

One may also concern that for a genome showing overall high level of CPB, it may be possible that only a few genes with strong ICP caused the overall bias pattern. To test this possibility, we performed 1000 iterations of gene re-sampling for M. acetivorans data. Each time only 50% of the genes were re-sampled and used to assess the CPB level. Significantly biased codon pairs detected in original data were then evaluated one by one for their bias level among all resampling datasets.

3. Results and discussion

3.1. CPB pattern in M. acetivorans largely follow an ICP rule

To better present our findings on synonymous codon arrangement pattern in archaea, we chose an intensively studied species, M. acetivorans, as an example. Belonging to the group Euryarchaeota and subgroup Methanomicrobia, M. acetivorans has by far the largest known archaeal genome (5.75 Mb), containing a total of 4540 protein-coding genes [23]. For each of 18 degenerate codon families, the actual numbers of all possible synonymous codon pairs in the M. acetivorans genome were calculated and compared with their expected values (see the Materials and methods section). Table 1A provides the SD results obtained for the glycine codon family (GGN) and all detailed data is available in Table S2. Apparently, the overall pairing pattern for glycine codons is not neutral. On the diagonal line, four identical codon pairs occurred 4893, 3354, 14347 and 6735 times, respectively (Table S2). These numbers were significantly higher than expected (>3 SDs, Table 1A), suggesting that the arrangement pattern of glycine codons is clearly biased towards ICP in M. acetivorans genome. Meanwhile, many non-identical codon pairs occurred significantly less than expected (<−3 SDs, Table 1A).

Interestingly, except for the four ICPs, no other positively biased (>3 SDs) pairs were observed (Table 1A). This is not fully consistent with the previously reported CCP rule in eukaryotes [15]. According to the Genomic tRNA Database [24], three different isoaccepting tRNAs (tRNA_{GCG}, tRNA_{ GCC}, and tRNA_{GCA}) are responsible for decoding four glycine codons (GGN) in the M. acetivorans genome. In this case, tRNA_{GCG} is expected to read both GGC and GGT codons; while tRNA_{ GCC} would read the GGC codon, and tRNA_{GCA} would read the GGA codon [18]. If the CCP rule were obeyed, one would expect the GGC–GGT
Table 1
Standard deviations (SDs) that are calculated away from expectations for all codon pairs of Gly codon family in Methanosarcina acetivorans.

<table>
<thead>
<tr>
<th>Gly codon</th>
<th>GGC</th>
<th>GTT</th>
<th>GGA</th>
<th>GGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGC</td>
<td>+9.55</td>
<td>−1.16</td>
<td>−6.50</td>
<td>+0.03</td>
</tr>
<tr>
<td>GTT</td>
<td>−1.64</td>
<td>+10.35</td>
<td>−0.60</td>
<td>−6.33</td>
</tr>
<tr>
<td>GGA</td>
<td>−3.74</td>
<td>−1.62</td>
<td>+6.73</td>
<td>−3.23</td>
</tr>
<tr>
<td>GGG</td>
<td>−2.93</td>
<td>−5.61</td>
<td>−1.50</td>
<td>+9.16</td>
</tr>
</tbody>
</table>

B) 12% of genes subject to CUB were excluded.

<table>
<thead>
<tr>
<th>Gly codon</th>
<th>GGC</th>
<th>GTT</th>
<th>GGA</th>
<th>GGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGC</td>
<td>+6.31</td>
<td>−3.36</td>
<td>−7.36</td>
<td>+0.22</td>
</tr>
<tr>
<td>GTT</td>
<td>−4.64</td>
<td>+8.78</td>
<td>−0.74</td>
<td>−5.83</td>
</tr>
<tr>
<td>GGA</td>
<td>−4.41</td>
<td>−1.81</td>
<td>+8.59</td>
<td>−1.38</td>
</tr>
<tr>
<td>GGG</td>
<td>−2.28</td>
<td>−5.64</td>
<td>+0.32</td>
<td>+10.79</td>
</tr>
</tbody>
</table>

C) The beginning 50 codons of all coding genes were excluded.

