Gene 590 (2016) 29–34

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Old genes experience stronger translational selection than young genes

Hongyan Yin [a,b], Lina Ma [a], Guangyu Wang [a,b], Mengwei Li [a,b], Zhang Zhang [a,*]

[a] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, Beijing 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 27 May 2016
Accepted 29 May 2016
Available online 31 May 2016

Keywords:
Translational selection
Gene age
Codon usage bias
Singleton
Duplicate
Housekeeping gene

ABSTRACT

Selection on synonymous codon usage for translation efficiency and/or accuracy has been identified as a widespread mechanism in many living organisms. However, it remains unknown whether translational selection associates closely with gene age and acts differentially on genes with different evolutionary ages. To address this issue, here we investigate the strength of translational selection acting on different aged genes in human. Our results show that old genes present stronger translational selection than young genes, demonstrating that translational selection correlates positively with gene age. We further explore the difference of translational selection in duplicates vs. singletons and in housekeeping vs. tissue-specific genes. We find that translational selection acts comparably in old singletons and old duplicates and stronger translational selection in old genes is contributed primarily by housekeeping genes. For young genes, contrastingly, singletons experience stronger translational selection than duplicates, presumably due to redundant function of duplicated genes during their early evolutionary stage. Taken together, our results indicate that translational selection acting on a gene would not be constant during all stages of evolution, associating closely with gene age.

1. Introduction

Codon usage bias (CUB) is a ubiquitous feature in a large variety of species and is shaped by a complex balance from natural selection and mutational bias (Bulmer, 1991; Duret, 2002; Hershberg and Petrov, 2008). A significant positive correlation between CUB and gene expression level is indicative of translational selection as biased codon usage arises considerably from selection for translational efficiency and/or accuracy (Akashi and Eyre-Walker, 1998; Lavner and Kotlar, 2005; Plotkin and Kudla, 2011). As genes are created by different mechanisms and possess different evolutionary origins (Ohno, 1970; Long et al., 2003; Zhang, 2003), it is likely that genes with different evolutionary ages experience diverse selection on synonymous codon usage. Therefore, detection of the strength of translational selection acting on different aged genes is of great significance for deciphering the underlying forces in molecular sequence evolution.

As the efficacy of translational selection is influenced by population size, over the past years, translational selection has been extensively documented in prokaryotes (Ikemura, 1981; Stoletzki and Eyre-Walker, 2007; Supek et al., 2010) and unicellular eukaryotes (Sharp and Li, 1986; Percudani et al., 1997; Akashi, 2003). Albeit it is argued that mutation plays a significant role in shaping biased codon usage (Chen et al., 2004; Rao et al., 2011) and that translational selection might be too weak to be detectable in species with small effective population sizes (Hershberg and Petrov, 2008; Doherty and McInerney, 2013), evidence has accumulated that translational selection is a widespread mechanism identified in many living organisms (dos Reis et al., 2004; Doherty and McInerney, 2013), including *Drosophila melanogaster* (Akashi, 1994; Dunn et al., 2001), *Caenorhabditis elegans* (Duret, 2000; Zhou et al., 2010), plants (Morton and Wright, 2007; Gu et al., 2012; Liu, 2012) and vertebrates (Musto et al., 2001; Urrutia and Hurst, 2003; Plotkin et al., 2004; dos Reis and Wernisch, 2009; Doherty and McInerney, 2013). Moreover, our recent study has further demonstrated that translational selection in human is more pronounced in housekeeping (HK) genes than tissue-specific (TS) genes (Ma et al., 2014), indicating that translational selection might be related to gene age given the antiquity of HK genes. Regarding gene age, on the other hand, studies have shown that in human young genes experience more variable selection pressure (Vishnoi et al., 2010) and evolve faster than old ones (Alba and Castresana, 2005; Wolf et al., 2009; Zhou et al., 2014). Additionally, young human genes are found to be most likely to present distinct expression patterns in temporal and spatial context (Long et al., 2013) and less likely to be essential than old genes (Chen et al., 2012).

Despite this, few efforts have been devoted to detecting translational selection with a consideration of gene age. Accordingly, we hypothesize that translational selection acts differentially on human genes associating closely with their different evolutionary ages. To test this hypothesis, here we investigate the strength of translational selection in human by associating with gene age. As genes have different evolutionary origins as well

Abbreviations: HK gene, Housekeeping gene; TS gene, Tissue-specific gene; EIG, Expression-invariable gene; CUB, Codon usage bias; RSCU, Relative synonymous codon usage.

* Corresponding author at: CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, Beijing 100101, China.
E-mail address: zhangzhang@big.ac.cn (Z. Zhang).

http://dx.doi.org/10.1016/j.gene.2016.05.041
0378-1119/

as diverse expression patterns, we further examine translational selection in duplicates vs. singletons and in HK vs. TS genes, respectively.

2. Materials and methods

2.1. Gene age

Gene age annotations were collected from (Domazet-Loso and Tautz, 2008; Neme and Tautz, 2013) coupled with orthology relationship in PANTHER (Mi et al., 2013), where evolutionary age for any given human gene was defined based on the presence of a homolog in a wide range of species from single-celled organisms to Primates. Considering the possibility of horizontal gene transfer among single-celled organisms, we determined the evolutionary age of human genes with orthologs in more than two single-celled organisms. Totally, we categorized human genes into five age groups (Table S1 and Fig. S1) and obtained the corresponding divergence time from TimeTree (Hedges, et al., 2015), including old (single-celled organisms about 1,300 million years ago; 13,204 genes), medium-old (from Holozoa to Eumetazoa about 900 million years ago; 2799 genes), medium (from Bilateria to Olfactores about 730 million years ago; 602 genes), medium-young (from Vertebrate to Amniota about 320 million years ago; 1378 genes) and young (from Mammalia to Primates; 743 genes).

