# Evolution of Synonymous Codon Usage Bias in West African and Central African Strains of Monkeypox Virus

Sudeesh Karumathil[1], Nimal T Raveendran[2], Doss Ganesh[3], Sampath Kumar NS[4], Rahul R Nair[1] and Vijaya R Dirisala[4]

[1]Centre for Evolutionary Ecology, Aushmath Biosciences, Coimbatore, India. [2]Amrita Centre for Nanosciences, Amrita Vishwa Vidyapeetham, Kochi, India. [3]Department of Plant Biotechnology, School of Biotechnology, Madurai Kamaraj University, Madurai, India. [4]Department of Biotechnology, Vignan's University, Guntur, India.

**ABSTRACT:** The evolution of bias in synonymous codon usage in chosen monkeypox viral genomes and the factors influencing its diversification have not been reported so far. In this study, various trends associated with synonymous codon usage in chosen monkeypox viral genomes were investigated, and the results are reported. Identification of factors that influence codon usage in chosen monkeypox viral genomes was done using various codon usage indices, such as the relative synonymous codon usage, the effective number of codons, and the codon adaptation index. The Spearman rank correlation analysis and a correspondence analysis were used for correlating various factors with codon usage. The results revealed that mutational pressure due to compositional constraints, gene expression level, and selection at the codon level for utilization of putative optimal codons are major factors influencing synonymous codon usage bias in monkeypox viral genomes. A cluster analysis of relative synonymous codon usage values revealed a grouping of more virulent strains as one major cluster (Central African strains) and a grouping of less virulent strains (West African strains) as another major cluster, indicating a relationship between virulence and synonymous codon usage bias. This study concluded that a balance between the mutational pressure acting at the base composition level and the selection pressure acting at the amino acid level frames synonymous codon usage bias in the chosen monkeypox viruses. The natural selection from the host does not seem to have influenced the synonymous codon usage bias in the analyzed monkeypox viral genomes.

**KEYWORDS:** Monkeypox viruses (MPXV), synonymous codon usage bias (SCUB), mutational pressure, selection pressure

## Introduction

Molecular evolution is a broad term reflecting changes in various genomic parameters due to alterations in the nucleotide and the dinucleotide compositions that lead to an accumulation of mutations over time.[1] Because the genetic code is degenerate, more than one codon can encode a particular amino acid; however, the usage of these "synonymous codons" for a given amino acid is not uniform.[2] In a given amino acid, a subset of codons may be used more frequently than others are, and such a subset is referred to as "preferred codons."[3] Synonymous codon usage bias (SCUB) is species specific and varies within and between genomes.[4] This nonuniform usage of synonymous codons (ie, SCUB) can be significant in highly expressed genes.[5] Thus, an understanding of SCUB is critical as it reveals the various forces that frame genomic evolution.[6]

The mutational pressure, which is due to base compositional constraints, and the selection pressure, which increases the translational speed and accuracy, have been identified as 2 important forces causing SCUB in various lineages, such as plants, mammals, macro-invertebrates, bacteria, fungi, and viruses.[7–11] Selection pressure favors codons having abundant transfer RNAs, particularly in highly expressed genes.[12–15] Furthermore, synonymous codon choices for protein formation have been found to affect secondary structure and protein folding,[16] and messenger RNAs (mRNAs) and protein structures have been found to cause selection pressure.[17–19] For instance, a significant species-specific correlation was noticed between the usage of AAC (asparagine) and the C-terminal regions of β-sheet segments in *Escherichia coli* as selection for translational efficiency favors downstream asparagine (AAC) residues that are essential for the formation of the β-sheet.[19] Similarly, a significant correlation was found to exist between GAU (aspartic acid) and the N-termini of α-helices in humans as selection acts on co-translational protein folding in eukaryotes.[19] In an another important study, selection on synonymous codon usage (SCU) facilitated the optimization of the characteristics of mRNA secondary structures as a specific codon usage pattern was observed in the nucleotide sequence of repetitive units of silk fibroin mRNA.[17] However, in another study, the mutational pressure was found to frame the overall nucleotide composition in genomes through GC « » AT changes,[20] and intrinsic bias in dinucleotide frequencies may have had an influence on SCUB[6] as such bias can be extreme.[20] For instance, the CpG (C-phosphate-G) content is underrepresented in many vertebrates owing to the methylation of cytosine residues,[21] and the

TpA (T-phosphate-A) content is restricted in many organisms due to the susceptibility of uracil in UpA to RNase[22] and low thermal stability.[23]

The quantification of SCUB and the identification of its causative factors in zoonotic viral genomes are crucial in understanding viral evolution and ecology.[6] Detailed analyses of trends and SCUB-associated factors are essential if the mechanisms of viral infection and immune response are to be revealed.[20] Greater emphasis on understanding the various factors contributing to codon usage patterns is, therefore, more important than merely understanding viral SCUB.[24–28] The survival, fitness, and evolution of viruses depend strongly on SCUB coactions between viruses and hosts because replication and translation of viral genomes are host associated.[20] Few studies have been undertaken to reveal the major forces and trends associated with viral DNA SCUB.[20,29,30] Substantial differences between the SCUB in a virus and that in its host will have an effect on viral replication and protein synthesis,[31] as evidenced in human papillomaviruses.[32]

Monkeypox viruses (MPXVs) belong to the genus *Orthopoxvirus* of the family Poxviridae.[33] The family Poxviridae consists of large double-stranded DNA viruses capable of replicating in the cytoplasm of vertebrate and invertebrate cells.[33,34] Monkeypox viruses cause human diseases similar to the eradicated smallpox caused by the variola virus (VAR).[33] By 1977, smallpox was reported to have been eradicated and vaccination was stopped.[35] As a result, closely related zoonotic viruses such as MPXVs infected unvaccinated human populations and caused a fatal illness (human monkeypox), but with a very low human-to-human transmission rate.[36,37] Although human monkeypox is clinically similar to smallpox, regarding the case fatality rates (CFRs), smallpox was reported to be severe than human monkeypox, with the former having a CFR of 30%[38] and the latter a CFR of 10%.[36] A recent outbreak investigation conducted in the Bokungu Health Zone of the Democratic Republic of the Congo (DRC) from July 1 to December 8, 2013, revealed a 600-fold increase in the number of human monkeypox cases.[39]

Rodents are the major animal reservoirs for MPXVs.[40–42] The viral transmission to humans takes place through direct contact with animals.[43] Wounds in the skin are the major route through which infection happens while handling infected animals.[41] In some cases, respiratory transmission from animal to human and then from human to human has occurred.[41,44] The incubation period is 10 to 14 days.[43] After the incubation period, the prodromal period lasts for 2 days, and in this phase, the infected individual may experience fever, chills, malaise, headache, backache, sore throat, shortness of breath, and swollen lymph nodes.[45,46] A clinical feature that can be used to differentiate between human monkeypox and human smallpox infections is the presence of enlarged lymph nodes in the submandibular, cervical, or inguinal regions in the former.[35] The infected individual becomes most contagious subsequent to the development of a progressive maculopapular rash (0.2-1.0 cm) after the prodromal period.[45,46] The spread of the lesions over the body follows a centrifugal pattern, and in certain cases, dyspigmented scars may develop from the lesions.[43] In general, during a 2- to 4-week time period, the lesions over the body progressively undergo several changes from macules to papules, vesicles, and pustules, followed by scabbing and desquamation.[35,43]

Human monkeypox is endemic to the DRC, and infections take place throughout the Congo Basin.[39] Different isolates of MPXVs from West Africa and the Congo Basin have been proven to be genetically distinct, and substantial differences in virulence between them have been reported.[47] For instance, MPXV-ZAI-V79 isolated from the Congo Basin is thought to be more virulent than MPXV-COP-58 isolated from West Africa[47] as no mortalities were reported during the West African isolate MPXV outbreaks in the United States in 2003.[47] However, high virulence (>90%) and fatalities have been reported in the Congo Basin, and D10L, D14L, B10R, B14R, and B19R have been identified as possible candidate loci for virulence.[47] Although genetic analyses revealed that MPXVs are not the immediate ancestors of the VAR because considerable differences were found between MPXVs and the VAR in the terminal genomic regions encoding virulence and host range factors, the possibility of an MPXV evolving into a highly virulent VAR-like virus with significant human-to-human transmission rates cannot be ignored.[37]

In this study, extensive analyses of SCUBs in 13 representative MPXV genomes isolated from different African regions were conducted to unravel patterns and factors associated with MPXV diversification. The size of the double-stranded DNA genome of an MPXV is ≈200 kb, comprising ≈190 nonoverlapping open reading frames (ORFs) that contain ≥180 nucleotides.[48] A typical monkeypox genome contains a central conserved region (≈56 000 to 120 000 nucleotides long), with variable regions to the left and the right, as well as an inverted terminal region (ITR) with tandem repeats.[33] The central conserved region contains genes with the codes for the replication machinery.[48] The ITR in the MPXV genome represents a global repeat[49,50] and accounts for almost 1% of the total genome size.[50,51] At least 4 ORFs are included in the ITR of the MPXV genomes.[52,53] The ORFs in the ITR take part in the virus-host interactions.[48,54]

As differences in virulence regarding location have been reported,[47] an objective of this study was to reveal associations between virulence and various trends associated with SCUB in MPXV genomes. The results of this research should contribute to an understanding of the coaction between the genome-wide neutral mutational and selection pressures, which, in turn, increases our understanding of viral DNA evolution, as well as the interactions between the viruses and their hosts. Most importantly, the results of SCUB analyses of viral genomes should have important applications in studies related to the genetic engineering of viral genome sequences.[20]

**Table 1.** Details of examined monkeypox virus strains.

| S. NO. | STRAIN | ACCESSION NO. | ISOLATION | NO. OF CHOSEN CODING SEQUENCES | LENGTH |
|---|---|---|---|---|---|
| 1 | Congo-2003-358 | DQ011154.1 | Congo | 158 | 160 929.0 |
| 2 | COP-58 | AY753185.1 | West Africa | 152 | 156 321.0 |
| 3 | DRC Yandongi-1985 | KC257460.1 | Congo: Yandongi | 157 | 159 768.0 |
| 4 | Liberia-1970-184 | DQ011156.1 | Liberia | 161 | 161 544.0 |
| 5 | MPXV-WRAIR7-61 | AY603973.1 | West Africa | 152 | 156 414.0 |
| 6 | Sierra Leone | AY741551.1 | West Africa | 151 | 155 874.0 |
| 7 | Sudan-2005-01 | KC257459.1 | Sudan: Nuria | 169 | 171 372.0 |
| 8 | USA-2003-039 | DQ011157.1 | USA | 160 | 161 013.0 |
| 9 | USA-2003-044 | DQ011153.1 | USA | 160 | 161 013.0 |
| 10 | V79-I-005 | HQ857562.1 | Zaire | 159 | 160 967.0 |
| 11 | Zaire-1979-005 (cidofovir resistant) | HM172544.1 | Zaire | 157 | 156 474.0 |
| 12 | Zaire-1979-005 | DQ011155.1 | Zaire | 161 | 161 664.0 |
| 13 | Zaire-96-I-16 | NC_003310.1 | Zaire | 158 | 160 944.0 |

## Materials and Methods

### Sequence data

The complete genomes of 13 representative MPXVs (Table 1) were retrieved from the National Center for Biotechnology Information. Details such as accession numbers, the region of isolation, the number of coding sequences (CDSs) selected, and the sizes of the genomes were also provided (Table 1). The integrity of full-length coding sequences without introns was confirmed by checking for the presence of proper initiation and termination codons.[55] To avoid sampling errors and stochastic variations, we chose CDSs having more than 300 nucleotides for analysis (Table 1).[8] Information regarding the ITRs of the MPXV genomes was obtained from GenBank, and for the calculation of the codon usage in an ITR, the orientation was changed in such a way as to maintain the corresponding amino acid sequences intact and thereby avoid any miscalculation of the codon usage.

### Measures of SCU

The effective number of codons (ENC) is a commonly employed index for measuring SCUB independently of the length of the CDS.[56] The ENC values vary from 20 to 61. In any given gene, if only one codon is used to encode one particular amino acid, the ENC value will be 20 (extreme SCUB). If all synonymous codons of a particular amino acid are used equally, the ENC value will be 61 (almost no SCUB). The compositions of the G and the C nucleotides were calculated for the first, second, and third codon positions. Expected ENC values were calculated using the $GC_3$ (GC composition at the third codon position) values.[56] An ENC versus $GC_3$ plot can be used to distinguish between the 2 major evolutionary forces, the mutational pressure and the translational selection, for the observed SCU patterns by displaying gene groupings along the expected ENC curve. This is true because these 2 major evolutionary forces are the ones that contribute to SCUB. Even though, in some cases, genetic drift can be considered as a factor shaping codon usage; the ENC versus $GC_3$ plot will only give an indication of the influences of the mutational pressure and the selection pressure. In this research, ENC values were calculated according to the following equation[56]:

$$ENC = 2 + \frac{9}{F2} + \frac{1}{F3} + \frac{5}{F4} + \frac{3}{F6}$$

where $F2$, $F3$, $F4$, and $F6$ are the average homozygosity values for 4 different synonymous family types and were estimated using the codon frequencies squared. The average homozygosity for each amino acid was calculated according to the following equation[56]:

$$F = \frac{n\left(\sum_{i=1}^{k} p_i^2 - 1\right)}{n-1}$$

where $k$ is the number of alleles squared. The expected ENC versus $GC_3$ curve was plotted using $GC_3$ values ranging from 0% to 100% in intervals of 10% and their corresponding expected ENC values, which, under no selection, can be calculated using the following equation[56]:

$$E(ENC) = 2 + s + \left\{ 29 \, \middle| \, \left[ s^2 + (1-s)^2 \right] \right\}$$

where $s = GC_3$.