<table>
<thead>
<tr>
<th>Gly codon</th>
<th>GGC</th>
<th>GTT</th>
<th>GGA</th>
<th>GGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGC</td>
<td>+11.47</td>
<td>−1.00</td>
<td>−5.40</td>
<td>+1.45</td>
</tr>
<tr>
<td>GTT</td>
<td>−2.46</td>
<td>+7.42</td>
<td>−2.80</td>
<td>−7.45</td>
</tr>
<tr>
<td>GGA</td>
<td>−3.36</td>
<td>−3.33</td>
<td>+6.03</td>
<td>−2.77</td>
</tr>
<tr>
<td>GGG</td>
<td>−1.13</td>
<td>−6.46</td>
<td>−0.36</td>
<td>+10.14</td>
</tr>
</tbody>
</table>

D) Both the genes subject to CUB and the beginning 50 codons of remaining genes were excluded.

<table>
<thead>
<tr>
<th>Gly codon</th>
<th>GGC</th>
<th>GTT</th>
<th>GGA</th>
<th>GGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGC</td>
<td>8.34</td>
<td>−3.13</td>
<td>−6.38</td>
<td>1.68</td>
</tr>
<tr>
<td>GTT</td>
<td>−5.13</td>
<td>5.62</td>
<td>−2.75</td>
<td>−7.05</td>
</tr>
<tr>
<td>GGA</td>
<td>−4.12</td>
<td>−3.50</td>
<td>8.02</td>
<td>−0.94</td>
</tr>
<tr>
<td>GGG</td>
<td>−0.58</td>
<td>−6.23</td>
<td>1.35</td>
<td>11.42</td>
</tr>
</tbody>
</table>

Note: Pairs with significant positive deviations (>3 SDs) from expectation are shown in bold. All expected co-tRNA codon pairs are shaded (Grosjean et al., 2010).

In addition, we also evaluated the two-fold-degenerate codon families as well as the Ile codon family (Table S2). For Asn, Asp, Cys, His, Phe, Tyr and Ser2 codon families, only one isoaccepting tRNA (G34-tRNA) was present in the M. acetivorans genome and it read both C- and T-ending codons in these families. According to the CCP rule observed in eukaryote, all four possible codon pairs in each of these families should be equally favored and thus show no bias. However, this was again not the case. ICPs in these families occurred significantly more often than expected while non-identical codon pairs were significantly disfavored (Table S2). For Glu, Glu, Lys, Arg2, and Leu2 codon families, two isoaccepting tRNAs (C34-tRNA and U34-tRNA) are present and each is expected to read a codon [18]. The calculated results for these families appear to obey the CCP rule, with ICPs occurring much more frequently than expected. Finally, for the Ile codon family, two different isoaccepting tRNAs are available. The G34-tRNA would read both ATC and ATT codons and the other tRNA would read the ATA codon. Again, CCPs (ATC–ATT and ATT–ATC) were not overrepresented in this family, with only ICPs occurring significantly more often than expected (Table S2). When summarizing the synonymous codon pairing patterns for all 18 degenerate codon families, we found that more than 91% of ICPs (54 out of a maximum of 59) were significantly enriched in M. acetivorans genome, whereas none of the non-identical CCPs (0 out of 32) was enriched (Table 2, Table S2).

Therefore, we concluded that instead of the CCP rule, the ICP rule seems to better represent the codon pairing pattern in Methanosarcina acetivorans.

3.2. ICP is a major codon pairing pattern among many investigated archaea

To explore whether ICP is a general codon pairing pattern among archaea, we thoroughly analyzed 122 archaea genomes and divided them into 13 subgroups, based on their phylogenetic relationships. Table 3 summarized the obtained results for each subgroup. (For more detailed results of each species, please refer to Fig. S1).

