

Article

Codon Pattern and Compositional Constraints Determination of Genes Associated with Chronic Periodontitis

Rekha Khandia ^{1,*} , Megha Pandey ², Igor Vladimirovich Rzhepakovsky ³ , Azmat Ali Khan ⁴  and Isabel Legaz ^{5,*} 

¹ Department of Biochemistry and Genetics, Barkatullah University, Bhopal 462026, India

² Translational Medicine Center, All India Institute of Medical Sciences, Bhopal 462020, India

³ Medical and Biological Faculty, North Caucasus Federal University, 355009 Stavropol, Russia

⁴ Pharmaceutical Biotechnology Laboratory, Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia

⁵ Department of Legal and Forensic Medicine, Biomedical Research Institute (IMIB), Regional Campus of International Excellence “Campus Mare Nostrum”, Faculty of Medicine, University of Murcia, E-30120 Murcia, Spain

* Correspondence: bu.rekha.khandia@gmail.com or rekha.khandia@bubhopal.ac.in (R.K.); isalegaz@um.es (I.L.)

Abstract: Genome-wide association studies showed the relationship of *NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5* with chronic periodontitis. The study’s objective was to investigate different molecular patterns and evolutionary forces acting on the mentioned genes. The investigation of molecular patterns encompasses the study of compositional parameters, expression profile, physical properties of genes, codon preferences, degree of codon bias, determination of the most influential codons, and assessment of actions of evolutionary forces, such as mutations and natural selection. The overall compositional analysis revealed the dominance of A and G nucleotides compared to T and C. A relatively low codon usage bias is observed. The CTG codon is the most overused codon, followed by TCC. The genes, *AP3B2* and *SIGLEC5*, preferred GC-ending codons, while *NIN*, *ABHD12B*, and *WHAMM* preferred AT-ending codons. The presence of directional mutational force and natural selection was found to operate codon usage in genes envisaged, and selective forces were dominant over mutational forces. Apart from mutation and selection forces, compositional constraints also played imperative roles. The study enriched our knowledge of specific molecular patterns associated with the set of genes significantly associated with chronic periodontitis. Further studies are warranted to identify more genetic signatures associated with the disease.

Keywords: chronic periodontitis; inflammatory disease; codon usage; genetics relationship of periodontitis



Citation: Khandia, R.; Pandey, M.; Rzhepakovsky, I.V.; Khan, A.A.; Legaz, I. Codon Pattern and Compositional Constraints Determination of Genes Associated with Chronic Periodontitis. *Genes* **2022**, *13*, 1934. <https://doi.org/10.3390/genes13111934>

Academic Editors: Elżbieta Kaczmarek and Bogna Grygiel-Gorniak

Received: 20 September 2022

Accepted: 18 October 2022

Published: 24 October 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A study on the global burden of diseases, injuries, and risk factors [1] demonstrated that dental caries in permanent teeth and periodontitis were the primary and 11th most prevalent reasons for ailments worldwide. The etiology of this disease needs to be characterized as it confers both health and financial burdens [2,3]. Genetic factors contribute to oral health and the heritability of dental caries to an extent as high as 50% [4]. Characterizing these genetic factors will help us understand the disease and develop countermeasures to fight these diseases. Chronic periodontitis (CP) negatively influences glycemic control, and diabetes coupled with CP enhances the risk of cardiorenal mortality [5]. The CP patients demonstrate significant endothelial dysfunction [6], and studies revealed a close association between periodontitis and cardiovascular diseases [7], liver diseases [8], rheumatoid arthritis [9], alopecia areata [10], anxiety-like behaviors [11], respiratory diseases [12], and neuropathy in diabetic patients [13]. In addition, studies have shown that

CP enhances the risk of preeclampsia in pregnant women [14]. Chronic periodontitis (CP) involves host inflammatory/immune response against the subgingival biofilm [15]. A study encompassing a total of 3160 study subjects revealed that the *NIN* and *SIGLEC5* genes are associated with CP susceptibility [16]. Gene-centric association analysis based on the Gene Ontology, Ingenuity, KEGG, Panther, Reactome, and Biocarta databases for gene set enrichment analyses that included the study of 18,307 genes; 4 genes (*NIN*, *ABHD12B*, *WHAMM*, and *AP3B2*) showed significant association with CP [17].

Amino acids are the protein molecule's basic unit, and a codon encodes for an amino acid. Except for methionine and tryptophan, all the amino acids are encoded by two or more codons, and such codons are called synonymous codons. In any organism, all the synonymous codons are not used equally, and there is unequal usage of a codon in any gene, tissue, or organism and is called codon usage bias. Codon bias can be gene-specific [18], species-specific [19], tissue-specific [20,21], or function-specific [22,23]. Codon bias is often determined by compositional, mutational, or selectional forces [24]. Codon bias plays a crucial determinant of gene expression levels mainly via influencing transcription [25], and reports suggest more substantial bias in the case of highly expressed genes due to selection pressure [26].

Periodontitis affects 11% of the world's population and poses a tremendous public health challenge [27]. Studies have been conducted on the association of CP with various disorders; however, no work has been reported on codon usage studies in genes related to CP. The present study encompasses the study of 30 transcripts belonging to five genes, *NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5*. The function of the *NIN* gene includes correct positioning and anchoring of microtubule minus ends. *ABHD12B* regulates immune and neurological processes, *WHAMM* participates in Arp2/3-mediated actin polymerization, *AP3B2* is involved in neurotransmitter release, and *SIGLEC5* performs sialic-acid dependent cell binding. Since the underlying molecular mechanisms of periodontitis are still unclear, the effects of the *NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5* genes on periodontitis have not been elucidated [16].

The study provides insight into various factors influencing codon bias. The study also elucidates the over- or under-represented codons and the codons having maximum influence in codon usage. Additionally, the study will help identify molecular signatures associated with the CP-associated genes and design synthetic gene constructs for gene therapy purposes to manipulate gene expression levels, where an aberrant gene expression level is responsible for ailments.

2. Materials and Methods

2.1. Sequence Retrieval

In this study, the coding transcripts of the genes (S1) associated with CP were retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (<http://www.ncbi.nlm.nih.gov> accessed on 20 March 2022). Upon investigating the literature, we found only five genes responsible for CP (*NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5*). Therefore, we considered all the protein-encoding transcripts for the mentioned genes. The criteria set for the transcripts were that they must be initiated with a start codon and ended with a stop codon and should be devoid of bases other than A, T, C, and G. There were eight, seven, six, five, and four numbers of transcripts that fulfilled the criteria set for *NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5* genes, respectively.