The relative SCU (RSCU), which is the ratio of the observed codon frequency to the expected codon frequency, provided all synonymous codons of that particular amino acid have uniform usage, is another important index for measuring SCUB.[3,12] The RSCU values greater than 1 denote codons used more frequently than their synonymous counterparts, whereas the RSCU values less than 1 represent codons used less frequently; codons with an RSCU value of 1 denote no bias.[3]

The codon adaptation index (CAI) assesses the significance of selection in shaping the observed patterns of the SCU of a gene[5] using a reference set of highly expressed genes from a particular species. The CAI indicates the level of gene expression[5,10,11] by calculating a score for each gene. The CAI values from 0.75 to 1.0 indicate a high level of gene expression.[5] Although the CAI is independent of gene length, the CAI of short genes may be affected by sampling bias.[5] We used the *Homo sapiens* general codon usage table as a reference set because the CAI is a good indicator of viral gene adaptation to the host.[5]

Protein hydrophobicity and aromaticity (ie, frequency of aromatic amino acids such as Phe, Trp, and Tyr) were calculated.[57] A correspondence analysis of RSCU (COA-RSCU) has been generally adopted to identify intragenomic variations while avoiding the influence of the amino acid's composition.[8,11] In a COA-RSCU, each CDS is represented as a 59-dimensional vector,[58] wherein each dimension corresponds to the RSCU value of a particular codon.[58] The COA-RSCU partitions the total variation in codon usage across 59 orthogonal axes with 41 degrees of freedom.[8] The first axis of the COA-RSCU (axis 1) accounts for most of variations, whereas subsequent axes capture decreasing amounts of variance.[8]

Putative optimal codons were identified by applying the $\chi^2$ test to a $2 \times 2$ matrix having 1 degree of freedom. We chose 10% of the genes lying on the left and the right extremes of axis 1 of the COA-RSCU to form 2 data sets as axis 1 of the COA-RSCU accounts for most of the variations in the RSCU. The first row of this matrix contains the observed codon frequencies from the 2 data sets, whereas the second row contains the total number of synonymous alternatives of that particular codon.[8] Codons whose frequencies of usage were significantly higher ($P < .05$) in one data set than in the other data set were defined as putative optimal codons.

### Cluster analysis

A cluster analysis of the RSCU values was performed to reveal the relationship between the SCUB and other factors based on groupings of the codon usage.[7] In the cluster analysis, a $13 \times 59$ matrix, in which rows and columns corresponded to the 13 MPXV strains and the pooled RSCU values of the 59 codon species, respectively, was generated. Clustering of the MPXVs based on RSCU values was found to have occurred using unweighted pair-group average clustering and Euclidean distances.

### Statistical analysis

The nonparametric Spearman rank correlation was adopted for all correlation analyses between the various codon usage indices and the other parameters as it does not hold any assumptions regarding the distribution of underlying data.[8,55] The Mann-Whitney 2-sample test was used to analyze the intergenomic differences in the ENC values. PAST software version 2.12 was used for the Spearman rank correlation analysis.[59] CodonW (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::codonw) was used to compute the values of the ENC, hydrophobicity, and aromaticity.[60] MEGA version 5.2.2 was used to calculate the compositions of the nucleotides.[61] DAMBE version 5.3.31 was employed to determine the RSCU values,[62] and the CAI values were computed using ACUA 1.0.[63] The level of significance was taken as .05.

## Results

### Effect of base composition on SCUB

The overall and the wobble base contents were estimated in all 13 examined MPXV genomes. Overall, the AT content was found to be higher than the GC content. Among the individual nucleotide compositions, the A content was higher than the T, G, and C contents and varied by $35.26 \pm 0.053$; thus, it was overrepresented in the protein-coding genes (PCGs) of all genomes. In all examined genomes, the C content was observed to be the least among all other nucleotide contents and to vary by $15.52 \pm 0.025$; thus, it was underrepresented in the PCGs of all genomes. Moreover, the GC content was observed to vary by $33.74 \pm 0.065$ in all genomes. Because the base changes that occur at the third site of synonymous codons for a given amino acid are neutral, the third site of a codon is commonly known as "the silent site." Interestingly, the $T_3$ content was higher than the contents of other silent bases ($A_3$, $G_3$, and $C_3$) and was found to vary by $38.23 \pm 0.082$; the GC composition at silent sites ($GC_3$) was found to vary by $29.12 \pm 0.080$.

A Spearman rank correlation analysis revealed complex correlations between the overall and the silent base compositions, indicating the presence of compositional constraints in all genomes. The existence of positive correlations between homogeneous nucleotide contents and negative correlations between heterogeneous nucleotide contents implies that mutational pressure due to compositional constraints might play a crucial role in shaping the codon usage.[64] In the case of viral genomes, the positively correlated heterogeneous contents and the negatively correlated homogeneous contents indicate natural selection by the host.[24] In this study, significant positive correlations were found between A and $A_3$, T and $T_3$, G and $G_3$, and C and $C_3$. The most heterogeneous base contents were found for significant negative correlations (Table 2). The $G_3$, $C_3$, and $GC_3$ contents were found to have significant positive correlations with the overall GC content. No correlations were observed

**Table 2.** Spearman rank correlation analysis between overall and silent base compositions.

| STRAINS | BASES | $A_3$ | $T_3$ | $G_3$ | $C_3$ | $GC_3$ |
|---|---|---|---|---|---|---|
| Congo-2003-358 | A | 0.536** | −0.147 | −0.408** | −0.155 | −0.395** |
| | T | −0.150 | 0.537** | −0.087 | −0.344** | −0.319** |
| | G | −0.283** | −0.070 | 0.458** | 0.074 | 0.358** |
| | C | −0.154 | −0.371** | 0.098 | 0.526** | 0.458** |
| | GC | −0.298** | −0.270** | 0.381** | 0.361** | 0.526** |
| C0P-58 | A | 0.538** | −0.148 | −0.409** | −0.170* | −0.396** |
| | T | −0.123 | 0.558** | −0.149 | −0.324** | −0.335** |
| | G | −0.293** | −0.089 | 0.505** | 0.055 | 0.362** |
| | C | −0.159 | −0.334** | 0.097 | 0.508** | 0.435** |
| | GC | −0.310** | −0.267** | 0.422** | 0.343** | 0.526** |
| DRC Yandongi-1985 | A | 0.556** | −0.171* | −0.409** | −0.175* | −0.392** |
| | T | −0.144 | 0.578** | −0.134 | −0.337** | −0.351** |
| | G | −0.283** | −0.091 | 0.487** | 0.076 | 0.366** |
| | C | −0.157 | −0.346** | 0.089 | 0.538** | 0.450** |
| | GC | −0.306** | −0.278** | 0.407** | 0.371** | 0.539** |
| Liberia-1970-184 | A | 0.544** | −0.185* | −0.356** | −0.158 | −0.365** |
| | T | −0.132 | 0.553** | −0.147 | −0.325** | −0.338** |
| | G | −0.272** | −0.093 | 0.470** | 0.060 | 0.365** |
| | C | −0.152 | −0.294** | 0.067 | 0.483** | 0.405** |
| | GC | −0.291** | −0.253** | 0.389** | 0.325** | 0.516** |
| MPXV-WRAI7-61 | A | 0.538** | −0.151 | −0.410** | −0.170* | −0.394** |
| | T | −0.124 | 0.555** | −0.153 | −0.319** | −0.333** |
| | G | −0.293** | −0.087 | 0.508** | 0.055 | 0.363** |
| | C | −0.158 | −0.333** | 0.101 | 0.506** | 0.435** |
| | GC | −0.310** | −0.264** | 0.424** | 0.343** | 0.525** |
| Sierra Leone | A | 0.528** | −0.159 | −0.386** | −0.182 | −0.386** |
| | T | −0.120 | 0.584** | −0.168* | −0.340** | −0.364** |
| | G | −0.286** | −0.079 | 0.484** | 0.059 | 0.356** |
| | C | −0.136 | −0.327** | 0.085 | 0.503** | 0.426** |
| | GC | −0.295** | −0.252** | 0.401** | 0.347** | 0.519** |
| Sudan 2005-01 | A | 0.524** | −0.190* | −0.366** | −0.132 | −0.346** |
| | T | −0.107 | 0.570** | −0.131 | −0.361** | −0.364** |
| | G | −0.277** | −0.039 | 0.456** | 0.047 | 0.327** |
| | C | −0.165* | −0.338** | 0.081 | 0.516** | 0.439** |
| | GC | −0.318** | −0.253** | 0.374** | 0.359** | 0.521** |
| USA-2003-039 | A | 0.536** | −0.184* | −0.362** | −0.161* | −0.369** |
| | T | −0.132 | 0.569** | −0.148 | −0.327** | −0.339** |

*(Continued)*

**Table 2.** (Continued)

| STRAINS | BASES | $A_3$ | $T_3$ | $G_3$ | $C_3$ | $GC_3$ |
|---|---|---|---|---|---|---|
| | G | −0.278** | −0.117 | 0.491** | 0.059 | 0.386** |
| | C | −0.151 | −0.287** | 0.063 | 0.489** | 0.398** |
| | GC | −0.292** | −0.275** | 0.407** | 0.329** | 0.532** |
| USA-2003-044 | A | 0.536** | −0.184* | −0.362** | −0.161* | −0.369** |
| | T | −0.132 | 0.569** | −0.148 | −0.327** | −0.339** |
| | G | −0.278** | −0.117 | 0.491** | 0.059 | 0.386** |
| | C | −0.151 | −0.287** | 0.063 | 0.489** | 0.398** |
| | GC | −0.292** | −0.275** | 0.407** | 0.329** | 0.532** |
| V79-I-005 | A | 0.529** | −0.167* | −0.391** | −0.137 | −0.368** |
| | T | −0.126 | 0.573** | −0.128 | −0.366** | −0.362** |
| | G | −0.284** | −0.089 | 0.476** | 0.063 | 0.366** |
| | C | −0.133 | −0.351** | 0.073 | 0.519** | 0.439** |
| | GC | −0.285** | −0.280** | 0.384** | 0.347** | 0.527** |
| Zaire-1979-005 (cr) | A | 0.534** | −0.139 | −0.405** | −0.167 | −0.390** |
| | T | −0.131 | 0.538** | −0.098 | −0.347** | −0.334** |
| | G | −0.296** | −0.081 | 0.465** | 0.086 | 0.365** |
| | C | −0.148 | −0.366** | 0.096 | 0.522** | 0.452** |
| | GC | −0.311** | −0.281** | 0.391** | 0.374** | 0.540** |
| Zaire-1979-005 | A | 0.534** | −0.149 | −0.407** | −0.149 | −0.385** |
| | T | −0.139 | 0.544** | −0.105 | −0.345** | −0.334** |
| | G | −0.287** | −0.092 | 0.478** | 0.074 | 0.370** |
| | C | −0.149 | −0.354** | 0.088 | 0.520** | 0.444** |
| | GC | −0.301** | −0.283** | 0.397** | 0.359** | 0.535** |
| Zaire-96-I-16 | A | 0.536** | −0.167* | −0.397** | −0.146 | −0.381** |
| | T | −0.131 | 0.571** | −0.131 | −0.356** | −0.353** |
| | G | −0.293** | −0.075 | 0.473** | 0.070 | 0.362** |
| | C | −0.135 | −0.351** | 0.079 | 0.518** | 0.440** |
| | GC | −0.290** | −0.270** | 0.387** | 0.347** | 0.524** |

*$P<.05$.
**$P<.01$.

between $G_3$ and C, $T_3$ and A, and vice versa. These noncorrelations did not reveal any SCUB characteristics. The correlation analyses of nucleotide contents did not reveal the role of natural selection by the host. These results suggest that mutational pressure due to compositional constraints shapes the SCUB in MPXV genomes to a large extent.

