

Research Article

Comprehensive Analysis and Comparison on the Codon Usage Pattern of Whole *Mycobacterium tuberculosis* Coding Genome from Different Area

Li Gun , Ren Yumiao, Pan Haixian, and Zhang Liang

Department of Biomedical Engineering, School of Electronic Information Engineering, Xi'an Technological University, Xi'an, China

Correspondence should be addressed to Li Gun; ligun@xatu.edu.cn

Received 13 December 2017; Revised 25 February 2018; Accepted 28 March 2018; Published 8 May 2018

Academic Editor: Sankar Subramanian

Copyright © 2018 Li Gun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Phenomenon of unequal use of synonymous codons in *Mycobacterium tuberculosis* is common. Codon usage bias not only plays an important regulatory role at the level of gene expression, but also helps in improving the accuracy and efficiency of translation. Meanwhile, codon usage pattern of *Mycobacterium tuberculosis* genome is important for interpreting evolutionary characteristics in species. In order to investigate the codon usage pattern of the *Mycobacterium tuberculosis* genome, 12 *Mycobacterium tuberculosis* genomes from different area are downloaded from the GeneBank. The correlations between G_3 , GC_{12} , whole GC content, codon adaptation index, codon bias index, and so on of *Mycobacterium tuberculosis* genomes are calculated. The ENC-plot, relationship between $A_3/(A_3 + T_3)$ and $G_3/(G_3 + C_3)$, GC_{12} versus GC_3 plot, and the RSCU of overall/separated genomes all show that the codon usage bias exists in all 12 *Mycobacterium tuberculosis* genomes. Lastly, relationship between CBI and the equalization of ENC shows a strong negative correlation between them. The relationship between protein length and GC content (GC_3 and GC_{12}) shows that more obvious differences in the GC content may be in shorter protein. These results show that codon usage bias existing in the *Mycobacterium tuberculosis* genomes could be used for further study on their evolutionary phenomenon.

1. Introduction

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis* that most often affect the lungs. Today, TB is still an unsolved problem. According to the WHO, in Oct, 2017, fact sheet pointed out that the TB is one of the top 10 causes of death worldwide [1]. It spreads from person to person through the air. One person may be infected by TB when only inhaling a few of TB germs. So, many scientists studied the characteristics including codon usage pattern [2], promoter architecture of toxin-antitoxin systems [3], drug-resistance [4], molecular epidemiology [5], protein function and immunogenicity [6], and drugs for the treatment [7] of TB. Typically, Sheen et al. studied the multiple genomes of *Mycobacterium tuberculosis* and found out some specific novel genes and mutations associated with pyrazinamide resistance [8]. Sun et al. and Khrustalev et al. characterize the mutations in streptomycin-resistant *Mycobacterium tuberculosis* [9, 10]. Recently, bioinformatics tools are widely used for analyzing the TB [11].

One amino acid is often encoded by more than one codon because the genetic code is redundant. This phenomenon is also known as synonymous codon usage. Many factors may affect the codon usage in *Mycobacterium tuberculosis*, such as mutation pressure [12], gene length [13], and natural selection [14]. Most previous studies deal with the relatively smaller set of *Mycobacterium tuberculosis* isolates [15, 16] or the isolates from one area [17, 18]. In this paper, genetic diversity of *Mycobacterium tuberculosis* from 12 areas is compared and analyzed. So, the objectives of this study aim to deal with the following issues: (1) correlation between codon preference parameters such as the codon usage index (CBI), the effective number of codons (ENC), the overall GC content, the codon adaptation index (CAI), and the Frequency of Optimal Codons (FOP); (2) mutation and evolution characteristics via the relationship between ENC and the GC_3 of the whole 12 *Mycobacterium tuberculosis* genomes; (3) phylogenetic analysis via RSCU values of the separated genomes for *Mycobacterium tuberculosis*.

TABLE 1: Basic characteristics of sequences for *Mycobacterium tuberculosis*.

Accession number	Number of total genes	Number of selected genes	Country
LWDR01000001.1	2048	1166	Argentina
CP011510.1	1973	1138	China
CP017596.1	2062	1154	Colombia
AP017901.1	2141	933	Japan
CP012506.2	2039	1018	Kazakhstan
CP010968.1	1985	1083	Malaysia
JKXT01000001.1	1841	1002	Mali
NZ_CM002052.1	2173	1089	Panama
NZ_CP023622.1	2159	1110	Peru
CP008981.1	1857	1028	South Korea
NZ_CM001226.1	2214	1082	Sweden
AP018033.1	2028	1162	Vietnam

2. Materials and Methods

The sequences of *Mycobacterium tuberculosis* (accession numbers in NCBI are shown in the Table 1) from Peru, Argentina, China, and so on in Homo sapiens are downloaded and examined. Specifically, coding sequences within *Mycobacterium tuberculosis* genome are selected according to the following: (1) containing more than 300 bases, (2) starting with the start codon, (3) number of bases being a multiple of 3, and (4) having no stop codon in the inner sequence. Thus, a total of 12965 coding sequences for *Mycobacterium tuberculosis* are selected (originally the total of coding sequences is 24520).