2.2. Compositional Analysis

The overall nucleotide composition of the transcripts was determined. The nucleotide composition at the first and third codon positions (%A, %T, %C, %G, %A3, %T3, %C3, %G3) were determined. Overall %AT and %AT3 and %GC and %GC3 were also determined. Analyses of %GC3 (GC composition at the third codon position) and %GC12 (composition at the first and second codon) were also done. %GC3 helps evaluate the compositional and codon bias [28].

2.3. Relative Synonymous Codon Usages (RSCU)

RSCU is the ratio of the observed frequency of codons to the expected frequency of the codon [29]. All transcripts' RSCU values were calculated using the CAIcal server [30]. RSCU values above 1.6 and below 0.6 were considered overrepresented and underrepresented, respectively. The values between 0.6 and 1.6 are taken as randomly used, and codons are called unbiased [31].

2.4. Codon Adaptation Index (CAI)

CAI is a measure to estimate protein expression levels and determine the effect of natural selection that helps shape the codon usage pattern in transcripts [32]. If all the synonymous codons are used equally, the CAI value will be 0, while in the case of the strongest bias, the value of 1 will be obtained [33]. Higher CAI values show both high expression as well as high bias. CAI was calculated using the formula given by Sharp and Li (1987) [31].

2.5. Effective Number of Codons (ENc)

ENc is a commonly used indicator of codon usage bias [34]. The values of ENc range between 20–61. It is a non-directional measure of codon bias, and the value 20 shows the highest bias, while 61 shows no bias. The value 20 indicates that only one codon will be used for a single amino acid, while the value 61 indicates that all the synonymous codons are equally used. Generally, genes with ENc values below 35 are considered highly biased, while ENc above 50 suggests low bias [35,36].

2.6. PR2-Bias Plot Analysis

The intra-strand base composition theoretically should be $A = T$ and $G = C$ which is called the type 2 parity rule or PR2, and violation of this rule is expected in amino acids encoded by four codons [37]. It is plotted with AT-bias value $[A3/(A3+ T3)]$ as the ordinate and GC-bias $[G3/(G3 + C3)]$ as the abscissa. In this plot, the center is 0.5, which is the place where $A = T$ and $G = C$ (PR2), indicating no bias. The center shows no bias between the two strands of DNA for mutation and selection rates [36]. A value less than 0.5 indicates a preference for pyrimidine over purine [38].

2.7. Neutrality Plot

Regression analysis between %GC3 and %GC12 (mean %GC1 and %GC2) is a standard method to explore the effects of natural selection and mutation pressure [39]. It determines which evolutionary force majorly influences codon usage [40]. %GC3 and %GC12 are plotted on lateral and vertical axes to produce a scatter plot serving as the neutrality plot. When the slope value is 1, it is suggestive of codon usage bias solely driven by neutrality. The effects of other forces are determined by its deviation from neutrality [41].

2.8. Principal Component Analysis (PCA)

PCA is a commonly employed multivariate statistical technique to analyze the trends in codon usage patterns. Here, we used PCA to obtain significant trends in the transcripts of genes associated with enhanced risk of CP. The method reduces the original variables to a lower number of principal components without losing its information. Each transcript is represented as a 59-dimensional vector (the number of possible codons excluding stop codons, methionine, and tryptophan encoded by single codons). Finally, we took the top two principal components with the most significant variance. We performed PCA analysis using Origin18 software (<https://www.originlab.com/demodownload.aspx>, accessed on 19 September 2022).

2.9. Phylogenetic Analysis

To explain any phylogenetic relationship between the gene transcripts, phylogenetic analysis was conducted using the hierarchical clustering method and Ward's algorithm

based on RSCU values with 1000 bootstrap values. PAST4 [42] and Molecular Evolutionary Genetics Analysis 10 (MEGA-X) (version 10.1.8) were used for the study.

3. Results

3.1. Nucleotide Composition Affects the Codon Usage Bias

Reports suggest the impact of compositional constraints on codon bias [43]; therefore, overall nucleotide composition and AT and GC composition at various codon positions were determined for all the 30 transcripts envisaged. Compositional features of the transcripts reveal the dominance of nucleotide A. Average nucleotide composition analysis revealed that percent composition A was highest (27.81 ± 4.67), followed by nucleotide G (26.93 ± 2.17). Percent composition for nucleotide T was least (21.24 ± 2.42). Percent composition at the third codon position was highest for nucleotide G (30.18 ± 4.91) followed by C (25.38 ± 8.6). The results indicate that though the average composition of A nucleotide is highest, the composition of nucleotide G was highest at the third codon position. On the other hand, the percent composition was least for nucleotide T, both overall and at the third codon position. The percent composition of the %A3 and %T3 were almost similar (22.32 ± 6.96 and 22.1 ± 5.39 for A3 and T3, respectively). The GC and AT compositions were nearly equal (50.95 ± 5.24 and 49.0 ± 5.24 for GC and AT compositions, respectively). The results indicated that both the GC and AT contents are used equally in the transcripts. However, the GC and AT compositions at the third codon positions differed, and %GC3 (55.57 ± 11.07) was higher than %AT (44.42 ± 11.07). When the %GC compositions at all three codon positions were evaluated, it was observed that %GC3 composition is the most variable, while %GC2 composition is the least (Figure 1A).

3.2. RSCU Analysis

In most sequenced genomes, synonymous codons are not used in equal frequencies. This phenomenon is termed codon-usage bias. RSCU is a measure of codon bias [44]. Codons with RSCU values >1.6 and <0.6 are called overrepresented and underrepresented, respectively, while values between 0.6 and 1.6 are called randomly used codons [45]. The analysis revealed that codon CTG varied largely, and its RSCU values ranged between 1.03 (random usage) and 5.16 (extremely biased usage). In contrast, codon ACG varied the least with RSCU values ranging between 0 and 0.69. Out of six CG-containing codons, TCG, CCG, ACG, GCG, CGA, and CGT, the codons ending with CG (TCG, CCG, ACG, GCG) had more bias than all CG-containing codons and were underrepresented in more than 75% of genes envisaged. Out of six AT-containing codons, only one codon, CTA, was underrepresented in more than 75% of genes. The RSCU values of envisaged genes are given in Table 1.