## Quantification of SCUB

The ENC versus $GC_3$ plots were developed to quantify the SCUB (Figure 1). The ENC values were found to vary by 47.00 ± 0.078. The calculated ENC values of all genes were found to be greater than 35, suggesting a weak codon bias in all examined MPXV genomes. The ENC values were approximately normally distributed, and the Mann-Whitney 2-sample test revealed no significant intergenomic differences in the ENC values ($P>.05$). In the plots, most of genes were found to lie on or just below the expected $GC_3$ curve, suggesting that the SCUB was shaped mainly by GC compositional constraints. However, a considerable number of genes were grouped far below the expected $GC_3$ curve, suggesting that other factors also influenced the SCUB in the MPXV genomes.
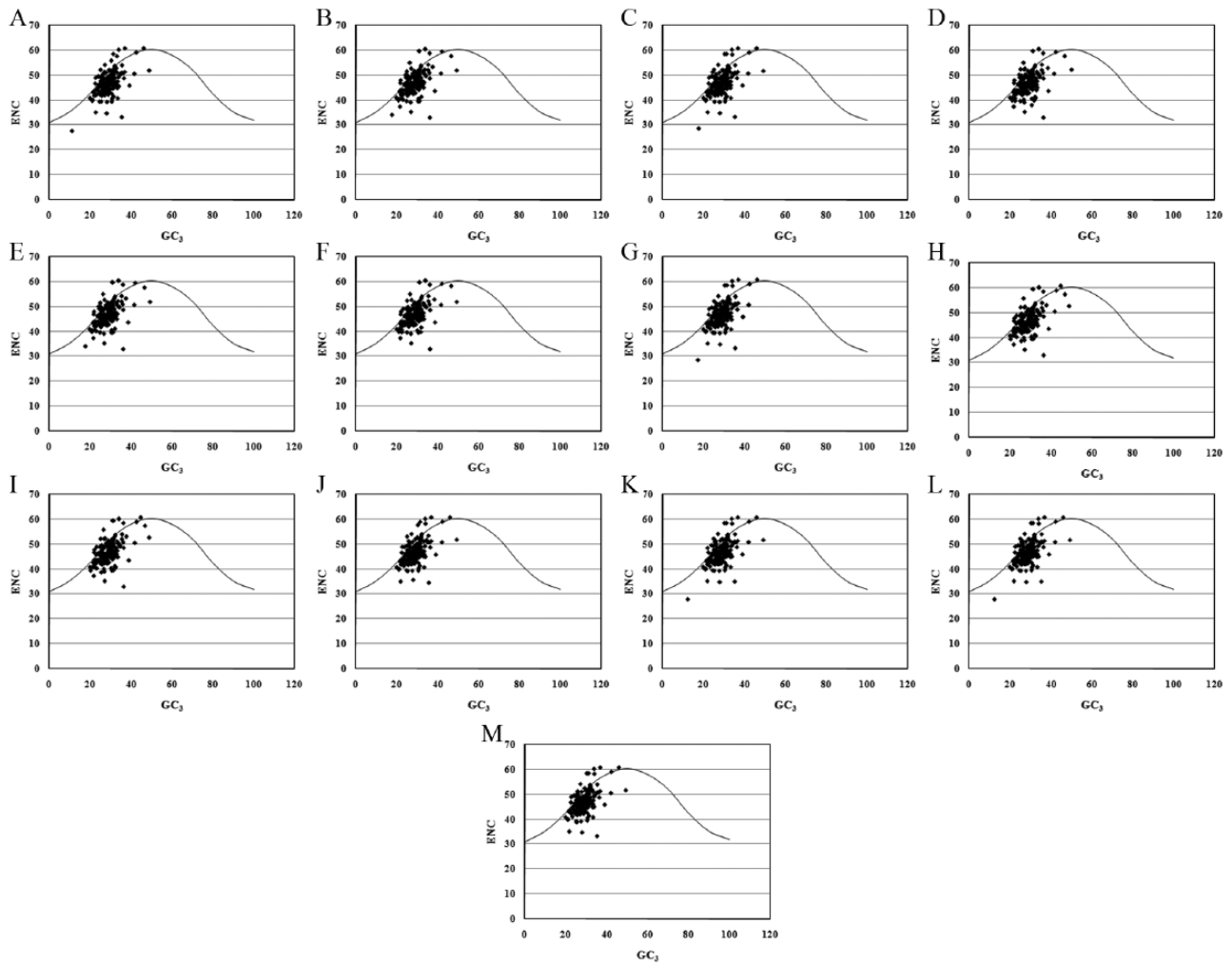
**Figure 1.** Mutational pressure versus selection pressure in MPXV genomes. ENC versus GC$_3$ plots for (A) Congo-2003-358, (B) COP-58, (C) DRC Yandongi-1985, (D) Liberia-1970-184, (E) MPXV-WRAIR7-61, (F) Sierra Leone, (G) Sudan-2005-01, (H) USA-2003-039, (I) USA-2003-044, (J)V79-I-005, (K) Zaire-1979-005 (cr), (L) Zaire-1979-005, and (M) Zaire-96-I-16. ENC indicates effective number of codons; MPXV, monkeypox viruses.

Neutrality plots[65] revealed no significant correlations between GC$_3$ and GC$_{12}$ (the G and the C contents at the first and the second codon positions) as the slope of the scatterplot approached 0, which is an indication that other major factors, such as selection, also have an influence on the SCUB in the MPXV genomes (Figure 2). The association between purines (A and G) and pyrimidines (C and T) was analyzed using a PR2 bias plot, and the A and the T contents were found to be used more than the C and the G contents (Figure 3). The PR2 bias plots clearly exhibited deviations from Chargaff's second parity rule[66] as most of the genes were localized far from the origin of the axis (Figure 3). The values of the PCG in all analyzed MPXV genomes (Table 1) had CAI values greater than 0.50; this indicated good host adaptation as the CAI values were calculated based on the *Homo sapiens* general codon usage. Significant positive correlations were found between the ENC and the CAI ($P < .05$), indicating that the level of gene expression had a large influence on the SCUB. The ENC was also positively correlated with the GC$_3$ values ($P < .01$) and with the hydrophobicity scores ($P < .05$), revealing their crucial roles in shaping the SCUB in MPXV genomes.

## Qualitative evaluation of SCUB

The codons with RSCU values greater than 1.0 are considered to be preferred as such codons are used more often than those with RSCU values less than 1.0.[3] In all synonymous amino acid families (6-fold, 4-fold, 3-fold, and 2-fold degenerate amino acids), A/T-ending codons were found to be used more frequently than G/C-ending codons (Table 3). In contrast, the human cells (host) use G/C-ending codons more frequently than A/T-ending codons.[67,68] The AGA that codes Arg is the only A-ending codon preferred in human cells.[67,68] In MPXV genomes, GAC (D), GGG (G), GGC (G), CAC (H), ATC (I), AAG (K), CTC (L), CTG (L), AAC (N), CAG (Q), AGG (R), CGC (R), AGC (S), TCC (S), ACC (T), ACG (T), GTG (V), GTC (V), and TAC (Y) were noted to be rare (RSCU < 0.66). In the host genome, the rare codons were reported to be GCG, CGA, AAT, GAT, TGT, CAA, GAA, GGT, CAT, ATA, TTA, AAA, TTT, CCG, TCG, ACG, TAT, and GTA.[67,68] Strand-specific codon biases were observed in all MPXV genomes for the amino acid Ile; ie, in positive strands, all strains preferred ATA, whereas in negative strands, all strains preferred ATT
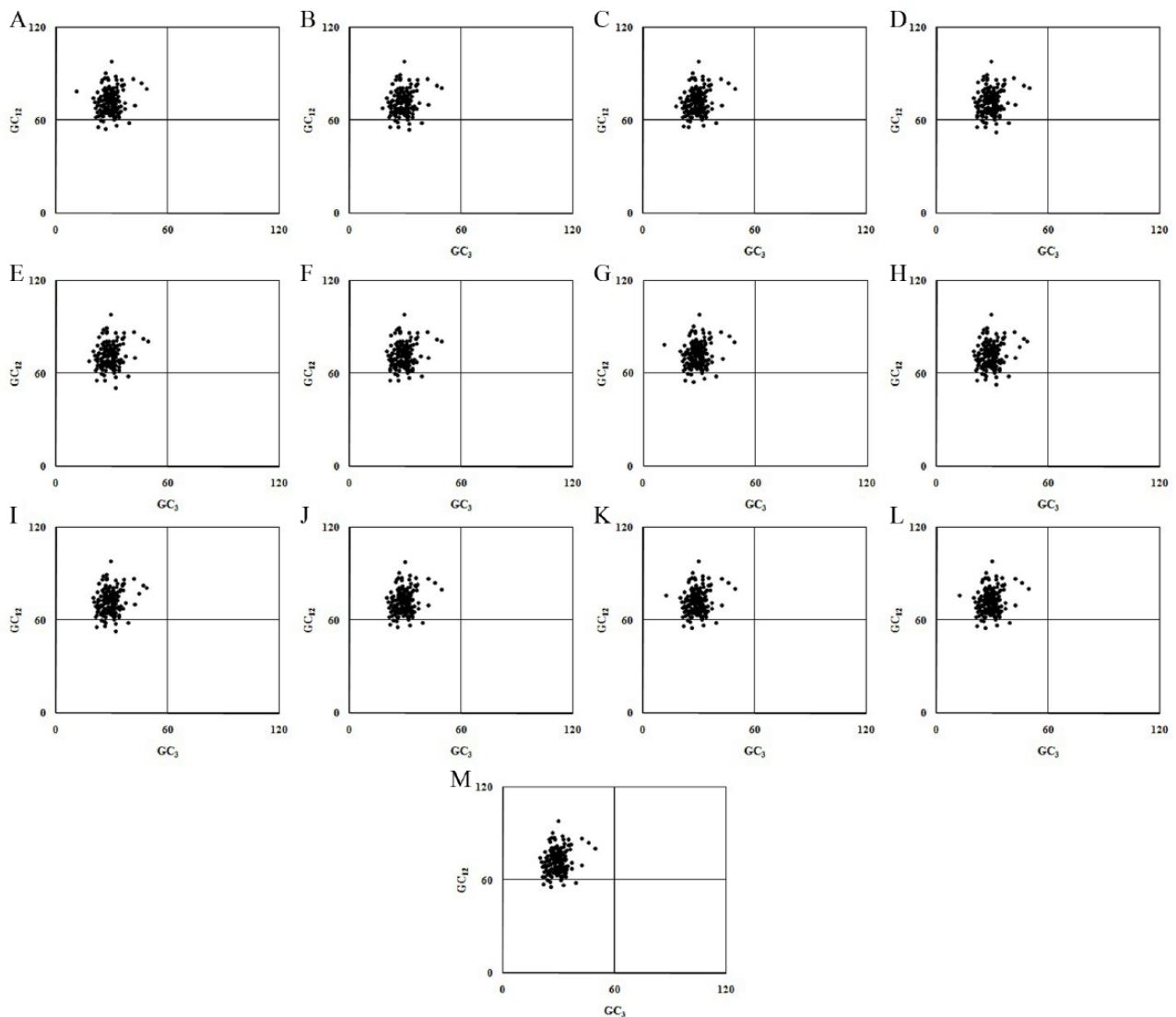
**Figure 2.** Influence of GC in shaping SCUB in MPXV genomes. Neutrality plots for (A) Congo-2003-358, (B) COP-58, (C) DRC Yandongi-1985, (D) Liberia-1970-184, (E) MPXV-WRAIR7-61, (F) Sierra Leone, (G) Sudan-2005-01, (H) USA-2003-039, (I) USA-2003-044, (J)V79-I-005, (K) Zaire-1979-005 (cr), (L) Zaire-1979-005, and (M) Zaire-96-I-16. MPXV indicates monkeypox viruses; SCUB, synonymous codon usage bias.

(Table 4). The amino acids Arg, Thr, and Val also exhibited strand-specific bias, but not in all strains (Table 4). Interestingly, positive strand–encoded genes preferentially used A-ending codons, whereas negative strand–encoded genes preferred T-ending codons. However, in the negative strand–encoded genes of the DRC Yandongi-1985 and the Sudan-2005-01 strains, the amino acid Val preferred both GTT and GTA.

Bias in the dinucleotide frequency analysis demonstrated that AT was overrepresented in all genomes, whereas GC was underrepresented. The ρ values of the dinucleotides were calculated by taking the ratio of the observed to the expected dinucleotide frequency and, in all genomes except GC, were found to be very close to 1. The most biased dinucleotides were ρAT, ρGA, and ρTC. The $\chi^2$ test revealed that the dinucleotide frequencies were not randomly distributed ($P < .05$).