2.1. Effective Number of Codons (ENC) Analysis. ENC analysis can be used to quantify the absolute codon usage bias in coding sequences. ENC was calculated using the following formula [19]:

$$ENC^{\text{calculated}} = 2 + \frac{9}{\bar{f}_2} + \frac{1}{\bar{f}_3} + \frac{5}{\bar{f}_4} + \frac{3}{\bar{f}_6}, \quad (1)$$

where \bar{f}_k ($k = 2, 3, 4, 6$) is the mean of f_k values for the k -fold degenerate amino acids; \bar{f}_k denotes the average homozygosity for the amino acid class whose degree of codon degeneracy is k . The coefficients 9, 1, 5, and 3 denote the number of amino acids belonging to different classes. Here, f_k is estimated using the formula: $f_k = (nS - 1)/(n - 1)$, where n is the total number of occurrences of the codons for that amino acid and $S = \sum (n_i/n)^2$ ($i = 1 \dots k$), and n_i here is the total number of occurrences of the i th codon for that amino acid. Genes for which the codon choice is only constrained by a mutation bias will lie on or just below the curve of the expected ENC values [20]. Therefore, to elucidate the relationship between GC₃ and ENC values, the expected ENC values for different GC₃ were calculated as follows [21]:

$$ENC^{\text{expectation}} = 2 + s + \frac{29}{s^2 + (1 - s)^2}, \quad (2)$$

where s represents the given GC₃%.

2.2. RSCU Value Analysis. The RSCU values for all genes in *Mycobacterium tuberculosis* are calculated to determine the characteristics of synonymous codon usage without the confounding influence of amino acid composition and coding sequence size of different gene samples following a previously described method [22]. The RSCU index was calculated as follows:

$$RSCU = \frac{g_{ij}}{\sum_j n_i} n_i, \quad (3)$$

where g_{ij} is the observed number of the i th codon for the j th amino acid, which has n_i kinds of synonymous codons. RSCU values represent the ratio between the observed usage frequency of one codon in a gene sample and the expected usage frequency in the synonymous codon family, given that all codons for the particular amino acid are used equally. The synonymous codons with RSCU values > 1.0 have positive codon usage bias and were defined as abundant codons, whereas those with RSCU values < 1.0 have negative codon usage bias and were defined as less-abundant codons.

2.3. Codon Adaptation Index (CAI) Calculation. The CAI value is generally used to predict the expression level of an inbred gene (it is more suitable for unicellular organisms) and can be used to predict the expression level of foreign genes. The CAI value can be calculated via the following equation [23]:

$$CAI = \left(\prod_{K=1}^L w_K \right)^{1/L}, \quad (4)$$

where L refers to the number of codons used in the gene. The value of CAI is between 0 and 1; the larger the value, the stronger the codon usage bias. w_{ij} is the relative adaptiveness of a codon, which can be calculated via

$$w_{ij} = \frac{RSCU_{ij}}{RSCU_{i\max}}. \quad (5)$$

In (5), the $RSCU_{ij}$ denotes RSCU value of the codon for the i th amino acid. And the $RSCU_{i\max}$ is the RSCU value for the most frequently used codon for the specific amino acid. The value of the CAI falls between 0 and 1; larger value denotes the larger bias of the codon usage.

2.4. Calculation for the Codon Bias Index (CBI). The codon bias index reflects the presence of components with high codon usage in a particular gene. The value of CBI has been widely used for it can more clearly describe the foreign gene expression in the host. The CBI calculation method can be expressed by the following formula [24]:

$$CBI = \frac{N_{\text{opt}} - N_{\text{ran}}}{N_{\text{tot}} - N_{\text{ran}}}, \quad (6)$$

where the N_{opt} represents the total number of occurrences of the superior codon in the gene; in this work, the superior codons are the codons whose RSCU value is more than 1.6.

N_{ran} represents the sum of the number of occurrences of the superior codon when all the synonymous codons are random in a certain protein; N_{tot} represents the occurrence number of the amino acid corresponding to the superior codon in the gene. In this work, the CBI of genes for *Mycobacterium tuberculosis* genome is calculated and analyzed. Accordingly, the Frequency of Optimal Codons (FOP) of *Mycobacterium tuberculosis* genome, which is the weighted average of the RSCU of superior codons (RSCU value > 1.6), is calculated similarly.

2.5. Parity Rule 2 (PR2) Analysis. The parity rule 2 (PR2) plot is usually used to estimate the impact of mutation and selection pressure on codon usage of genes. In the PR2 plot, the abscissa is $[G_3/(G_3 + C_3)]$ denoting the GC-bias at the third codon position in entire coding sequences, while the ordinate is the $A_3/(A_3 + U_3)$ which denotes the AU-bias at the third codon position of entire coding sequences [24]. In this work, the effects of mutation pressure and natural selection on the codon usage of genes in *Mycobacterium tuberculosis* will be analyzed via the PR2 analysis.