Several important conclusions regarding the codon pairing pattern in archaea can be made. First, the total number of significantly favored codon pairs (deviation > 3 SDs) was extremely variable among archaeal species, ranging from 0 in Nanoarchaeum equitans (an obligatory symbiont) to 78 in Methanoseta harundinacea (Fig. S1). Even in a specific subgroup, species showed variable values regarding the total numbers of significantly biased codon pairs (Table 3). For example, the number varied from 29 to 62 in Thermoproteales, and from 4 to 48 in Desulfurococcales, etc. This suggests to us that the bias level of the codon pairing pattern could be quickly relaxed or strengthened during species evolution, although close species usually maintain similar levels of CPB. Second, for many archaeal genomes, as in M. acetivorans, a high percentage of significantly favored codon pairs are those identical ones (Table 3, Fig. S1). For 53 out of 122 examined archaeal genomes, only ICPs were detected to be significantly favored. Third, the main ICP pattern observed in archaea is not the same as the previously reported CCP pattern in eukaryotes. In non-Methanococcus-like archaea, generally three different tRNAs (G34-tRNA, C34-tRNA, and U34-tRNA) exist for each codon box and the G34-tRNA would read both U3 and C3 codons [18]. In Methanococcus-like archaea, generally two different tRNAs (C34-tRNA and U34-tRNA) exist for each codon box and the G34-tRNA would still read both U3 and C3 codons, while U3-tRNA would read both A3 and G3 codons [18]. Therefore, both C3, U3, codons and A3, G3 codons can often be regarded as co-tRNA codons. We ask, do these codons successively pair and the GGT–GGC pair to be significantly favored. However, the calculated deviations for these two codon pairs are −1.16 SD and −1.64 SD, respectively (Table 1A).

Besides glycine codon family, what about the cases of the other four-fold-degenerate codon families (Ala, Pro, Thr, Val, Arg4, Leu4, and Ser4)? Table S2 shows detailed codon co-occurrence data for all these codon families. As for the glycine codon family, three different isoaccepting tRNAs (G34-tRNA, C34-tRNA, and U34-tRNA, respectively) are present in these families and the same decoding strategies are applicable [18,24]. Among a total of 28 ICPs in these families, 25 were significantly favored (>3 SDs) and the remaining three were also positively favored (>2 SD). However, among a total of 14 assumed co-tRNA but non-identical codon pairs (NNC–NNT or NNT–NNC), none was significantly favored. Instead, eight were even significantly disfavored.
appear more frequently than expected in archaea? For all 122 archaeal genomes examined in this study, as shown in Fig. S1 and S2, these non-identical CCPs are rarely found to be significantly favored. Therefore, we concluded that instead of CCP, ICP is a major codon pairing pattern.

3.3. Neither known selectional strategies nor mutational force alone caused the observed ICP bias pattern in M. acetivorans genome

It has long been known that an organism's codon usage pattern is determined by both mutational and selectional influences on synonymous codon choices [1,2,25]. We next used M. acetivorans data to assess whether the observed codon pairing pattern was caused by known selectional strategies and/or by random mutation.

Translational selection can influence synonymous codon usages by both CUB and Ramp strategies [1-9,16]. In the former case, high-expression genes preferably use many optimal codons, which would likely increase successive pairing number of these optimal codons. In the other case, the Ramp strategy would drive coding genes to use non-optimal codons at start regions, which also could increase the co-occurrence number of non-optimal codons. The lack of global expression data for M. acetivorans genomes hampered us to remove all genes with high expression level. Fortunately, a recent work estimated that in M. acetivorans, about 514 (12%) genes were influenced by CUB strategy [14]. Therefore, we removed all these genes from the dataset to assess the influences of CUB strategy on CPB pattern. Also, we deleted the beginning 50 codons of all genes in M. acetivorans to test the influences of Ramp strategy on CPB pattern (see the Materials and methods section). As shown in Table 1B, C and D, for glycine codons, the revealed codon pairing pattern is not altered at all. For the remaining 17 amino acids, the assessed CPB patterns had little changes no matter which dataset was used (Table S2), suggesting that CUB and Ramp strategies, either separated or combined, exert little influence on the observed ICP pattern in M. acetivorans.