Putative optimal codons were chosen based on the $\chi^2$ analysis of the 2 data sets formed by selecting 10% of the genes located at the 2 extremes of COA axis 1. All putative optimal codons were found to end in A/T (Table 5). The SCUBs of strains having threshold fitness or "good fitness"[24] were hypothesized to be shaped due to natural selection by the host.[24] However, the presence of A/T-ending putative optimal codons in the MPXV genomes, as found in this study, can be explained largely by the high AT content in the respective genomes. Natural selection by the host, if it existed, would have resulted in particular codon usage patterns in which amino acids would have preferentially used any nucleotide-ending codons.[24]

*Various factors influencing SCUB*

The COA partitioned the total number of SCU variations into 59 axes. Among the 59 axes, axes 1 to 5 accounted for approximately 10.42%, 8.43%, 7.13%, 5.66%, and 4.55% of the total SCU variations, respectively (Supplementary Figure 1). In all the strains isolated from various regions of Central Africa, $E_3$ and $GC_3$ had a high positive correlation with axis 1 ($P < .01$).
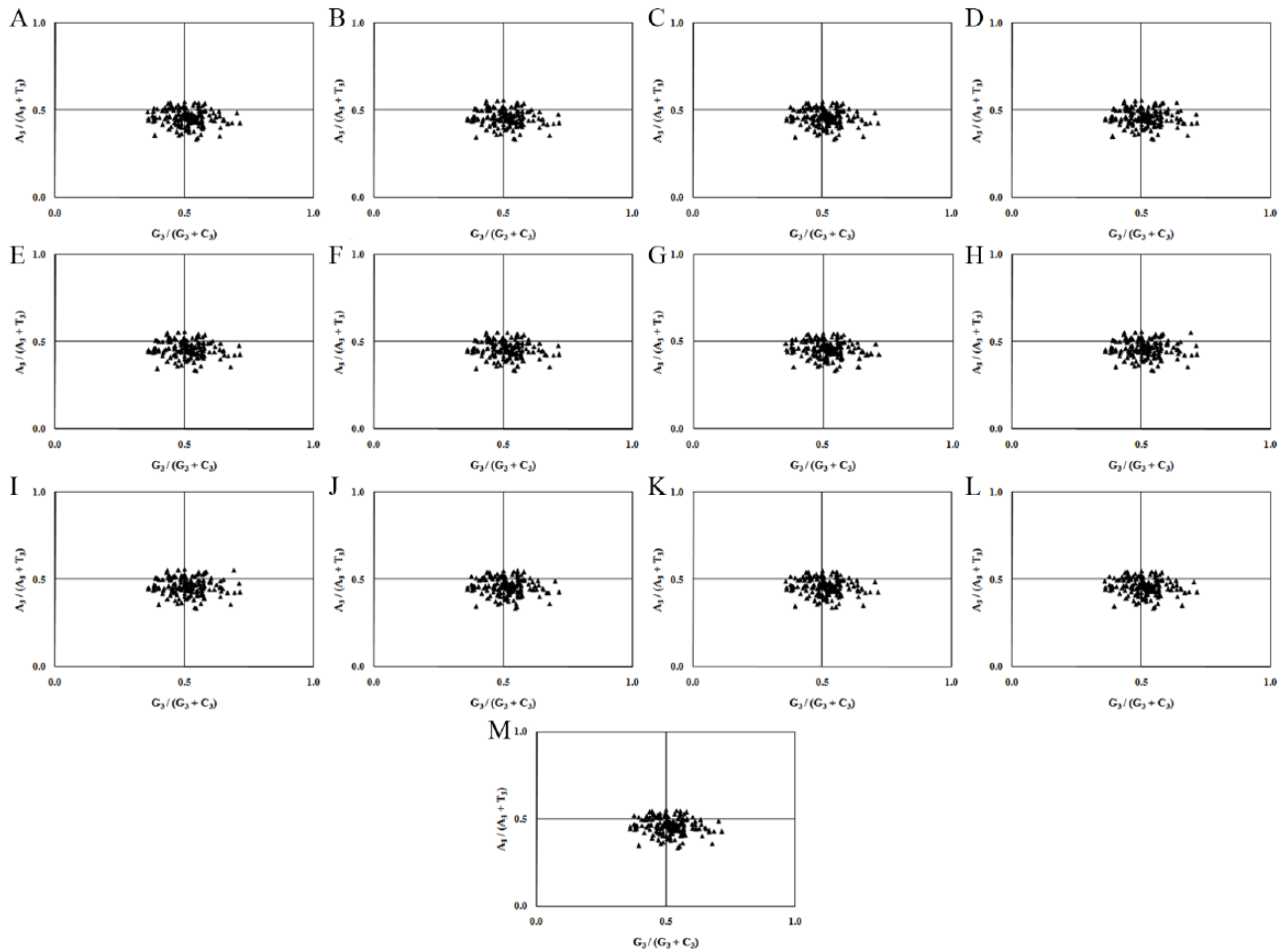
**Figure 3.** Deviation from parity rule 2 in MPXV genomes. PR2 bias plots for (A) Congo-2003-358, (B) COP-58, (C) DRC Yandongi-1985, (D) Liberia-1970-184, (E) MPXV-WRAIR7-61, (F) Sierra Leone, (G) Sudan-2005-01, (H) USA-2003-039, (I) USA-2003-044, (J)V79-I-005, (K) Zaire-1979-005 (cr), (L) Zaire-1979-005, and (M) Zaire-96-I-16. MPXV indicates monkeypox viruses.

The index indicating the level of gene expression (ie, CAI) had a higher positive correlation with axis 1 ($P < .01$) in all strains than the other proposed gene expression index, ENC, did (Table 6). The lengths of the coding sequences were weakly correlated with axis 1 for Central African strains such as V79-I-005 and Zaire-1979-005 (cr) ($P < .05$). The $T_3$ content exhibited a significant negative correlation with axis 1 ($P < .01$) in all Central African strains (Table 6).

In strains isolated from West Africa, the $A_3$ content was highly negatively correlated with axis 1, whereas it was not correlated with axis 1 in strains from Central Africa. High positive correlation was observed between axis 1 and the $G_3$ content ($P < .01$) for all West African strains. Similarly, $G_3$ positively correlated with axis 1 in strains isolated from the United States. However, no correlation between $G_3$ and axis 1 was observed in Central African and North African strains (Table 6). The CAI and the ENC also correlated highly with axis 1 in strains from West Africa and the United States (Table 6). That $GC_3$ was significantly correlated with the first principal axis (ie, axis 1 in all strains) strongly suggests that nucleotide compositional constraints play an important role in shaping the SCUB across all MPXV genomes. Furthermore, high positive correlation with CAI ($P < .01$) revealed that the level of gene expression

might also influence the SCUB across the examined MPXV genomes.

A correlation analysis between the dinucleotide content and the various COA axes did not reveal any true SCUB features, although some correlations did exist (Table 7). A cluster analysis of the pooled RSCU values of the PCG for each strain revealed 2 major clusters (Figure 4). More virulent Central African strains formed the upper cluster, and less virulent West African strains formed the lower cluster, indicating the presence of SCUB variations based on epidemic region and virulence.

## Discussion

In this study, trends associated with the SCUB and with various factors influencing its diversification in selected MPXV genomes were investigated in detail. Studies related to the evolution of MPXV genomes are highly important as MPXVs can be used as potential bioterrorism agents.[69] The mean ENC values of all examined MPXV genomes were greater than 40, indicating weak SCUB. The weak MPXV bias may be attributed to the ability of an MPXV to suppress antiviral CD4[+] and CD8[+] T-cell responses by inhibiting antiviral T-cell activation and inflammatory cytokine production without involving major histocompatibility complex molecules as this mechanism would reduce competition

**Table 3.** Overall relative synonymous codon usage values of protein-coding genes in examined monkeypox virus.

| AA | CODON | STRAINS | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | CONGO-2003-358 | COP-58 | DRC YANDONGI-1985 | LIBERIA-1970-184 | MPXV-WRAIR7-61 | SIERRA LEONE | SUDAN-2005-01 | USA-2003-039 | USA-2003-044 | V79-I-005 | ZAIRE-1979-005 | ZAIRE-1979-005 | ZAIRE-96-I-16 |
| A | GCT | 1.405 | 1.391 | 1.400 | 1.403 | 1.391 | 1.396 | 1.420 | 1.404 | 1.404 | 1.418 | 1.408 | 1.411 | 1.411 |
| A | GCG | 0.690 | 0.693 | 0.689 | 0.689 | 0.693 | 0.689 | 0.680 | 0.687 | 0.687 | 0.692 | 0.675 | 0.686 | 0.684 |
| A | GCC | 0.634 | 0.650 | 0.632 | 0.641 | 0.650 | 0.648 | 0.630 | 0.648 | 0.648 | 0.625 | 0.637 | 0.631 | 0.631 |
| A | GCA | 1.271 | 1.266 | 1.279 | 1.267 | 1.266 | 1.266 | 1.270 | 1.260 | 1.260 | 1.264 | 1.280 | 1.272 | 1.275 |
| C | TGT | 1.588 | 1.603 | 1.591 | 1.597 | 1.603 | 1.603 | 1.580 | 1.603 | 1.603 | 1.589 | 1.590 | 1.591 | 1.588 |
| C | TGC | 0.412 | 0.397 | 0.409 | 0.403 | 0.397 | 0.397 | 0.420 | 0.397 | 0.397 | 0.411 | 0.410 | 0.409 | 0.412 |
| D | GAT | 1.530 | 1.527 | 1.524 | 1.524 | 1.526 | 1.524 | 1.520 | 1.526 | 1.526 | 1.522 | 1.528 | 1.527 | 1.524 |
| D | GAC | 0.470 | 0.473 | 0.476 | 0.476 | 0.474 | 0.476 | 0.480 | 0.474 | 0.474 | 0.478 | 0.472 | 0.473 | 0.476 |
| E | GAG | 0.544 | 0.540 | 0.545 | 0.543 | 0.541 | 0.540 | 0.550 | 0.543 | 0.543 | 0.546 | 0.542 | 0.543 | 0.546 |
| E | GAA | 1.456 | 1.460 | 1.455 | 1.457 | 1.459 | 1.460 | 1.450 | 1.457 | 1.457 | 1.454 | 1.458 | 1.457 | 1.454 |
| F | TTT | 1.425 | 1.427 | 1.429 | 1.422 | 1.426 | 1.425 | 1.420 | 1.424 | 1.424 | 1.426 | 1.426 | 1.423 | 1.423 |
| F | TTC | 0.575 | 0.573 | 0.571 | 0.578 | 0.574 | 0.575 | 0.580 | 0.576 | 0.576 | 0.574 | 0.574 | 0.577 | 0.577 |
| G | GGT | 1.303 | 1.340 | 1.302 | 1.341 | 1.340 | 1.337 | 1.300 | 1.340 | 1.340 | 1.288 | 1.321 | 1.306 | 1.304 |
| G | GGG | 0.297 | 0.290 | 0.298 | 0.287 | 0.290 | 0.290 | 0.300 | 0.287 | 0.287 | 0.301 | 0.293 | 0.299 | 0.298 |
| G | GGC | 0.329 | 0.307 | 0.328 | 0.303 | 0.307 | 0.304 | 0.330 | 0.304 | 0.304 | 0.337 | 0.331 | 0.336 | 0.337 |
| G | GGA | 2.071 | 2.064 | 2.071 | 2.069 | 2.064 | 2.069 | 2.070 | 2.069 | 2.069 | 2.074 | 2.056 | 2.060 | 2.060 |
| H | CAC | 0.512 | 0.517 | 0.515 | 0.513 | 0.517 | 0.514 | 0.530 | 0.509 | 0.509 | 0.513 | 0.516 | 0.516 | 0.514 |
| H | CAT | 1.488 | 1.483 | 1.485 | 1.487 | 1.483 | 1.486 | 1.470 | 1.491 | 1.491 | 1.487 | 1.484 | 1.484 | 1.486 |
| I | ATT | 1.217 | 1.217 | 1.220 | 1.217 | 1.218 | 1.213 | 1.200 | 1.216 | 1.216 | 1.219 | 1.221 | 1.217 | 1.219 |
| I | ATA | 1.209 | 1.221 | 1.209 | 1.218 | 1.222 | 1.220 | 1.220 | 1.219 | 1.219 | 1.208 | 1.205 | 1.210 | 1.207 |
| I | ATC | 0.573 | 0.562 | 0.572 | 0.565 | 0.560 | 0.567 | 0.580 | 0.566 | 0.566 | 0.573 | 0.573 | 0.573 | 0.574 |
| K | AAA | 1.381 | 1.386 | 1.381 | 1.384 | 1.387 | 1.387 | 1.380 | 1.383 | 1.383 | 1.382 | 1.383 | 1.383 | 1.381 |
| K | AAG | 0.619 | 0.614 | 0.619 | 0.616 | 0.613 | 0.613 | 0.620 | 0.617 | 0.617 | 0.618 | 0.617 | 0.617 | 0.619 |
| L | CTA | 1.689 | 1.704 | 1.687 | 1.692 | 1.704 | 1.705 | 1.260 | 1.686 | 1.686 | 1.685 | 1.679 | 1.685 | 1.683 |
| L | CTC | 0.557 | 0.553 | 0.563 | 0.558 | 0.553 | 0.550 | 0.420 | 0.561 | 0.561 | 0.557 | 0.562 | 0.556 | 0.557 |
| L | CTG | 0.660 | 0.649 | 0.651 | 0.653 | 0.649 | 0.650 | 0.480 | 0.653 | 0.653 | 0.658 | 0.649 | 0.657 | 0.662 |
| L | CTT | 1.094 | 1.093 | 1.099 | 1.097 | 1.093 | 1.095 | 0.800 | 1.099 | 1.099 | 1.100 | 1.111 | 1.102 | 1.098 |
| L | TTA | 1.192 | 1.206 | 1.195 | 1.197 | 1.207 | 1.208 | 1.820 | 1.199 | 1.199 | 1.198 | 1.204 | 1.197 | 1.196 |
| L | TTG | 0.808 | 0.794 | 0.805 | 0.803 | 0.793 | 0.792 | 1.210 | 0.801 | 0.801 | 0.802 | 0.796 | 0.803 | 0.804 |
| N | AAC | 0.580 | 0.578 | 0.580 | 0.576 | 0.577 | 0.577 | 0.580 | 0.576 | 0.576 | 0.579 | 0.587 | 0.581 | 0.580 |
| N | AAT | 1.420 | 1.422 | 1.420 | 1.424 | 1.423 | 1.423 | 1.420 | 1.424 | 1.424 | 1.421 | 1.413 | 1.419 | 1.420 |
| P | CCA | 1.505 | 1.511 | 1.493 | 1.522 | 1.509 | 1.516 | 1.510 | 1.511 | 1.511 | 1.494 | 1.502 | 1.5 | 1.496 |