2.6. Neutral Evolution and Clustering Analysis. The neutrality plot is the so-called neutral evolution analysis, which could be performed to determine and compare the extent of influence of mutation pressure and natural selection. In this work, the neutral evolution analysis of the codon usage pattern of *Mycobacterium tuberculosis* is conducted and analyzed by plotting the GC_{12} values of the synonymous codons against the GC_3 values. In the present study, clustering analysis was used to analyze the difference in codon usage patterns among *Mycobacterium tuberculosis* via RSCU of coding sequences. In the clustering analysis progress, in order to minimize the effect of amino acid composition on codon usage, the three stop codons and two particular amino acid codons, AUG and UGG, are excluded. The pdist and dendrogram function of Matlab are used to build the tree. The function could generate a dendrogram plot of the hierarchical binary cluster tree. Others, such as the relationship between ENC and protein length, relationship between CBI and protein length, relationship between overall GC content and protein length, histogram of CBI, FOP, and overall GC content, as well as the protein versus GC_{12}/GC_{3s} are all explored.

3. Results and Discussion

Codon usage pattern belonging to compositional characteristics such as the T_3 , C_3 , A_3 , G_3 , GC_{12} , and GC_3 and overall GC content of 12 *Mycobacterium tuberculosis* genomes from different area are calculated. The results show that the average value of overall GC content is 65.36%, average value of GC_3 is 79.18%, and the average T_3 (11.38%) and A_3 (9.43%) of 12 *Mycobacterium tuberculosis* genomes are fewer than GC_3 . Furthermore, the average ENC, FOP, CBI, and CAI are 41.25, 0.64, 0.39, and 0.45.

ENC is an important index to measure the codon usage bias in a genome and play a major role in their codon usage profile. In order to investigate the synonymous codon usage

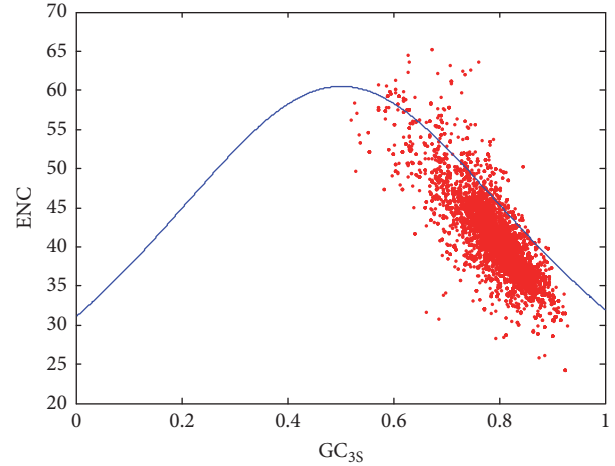


FIGURE 1: Plot of ENC versus GC_3 within *Mycobacterium tuberculosis* genomes.

pattern of the *Mycobacterium tuberculosis* genomes, the ENC versus GC_3 is plotted and the result is shown in Figure 1. Every point represents one coding gene in *Mycobacterium tuberculosis*. It can be seen that most dots are under the expected curve. Many scientists tended to think that the ENC value is belonging to a range of 21 to 61, but when the f_k is calculated via $(n \sum (n_i/n)^2 - 1)/(n - 1)$, all ENC values fell in the range from 20 to infinite; here, the value 20 denotes the codon usage bias using only one possible synonymous codon to correspond to an amino acid, and the larger value indicates that there is less bias of using all possible synonymous codons, and the frequency of using all possible synonymous codons may tend to be equal.

The parity rule states that if there is no mutation in genes, or no bias on the codon selection effect, the base content should obey the laws $A = T$ and $G = C$. This method is usually used to analyze the PR2 bias of the third place codon via comparing the $A_3/(A_3 + T_3)$ and the values of $G_3/(G_3 + C_3)$. The PR2 bias plot of 12 *Mycobacterium tuberculosis* genomes is shown in Figure 2. The distance between dot and the center denotes the degree and direction of the PR2 bias. From Figure 2, in most genes, A_3 are less than T_3 . Most C_3 content of the genes are more than the occurrence rate of T_3 in these genes.

Usually, the neutrality plot is used to analyze the directional mutation pressure versus natural selection of a certain genome, which could also reveal the relationship between the GC_{12} and the GC_3 with the GC_{12} as the ordinate and the GC_3 as the abscissa. The neutrality plot of 12 *Mycobacterium tuberculosis* genomes is shown in Figure 3. Each point shown in Figure 3 represents one separate gene for the *Mycobacterium tuberculosis*. The result shows that the distribution of most GC_{12} rate is between 0.5 and 0.65 and most of the GC_3 rate is during 70%–90%. In addition, GC_3 is more than GC_{12} in most of the genes. Meanwhile, the correlation analysis in Table 2 shows that there is a weak correlation with the value of -0.015 . Form these data performance, it can be seen that the natural selection may be a greater impact factor on codon preference in *Mycobacterium tuberculosis* genomes.

TABLE 2: Correlation of several basic parameters in *Mycobacterium tuberculosis* genomes.