To test whether mutation alone could cause the codon pairing pattern to reach the bias level observed, we repeatedly shuffled the synonymous codon orders within each gene 1000 times. After each shuffle, we calculated the co-occurrence numbers for all synonymous codon pairs. The results demonstrated that after shuffling, the total number of ICPs in a codon family was often significantly reduced. For example, in the glycine codon family, the actual occurrence number for four ICPs collectively was 29,329; but for the shuffled data, the mean was only 28,744 with a SD of 130. The difference between these values looks slight but is highly significant ($P < 0.00001$, t-test), with shuffled numbers never reaching the observed number (Fig. 1A). In M. acetivorans, significant reductions on the total number of ICPs were also detected in 10 other degenerate codon families (Table S3). This suggests that ICPs in these codon families probably have been enriched by forces other than random mutation, e.g., translational selection.

When same codon permutation procedures were applied to other 121 archaeal genomes, the results showed that for many species with high CPB levels of ICP bias, significant decrease on the total number of ICPs was often observed in many codon families (Table S3). However, significant reductions were detected less frequently for species showing weak level of ICP bias. This general trend indicates that within-gene codon orders are more likely arranged randomly in species showing overall low CPB bias level.

3.4. A majority of genes collectively contribute to the biased ICP pattern

One intriguing question is: how many genes in a genome, e.g. M. acetivorans, need to be influenced so as to reach a high bias level of codon pairing at whole genome level? One may suspect that only a small percent of genes showing strong CPB pattern (due to CUB selection or other reasons) would be able to cause the observed CPB pattern at whole genomic level (e.g. about 12% genes in M. acetivorans genome showed significant CUB). To test this concern, we performed 1000 iterations of gene re-sampling using M. acetivorans data. Each time only 50% of the genes were re-sampled and used to assess the CPB level. Among a total of 4540 coding genes, we randomly picked half to assess the codon co-occurrence bias. If only a minority of genes contributed to the observed ICP bias pattern, one would expect that at least in some cases, the re-sampled data would not show a significant level of biased ICP bias. However, we found that this was not the case. Taking glycine as an example, among 1000 iterations of re-sampling, the SD values were found >3 in 993, 1000, 944 and 998 cases for the GCC–GCG, GCT–GCT, GCA–GAA and GCG–GGG pairs, respectively. Similar consistent results were also observed for other 50 identified biased ICPs (Fig. S3). Even if we only resample 20% genes for 1000 times, the originally identified CPB bias towards ICP can still be steadily detected (data not shown but available upon request). Therefore, such results indicate to us that in the M. acetivorans genome, a majority of genes have been influenced so as to contribute the overall biased ICP pattern.

3.5. Why did this CPB pattern develop?

M. acetivorans (and many other archaeal species) appears to have tuned its translation such that the majority of genes are translated efficiently. In most archaea, synonymous codon pairing was significantly biased towards ICP. Mutational alone is unlikely to cause synonymous codon orders to reach the observed bias level in its genome. The currently known CUB and Ramp strategies of translational selection only influence some or part of the coding genes. After removing these genes or regions, the observed ICP pattern still held true. Collectively these results suggest that selection force is likely involved and that it is different from that in CUB and Ramp strategies. In their original study, Cannarozzi et al. [15] proposed a tRNA recycle model which stated that by forming co-tRNA codon pairs, a tRNA molecule used in the former codon position could be reused by ribosomes to read the next codon, thus avoiding waiting times in changing isoaccepting tRNAs, improving translational elongation rate, and reducing global ribosomal costs.