**Table 3.** (Continued)

| AA | CODON | CONGO-2003-358 | COP-58 | DRC YANDONGI-1985 | LIBERIA-1970-184 | MPXV-WRAIR7-61 | SIERRA LEONE | SUDAN-2005-01 | USA-2003-039 | USA-2003-044 | V79-I-005 | ZAIRE-1979-005 | ZAIRE-1979-005 | ZAIRE-96-I-16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | STRAINS | | | | | | | |
| P | CCC | 0.488 | 0.471 | 0.487 | 0.474 | 0.471 | 0.464 | 0.490 | 0.478 | 0.478 | 0.488 | 0.488 | 0.488 | 0.489 |
| P | CCT | 1.346 | 1.340 | 1.347 | 1.342 | 1.340 | 1.341 | 1.320 | 1.347 | 1.347 | 1.350 | 1.350 | 1.344 | 1.348 |
| P | CCG | 0.662 | 0.678 | 0.673 | 0.662 | 0.680 | 0.678 | 0.680 | 0.664 | 0.664 | 0.669 | 0.660 | 0.668 | 0.667 |
| Q | CAA | 1.458 | 1.450 | 1.452 | 1.45 | 1.450 | 1.451 | 1.450 | 1.457 | 1.457 | 1.456 | 1.454 | 1.458 | 1.460 |
| Q | CAG | 0.542 | 0.550 | 0.548 | 0.55 | 0.550 | 0.549 | 0.550 | 0.543 | 0.543 | 0.544 | 0.546 | 0.542 | 0.540 |
| R | AGA | 1.705 | 1.707 | 1.706 | 1.702 | 1.707 | 1.707 | 3.240 | 1.701 | 1.701 | 1.708 | 1.711 | 1.708 | 1.708 |
| R | AGG | 0.295 | 0.293 | 0.294 | 0.298 | 0.293 | 0.293 | 0.590 | 0.299 | 0.299 | 0.292 | 0.289 | 0.292 | 0.292 |
| R | CGA | 1.474 | 1.510 | 1.478 | 1.478 | 1.510 | 1.508 | 0.790 | 1.476 | 1.476 | 1.485 | 1.496 | 1.49 | 1.472 |
| R | CGC | 0.477 | 0.480 | 0.483 | 0.483 | 0.480 | 0.479 | 0.260 | 0.473 | 0.473 | 0.474 | 0.459 | 0.474 | 0.484 |
| R | CGG | 0.409 | 0.403 | 0.409 | 0.401 | 0.403 | 0.403 | 0.220 | 0.405 | 0.405 | 0.406 | 0.429 | 0.411 | 0.412 |
| R | CGT | 1.640 | 1.607 | 1.631 | 1.638 | 1.607 | 1.610 | 0.900 | 1.645 | 1.645 | 1.635 | 1.616 | 1.625 | 1.632 |
| S | AGC | 0.498 | 0.499 | 0.495 | 0.497 | 0.500 | 0.498 | 0.410 | 0.499 | 0.499 | 0.500 | 0.496 | 0.497 | 0.501 |
| S | AGT | 1.502 | 1.501 | 1.505 | 1.503 | 1.500 | 1.502 | 1.230 | 1.501 | 1.501 | 1.500 | 1.504 | 1.503 | 1.499 |
| S | TCA | 1.123 | 1.112 | 1.118 | 1.134 | 1.113 | 1.112 | 1.240 | 1.128 | 1.128 | 1.118 | 1.122 | 1.123 | 1.122 |
| S | TCC | 0.627 | 0.629 | 0.631 | 0.621 | 0.628 | 0.629 | 0.680 | 0.617 | 0.617 | 0.624 | 0.630 | 0.625 | 0.622 |
| S | TCG | 0.508 | 0.512 | 0.508 | 0.509 | 0.512 | 0.513 | 0.560 | 0.515 | 0.515 | 0.512 | 0.517 | 0.513 | 0.507 |
| S | TCT | 1.741 | 1.747 | 1.743 | 1.736 | 1.747 | 1.746 | 1.870 | 1.74 | 1.740 | 1.746 | 1.731 | 1.740 | 1.749 |
| T | ACC | 0.550 | 0.572 | 0.561 | 0.566 | 0.572 | 0.569 | 0.580 | 0.566 | 0.566 | 0.543 | 0.556 | 0.555 | 0.558 |
| T | ACA | 1.456 | 1.428 | 1.431 | 1.434 | 1.428 | 1.424 | 1.450 | 1.434 | 1.434 | 1.429 | 1.452 | 1.448 | 1.434 |
| T | ACG | 0.541 | 0.545 | 0.551 | 0.540 | 0.545 | 0.547 | 0.550 | 0.543 | 0.543 | 0.555 | 0.545 | 0.546 | 0.548 |
| T | ACT | 1.453 | 1.455 | 1.457 | 1.460 | 1.455 | 1.460 | 1.420 | 1.458 | 1.458 | 1.473 | 1.447 | 1.452 | 1.460 |
| V | GTT | 1.377 | 1.371 | 1.372 | 1.371 | 1.371 | 1.371 | 1.370 | 1.373 | 1.373 | 1.370 | 1.376 | 1.374 | 1.374 |
| V | GTG | 0.589 | 0.592 | 0.589 | 0.594 | 0.592 | 0.596 | 0.600 | 0.596 | 0.596 | 0.588 | 0.584 | 0.587 | 0.589 |
| V | GTC | 0.531 | 0.522 | 0.534 | 0.533 | 0.522 | 0.523 | 0.540 | 0.528 | 0.528 | 0.535 | 0.531 | 0.533 | 0.533 |
| V | GTA | 1.502 | 1.515 | 1.505 | 1.503 | 1.515 | 1.510 | 1.490 | 1.504 | 1.504 | 1.507 | 1.509 | 1.505 | 1.504 |
| Y | TAC | 0.555 | 0.571 | 0.554 | 0.573 | 0.576 | 0.561 | 0.570 | 0.571 | 0.571 | 0.556 | 0.555 | 0.553 | 0.556 |
| Y | TAT | 1.445 | 1.429 | 1.446 | 1.427 | 1.424 | 1.439 | 1.430 | 1.429 | 1.429 | 1.444 | 1.445 | 1.447 | 1.444 |

Abbreviations: AA, amino acid; MPXV, monkeypox viruses.

between the virus and the host, leading to efficient dissemination in the host.[70] Monkeypox virus infection effectively inhibits the genes involved in stimulating innate immunity, thereby suppressing the expressions of proteins such as TNF-α, IL-1α/β, CCL5, and IL-6.[71] Thus, these findings form the basis for the observed weak SCUB of the PCG across all examined MPXV genomes.

The SCUBs of all mammalian genomes are comparable, and all human viruses share this pattern of codon usage with the human host.[72] This sharing reveals the need for human viruses to adapt their codon usage to the host if the infection is to be successful, whereas in other mammalian viruses, adaptation is not a prerequisite for infecting the host.[72] Two possible

**Table 4.** Codons exhibiting strand-specific bias in examined monkeypox virus genomes.

| AA | STRANDS | CONGO-2003-358 | COP-58 | DRC YANDONGI-1985 | LIBERIA-1970-184 | MPXV-WRAIR7-61 | SIERRA LEONE | SUDAN-2005-01 | USA-2003-039 | USA-2003-044 | V79-I-005 | ZAIRE-1979-005 | ZAIRE-1979-005 | ZAIRE-96-I-16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | All | ATT | ATA | ATT | ATA | ATA | ATA | ATA | ATA | ATA | ATT | ATT | ATT | ATT |
|  | + | ATA | ATA | ATA | ATA | ATA | ATA | ATA | ATA | ATA | ATA | ATA | ATA | ATA |
|  | − | ATT | ATT | ATT | ATT | ATT | ATT | ATT | ATT | ATT | ATT | ATT | ATT | ATT |
| R | All | AGA | — | AGA | AGA | — | — | — | AGA | AGA | AGA | AGA | AGA | AGA |
|  | + | AGA | — | AGA | AGA | — | — | — | AGA | AGA | AGA | AGA | AGA | AGA |
|  | − | CGT | — | CGT | CGT | — | — | — | CGT | CGT | CGT | CGT | CGT | CGT |
| T | All | ACA | — | ACT | ACT | — | — | ACA | — | — | — | ACA | ACT | — |
|  | + | ACA | — | ACA | ACA | — | — | ACA | — | — | — | ACA | ACA | — |
|  | − | ACT | — | ACT | ACT | — | — | ACT | — | — | — | ACT | ACT | — |
| V | All | GTA | — | GTA | GTA | — | — | GTA | GTA | GTA | GTA | GTA | GTA | — |
|  | + | GTA | — | GTA | GTA | — | — | GTA | GTA | GTA | GTA | GTA | GTA | — |
|  | − | GTT | — | GTT and GTA | GTT | — | — | GTT and GTA | GTT | GTT | GTT | GTT | GTT | — |

Abbreviations: AA, amino acid; MPXV, monkeypox viruses.

**Table 5.** Identified putative optimal codons in examined monkeypox virus genomes.

| S. NO. | CONGO-2003-358 | COP-58 | DRC YANDONGI-1985 | LIBERIA-1970-184 | MPXV-WRAIR7-61 | SIERRA LEONE | SUDAN 2005-01 | USA-2003-039 | USA-2003-044 | V79-I-005 | ZAIRE-1979-005 | ZAIRE-1979-005 | ZAIRE-96-I-16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A (GCT) (GCA) | C (TGT) | A (GCA) | F (TTT) | C (TGT) | C (TGT) | A (GCT) (GCA) | D (GAT) | D (GAT) | A (GCA) | A (GCA) | A (GCA) | A (GCA) |
| 2 | C (TGT) | D (GAT) | C (TGT) | G (GGA) | D (GAT) | D (GAT) | F (TTT) | F (TTT) | F (TTT) | F (TTT) | G (GGA) | F (TTT) | C (TGT) |
| 3 | D (GAT) | E (GAA) | F (TTT) | I (ATA) | F (TTT) | F (TTT) | G (GGT) (GGA) | G (GGA) | G (GGA) | G (GGA) | I (ATA) | G (GGA) | F (TTT) |
| 4 | G (GGA) | F (TTT) | G (GGT) (GGA) | L (CTA) | G (GGT) (GGA) | G (GGT) (GGA) | I (ATA) | I (ATA) | I (ATA) | I (ATA) | L (CTA) | I (ATA) | G (GGT) (GGA) |
| 5 | Y (TAT) | G (GGT) GGA) | I (ATA) | P (CCT) | T (ACA) | P (CCT) | L (CTA) | L (CTA) | L (CTA) | L (CTA) | Y (TAT) | L (CTA) | V (GTT) |
| 6 | — | T (ACA) | L (CTA) | T (ACA) | V (GTT) | T (ACA) | — | T (ACA) | T (ACA) | Y (TAT) | — | Y (TAT) | Y (TAT) |
| 7 | — | V (GTT) | Y (TAT) | Y (TAT) | Y (TAT) | V (GTT) | — | — | Y (TAT) | — | — | — | — |
| 8 | — | Y (TAT) | — | — | — | Y (TAT) | — | — | — | — | — | — | — |

Abbreviation: MPXV, monkeypox viruses.