Title	GC	GC _{3s}	CAI	CBI	ENC	FoP
GC _{3s}	0.536					
CAI	-0.37	-0.588				
CBI	0.263	0.618	-0.325			
ENC	-0.474	-0.809	0.553	-0.685		
FoP	0.28	0.46	-0.134	0.839	-0.612	
GC ₁₂	0.832	-0.015	-0.056	-0.088	-0.034	0.033

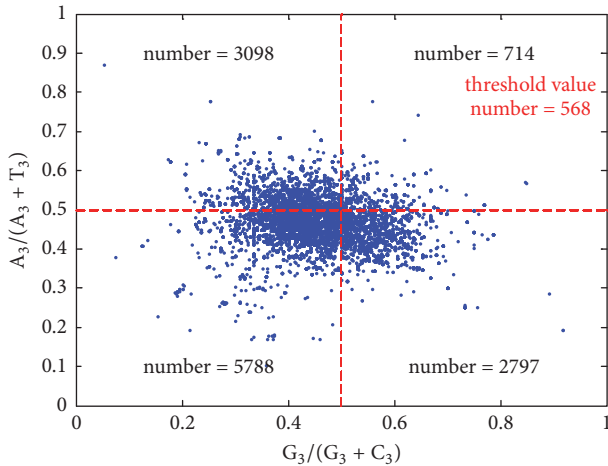


FIGURE 2: PR2-bias plot analysis of *Mycobacterium tuberculosis* genomes.

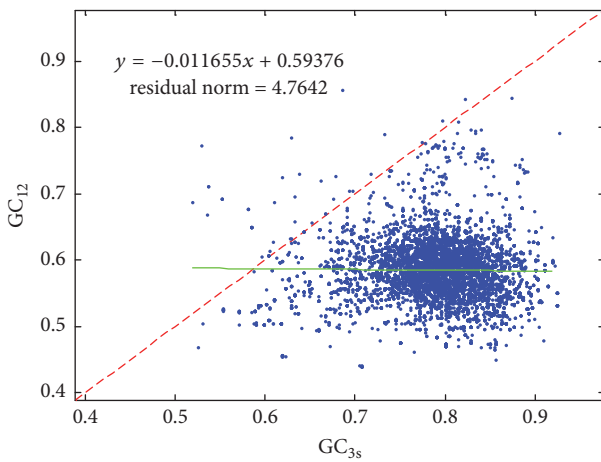


FIGURE 3: Neutrality plot of GC12 versus GC3s for 12 *Mycobacterium tuberculosis* genomes.

RSCU value, which is an important parameter for evaluating the bias of the synonymous codon, represents the ratio between the occurrence frequency of one codon and the expected usage frequency in a gene sample. Codons whose RSCU values are more than 1.0 would be regarded as having positive codon usage bias. On the contrary, those RSCU values less than 1.0 are defined as the less-abundant codons. The overall RSCU value can intuitively reflect the preference

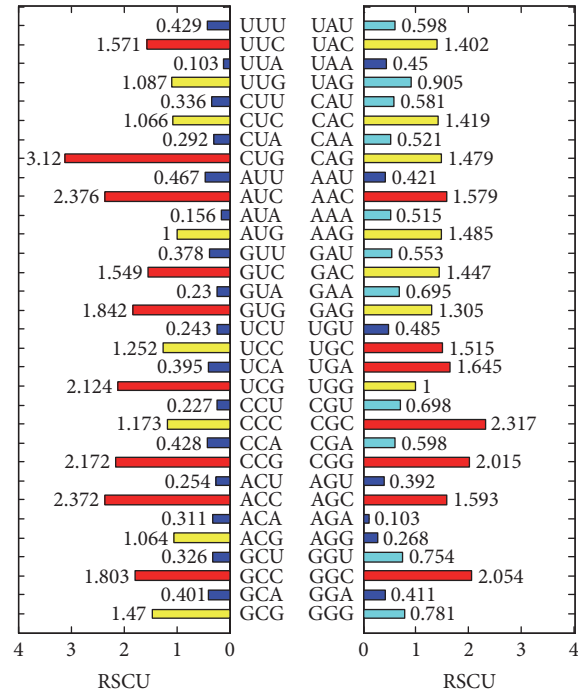


FIGURE 4: Overall RSCU of the 12 *Mycobacterium tuberculosis* genomes.

for codon usage in a certain genome. The overall RSCU values of 12 *Mycobacterium tuberculosis* genomes are calculated and the results are shown in Figure 4. The codon usage bias shows that the red bars in those RSCU values are more than 1.5, which could be regarded as the abundant codons; these codons are UUC, CUG, AUC, GUC, GUG, UCG, CCG, ACC, AAC, UGC, UGA, CGC, CGG, AGC, and GGC. Among them, CUG has the highest RSCU value of 3.12. The blue bars in Figure 4 are showing the opposite (RSCU values are less than 0.5) characteristics which could be called the less-abundant codons. Generally, RSCU values for the codons ending with the A or T are smaller than the RSCU value of the codons ending with G or C. There are three stop codons (UGA, UAA, and UAG), and no corresponding amino acid corresponding to stop codons. In Figure 4, RSCU values for terminal codons are UGA = 1.645, UAA = 0.45, and UAG = 0.905, and sum of them is 3 (there are 3 stop codons). On the other hand, the RSCU value 1.645 for UGA here denotes that 54.8% (1.645/3) of stop codons are UGA, which is more than the sum of UAA and UAG.