CAI is a geometric mean of the RSCU values of codons of genes and is measured using a trained codon usage table [46]. ENc is also a non-directional measure of codon usage bias [36]. CAI and ENc values were correlated with overall %GC and %GC composition at all three codon positions. The correlation analysis revealed that CAI had a significant negative correlation with ENc, which indicated that with the increasing bias, gene expression also increased in the case of transcripts associated with CP (since ENc is a non-directional measure of codon bias). Furthermore, CAI showed a positive correlation with %GC, %GC2, and %GC3, while ENc showed a negative correlation (Table 2). The results are indicative of the effects of compositional constraints on codon usage.

Table 1. RSCU of NIN, ABHD12B, WHAMM, AP3B3, and SIGLEC6 genes.

Codon	Amino Acid	NIN	ABHD12B	WHAMM	AP3B3	SIGLEC6
TTT	F	1.395	1.524	1.267	0.571	0.566
TTC		0.605	0.476	0.733	1.429	1.434
TTA		0.790	0.816	0.859	0.068	0.114
TTG		0.895	0.893	1.284	0.239	0.321
CTT	L	1.019	1.106	0.936	0.329	0.712
CTC		0.872	0.553	0.779	0.977	1.594
CTA		0.468	0.512	0.709	0.318	0.093
CTG		1.956	2.120	1.432	4.070	3.167
ATT	I	1.194	1.669	1.258	1.296	0.335
ATC		0.786	1.038	0.842	1.348	2.248
ATA		1.021	0.293	0.901	0.356	0.418
GTT		0.806	1.425	1.120	0.390	0.553
GTC	V	0.851	1.069	0.808	0.827	1.364
GTA		0.704	0.315	0.438	0.422	0.298
GTG		1.639	1.190	1.634	2.361	1.786
TCT		1.584	1.861	1.722	0.835	0.646
TCC	S	1.068	0.233	1.179	1.508	2.378
TCA		0.800	0.297	1.013	0.684	0.402
TCG		0.076	0.630	0.260	0.316	0.629
AGT		1.133	1.796	0.881	1.171	0.090
AGC	P	1.339	1.184	0.944	1.486	1.856
CCT		1.043	0.873	1.083	1.183	0.807
CCC		0.979	1.660	0.645	1.827	1.816
CCA		1.534	1.335	1.542	0.872	1.035
CCG	T	0.443	0.132	0.729	0.118	0.343
ACT		1.071	0.749	1.130	0.597	0.176
ACC		0.990	0.783	1.209	2.507	2.299
ACA		1.447	2.153	1.593	0.726	0.854
ACG	A	0.491	0.315	0.068	0.171	0.672
GCT		1.076	0.902	1.695	0.923	1.236
GCC		1.060	1.280	0.986	1.984	2.031
GCA		1.347	1.634	0.534	0.540	0.599
GCG	Y	0.517	0.184	0.785	0.554	0.135
TAT		1.247	1.449	1.250	0.370	0.720
TAC		0.753	0.551	0.750	1.630	1.280
CAT		1.143	0.632	0.794	0.444	0.628
CAC	H	0.857	1.368	1.206	1.556	1.372
CAA		0.657	0.429	0.749	0.381	0.658
CAG	Q	1.343	1.571	1.251	1.619	1.342

Table 1. *Cont.*

Codon	Amino Acid	NIN	ABHD12B	WHAMM	AP3B3	SIGLEC6
AAT	N	1.140	0.927	1.087	0.617	0.555
AAC		0.860	1.073	0.913	1.383	1.445
AAA	K	1.023	0.803	1.356	0.543	0.549
AAG		0.977	1.197	0.644	1.457	1.451
GAT	D	1.086	1.042	1.229	0.702	0.862
GAC		0.914	0.958	0.771	1.298	1.139
GAA	E	1.072	1.180	1.273	0.424	0.309
GAG		0.928	0.820	0.727	1.576	1.691
TGT	C	1.021	0.973	1.196	1.464	0.687
TGC		0.979	1.027	0.804	0.536	1.313
CGT	R	0.594	1.005	0.560	0.829	0.196
CGC		0.344	0.479	0.691	1.249	1.321
CGA		0.628	0.123	0.501	0.715	0.534
CGG		0.929	0.882	1.088	1.706	0.677
AGA		1.752	2.353	2.009	0.447	1.805
AGG		1.753	1.157	1.151	1.053	1.467
GGT		0.968	0.628	1.034	0.406	0.542
GGC		G	0.988	0.946	1.547	2.069
GGA	1.036		1.135	0.770	0.698	0.935
GGG	1.009		1.291	0.649	0.828	1.346

Table 2. Correlation analysis between overall GC composition and GC composition at all the three codon positions along with CAI and ENC.

	CAI	ENC	%GC	%GC1	%GC2	%GC3
CAI		***	***	NS	*	***
ENC	<i>-0.945</i>		***	NS	*	***
%GC	0.808	<i>-0.819</i>		NS	***	***
%GC1	<i>-0.028</i>	<i>-0.038</i>	0.328		NS	NS
%GC2	0.414	<i>-0.463</i>	0.667	<i>-0.134</i>		*
%GC3	0.897	<i>-0.864</i>	0.916	0.278	0.362	

Upper diagonal shows the *p*-value, and the lower diagonal shows the Pearson correlation coefficient. The negative correlation coefficient is shown in *italic* font, while the positive correlation coefficient is in straight. Level of significance *** *p* < 0.001; * *p* < 0.05.

3.3. At the Third Codon Position, G and T Are Preferred

Parity analysis gives information regarding the skew of A/T and C/G nucleotides at the third codon position. In the presence of only mutational force, CG and AT will be equally used, while unequal usage is not entirely explained by natural selection [47]. The average GC bias (%G3/%G3 + %C3) was 0.551 ± 0.07 , and AT (%A3/%A3 + %T3) bias was 0.495 ± 0.065 . The average value indicated that G is preferred over C and T over A (Figure 1B).

3.4. Selection Is Dominant Force Affecting Codon Usage Evidenced by ENc-GC3 Plot

A significant positive correlation has been observed between the %GC12 and %GC3 contents ($r = 0.466$, $p < 0.01$), which indicates that the directional mutational forces are working on all GC contents at all three codon positions [36]. The ENc-GC3 plot is constructed to assess compositional, selectional, and mutational forces acting on CP-related genes. The expected Nc (ENc) is presented as a curve in Figure 1C [35]. The GC3 content ranged between 41.70% and 77.46%, and the ENc value ranged between 44.05 and 57.51, pointing toward the low codon bias. The ENc-GC3 analysis is considered an alternative to the multivariate statistical analysis method to reveal synonymous codon usage and determine the forces involved in shaping codon usage. The ENc line suggests composition-biased mutation pressure on codon usage, and selection pressure is zero if all the data points are present over the expected line [48]. In the case of translational selection, the data point will lie too far from the curve [49]. In the present study, the role of selection pressure appears to be significant (Figure 1C).