**Table 6.** Spearman rank correlation analysis between various correspondence analysis axes and important codon usage indices.

| STRAINS | AXES | A3 | T3 | G3 | C3 | GC | GC$_3$ | ENC | CAI | GRAVY | AROMATICITY | LENGTH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Congo-2003-358 | Axis 1 | -0.123 | -0.337** | 0.094 | 0.457** | 0.210** | 0.418** | 0.167* | 0.244** | -0.100 | -0.126 | 0.134 |
| | Axis 2 | -0.343** | 0.054 | 0.155 | 0.236** | 0.244** | 0.299** | 0.041 | 0.351** | -0.010 | -0.063 | 0.052 |
| | Axis 3 | -0.198** | -0.142 | 0.258** | 0.119 | 0.048 | 0.282** | 0.120 | 0.165* | -0.019 | 0.082 | 0.102 |
| | Axis 4 | 0.340** | -0.093 | -0.294** | -0.145 | -0.218** | -0.257** | -0.283** | -0.117 | 0.018 | -0.001 | -0.062 |
| | Axis 5 | -0.069 | 0.083 | -0.027 | 0.136 | 0.071 | 0.033 | -0.043 | 0.098 | 0.120 | 0.027 | -0.060 |
| COP-58 | Axis 1 | -0.257** | -0.341* | 0.214** | 0.514** | 0.291** | 0.532** | 0.237** | 0.317** | -0.087 | -0.103 | 0.135 |
| | Axis 2 | -0.326** | 0.060 | 0.173* | 0.201** | 0.194* | 0.269** | 0.105 | 0.310** | -0.001 | -0.055 | 0.137 |
| | Axis 3 | -0.119 | -0.110 | 0.166* | 0.053 | -0.040 | 0.159 | 0.091 | 0.161* | -0.072 | 0.091 | 0.170* |
| | Axis 4 | 0.180* | 0.093 | -0.247** | -0.040 | -0.063 | -0.200** | -0.160* | 0.026 | 0.077 | -0.016 | -0.002 |
| | Axis 5 | -0.345** | -0.118 | 0.244** | 0.399** | 0.356** | 0.438** | 0.318** | 0.286** | 0.175* | -0.056 | 0.003 |
| DRC Yandongi-1985 | Axis 1 | -0.120 | -0.362* | 0.109 | 0.447** | 0.218** | 0.431** | 0.157 | 0.258** | -0.105 | -0.123 | 0.116 |
| | Axis 2 | -0.334** | 0.078 | 0.142 | 0.229** | 0.238** | 0.275** | 0.035 | 0.365** | -0.041 | -0.072 | 0.101 |
| | Axis 3 | -0.153 | -0.087 | 0.165* | 0.053 | -0.001 | 0.183* | 0.074 | 0.169* | -0.048 | 0.124 | 0.150 |
| | Axis 4 | 0.398** | -0.026 | -0.364** | -0.200** | -0.282** | -0.354** | -0.293** | -0.173* | 0.021 | -0.014 | -0.033 |
| | Axis 5 | 0.094 | -0.115 | -0.040 | -0.054 | -0.070 | -0.033 | -0.006 | -0.062 | -0.126 | -0.065 | 0.014 |
| Liberia-1970-184 | Axis 1 | -0.213** | -0.329* | 0.214** | 0.433** | 0.248** | 0.486** | 0.179* | 0.348** | -0.165* | -0.104 | 0.110 |
| | Axis 2 | -0.322** | 0.051 | 0.183* | 0.191* | 0.213** | 0.285** | 0.024 | 0.321** | -0.002 | -0.066 | 0.043 |
| | Axis 3 | -0.025 | -0.171* | 0.127 | 0.025 | -0.030 | 0.128 | 0.051 | 0.084 | -0.097 | 0.054 | 0.107 |
| | Axis 4 | 0.199* | -0.047 | -0.022 | -0.229** | -0.204** | -0.164* | -0.090 | -0.342** | -0.092 | 0.011 | 0.103 |
| | Axis 5 | 0.433** | 0.105 | -0.283** | -0.366** | -0.382** | -0.465** | -0.298** | -0.339** | 0.038 | 0.056 | 0.036 |
| MPXV-WRAIR7-61 | Axis 1 | -0.256** | -0.341** | 0.213** | 0.515* | 0.291* | 0.532* | 0.237* | 0.318* | -0.091 | -0.103 | 0.135 |
| | Axis 2 | -0.327** | 0.058 | 0.168* | 0.199* | 0.195* | 0.268* | 0.106 | 0.311* | -0.002 | -0.054 | 0.137 |
| | Axis 3 | -0.118 | -0.108 | 0.161* | 0.052 | -0.040 | 0.156 | 0.090 | 0.163* | -0.071 | 0.091 | 0.168* |
| | Axis 4 | 0.173* | 0.091 | -0.247** | -0.031 | -0.052 | -0.193* | -0.154 | 0.034 | 0.073 | -0.015 | -0.003 |

*(Continued)*

**Table 6.** (Continued)

| STRAINS | AXES | A3 | T3 | G3 | C3 | GC | GC$_3$ | ENC | CAI | GRAVY | AROMATICITY | LENGTH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | Axis 5 | -0.347** | -0.122 | 0.244** | 0.402* | 0.355* | 0.440* | 0.321* | 0.286* | 0.170* | -0.057 | 0.010 |
| Sierra Leone | Axis 1 | -0.242* | -0.334* | 0.205** | 0.509** | 0.271** | 0.528** | 0.217** | 0.319** | -0.083 | -0.093 | 0.128 |
|  | Axis 2 | -0.312* | 0.067 | 0.176* | 0.179* | 0.198* | 0.256** | 0.083 | 0.305** | -0.007 | -0.056 | 0.117 |
|  | Axis 3 | -0.131* | -0.096 | 0.176* | 0.050 | -0.049 | 0.161* | 0.082 | 0.153 | -0.044 | 0.107 | 0.161* |
|  | Axis 4 | 0.061 | 0.077 | -0.158 | 0.058 | 0.026 | -0.084 | -0.072 | 0.092 | 0.098 | 0.007 | -0.006 |
|  | Axis 5 | -0.403* | -0.128 | 0.244** | 0.463** | 0.393** | 0.490** | 0.333** | 0.363** | 0.143 | -0.099 | -0.030 |
| Sudan-2005-01 | Axis 1 | -0.132* | -0.326** | 0.115 | 0.426** | 0.202** | 0.414** | 0.187* | 0.281** | -0.126 | -0.076 | 0.125 |
|  | Axis 2 | -0.313** | 0.073 | 0.116 | 0.226** | 0.196* | 0.266** | 0.036 | 0.290** | -0.042 | -0.026 | 0.081 |
|  | Axis 3 | -0.178* | -0.043 | 0.165* | 0.065 | 0.014 | 0.169* | 0.072 | 0.287** | -0.090 | 0.077 | 0.084 |
|  | Axis 4 | -0.071 | 0.003 | 0.216** | -0.063 | -0.056 | 0.083 | 0.173* | -0.250** | 0.018 | 0.078 | 0.043 |
|  | Axis 5 | 0.295** | -0.022 | -0.179* | -0.250** | -0.351** | -0.272** | -0.159 | -0.334** | -0.060 | 0.073 | 0.018 |
| USA-2003-039 | Axis 1 | -0.213** | -0.325** | 0.203** | 0.466** | 0.272** | 0.494** | 0.199* | 0.326** | -0.142 | -0.101 | 0.131 |
|  | Axis 2 | -0.341** | 0.018 | 0.224** | 0.213** | 0.231** | 0.333** | 0.075 | 0.337** | 0.022 | -0.050 | 0.018 |
|  | Axis 3 | -0.039 | -0.223** | 0.159 | 0.055 | 0.028 | 0.168* | 0.110 | 0.111 | -0.105 | 0.047 | 0.114 |
|  | Axis 4 | 0.333** | 0.157 | -0.252** | -0.308** | -0.345** | -0.419** | -0.309** | -0.226** | 0.050 | 0.064 | 0.024 |
|  | Axis 5 | -0.183* | 0.026 | 0.015 | 0.218** | 0.177* | 0.153 | 0.112 | 0.305** | 0.092 | 0.019 | -0.093 |
| USA-2003-044 | Axis 1 | -0.213** | -0.325** | 0.203** | 0.466** | 0.272** | 0.494** | 0.199* | 0.326** | -0.142 | -0.101 | 0.131 |
|  | Axis 2 | -0.341** | 0.018 | 0.224** | 0.213** | 0.231** | 0.333** | 0.075 | 0.337** | 0.022 | -0.050 | 0.018 |
|  | Axis 3 | -0.039 | -0.223** | 0.159 | 0.055 | 0.028 | 0.168* | 0.110 | 0.111 | -0.105 | 0.047 | 0.114 |
|  | Axis 4 | 0.333** | 0.157 | -0.252** | -0.308** | -0.345** | -0.419** | -0.309** | -0.226** | 0.050 | 0.064 | 0.024 |
|  | Axis 5 | -0.183* | 0.026 | 0.015 | 0.218** | 0.177* | 0.153 | 0.112 | 0.305** | 0.092 | 0.019 | -0.093 |

**Table 6.** (Continued)

| STRAINS | AXES | A3 | T3 | G3 | C3 | GC | GC$_3$ | ENC | CAI | GRAVY | AROMATICITY | LENGTH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V79-I-005 | Axis 1 | -0.113 | -0.344** | 0.109 | 0.419** | 0.193* | 0.395** | 0.190* | 0.229** | -0.110 | -0.105 | 0.165* |
| | Axis 2 | -0.345** | 0.022 | 0.168* | 0.249** | 0.256** | 0.320** | 0.019 | 0.402** | -0.034 | -0.092 | 0.032 |
| | Axis 3 | -0.123 | -0.108 | 0.158 | 0.038 | -0.022 | 0.168* | 0.054 | 0.159* | -0.055 | 0.125 | 0.152 |
| | Axis 4 | 0.303** | 0.134 | -0.271** | -0.242** | -0.245** | -0.370** | -0.266** | -0.201** | 0.089 | 0.060 | -0.005 |
| | Axis 5 | -0.272** | -0.106 | 0.185* | 0.333** | 0.331** | 0.350** | 0.194* | 0.319** | 0.146 | -0.082 | -0.099 |
| Zaire-1979-005 | Axis 1 | -0.106 | -0.365** | 0.123 | 0.432** | 0.223** | 0.414** | 0.224** | 0.243** | -0.117 | -0.126 | 0.151 |
| | Axis 2 | -0.371** | 0.032 | 0.184* | 0.268** | 0.266** | 0.336** | 0.048 | 0.398** | -0.033 | -0.079 | 0.060 |
| | Axis 3 | -0.153 | -0.146 | 0.212** | 0.082 | 0.010 | 0.220** | 0.112 | 0.173* | -0.054 | 0.125 | 0.144 |
| | Axis 4 | 0.351** | -0.076 | -0.296** | -0.169* | -0.220** | -0.277** | -0.287** | -0.132 | 0.034 | -0.003 | -0.043 |
| | Axis 5 | 0.017 | 0.173* | -0.055 | -0.054 | -0.022 | -0.120 | -0.090 | -0.027 | 0.094 | 0.020 | 0.012 |
| Zaire-1979-005 | Axis 1 | -0.130 | -0.327** | 0.116 | 0.416** | 0.207** | 0.405** | 0.210* | 0.245** | -0.113 | -0.114 | 0.166* |
| | Axis 2 | -0.368** | 0.028 | 0.198* | 0.258** | 0.273** | 0.338** | 0.042 | 0.389** | -0.023 | -0.079 | 0.065 |
| | Axis 3 | -0.134 | -0.105 | 0.162* | 0.053 | -0.028 | 0.176* | 0.064 | 0.162* | -0.062 | 0.126 | 0.112 |
| | Axis 4 | 0.372** | -0.083 | -0.316** | -0.170* | -0.220** | -0.290** | -0.286** | -0.153 | 0.026 | -0.010 | -0.038 |
| | Axis 5 | -0.039 | -0.184* | 0.076 | 0.086 | 0.046 | 0.155 | 0.100 | 0.051 | -0.105 | -0.034 | -0.018 |
| Zaire-96-I-16 | Axis 1 | -0.133 | -0.357** | 0.123 | 0.462** | 0.220** | 0.437** | 0.188* | 0.248** | -0.113 | -0.106 | 0.152 |
| | Axis 2 | -0.342** | 0.057 | 0.157* | 0.211** | 0.214** | 0.292** | 0.033 | 0.400** | -0.041 | -0.049 | 0.070 |
| | Axis 3 | -0.054 | -0.157 | 0.135 | 0.018 | -0.042 | 0.141 | 0.056 | 0.114 | -0.093 | 0.123 | 0.154 |
| | Axis 4 | -0.272** | -0.107 | 0.262** | 0.181* | 0.222** | 0.321** | 0.241** | 0.161* | -0.105 | -0.063 | 0.014 |
| | Axis 5 | 0.286** | 0.149 | -0.171* | -0.402** | -0.368** | -0.394* | -0.221** | -0.342** | -0.147 | 0.088 | 0.099 |

*$P < .05$.
**$P < .01$.