The value of RSCU is usually used as a measurement of the codon usage bias. The overall RSCU values (Figure 4), to a certain degree, can reflect the characteristics of codon usage, but when it comes to the codon usage of individual sequence for a particular genome, it can be seen that there is a great difference of RSCU values from different genes even of one genome. RSCU values of 12 *Mycobacterium tuberculosis* genomes are separately shown in Figure 5. There is some difference between all genomes especially in the usage of CUU, GUG, ACC, GAU, CGG, GGC, and so on.

The CBI value can more clearly describe the foreign gene expression in the host. The CBI values of 12 separated

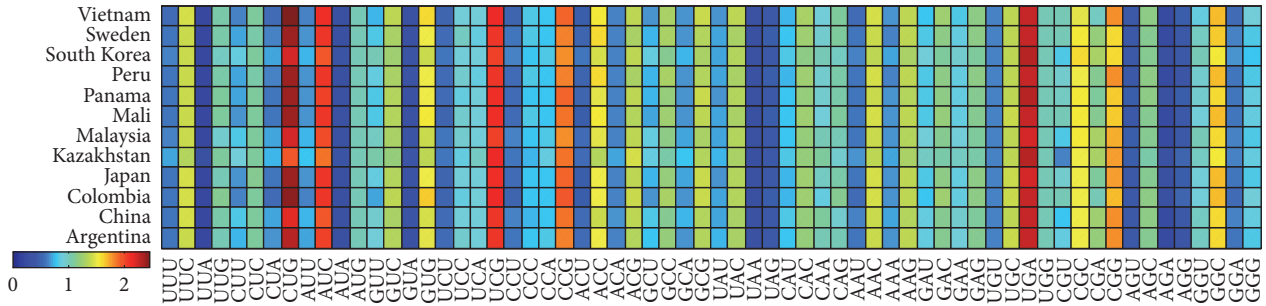


FIGURE 5: Heatmap of RSCU values for 12 separated *Mycobacterium tuberculosis* genomes.

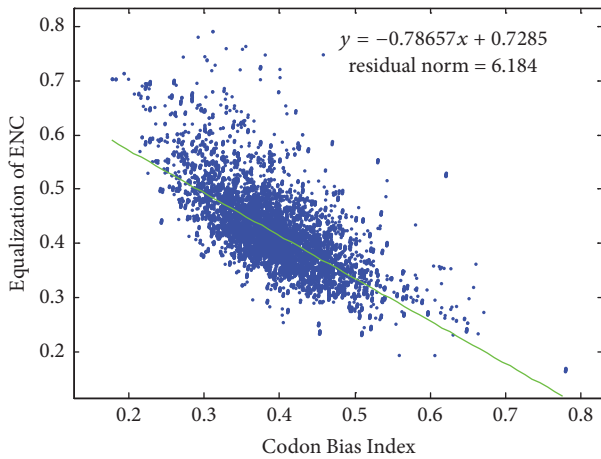


FIGURE 6: Overall relationship between CBI and the equalization of ENC.

Mycobacterium tuberculosis genomes are calculated via (6). Meanwhile, in order to explore the relationship between the CBI and the ENC, the ENC values are equalized to the range of the CBI via the following equation:

$$E.ENC = \frac{ENC - ENC_{min}}{ENC_{max} - ENC_{min}} (CBI_{max} - CBI_{min}) + CBI_{min} \quad (7)$$

Most previous researchers pointed out that the smaller ENC value may denote higher codon usage bias (CBI). The phenomenon is shown in Figure 6, and here the negative relationship between them is digitized by the linear fitting results. The results show that the negative correlation coefficient is -0.78657 .

Proteins in *Mycobacterium tuberculosis* could be analyzed via many methods, mass spectrometry, codon usage, and so on [25, 26]. Codon usage is linked to nucleic acids and proteins, especially the GC content. In Figure 7, relationship between protein length and GC content (GC_3 and GC_{12}) is described. Most proteins are concentrated within five hundred amino acids in length; it can also be seen from Figure 7 that when encoding the same length of protein, most of the GC_3 content is higher than the GC_{12} content. As the length of the code increases, the GC content of either GC_{12} or GC_3 tends to increase slightly. But within the lengths

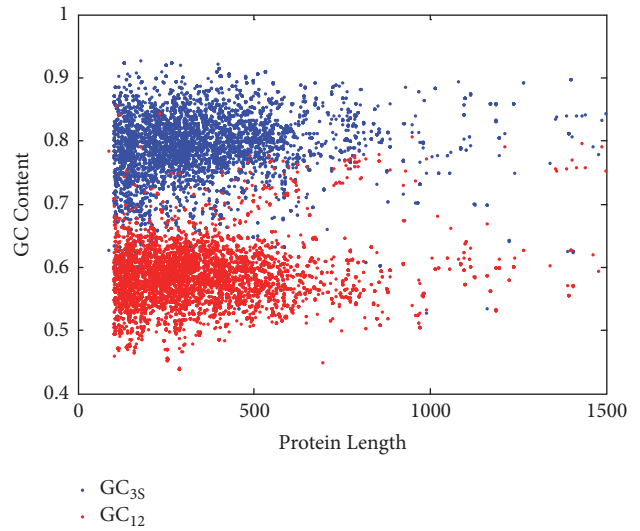


FIGURE 7: Relationship between protein length and GC content (GC_{35} and GC_{12}).

between 1000 and 2000 nucleotide residues, the GC-contents (GC_{12} and GC_3) are not obviously varied with the gene length (Okazaki fragment in bacteria is between 1000 and 2000 nucleotide residues; its corresponding protein length is between 333 and 667 or so). Here, only proteins of less than 1500 amino acids in length are considered.