3.5. Selection Is Dominant Force Affecting Codon Usage Evidenced by Neutrality Plot

A neutrality plot was constructed between %GC3 and %GC12. %GC12 is the average value of %GC composition at the first and second codon positions. The neutrality plot (Figure 1D) shows the equilibrium between mutation and selection. The equilibrium near zero shows high neutrality or mutation forces, while values near one show high selection forces [50]. The model depicted that 21.77% of the variation in %GC12 results from GC3 component variation [26].

The neutrality plot reveals the impact on mutational and selection forces on codon usage [41]. To determine what major force is driving codon usage bias in envisaged genes, *NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5*, the neutrality plot was generated. The slope of the estimated equation revealed that mutation pressure contributed a minor role of 15%, while selection (85%) played the main role in determining codon usage (Figure 1D).

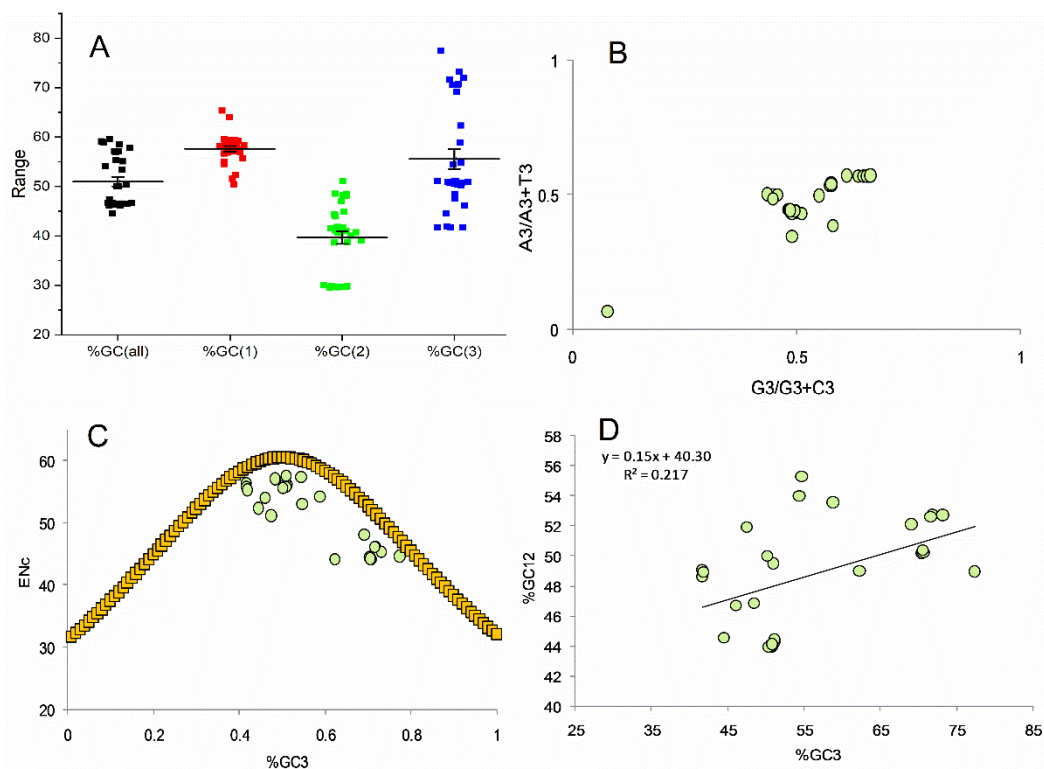


Figure 1. (A). Overall %GC composition and %GC composition at all the three codon positions (B). Parity plot analysis (C). ENc-GC3 analysis (D). Neutrality plot.

3.6. Principal Component Analysis (PCA) Revealed the Effect of Composition, Gene Expression, and Protein Properties on Codon Usage

PCA is a method to reduce the number of variables without compromising the information. The PCA done for 30 transcripts revealed that the first two PCs contributed 49.93% and 14.80% variation, respectively. PCA biplot analysis reveals the effect of each codon based on its loading values. The length of the arrow is proportional to its contribution to variation. Therefore, the lengthiest the arrow, the more contribution towards variation [51]. In the present study, out of 59 codons, codon CTG contributed to maximum variation, followed by TCC (Figure 2). GC-ending codons were preferred by *AP3B2* and *SIGLEC5* gene transcripts, while AT-ending codons were preferred by *NIN*, *ABHD12B*, and *WHAMM*.

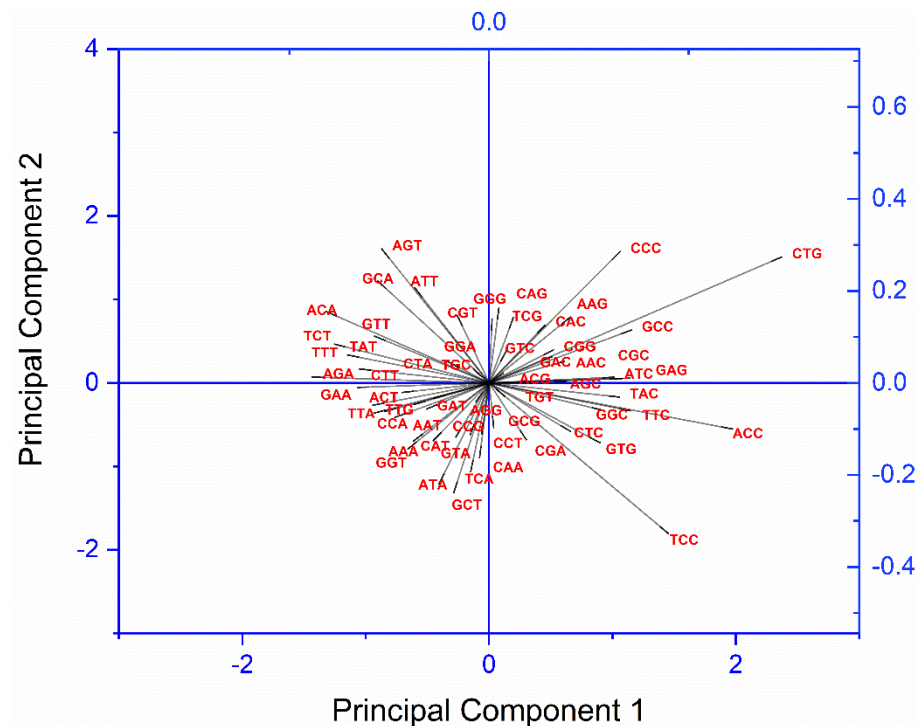


Figure 2. PCA biplot analysis revealed that CTG codon is maximally influencing the overall codon usage.