A precondition for this tRNA recycle model is worth mentioning: That is, in a cell, the total tRNAs should be in short supply compared to ribosomes. If sufficient tRNA molecules are available for each ribosome to read synonymous codons, there is no benefit to recycle a used tRNA. On the contrary, if a shortage is in fact the case, and on average a working ribosome is only surrounded by an incomplete set of tRNAs, then more time would elapse when a codon is to be read by a currently unavailable tRNA, and the ribosomal A site would be at risk of being occupied by near-cognate aa-tRNAs [26]. Available evidence seems to support such a
“shortage” scenario. In bacterial cells, only 13 tRNA molecules on average are available for each ribosome under normal growth conditions, which is far less than a complete set [27]. A similar reasoning for tRNA shortages in yeast and bacteria has also been used in a recent study [28].

It is well known that although sharing a common genetic code and tRNA sets, the three life domains have developed their unique translation-apparatus, such as the composition of ribosomes and different tRNA modification systems [18]. Acknowledging the proposed tRNA recycle model in eukaryotes [15] but not favoring the CCP rule, we wonder why identical codon pairs are mainly selected for in archaea? Next, we will discuss three possible mechanisms that could explain this ICP pattern.

In the first case, we propose that although a tRNA could read more than one codon (e.g., G34-tRNA would read both C3 and U3 codons in many codon boxes), wobble base-modification status may differentiate its interaction with different codons. It has been shown that in prokaryotes and eukaryotes (unclear presently for archaea), G34-tRNA could remain unmodified in some cases while modified into queuosine (Q) in other cases [29–32]. Unmodified G34 showed a higher reading affinity with the C3 codon, but modified Q34 could read U3 more efficiently [33,34]. Discrepant reading affinities were also detected between codons and variously modified U34-tRNAs in prokaryotes [35,36]. Discrepant reading affinities were also detected between codons and variously modified U34-tRNAs in prokaryotes [35,36].

In the second case, we consider conformational changes of a wobble tRNA species in reading different codons. Wobble occurs because the conformation of the tRNA anticodon permits flexibility in order to reach a stable and high-affinity interaction with its cognate codon. For a wobble tRNA to read two different codons, (e.g., G34-tRNA to read C3 and U3 codons or U34-tRNA to read A3 and G3 codons), the wobble base would use different conformations to form stable interactions with cognate codons. In the tRNA recycle model [15], if a reused tRNA molecule from a previous codon needed to change conformation (a time-consuming proposition) to read a co-tRNA but a non-identical codon, then selection would not favor these different co-tRNA codons occurring successively, but instead favor only identical codons. Unfortunately, there is no evidence to support the belief that a used tRNA would maintain its pairing conformation after it exits the ribosome.

In the third case, we considered the combined influences of translational selection and mutation on synonymous codon orders. As suggested by the co-tRNA rule, translational selection would influence codon choice to have more co-tRNA codons (e.g., C3 and T3 codons read by G34 tRNA) and to pair together. However, local mutational bias (GC biased or AT biased) would then drive non-identical codon pairs (C3–T3 or T3–C3) to become (C3–C3 or T3–T3). If mutated to A3 or G3, then the two codons are not co-tRNA codons anymore (since A3 and G3 codons are usually read by different tRNAs) and they would be subject to co-tRNA selection again. Thus
in this scenario, co-tRNA selection and local mutation act together to shape codon orders in coding genes and finally result in the observed biased ICP pattern.

There is currently no scientific support regarding any of the aforementioned three explanations. However, the biased pattern for ICP is indeed true in both the archaea and bacterial domains. The mechanism behind this synonymous codon co-occurrence pattern is worth further exploration.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.04.008.

Acknowledgments

We thank Professor Glenn R. Björk and three anonymous reviewers for critical reading of the manuscript. This work was supported by the National Natural Science Foundation of China (30930008, 31000105 and 31170210), National Postdoctoral Science Foundation of China (20090461092, 201003570), Nanjing University Start Grant, and Postgraduate Students Innovation Project of Jiangsu Province (CXXZ11_0038).

References