Codon bias has a very broad significance for exploring a genome. From the molecular level, codon bias can also explain the basic phenomenon of evolutionary process in biology [27]. Sometimes, the differences between RSCU values of different genomes can be used to describe the evolutionary distance. Numerical closeness of RSCU values between coding sequences shows the closeness of relationship [28]. Genetic diversity of *Mycobacterium tuberculosis* from different area may have different frequency of ancestral strains [29]. Diversity and disease pathogenesis for *Mycobacterium tuberculosis* from different area may show difference too [30, 31]. When the evolutionary relationship of 12 separated *Mycobacterium tuberculosis* is concerned, samples from different countries have subtle differences. Genetic relationship of them is shown in Figure 8.

Mycobacterium tuberculosis exists in most countries and exploring its codon usage bias is useful for understanding

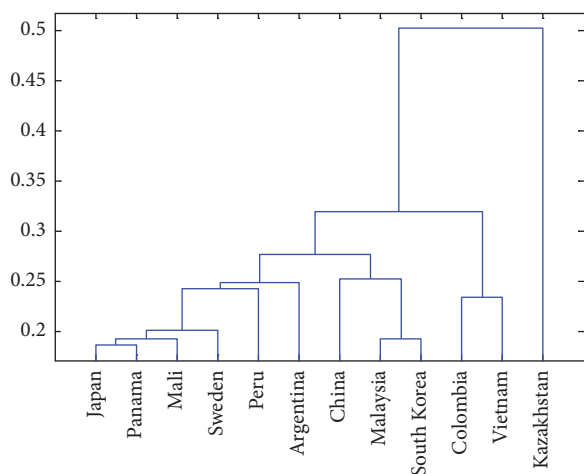


FIGURE 8: Phylogenetic tree of 12 *Mycobacterium tuberculosis* genomes.

genetic characteristics and geographical differences of whole genomes. Although many scientists endeavor to study the *Mycobacterium tuberculosis* from the genetic perspective [32, 33] for its very important for defending *tuberculosis* [34], its codon usage bias is not very clear until now. In this study, not only are traditional methods, such as ENC-plot, PR2 analysis, RSCU values, and phylogenetic analysis of 12 *Mycobacterium tuberculosis* genomes, studied, but also some new methods, CBI versus ENC and protein length versus GC_{3S}/GC_{12} , are all explored; the former relationship describing the relationship between CBI and ENC could reveal their negative correlation characteristics. Relationship between protein length and GC content (GC_{3S} and GC_{12}) can reflect the changes of GC content within different coding sequences during the evolutionary process. Previous studies indicated that genetic diversity exists in *Mycobacterium tuberculosis* isolated from different *tuberculosis* [35] when the samples are from different areas. Many studies also revealed that the host is an important factor that can affect the codon usage characteristics [36]. Overall RSCU values of overall 12 *Mycobacterium tuberculosis* genomes (Figure 4) and their separated genomes (Figure 5) all reveal strong codon usage bias. All these results combined with the relationship between CBI and ENC, protein length, and GC content (GC_3 and GC_{12}) show that the natural selection pressure is a little more important than mutation pressure. There are only 12 genomes of *Mycobacterium tuberculosis* analyzed in this work; although *Mycobacterium tuberculosis* exists in almost all countries. We selected 12 geographically representative genomes in order to analyze their regional differences. Codon usage differences of *Mycobacterium tuberculosis* from broader range and even from one area are not mentioned in this work; it is necessary to expand the scope of the genomes in the further study.

4. Conclusions

In this paper, the codon usage patterns of 12 *Mycobacterium tuberculosis* genomes, such as the ENC-plot, the

$A_3/(A_3 + T_3)$ versus $G_3/(G_3 + C_3)$ plot, the relationship GC_{12} versus GC_3 , the RSCU of overall/separated genomes, the relationship between CBI and the equalization of ENC, and the relationship between protein length and GC content (GC_{3S} and GC_{12}), and their phylogenetic relationship are all analyzed. The codon usage pattern and its influencing factors, especially, were identified for 12 *Mycobacterium tuberculosis* genomes. It is observed that codon usage patterns in *Mycobacterium tuberculosis* genomes are influenced by GC3 bias. Correlation between codon bias index and GC_3 ($=0.618$) shows that the GC3 bias of *Mycobacterium tuberculosis* genomes may also reflect its codon bias index. Majorities of the codons are G/C ended; this phenomenon combined with a negative correlation between overall GC-content and CAI may indicate the important role of compositional constraints and mutation pressure in shaping the codon usage bias in *Mycobacterium tuberculosis* genomes. We performed the comparative analysis of codon usage bias in *Mycobacterium tuberculosis* genomes from different area. The results would help further elucidate the underlying dynamics of genetic evolution in *Mycobacterium tuberculosis* genomes. These results showed that codon usage bias exists in the *Mycobacterium tuberculosis* genomes. All this information is important for explaining the function of *Mycobacterium tuberculosis* genomes and helps in understanding the evolutionary process of the *Mycobacterium tuberculosis*.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this study.