Furthermore, we performed correlation analysis of PC1 and PC2 with other variables (Table 3). The analysis revealed that both PC1 and PC2 are correlated with few of the compositional constraints, suggestive of the impact of composition on codon usage pattern. CAI values ranged between 0.706 and 0.839 with an average of 0.751 ± 0.04 , suggestive of good expression level and existence of codon bias [30]. PC1 showed significant correlation with CAI, and PC2 showed significant correlation with protein properties, and it indicated the effect of codon usage bias on both the protein expression and protein properties.

Table 3. Correlation analysis of PC1 and PC2 with compositional constraints, protein properties (GRAVY and AROMA), and gene expression.

Component	%GC(3)	CAI	GRAVY	AROMA	%A	%C	%T	%G
PC1 (r value)	0.974	0.929	0.345	0.131	<i>−0.688</i>	0.853	<i>−0.577</i>	0.515
Significance	***	***	NS	NS	***	***	***	***
PC2 (r value)	0.117	0.236	0.755	0.638	<i>−0.470</i>	0.002	0.474	0.479
Significance	NS	NS	***	***	**	NS	*	*
Component	%A3	%C3	%T3	%G3	%G + C	%G3 + C3	%A + T	%A3 + T3
PC1 (r value)	<i>−0.902</i>	0.887	<i>−0.837</i>	0.644	0.881	0.974	<i>−0.881</i>	<i>−0.974</i>
Significance	***	***	***	***	***	***	***	***
PC2 (r value)	<i>−0.404</i>	0.260	0.281	<i>−0.192</i>	0.200	0.117	<i>−0.200</i>	<i>−0.117</i>
Significance	NS	NS	NS	NS	NS	NS	NS	NS

Italics font showed negative correlation coefficient, while straight showed positive correlation coefficient; *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; NS: non-significant.

3.7. Phylogenetic Analysis

Phylogenetic analysis and the evolutionary relationship of 30 transcripts were carried out using MEGA 10 software. Phylogenetic analysis revealed that the genes could be separated mainly into two clusters. One cluster is comprised of *SIGLEC5* and *AP3B2*, while the second cluster is made up of *NIN*, *ABHD12B*, and *WHAMM* transcripts. In the second cluster, *ABHD12B* transcripts were separated from others showing distinct codon usage by these transcripts. From the tree, we can derive that each gene is phylogenetically distinct, and transcripts from a single gene clustered together suggest no common origin (Figure 3).

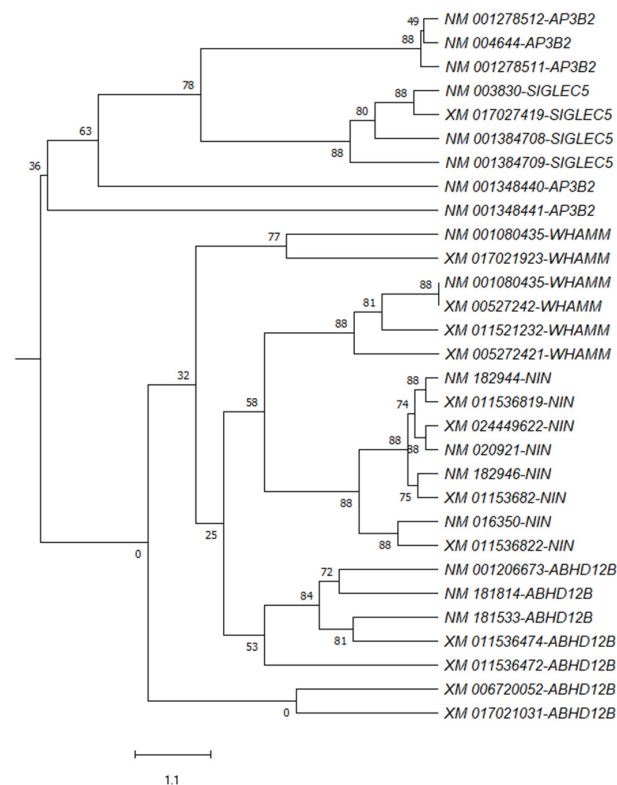


Figure 3. Classical clustering showed the transcripts might be grouped into two clusters. AP3B2 and SIGLEC5 transcripts clustered together, while the rest of the transcripts clustered separately.

4. Discussion

Periodontitis is a condition of gum infection and inflammation that may result in tooth decay and other serious health complications, including heart problems [52], lung diseases [12], and diabetes [5]. Periodontitis is a chronic inflammatory disorder of the tooth-supporting tissue, and age, race, obesity, and smoking status are a few causative factors. Genome-wide association studies conducted by [16,17] revealed the association of *NIN*, *ABHD12B*, *WHAMM*, *AP3B2*, and *SIGLEC5* with CP. Therefore, we are tempted to undertake molecular studies related to these genes. We performed various studies that revealed molecular signatures and different forces acting on these genes during evolution.

The present study's average ENc value was 52.35 ± 5.01 , indicating a comparatively lower codon usage bias in the envisaged genes. In other diseases associated with human disorders, such as Alzheimer's, ENc was 48.84 [52], and in Krabbe disease, a rare neurodegenerative disorder, ENc was 56.98 [22]. The results from other studies support our result, which shows that low codon bias might facilitate efficacious replication in vertebrates [43]. The GC content influences the codon usage and modulates the thermostability and bendability of DNA, consequently influencing the transcription process by keeping the DNA region transcribed in unwound status [53]. The genes associated with CP had higher A and G nucleotides than T and C in the present study. The GC component was highest at the third codon position while lowest at the second codon position. Our results concord with Uddin who found in genes associated with anxiety in humans [47]. Furthermore, a significant positive correlation between GC12 and GC3 content ($r = 0.466$, $p < 0.01$) is suggestive of the action of directional mutation force at all three codon positions and is in concordance with the earlier studies [54].