**Table 7.** Spearman correlation analysis between various correspondence analysis axes and dinucleotide contents.

| STRAINS | AXES | AA | AC | AG | AT | CA | CC | CG | CT | GA | GC | GG | GT | TA | TC | TG | TT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Congo-2003-358 | Axis 1 | 0.117 | 0.164* | 0.137 | 0.112 | 0.154 | 0.175* | 0.183* | 0.144 | 0.183* | 0.176* | 0.076 | 0.117 | 0.076 | 0.155 | 0.156 | 0.026 |
| | Axis 2 | 0.032 | 0.045 | 0.017 | −0.013 | 0.074 | 0.079 | 0.007 | 0.148 | 0.050 | 0.141 | 0.081 | 0.069 | −0.057 | 0.103 | 0.162* | −0.028 |
| | Axis 3 | 0.088 | 0.086 | 0.041 | 0.094 | 0.057 | 0.102 | 0.053 | 0.132 | 0.088 | 0.033 | 0.142 | 0.107 | 0.078 | 0.113 | 0.139 | 0.088 |
| | Axis 4 | −0.016 | −0.010 | −0.046 | −0.027 | −0.016 | −0.077 | −0.107 | −0.099 | −0.054 | −0.168* | −0.115 | −0.069 | −0.035 | −0.083 | −0.096 | −0.051 |
| | Axis 5 | −0.075 | −0.038 | −0.089 | −0.053 | −0.003 | −0.104 | −0.065 | −0.090 | −0.083 | −0.018 | −0.043 | −0.025 | −0.078 | −0.069 | −0.014 | −0.092 |
| COP-58 | Axis 1 | 0.109 | 0.156 | 0.104 | 0.098 | 0.152 | 0.175* | 0.186* | 0.169* | 0.177* | 0.210** | 0.082 | 0.128 | 0.058 | 0.164* | 0.197* | 0.013 |
| | Axis 2 | 0.105 | 0.136 | 0.076 | 0.076 | 0.161* | 0.150 | 0.074 | 0.211** | 0.126 | 0.176* | 0.172* | 0.149 | 0.034 | 0.175* | 0.229** | 0.042 |
| | Axis 3 | 0.172* | 0.176* | 0.106 | 0.175* | 0.157 | 0.128 | 0.104 | 0.156 | 0.148 | 0.060 | 0.210** | 0.181* | 0.160* | 0.161* | 0.194* | 0.144 |
| | Axis 4 | 0.009 | 0.043 | 0.015 | 0.010 | 0.068 | −0.034 | −0.036 | −0.035 | −0.007 | −0.014 | −0.040 | 0.004 | 0.004 | −0.033 | −0.002 | −0.026 |
| | Axis 5 | −0.081 | 0.038 | −0.023 | −0.028 | 0.088 | 0.088 | 0.048 | 0.008 | 0.007 | 0.136 | 0.086 | 0.048 | −0.063 | 0.021 | 0.089 | −0.088 |
| DRC Yandongi-1985 | Axis 1 | −0.109 | −0.118 | −0.080 | −0.167* | −0.118 | −0.123 | −0.144 | −0.093 | −0.071 | −0.055 | −0.066 | −0.116 | −0.168* | −0.128 | −0.067 | −0.173* |
| | Axis 2 | −0.017 | −0.002 | −0.018 | −0.013 | 0.031 | 0.035 | −0.024 | −0.005 | −0.005 | 0.039 | 0.035 | 0.041 | −0.040 | −0.005 | 0.068 | −0.030 |
| | Axis 3 | 0.001 | −0.020 | 0.008 | −0.008 | 0.014 | 0.022 | 0.015 | 0.030 | −0.011 | 0.038 | 0.046 | 0.027 | −0.026 | 0.045 | 0.018 | 0.005 |
| | Axis 4 | −0.117 | −0.124 | −0.117 | −0.135 | −0.132 | −0.125 | −0.138 | −0.110 | −0.107 | −0.155 | −0.128 | −0.165* | −0.142 | −0.123 | −0.156 | −0.143 |
| | Axis 5 | 0.101 | 0.046 | 0.042 | 0.043 | 0.029 | 0.071 | 0.051 | 0.033 | 0.033 | −0.003 | 0.031 | 0.070 | 0.064 | 0.050 | 0.013 | 0.008 |
| Liberia-1970-184 | Axis 1 | 0.100 | 0.129 | 0.102 | 0.088 | 0.135 | 0.144 | 0.166* | 0.132 | 0.174* | 0.188* | 0.064 | 0.086 | 0.037 | 0.134 | 0.156 | −0.011 |
| | Axis 2 | 0.022 | 0.053 | −0.007 | −0.014 | 0.077 | 0.071 | −0.009 | 0.135 | 0.046 | 0.090 | 0.082 | 0.068 | −0.057 | 0.101 | 0.155 | −0.039 |
| | Axis 3 | 0.128 | 0.114 | 0.071 | 0.103 | 0.076 | 0.074 | 0.063 | 0.108 | 0.103 | 0.008 | 0.130 | 0.107 | 0.103 | 0.104 | 0.115 | 0.095 |
| | Axis 4 | 0.144 | 0.071 | 0.103 | 0.105 | −0.012 | 0.065 | 0.114 | 0.124 | 0.104 | 0.012 | 0.042 | 0.050 | 0.145 | 0.092 | 0.021 | 0.150 |
| | Axis 5 | 0.094 | 0.044 | 0.032 | 0.088 | 0.015 | −0.103 | −0.045 | −0.016 | 0.030 | −0.113 | −0.081 | −0.006 | 0.110 | −0.021 | −0.038 | 0.070 |
| MPXV-WRAIR7-61 | Axis 1 | 0.109 | 0.157 | 0.105 | 0.099 | 0.153 | 0.176* | 0.186* | 0.169* | 0.178* | 0.211** | 0.083 | 0.128 | 0.058 | 0.165* | 0.197* | 0.012 |
| | Axis 2 | 0.105 | 0.137 | 0.076 | 0.078 | 0.162* | 0.150 | 0.074 | 0.211** | 0.126 | 0.176* | 0.172* | 0.149 | 0.034 | 0.175* | 0.230* | 0.042 |
| | Axis 3 | 0.171* | 0.175* | 0.104 | 0.174 | 0.156 | 0.127 | 0.103 | 0.154 | 0.147 | 0.059 | 0.209 | 0.179* | 0.160* | 0.160* | 0.193* | 0.143 |
| | Axis 4 | 0.005 | 0.044 | 0.013 | 0.009 | 0.068 | −0.033 | −0.035 | −0.037 | −0.008 | −0.011 | −0.038 | 0.004 | 0.001 | −0.036 | −0.001 | −0.031 |

**Table 7.** (Continued)

| STRAINS | AXES | AA | AC | AG | AT | CA | CC | CG | CT | GA | GC | GG | GT | TA | TC | TG | TT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Axis 5 | −0.076 | 0.044 | −0.016 | −0.022 | 0.092 | 0.095 | 0.054 | 0.014 | 0.014 | 0.142 | 0.090 | 0.052 | −0.056 | 0.027 | 0.094 | −0.082 |
| Sierra Leone | Axis 1 | 0.102 | 0.147 | 0.098 | 0.095 | 0.145 | 0.158 | 0.179* | 0.158 | 0.174* | 0.198* | 0.066 | 0.117 | 0.053 | 0.157 | 0.189* | 0.011 |
| | Axis 2 | 0.084 | 0.117 | 0.066 | 0.055 | 0.135 | 0.139 | 0.057 | 0.196* | 0.115 | 0.158 | 0.144 | 0.122 | 0.014 | 0.154 | 0.209** | 0.028 |
| | Axis 3 | 0.162* | 0.166* | 0.091 | 0.168* | 0.146 | 0.117 | 0.089 | 0.153 | 0.142 | 0.042 | 0.201** | 0.171* | 0.150 | 0.157 | 0.188* | 0.142 |
| | Axis 4 | −0.015 | 0.039 | −0.003 | 0.001 | 0.069 | −0.021 | −0.022 | −0.043 | −0.011 | 0.013 | −0.028 | 0.011 | −0.015 | −0.041 | 0.015 | −0.046 |
| | Axis 5 | −0.115 | 0.003 | −0.040 | −0.066 | 0.057 | 0.103 | 0.040 | −0.014 | −0.021 | 0.124 | 0.074 | 0.027 | −0.104 | 0.007 | 0.066 | −0.122 |
| Sudan-2005-01 | Axis 1 | 0.130 | 0.148 | 0.129 | 0.099 | 0.150 | 0.157 | 0.169* | 0.149 | 0.180* | 0.166* | 0.077 | 0.108 | 0.049 | 0.167* | 0.160* | 0.026 |
| | Axis 2 | 0.048 | 0.088 | 0.048 | 0.048 | 0.107 | 0.077 | 0.028 | 0.140 | 0.081 | 0.131 | 0.097 | 0.105 | 0.013 | 0.101 | 0.180* | −0.011 |
| | Axis 3 | 0.084 | 0.063 | 0.040 | 0.070 | 0.065 | 0.061 | 0.001 | 0.102 | 0.090 | 0.010 | 0.129 | 0.081 | 0.031 | 0.108 | 0.130 | 0.053 |
| | Axis 4 | 0.014 | 0.009 | 0.000 | 0.079 | −0.035 | −0.020 | 0.073 | 0.010 | 0.006 | 0.012 | 0.033 | 0.039 | 0.126 | 0.004 | 0.002 | 0.080 |
| | Axis 5 | 0.066 | 0.025 | 0.016 | 0.075 | −0.048 | 0.002 | 0.008 | −0.008 | 0.009 | −0.128 | −0.075 | −0.021 | 0.129 | −0.025 | −0.076 | 0.086 |
| USA-2003-039 | Axis 1 | 0.115 | 0.156 | 0.116 | 0.103 | 0.143 | 0.176* | 0.203** | 0.156 | 0.193* | 0.201** | 0.088 | 0.113 | 0.062 | 0.157 | 0.175* | 0.013 |
| | Axis 2 | −0.008 | 0.033 | −0.030 | −0.031 | 0.055 | 0.050 | −0.026 | 0.113 | 0.029 | 0.066 | 0.067 | 0.037 | −0.077 | 0.081 | 0.132 | −0.062 |
| | Axis 3 | 0.125 | 0.127 | 0.091 | 0.112 | 0.104 | 0.095 | 0.086 | 0.112 | 0.122 | 0.042 | 0.148 | 0.118 | 0.107 | 0.119 | 0.130 | 0.096 |
| | Axis 4 | 0.086 | 0.014 | 0.003 | 0.070 | −0.008 | −0.108 | −0.081 | −0.008 | 0.006 | −0.122 | −0.073 | −0.014 | 0.085 | −0.021 | −0.018 | 0.063 |
| | Axis 5 | −0.132 | −0.061 | −0.100 | −0.093 | 0.001 | −0.056 | −0.095 | −0.128 | −0.094 | −0.009 | −0.037 | −0.045 | −0.129 | −0.104 | −0.019 | −0.128 |
| USA-2003-044 | Axis 1 | 0.115 | 0.156 | 0.116 | 0.103 | 0.143 | 0.176* | 0.203** | 0.156 | 0.193* | 0.201** | 0.088 | 0.113 | 0.062 | 0.157 | 0.175* | 0.013 |
| | Axis 2 | −0.008 | 0.033 | −0.030 | −0.031 | 0.055 | 0.050 | −0.026 | 0.113 | 0.029 | 0.066 | 0.067 | 0.037 | −0.077 | 0.081 | 0.132 | −0.062 |
| | Axis 3 | 0.125 | 0.127 | 0.091 | 0.112 | 0.104 | 0.095 | 0.086 | 0.112 | 0.122 | 0.042 | 0.148 | 0.118 | 0.107 | 0.119 | 0.130 | 0.096 |
| | Axis 4 | 0.086 | 0.014 | 0.003 | 0.070 | −0.008 | −0.108 | −0.081 | −0.008 | 0.006 | −0.122 | −0.073 | −0.014 | 0.085 | −0.021 | −0.018 | 0.063 |
| | Axis 5 | −0.132 | −0.061 | −0.100 | −0.093 | 0.001 | −0.056 | −0.095 | −0.128 | −0.094 | −0.009 | −0.037 | −0.045 | −0.129 | −0.104 | −0.019 | −0.128 |
| V79-I-005 | Axis 1 | 0.149 | 0.181* | 0.152 | 0.145 | 0.187* | 0.196* | 0.214** | 0.162* | 0.205** | 0.197* | 0.108 | 0.142 | 0.106 | 0.192* | 0.184* | 0.064 |
| | Axis 2 | 0.018 | 0.043 | 0.010 | −0.030 | 0.068 | 0.053 | −0.017 | 0.127 | 0.046 | 0.111 | 0.066 | 0.050 | −0.078 | 0.078 | 0.147 | −0.063 |
| | Axis 3 | 0.143 | 0.141 | 0.092 | 0.160* | 0.122 | 0.112 | 0.080 | 0.149 | 0.145 | 0.047 | 0.186* | 0.149 | 0.134 | 0.146 | 0.181* | 0.132 |
| | Axis 4 | 0.034 | 0.013 | 0.000 | 0.039 | 0.020 | −0.112 | −0.050 | −0.057 | −0.016 | −0.077 | −0.074 | −0.010 | 0.045 | −0.042 | −0.036 | 0.013 |