Acknowledgments

This research work was supported by schoolmaster fund of Xi'an Technological University (no. XAGDXJJ14011).

References

- [1] . World Health Organization, "WHO position paper, Meningococcal A conjugate vaccine: Updated guidance, February 2015," *Vaccine*, 2017.
- [2] A. Pan, C. Dutta, and J. Das, "Codon usage in highly expressed genes of Haemophilus influenzae and Mycobacterium tuberculosis: Translational selection versus mutational bias," *Gene*, vol. 215, no. 2, pp. 405–413, 1998.
- [3] Z. Thakur, V. Saini, P. Arya, A. Kumar, and P. K. Mehta, "Computational insights into promoter architecture of toxin-antitoxin systems of Mycobacterium tuberculosis," *Gene*, vol. 641, pp. 161–171, 2018.
- [4] A. Poudel, B. Maharjan, C. Nakajima et al., "Characterization of extensively drug-resistant Mycobacterium tuberculosis in Nepal," *Tuberculosis*, vol. 93, no. 1, pp. 84–88, 2013.
- [5] C. O'Driscoll, J. Konjek, B. Heym et al., "Molecular epidemiology of Mycobacterium abscessus complex isolates in Ireland," *Journal of Cystic Fibrosis*, vol. 15, no. 2, pp. 179–185, 2016.
- [6] J. Deng, L. Bi, L. Zhou et al., "Mycobacterium Tuberculosis Proteome Microarray for Global Studies of Protein Function