RSCU analyses reveal the overrepresented, underrepresented, and randomly used codons in any given gene. Our analysis revealed that codon CTG is the most overrepresented codon in most transcripts. CTG codon is overrepresented in genes common to immunodeficiency and cancer [23]. Higher CTG representation might be understood based on exaggerated T→C transition rates, resulting in more CTG codons [55]. Out of six CG-containing codons, four CG-ending codons, i.e., TCG, CCG, ACG, GCG, were underrepresented in more than 75% of the gene. The results appear to be a consequence of the CpG dinucleotide depletion in most coding sequences. CpG depletion results from selective forces, including the involvement of CpG in gene silencing and X chromosome inactivation and spontaneous methylation of CpG dinucleotide into TpG [35]. Thus, overrepresentation of the CTG codon might be considered a result of selective forces. The CTG codon also influenced the overall codon usage the most. A significant correlation of CAI and protein properties (GRAVY and AROMA) with the first two PCs indicated that protein expression and physical properties are affected by changes in codon usage. Cluster analysis is a method where genes with common attributes are clustered together. All the tested gene transcripts were separated into two clusters; one contained *SIGLEC5* and *AP3B2* and *NIN*, *ABHD12B*, and *WHAMM* transcripts. It indicated that the codon usage pattern was similar for the *SIGLEC5* and *AP3B2* gene transcripts, and a slightly different pattern was followed by the *NIN*, *ABHD12B*, and *WHAMM* gene transcripts.

CP negatively influences glycemic control. Diabetes coupled with CP increases cardiorenal mortality risk [5]. The CP patients demonstrated significant endothelial dysfunction [56], and epidemiological studies revealed a close association between periodontitis and cardiovascular diseases [7], liver diseases [8], rheumatoid arthritis [57], alopecia areata [10], anxiety-like behaviors [11], respiratory diseases [12], and neuropathy in diabetic patients [58]. Our study provides information regarding molecular signatures, gene expression patterns, protein properties, codon usage, codon bias, information regarding codons influencing codon usage to various extents, and dinucleotide pattern. The information is also helpful in modulating the gene expression of the envisaged gene. Envisaged genes, *SIGLEC5*, *AP3B2*, *NIN*, *ABHD12B*, and *WHAMM*, in our study are associated with the CP, and gene product quantities influence the pathobiology. We have observed in our work that both the protein expression and physical properties are affected by changes in

codon usage; therefore, altering codon usage may also manipulate the gene expression pattern. A synthetic biology approach where a gene can be recoded to obtain the desired gene expression level might be helpful. Using the technology, both the upregulation and downregulation of gene expression are possible. The recoded gene may be inserted in CP patients, and the mutated copy of the gene might be exchanged using genome editing tools, such as CRISPR/Cas9 [59]. The codon usage studies, therefore, help combat many pancreatitis-associated disorders.

5. Conclusions

Periodontitis is a complex chronic inflammatory disorder caused by various factors acting simultaneously. Most of them are related to the immune fitness of the host. Further studies of genetic and epigenetic alterations in DNA might affect the genetic blueprint of the host immune response. Our study imparts partial knowledge of molecular signatures associated with the set of genes significantly associated with CP. Further studies are warranted to identify more genetic signatures associated with the disease.

6. Future Directions and Clinical Implications

CP is a chronic disorder and has shown an association with poor glycemic index, atherosclerosis, hair loss, liver inflammation, arthritis, respiratory problems, anxiety, and other mood disorders. Therefore, knowledge of codon usage and other molecular features in genes responsible for CP may help in gene augmentation through gene therapy with a recoded gene sequence so that gene expression levels may be modulated and brought to a normal level and CP-associated pathophysiology may be alleviated.

Author Contributions: Conceptualization, R.K. and M.P.; methodology, R.K.; software, M.P.; validation, I.V.R., M.P., A.A.K. and R.K.; formal analysis, I.V.R.; investigation, supervision, R.K., A.A.K. and I.L.; project administration, A.A.K., R.K. and I.L.; funding acquisition, A.A.K. and I.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by Researchers Supporting Project (RSP-2021/339), King Saud University, Riyadh, Saudi Arabia and project R-1007/2018, University of Murcia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Available upon request.

Acknowledgments: Institutes are thankful for their respective Universities and Institutes for providing research opportunity.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **2017**, *390*, 1211–1259. [\[CrossRef\]](#)
2. Righolt, A.J.; Jevdjevic, M.; Marcenes, W.; Listl, S. Global-, Regional-, and Country-Level Economic Impacts of Dental Diseases in 2015. *J. Dent. Res.* **2018**, *97*, 501–507. [\[CrossRef\]](#)
3. Vos, T.; Flaxman, A.D.; Naghavi, M.; Lozano, R.; Michaud, C.; Ezzati, M.; Shibuya, K.; Salomon, J.A.; Abdalla, S.; Aboyans, V.; et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2163–2196. [\[CrossRef\]](#)
4. Bretz, W.A.; Corby, P.M.; Schork, N.J.; Robinson, M.T.; Coelho, M.; Costa, S.; Melo Filho, M.R.; Weyant, R.J.; Hart, T.C. Longitudinal analysis of heritability for dental caries traits. *J. Dent. Res.* **2005**, *84*, 1047–1051. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Preshaw, P.M.; Alba, A.L.; Herrera, D.; Jepsen, S.; Konstantinidis, A.; Makrilakis, K.; Taylor, R. Periodontitis and diabetes: A two-way relationship. *Diabetologia* **2012**, *55*, 21–31. [\[CrossRef\]](#)
6. Moura, M.F.; Navarro, T.P.; Silva, T.A.; Cota, L.O.M.; Soares Dutra Oliveira, A.M.; Costa, F.O. Periodontitis and Endothelial Dysfunction: Periodontal Clinical Parameters and Levels of Salivary Markers Interleukin-1 β , Tumor Necrosis Factor- α , Matrix Metalloproteinase-2, Tissue Inhibitor of Metalloproteinases-2 Complex, and Nitric Oxide. *J. Periodontol.* **2017**, *88*, 778–787. [\[CrossRef\]](#)