2.2. Data collection

We collected RNA-Seq data for 32 tissues with FPKM (Fragments Per Kilobase of transcript per Million fragments sequenced) values for all human genes from Uhlen et al. (2015)). Expression levels with FPKM > 1 were used for further analysis. For a given tissue, genes were divided into five groups according to expression level (Table S2), namely, low (L), medium-low (ML), medium (M), medium-high (MH) and high (H). Followed by Yanai et al. (2005) and Zhao et al. (2014), we calculated τ, an index of tissue specificity, to estimate the variability of gene expression across different tissues. The τ value was defined as

$$\tau = \sum_{i=1}^{n}[1-\log_2 x(i)/\log_2 max(x)] \Big/ _{n-1}, \quad (1)$$

where n (=32) is the total number of tissues, $x(i)$ is expression level of a given gene in tissue i and $max(x)$ is the highest expression level across all examined tissues. HK genes were those that were expressed in all 32 tissues and had $\tau < 0.95$ (13,648 genes), and the rest were non-HK genes (5078 genes). Among non-HK genes, TS genes were defined as genes with $\tau > 0.95$ (3469 genes). Additionally, we strictly defined HK genes with $\tau < 0.4$, named expression-invariable genes (EIG; 1034 genes). The lists of HK genes, EIGs and TS genes were tabulated into Table S3. We obtained human duplicates (6922 genes; Table S4) from (Zhang et al., 2011) that were identified based on sequence alignment and chromosomal location. We extracted tRNA copy numbers for human from the genomic tRNA database (Chan and Lowe, 2009).

2.3. Codon usage analysis

We utilized a method named Codon Deviation Coefficient (Zhang et al., 2012) to estimate CUB, which outperformed existing methods including CAI (Carbone et al., 2003) and Effective Number of Codons (Wright, 1990; Sun et al., 2013). We calculated relative synonymous codon usage (RSCU) by CodonW (http://codonw.sourceforge.net/). The similarity between RSCU and tRNA abundance was calculated by cosine similarity, viz.,

$$\cos\theta = \sum_{i=1}^{n} x_i y_i \Big/ \sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} x_i^2}, \quad (2)$$

where n is the total number of codons, x_i is RSCU of a given codon i, and y_i is tRNA copy number of codon i.

3. Results and discussion

3.1. Translational selection correlates positively with gene age

To detect the strength of translational selection acting on different aged genes, we divide human genes according to their evolutionary origins into five different age groups (Table S1; see methods), viz., young ($n = 743$), medium-young ($n = 1378$), medium ($n = 602$), medium-old ($n = 2799$) and old ($n = 13{,}204$). Based on gene ontology enrichment analysis (Table S5), we find that different aged genes are likely to possess divergent functions. If translational selection is stronger in one gene, codon usage in this gene is likely to be more biased toward abundant tRNAs. To test this possibility, we calculate relative synonymous codon usage (RSCU) and examine the similarity between tRNA abundance and codon usage among different age groups (Fig. 1A). A higher similarity implies the consistency between tRNA abundance and codon usage. We find that, although the similarity between tRNA abundances and codon usage is slightly but significantly different across all five age groups (P-value < 0.05), it is still observed that the similarity increases as genes age; old genes show the highest similarity and young genes show the lowest similarity, demonstrating that old genes use codons more biased toward tRNA abundance. This result indicates that translational selection correlates positively with gene age, acting more prominently in old genes. It should be noted, however, that tRNA abundances may vary across different tissues (Dittmar et al., 2006). Therefore, to further validate the above result, we need to provide more evidence by directly factoring gene expression data.

As mentioned, highly expressed genes tend to have more biased codon usage to achieve higher translational efficiency and/or accuracy. In addition to evidence from tRNA abundance, translational selection is detectable by examining the correlation between CUB and gene expression level. Therefore, we examine the correlation between CUB and gene expression level in different age groups and across 32 tissues using contingency table chi-square test (Fig. 1B–G). Across all examined 32 tissues, it is consistently found that in old genes CUB correlates significantly with gene expression (Fig. 1G; P-value < 0.05). Taking adipose as an example, old genes present a significantly stronger positive correlation between CUB and gene expression (P-value < 10^{-3}; Fig. 1B), indicating stronger translational selection operating on old genes. Contrastingly, such pattern is absent in other age groups (P-value > 0.05; Fig. 1D-F), suggesting relatively weak translational selection. Similar results are obtained when we perform the above analysis on all examined tissues (Fig. 1G and Fig. S2). Collectively, these results indicate that translational selection correlates positively with gene age and is increasingly stronger as genes age. It is also suggested that translational selection, albeit relatively weak in contrast to prokaryotes and unicellular eukaryotes, is universally operative in higher eukaryotes, as testified by stronger translational selection in old genes across a diversity of human tissues.

3.2. Stronger translational selection in singletons

Gene duplication contributes considerably to gene birth (Zhang, 2003; Kaessmann, 2010). Evidence has accumulated that duplicates usually evolve slowly and possess elevated expression levels by comparison with singletons (Davis and Petrov, 2004; Jordan et al., 2004; Yang and Gaut, 2011). Accordingly, we investigate whether translational selection acts differentially in duplicates and singletons considering their different ages. To address this issue, we first examine the corresponding percentages of duplicates and singletons across all age groups. Interestingly, we find that singletons, albeit insignificant, account for slightly higher percentages of genes among all five age groups (Fig. 2A).

We further detect translational selection in duplicates and singletons, respectively, across all examined tissues (Fig. S3). For convenience, a tissue, in which translational selection is detected as testified by a significantly stronger correlation between CUB and gene expression (P-value < 0.05), is denoted as TTS (Tissue with significantly stronger