- and Immunogenicity," *Cell Reports*, vol. 9, no. 6, pp. 2317–2329, 2014.
- [7] M. AlMatar, H. AlMandea, I. Var, B. Kayar, and F. Köksal, "New drugs for the treatment of Mycobacterium tuberculosis infection," *Biomedicine & Pharmacotherapy*, vol. 91, pp. 546–558, 2017.
- [8] P. Sheen, D. Requena, E. Gushiken et al., "A multiple genome analysis of Mycobacterium tuberculosis reveals specific novel genes and mutations associated with pyrazinamide resistance," *BMC Genomics*, vol. 18, no. 1, article no. 769, 2017.
- [9] H. Sun, C. Zhang, L. Xiang et al., "Characterization of mutations in streptomycin-resistant Mycobacterium tuberculosis isolates in Sichuan, China and the association between Beijing-lineage and dual-mutation in gidB," *Tuberculosis*, vol. 96, pp. 102–106, 2016.
- [10] V. V. Khrustalev, M. Arjomandzadegan, E. V. Barkovsky, and L. P. Titov, "Low rates of synonymous mutations in sequences of Mycobacterium tuberculosis GyrA and KatG genes," *Tuberculosis*, vol. 92, no. 4, pp. 333–344, 2012.
- [11] S. Swaminathan, J. C. Sundaramurthi, A. N. Palaniappan, and S. Narayanan, "Recent developments in genomics, bioinformatics and drug discovery to combat emerging drug-resistant tuberculosis," *Tuberculosis*, vol. 101, pp. 31–40, 2016.
- [12] Q. Liu, "Analysis of codon usage pattern in the radioresistant bacterium Deinococcus radiodurans," *BioSystems*, vol. 85, no. 2, pp. 99–106, 2006.
- [13] M. Zhou, C. Tong, and J. Shi, "Analysis of Codon Usage Between Different Poplar Species," *Journal of Genetics and Genomics*, vol. 34, no. 6, pp. 555–561, 2007.
- [14] D. Lal, M. Verma, S. K. Behura, and R. Lal, "Codon usage bias in phylum Actinobacteria: relevance to environmental adaptation and host pathogenicity," *Research in Microbiology*, vol. 167, no. 8, pp. 669–677, 2016.
- [15] E. S. Hajimiri, M. Masoomi, N. Ebrahimzadeh et al., "High prevalence of Mycobacterium tuberculosis mixed infection in the capital of moderate tuberculosis incidence country," *Microbial Pathogenesis*, vol. 93, pp. 213–218, 2016.
- [16] M. Tadesse, G. Abebe, A. Bekele et al., "The predominance of Ethiopian specific Mycobacterium tuberculosis families and minimal contribution of Mycobacterium bovis in tuberculous lymphadenitis patients in Southwest Ethiopia," *Infection, Genetics and Evolution*, vol. 55, pp. 251–259, 2017.
- [17] D. Li, Y. Song, C.-L. Zhang, X. Li, X. Xia, and A.-M. Zhang, "Screening mutations in drug-resistant Mycobacterium tuberculosis strains in Yunnan, China," *Journal of Infection and Public Health*, vol. 10, no. 5, pp. 630–636, 2017.
- [18] G. Lopez-Avalos, G. Gonzalez-Palomar, M. Lopez-Rodriguez et al., "Genetic diversity of Mycobacterium tuberculosis and transmission associated with first-line drug resistance: a first analysis in Jalisco, Mexico," *Journal of Global Antimicrobial Resistance*, vol. 11, pp. 90–97, 2017.
- [19] F. Wright, "The "effective number of codons" used in a gene," *Gene*, vol. 87, no. 1, pp. 23–29, 1990.
- [20] S. Gajbhiye, P. K. Patra, and M. K. Yadav, "New insights into the factors affecting synonymous codon usage in human infecting Plasmodium species," *Acta Tropica*, vol. 176, pp. 29–33, 2017.
- [21] R. Prabha, D. P. Singh, S. Sinha, K. Ahmad, and A. Rai, "Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes," *Marine Genomics*, vol. 32, pp. 31–39, 2017.
- [22] J. Cristina, P. Moreno, G. Moratorio, and H. Musto, "Genome-wide analysis of codon usage bias in Ebolavirus," *Virus Research*, vol. 196, pp. 87–93, 2015.
- [23] P. E. Brandão, "The evolution of codon usage in structural and non-structural viral genes: The case of Avian coronavirus and its natural host Gallus gallus," *Virus Research*, vol. 178, no. 2, pp. 264–271, 2013.
- [24] M. Nath Choudhury, A. Uddin, and S. Chakraborty, "Codon usage bias and its influencing factors for Y-linked genes in human," *Computational Biology and Chemistry*, vol. 69, pp. 77–86, 2017.
- [25] M. S. Rabodoarivelo, M. Aerts, P. Vandamme, J. C. Palomino, V. Rasolofo, and A. Martin, "Optimizing of a protein extraction method for Mycobacterium tuberculosis proteome analysis using mass spectrometry," *Journal of Microbiological Methods*, vol. 131, pp. 144–147, 2016.
- [26] A. Banaei-Esfahani, C. Nicod, R. Aebersold, and B. C. Collins, "Systems proteomics approaches to study bacterial pathogens: application to Mycobacterium tuberculosis," *Current Opinion in Microbiology*, vol. 39, pp. 64–72, 2017.
- [27] T.-C. Wang and F.-C. Chen, "The evolutionary landscape of the Mycobacterium tuberculosis genome," *Gene*, vol. 518, no. 1, pp. 187–193, 2013.
- [28] N. Sweetline Anne, B. S. M. Ronald, T. M. A. S. Kumar, P. Kannan, and A. Thangavelu, "Molecular identification of Mycobacterium tuberculosis in cattle," *Veterinary Microbiology*, vol. 198, pp. 81–87, 2017.
- [29] E. C. Conceição, N. Rastogi, D. Couvin et al., "Genetic diversity of Mycobacterium tuberculosis from Pará, Brazil, reveals a higher frequency of ancestral strains than previously reported in South America," *Infection, Genetics and Evolution*, vol. 56, pp. 62–74, 2017.
- [30] D. F. Warner, A. Koch, and V. Mizrahi, "Diversity and disease pathogenesis in mycobacterium tuberculosis," *Trends in Microbiology*, vol. 23, no. 1, pp. 14–21, 2015.
- [31] M. N. Séraphin, R. Doggett, L. Johnston, J. Zabala, A. M. Gerace, and M. Lauzardo, "Association between Mycobacterium tuberculosis lineage and site of disease in Florida, 2009–2015," *Infection, Genetics and Evolution*, vol. 55, pp. 366–371, 2017.
- [32] A. Mitra, K. Angamuthu, and V. Nagaraja, "Genome-wide analysis of the intrinsic terminators of transcription across the genus Mycobacterium," *Tuberculosis*, vol. 88, no. 6, pp. 566–575, 2008.
- [33] L. Corrales-Garcia, E. Ortiz, J. Castañeda-Delgado, B. Rivas-Santiago, and G. Corzo, "Bacterial expression and antibiotic activities of recombinant variants of human β -defensins on pathogenic bacteria and *M. tuberculosis*," *Protein Expression and Purification*, vol. 89, no. 1, pp. 33–43, 2013.
- [34] P. Huberts and V. Mizrahi, "Cloning and sequence analysis of the gene encoding the DNA polymerase I from Mycobacterium tuberculosis," *Gene*, vol. 164, no. 1, pp. 133–136, 1995.
- [35] E. V. Mbugi, B. Z. Katale, K. K. Siame et al., "Genetic diversity of Mycobacterium tuberculosis isolated from tuberculosis patients in the Serengeti ecosystem in Tanzania," *Tuberculosis*, vol. 95, no. 2, pp. 170–178, 2015.
- [36] D. A. Aguilar-Ayala, J. C. Palomino, P. Vandamme, A. Martin, and J. A. Gonzalez-y-Merchand, "Genetic regulation of Mycobacterium tuberculosis in a lipid-rich environment," *Infection, Genetics and Evolution*, vol. 55, pp. 392–402, 2017.