7. Sanz, M.; Marco Del Castillo, A.; Jepsen, S.; Gonzalez-Juanatey, J.R.; D’Aiuto, F.; Bouchard, P.; Chapple, I.; Dietrich, T.; Gotsman, I.; Graziani, F.; et al. Periodontitis and cardiovascular diseases: Consensus report. *J. Clin. Periodontol.* **2020**, *47*, 268–288. [[CrossRef](#)] [[PubMed](#)]
8. Han, P.; Sun, D.; Yang, J. Interaction between periodontitis and liver diseases. *Biomed. Rep.* **2016**, *5*, 267–276. [[CrossRef](#)]
9. Bingham, C.O.; Moni, M. Periodontal disease and rheumatoid arthritis: The evidence accumulates for complex pathobiologic interactions. *Curr. Opin. Rheumatol.* **2013**, *25*, 345–353. [[CrossRef](#)]
10. Gil Montoya, J.A.; Cutando Soriano, A.; Jimenez Prat, J. Alopecia areata of dental origin. *Med. Oral.* **2002**, *7*, 303–308. [[PubMed](#)]
11. Varotto, B.L.R.; Martinez, R.C.R.; Gouveia, F.V.; Antunes, G.F.; de Campos Fabri, G.M.; Ballester, G.; Antequera, R.; de Siqueira, S.R.D.T.; Fonoff, E.T.; Teixeira, M.J.; et al. Increased Anxiety-Like Behavior in the Acute Phase of a Preclinical Model of Periodontal Disease. *Front. Neurol.* **2020**, *11*, 598851. [[CrossRef](#)]
12. Bansal, M.; Khatri, M.; Taneja, V. Potential role of periodontal infection in respiratory diseases—A review. *J. Med. Life* **2013**, *6*, 244–248.
13. Alvarenga, M.O.P.; Miranda, G.H.N.; Ferreira, R.O.; Saito, M.T.; Fagundes, N.C.F.; Maia, L.C.; Lima, R.R. Association between Diabetic Retinopathy and Periodontitis—A Systematic Review. *Front. Public Health* **2020**, *8*, 550614. [[CrossRef](#)] [[PubMed](#)]
14. Mahendra, J.; Mahendra, L.; Mugri, M.H.; Sayed, M.E.; Bhandi, S.; Alshahrani, R.T.; Balaji, T.M.; Varadarajan, S.; Tanneeru, S.; Rao, A.N.; et al. Role of Periodontal Bacteria, Viruses, and Placental mir155 in Chronic Periodontitis and Preeclampsia—A Genetic Microbiological Study. *Curr. Issues Mol. Biol.* **2021**, *43*, 831–844. [[CrossRef](#)]
15. Graves, D.T.; Oates, T.; Garlet, G.P. Review of osteoimmunology and the host response in endodontic and periodontal lesions. *J. Oral. Microbiol.* **2011**, *3*, 5304. [[CrossRef](#)] [[PubMed](#)]
16. Tong, H.; Wei, Z.; Yin, J.; Zhang, B.; Zhang, T.; Deng, C.; Huang, Y.; Zhang, N. Genetic susceptibility of common polymorphisms in NIN and SIGLEC5 to chronic periodontitis. *Sci. Rep.* **2019**, *9*, 2088. [[CrossRef](#)] [[PubMed](#)]
17. Rhodin, K.; Divaris, K.; North, K.E.; Barros, S.P.; Moss, K.; Beck, J.D.; Offenbacher, S. Chronic periodontitis genome-wide association studies: Gene-centric and gene set enrichment analyses. *J. Dent. Res.* **2014**, *93*, 882–890. [[CrossRef](#)]
18. Choudhury, M.N.; Chakraborty, S. Codon usage pattern in human SPANX genes. *Bioinformatics* **2015**, *11*, 454–459. [[CrossRef](#)]
19. Moura, G.R.; Pinheiro, M.; Freitas, A.; Oliveira, J.L.; Frommlet, J.C.; Carreto, L.; Soares, A.R.; Bezerra, A.R.; Santos, M.A.S. Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. *PLoS ONE* **2011**, *6*, e26817. [[CrossRef](#)]
20. Payne, B.L.; Alvarez-Ponce, D. Codon Usage Differences among Genes Expressed in Different Tissues of *Drosophila melanogaster*. *Genome Biol. Evol.* **2019**, *11*, 1054–1065. [[CrossRef](#)]
21. Plotkin, J.B.; Robins, H.; Levine, A.J. Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12588–12591. [[CrossRef](#)] [[PubMed](#)]
22. Alqahtani, T.; Khandia, R.; Puranik, N.; Alqahtani, A.M.; Almikhlaifi, M.A.; Alqahtany, M.A. Leucine encoding codon TTG shows an inverse relationship with GC content in genes involved in neurodegeneration with iron accumulation. *J. Integr. Neurosci.* **2021**, *20*, 905–918. [[CrossRef](#)] [[PubMed](#)]
23. Khandia, R.; Alqahtani, T.; Alqahtani, A.M. Genes Common in Primary Immunodeficiencies and Cancer Display Overrepresentation of Codon CTG and Dominant Role of Selection Pressure in Shaping Codon Usage. *Biomedicines* **2021**, *9*, 1001. [[CrossRef](#)]
24. Paul, P.; Malakar, A.K.; Chakraborty, S. Compositional bias coupled with selection and mutation pressure drives codon usage in *Brassica campestris* genes. *Food Sci. Biotechnol.* **2018**, *27*, 725–733. [[CrossRef](#)] [[PubMed](#)]
25. Zhou, Z.; Dang, Y.; Zhou, M.; Li, L.; Yu, C.-H.; Fu, J.; Chen, S.; Liu, Y. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6117–E6125. [[CrossRef](#)]
26. Khandia, R.; Ali Khan, A.; Alexiou, A.; Povetkin, S.N.; Verevkina, M.N. Codon Usage Analysis of Pro-Apoptotic Bim Gene Isoforms. *J. Alzheimer’s Dis.* **2022**, *86*, 1711–1725. [[CrossRef](#)]
27. Richards, D. Review finds that severe periodontitis affects 11% of the world population. *Evid. Based Dent.* **2014**, *15*, 70–71. [[CrossRef](#)]
28. Zhou, T.; Gu, W.; Ma, J.; Sun, X.; Lu, Z. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems* **2005**, *81*, 77–86. [[CrossRef](#)]
29. Sharp, P.M.; Li, W.H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [[CrossRef](#)]
30. Puigbò, P.; Bravo, I.G.; Garcia-Vallve, S. CAIcal: A combined set of tools to assess codon usage adaptation. *Biol. Direct* **2008**, *3*, 38. [[CrossRef](#)]
31. Sharp, P.M.; Li, W.H. The codon Adaptation Index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **1987**, *15*, 1281–1295. [[CrossRef](#)]
32. Naya, H.; Romero, H.; Carels, N.; Zavala, A.; Musto, H. Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. *FEBS Lett.* **2001**, *501*, 127–130. [[CrossRef](#)]
33. Stenico, M.; Lloyd, A.T.; Sharp, P.M. Codon usage in *Caenorhabditis elegans*: Delineation of translational selection and mutational biases. *Nucleic Acids Res.* **1994**, *22*, 2437–2446. [[CrossRef](#)] [[PubMed](#)]
34. Wright, F. The “effective number of codons” used in a gene. *Gene* **1990**, *87*, 23–29. [[CrossRef](#)]
35. Khandia, R.; Singhal, S.; Kumar, U.; Ansari, A.; Tiwari, R.; Dhama, K.; Das, J.; Munjal, A.; Singh, R.K. Analysis of Nipah Virus Codon Usage and Adaptation to Hosts. *Front. Microbiol.* **2019**, *10*, 886. [[CrossRef](#)] [[PubMed](#)]

36. Munjal, A.; Khandia, R.; Shende, K.K.; Das, J. *Mycobacterium lepromatosis* genome exhibits unusually high CpG dinucleotide content and selection is key force in shaping codon usage. *Infect. Genet. Evol.* **2020**, *84*, 104399. [[CrossRef](#)] [[PubMed](#)]
37. Sueoka, N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.* **1995**, *40*, 318–325. [[CrossRef](#)]
38. Wang, L.; Xing, H.; Yuan, Y.; Wang, X.; Saeed, M.; Tao, J.; Feng, W.; Zhang, G.; Song, X.; Sun, X. Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS ONE* **2018**, *13*, e0194372. [[CrossRef](#)]
39. Yu, X.; Liu, J.; Li, H.; Liu, B.; Zhao, B.; Ning, Z. Comprehensive analysis of synonymous codon usage patterns and influencing factors of porcine epidemic diarrhea virus. *Arch. Virol.* **2021**, *166*, 157–165. [[CrossRef](#)]
40. Liu, X.-Y.; Li, Y.; Ji, K.-K.; Zhu, J.; Ling, P.; Zhou, T.; Fan, L.-Y.; Xie, S.-Q. Genome-wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in *Cuscuta australis*. *Genomics* **2020**, *112*, 2695–2702. [[CrossRef](#)]
41. Guan, D.-L.; Ma, L.-B.; Khan, M.S.; Zhang, X.-X.; Xu, S.-Q.; Xie, J.-Y. Analysis of codon usage patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant selection constraints. *BMC Genom.* **2018**, *19*, 542. [[CrossRef](#)] [[PubMed](#)]
42. Hammer, O.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2001**, *4*, 4.
43. Jenkins, G.M.; Holmes, E.C. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* **2003**, *92*, 1–7. [[CrossRef](#)]
44. Plotkin, J.B.; Kudla, G. Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* **2011**, *12*, 32–42. [[CrossRef](#)] [[PubMed](#)]
45. Behura, S.K.; Severson, D.W. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS ONE* **2012**, *7*, e43111. [[CrossRef](#)]
46. Lee, S.; Weon, S.; Lee, S.; Kang, C. Relative codon adaptation index, a sensitive measure of codon usage bias. *Evol. Bioinform. Online* **2010**, *6*, 47–55. [[CrossRef](#)]
47. Uddin, A. Compositional Features and Codon Usage Pattern of Genes Associated with Anxiety in Human. *Mol. Neurobiol.* **2020**, *57*, 4911–4920. [[CrossRef](#)]
48. Tyagi, A.; Singh, N.K.; Gurtler, V.; Karunasagar, I. Bioinformatics analysis of codon usage patterns and influencing factors in *Penaes monodon* nudivirus. *Arch. Virol.* **2016**, *161*, 459–464. [[CrossRef](#)]
49. Nair, R.R.; Nandhini, M.B.; Monalisha, E.; Murugan, K.; Sethuraman, T.; Nagarajan, S.; Rao, N.S.P.; Ganesh, D. Synonymous codon usage in chloroplast genome of *Coffea arabica*. *Bioinformation* **2012**, *8*, 1096–1104. [[CrossRef](#)]
50. Chakraborty, S.; Nag, D.; Mazumder, T.H.; Uddin, A. Codon usage pattern and prediction of gene expression level in *Bungarus* species. *Gene* **2017**, *604*, 48–60. [[CrossRef](#)]
51. Pan, Y.; Li, X.; Yang, G.; Fan, J.; Tang, Y.; Zhao, J.; Long, X.; Guo, S.; Zhao, Z.; Liu, Y.; et al. Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. *J. Infect.* **2020**, *81*, e28–e32. [[CrossRef](#)] [[PubMed](#)]
52. Fuller, S.J.; Sohrabi, H.R.; Goozee, K.G.; Sankaranarayanan, A.; Martins, R.N. Alzheimer’s Disease and Other Neurodegenerative Diseases. In *Neurodegeneration and Alzheimer’s Disease*; Martins, R.N., Brennan, C.S., Eds.; Wiley: Hoboken, NJ, USA, 2019; pp. 11–42. ISBN 978-1-119-35678-3.
53. Schwartz, S.; Meshorer, E.; Ast, G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* **2009**, *16*, 990–995. [[CrossRef](#)] [[PubMed](#)]
54. Uddin, A.; Chakraborty, S. Codon Usage Pattern of Genes Involved in Central Nervous System. *Mol. Neurobiol.* **2019**, *56*, 1737–1748. [[CrossRef](#)] [[PubMed](#)]
55. Kliman, R.M.; Bernal, C.A. Unusual usage of AGG and TTG codons in humans and their viruses. *Gene* **2005**, *352*, 92–99. [[CrossRef](#)] [[PubMed](#)]
56. de Boer, S.P.M.; Cheng, J.M.; Rangé, H.; Garcia-Garcia, H.M.; Heo, J.H.; Akkerhuis, K.M.; Meilhac, O.; Cosler, G.; Pussinen, P.J.; van Geuns, R.-J.; et al. Antibodies to periodontal pathogens are associated with coronary plaque remodeling but not with vulnerability or burden. *Atherosclerosis* **2014**, *237*, 84–91. [[CrossRef](#)]
57. Ranade, S.B.; Doiphode, S. Is there a relationship between periodontitis and rheumatoid arthritis? *J. Indian Soc. Periodontol.* **2012**, *16*, 22–27. [[CrossRef](#)]
58. Menchaca-Díaz, R.; Bogarín-López, B.; Zamudio-Gómez, M.A.; Anzaldo-Campos, M.C. Severe periodontitis, edentulism and neuropathy in patients with type 2 diabetes mellitus. *Gac. Med. Mex.* **2012**, *148*, 34–41. [[PubMed](#)]
59. Zhang, Y.; Showalter, A.M. CRISPR/Cas9 Genome Editing Technology: A Valuable Tool for Understanding Plant Cell Wall Biosynthesis and Function. *Front. Plant Sci.* **2020**, *11*, 589517. [[CrossRef](#)]