*(Continued)*

**Table 7.** (Continued)

| STRAINS | AXES | AA | AC | AG | AT | CA | CC | CG | CT | GA | GC | GG | GT | TA | TC | TG | TT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Axis 5 | −0.153 | −0.069 | −0.098 | −0.125 | 0.012 | −0.012 | −0.074 | −0.074 | −0.084 | 0.048 | −0.010 | −0.065 | −0.180* | −0.046 | −0.005 | −0.157 |
| Zaire-1979-005 | Axis 1 | 0.136 | 0.175* | 0.142 | 0.127 | 0.177* | 0.197* | 0.212** | 0.151 | 0.208** | 0.203** | 0.099 | 0.125 | 0.086 | 0.176* | 0.178* | 0.048 |
| | Axis 2 | 0.039 | 0.066 | 0.033 | −0.009 | 0.086 | 0.086 | 0.013 | 0.158 | 0.068 | 0.138 | 0.085 | 0.088 | −0.056 | 0.107 | 0.178* | −0.032 |
| | Axis 3 | 0.137 | 0.122 | 0.074 | 0.148 | 0.109 | 0.106 | 0.072 | 0.153 | 0.133 | 0.064 | 0.184* | 0.134 | 0.120 | 0.151 | 0.182* | 0.118 |
| | Axis 4 | 0.009 | 0.003 | −0.019 | −0.005 | 0.001 | −0.074 | −0.086 | −0.086 | −0.035 | −0.140 | −0.099 | −0.045 | −0.011 | −0.069 | −0.078 | −0.038 |
| | Axis 5 | 0.010 | 0.010 | −0.013 | 0.024 | 0.042 | −0.073 | −0.002 | −0.033 | −0.014 | 0.023 | 0.018 | 0.035 | 0.013 | −0.011 | 0.031 | 0.002 |
| Zaire-1979-005 | Axis 1 | 0.146 | 0.185* | 0.154 | 0.146 | 0.190* | 0.208** | 0.218** | 0.168* | 0.215** | 0.205** | 0.109 | 0.143 | 0.106 | 0.194* | 0.192* | 0.061 |
| | Axis 2 | 0.046 | 0.069 | 0.039 | −0.005 | 0.090 | 0.087 | 0.017 | 0.161* | 0.075 | 0.147 | 0.098 | 0.090 | −0.052 | 0.110 | 0.183* | −0.025 |
| | Axis 3 | 0.109 | 0.098 | 0.049 | 0.122 | 0.081 | 0.070 | 0.038 | 0.116 | 0.102 | 0.019 | 0.143 | 0.105 | 0.096 | 0.115 | 0.142 | 0.089 |
| | Axis 4 | 0.018 | 0.010 | −0.011 | −0.007 | 0.003 | −0.063 | −0.080 | −0.074 | −0.028 | −0.137 | −0.096 | −0.045 | −0.009 | −0.062 | −0.080 | −0.029 |
| | Axis 5 | −0.017 | −0.013 | 0.007 | −0.037 | −0.045 | 0.070 | −0.001 | 0.027 | 0.012 | −0.025 | −0.012 | −0.038 | −0.029 | 0.006 | −0.032 | −0.013 |
| Zaire-96-I-16 | Axis 1 | 0.137 | 0.176* | 0.140 | 0.126 | 0.170* | 0.185* | 0.201** | 0.166* | 0.196* | 0.186* | 0.099 | 0.138 | 0.089 | 0.182* | 0.180* | 0.038 |
| | Axis 2 | 0.053 | 0.072 | 0.034 | 0.015 | 0.098 | 0.078 | −0.003 | 0.159 | 0.074 | 0.130 | 0.103 | 0.084 | −0.033 | 0.111 | 0.185* | −0.020 |
| | Axis 3 | 0.149 | 0.155 | 0.107 | 0.167* | 0.115 | 0.122 | 0.091 | 0.145 | 0.153 | 0.034 | 0.178* | 0.152 | 0.152 | 0.137 | 0.170* | 0.144 |
| | Axis 4 | −0.014 | −0.011 | 0.020 | −0.030 | −0.019 | 0.112 | 0.054 | 0.077 | 0.024 | 0.089 | 0.084 | 0.014 | −0.029 | 0.056 | 0.036 | 0.004 |
| | Axis 5 | 0.163* | 0.061 | 0.091 | 0.133 | −0.020 | −0.002 | 0.069 | 0.078 | 0.085 | −0.054 | 0.007 | 0.051 | 0.184* | 0.050 | 0.008 | 0.171* |

*P < .05.
**P < .01.
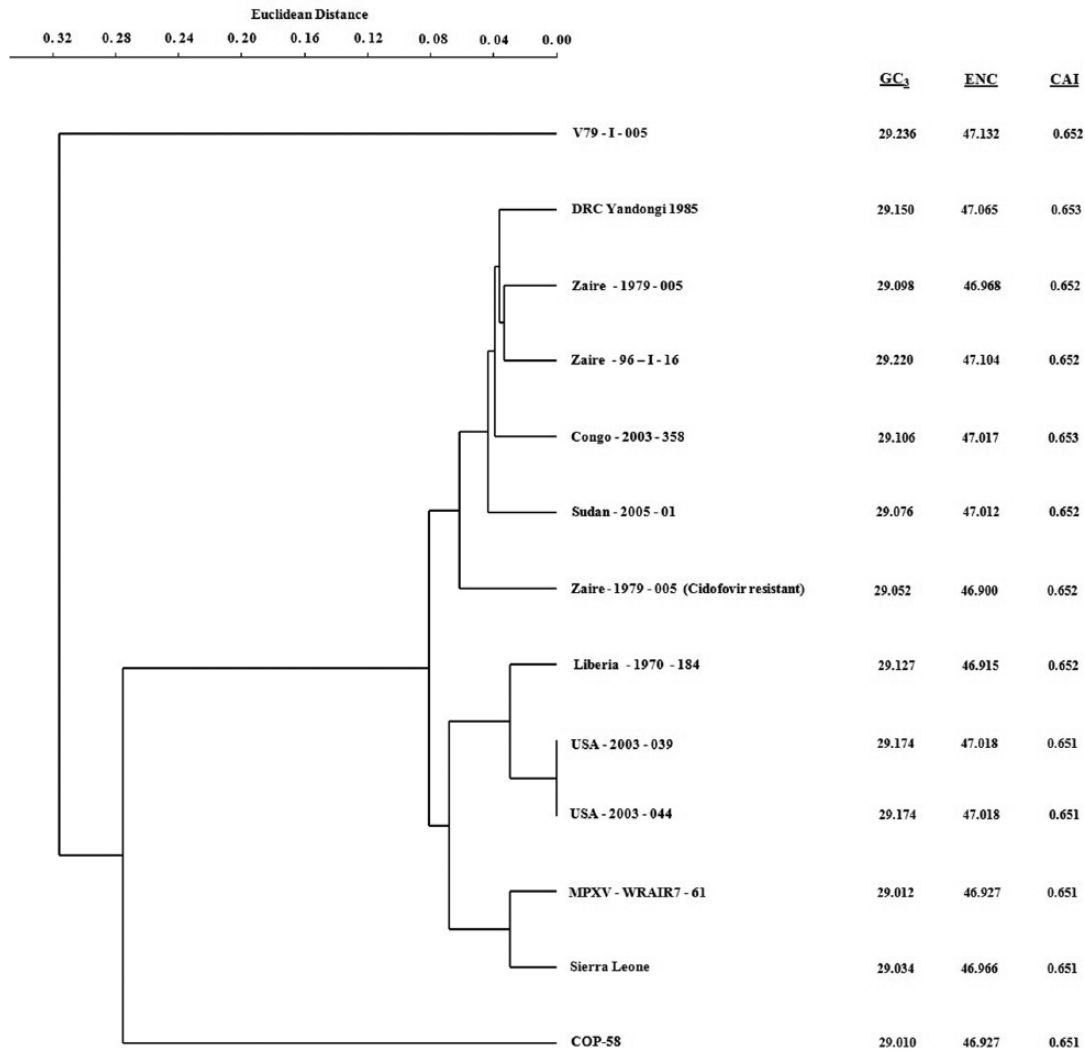
**Figure 4.** Relationship between synonymous codon usage bias and virulence. The cluster analysis grouped more virulent strains into one major cluster (upper cluster) and less virulent strains into another cluster (lower cluster). CAI indicates codon adaptation index; ENC, effective number of codons.

scenarios, which form the basis for developing this phenomenon, are coevolution of humans and viruses infecting humans and/or evolution of a human genome from a viral genome.[73]

Significant intragenomic variations in the ENC (SD > 4.0) and the $GC_3$ (SD > 4.0) values were observed in all the MPXV genomes used in this research. This heterogeneity in the base composition suggests that base compositional constraints play an important role in shaping SCUBs in MPXV genomes. A similar heterogeneity in the base composition was reported in herpesviruses belonging to the family Poxviridae.[7] Strand-specific codon usage was observed in MPXV genomes, whereas in the host genome, tissue-specific codon usage was reported; that is, in humans, the SCUBs of brain-specific, liver-specific, uterus-specific, testis-specific, ovary-specific, and vulva-specific genes were different from one another.[74] The SCUB in an MPXV may not be due to the GC composition as no correlations were observed between the $GC_3$ and the cumulative GC values at the first and the second codon positions. However, AT richness is directly linked with SCUB as most preferred codons were A/T ending. Gene length was weakly correlated with

different COA axes in some MPXV genomes, for example, the West African genomes COP-58, MPXV-VRAIR7-61, and Sierra Leone with axis 3, and the Central African genomes V79-I-005 and Zaire-1979-005 with axis 1. In addition, based on our analysis using axis 1 of the COA (the principal axis explaining most of the variations), we suggest that gene length may have a significant influence on SCUB only in Central African strains such as V79-I-005 and Zaire-1979-005.

All putative optimal codons were found to be A/T ending as MPXV genomes are AT rich and GC poor. In MPXV genomes, genome-specific preference toward a certain subset of codons was observed. Four codons (GGA, GGT, TAT, and TTT) were used as optimal codons in most MPXV genomes, although some exceptions occurred. The overrepresentation of AT contents and the underrepresentation of GC contents in the MPXV genomes seem to be the reason behind the use of A/T-ending codons, rather than natural selection, being preferred by the host. The weak codon bias of most genes across all examined MPXV genomes suggests that selection for translational accuracy and speed has less influence in dictating SCUB,

revealing an inability to act as expression vectors, as reported in herpesviruses, another class of large double-stranded DNA viruses.[7] However, the putative optimal codons identified in this study can be used for enhancing heterologous gene expression by increasing translational efficiency.[7,75-78] Axis 1 of the COA and the CAI exhibited significant positive correlations in all examined MPXV genomes ($P < .01$), indicating that gene expression levels have profound influences on SCUB.

Although no dinucleotide contents were found to be in high correlation with axis 1 of the COA in any of the examined MPXV genomes, AT dinucleotides were overrepresented, whereas GC dinucleotides were underrepresented in all genomes; AT, GA, and TC dinucleotides were most biased as their ρ values were greater than 1.10. Because GC dinucleotides possess the highest thermodynamic stacking energy,[23,79,80] viral genomes are always under selection pressure to decrease the GC dinucleotide frequency[20,79,81] to enhance viral genome replication and transcription.[79] Unmethylated GC in viral genomes stimulates immune responses in the host.[82] Hence, to reduce antiviral responses from the host, viral genomes contain fewer GC dinucleotides.[20] The Spearman rank correlation analysis revealed high positive correlations between $C_3$ and $GC_3$ and the principal axis (axis 1) of the COA and a significant negative correlation between $T_3$ and axis 1. These correlations suggest that base compositional constraints play a crucial role in dictating SCUB. Axis 1 was not correlated with aromaticity in any MPXV, indicating that aromatic amino acids do not have a special role in framing SCUB, which further reveals that all amino acids contribute to SCUB.

Protein hydrophobicity scores were weakly correlated with axis 1 in Liberia-1970-184. Moreover, Central African and West African MPXV genomes are genetically distinct.[47] Cluster analysis showed clustering of Central African strains and one North African MPXV strain (Sudan-2005-01) into an upper cluster with similar SCUBs, whereas other strains isolated from West Africa and the United States formed a lower cluster with similar SCUBs. However, the lower cluster revealed that the US-isolated MPXVs possessed similar SCUBs as they are in one clade close to Liberia-1970-184. Furthermore, Central African strains have been reported to be more virulent than West African strains.[47] Based on these results, we are able to postulate that a strong association exists between MPXV strain virulence and SCUB as more virulent strains formed one cluster exhibiting similar SCUBs, and less virulent strains formed another. Thus, we conclude that mutational pressure due to base compositional constraints, level of gene expression, and codon selection for utilization of putative optimal codons are major factors influencing the SCUB in MPXV genomes. Consequently, a balance exists between mutational pressure acting on nucleotide sequences and amino acid selection in MPXV genomes, which is similar to the finding in a report on hepatitis E viruses.[1] Generally, to conserve the protein sequence, purifying selection eliminates transversions at the third codon

positions in 2-fold degenerate amino acids. Among the 20 amino acids, most synonymous positions are in 2-fold degenerate amino acids. Hence, selection may act on an amino acid level to eliminate the possibility of nonsynonymous transversions in 2-fold degenerate amino acids. In addition, viral genomes have naturally evolved with a mechanism to tackle and escape host antiviral responses,[28] and according to the evolution rhetoric theory,[83] this mechanism may also act as a major selection pressure in framing the SCUB in MPXV genomes, as reported in hepatitis A viral genomes.[28] In this context, the multifactorial codon usage bias in MPXV genomes might have evolved as the result of a need to increase the efficiency of communication from the genome to the cell in transitional environments by keeping the message unmodified.[28,83]

## Author Contributions

RRN conceived the idea and designed the methodology. SK, NTR, and GD performed the analyses. SK, RRN, VRD, GD and SKNS interpreted the results. RRN wrote the manuscript. GD, SKNS and VRD offered critical comments. RRN and VRD developed the final draft. All authors read and approved the final manuscript.

## REFERENCES

1. Bouquet J, Cherel P, Pavio N. Genetic characterization and codon usage bias of full-length hepatitis E virus sequences shed new lights on genotypic distribution, host restriction and genome evolution. *Infect Genet Evol*. 2012;12:1842–1853. doi:10.1016/j.meegid.2012.07.021.
2. Sharp PM, Emery LR, Zeng K. Forces that influence the evolution of codon bias. *Phil Trans R Soc B*. 2010;365:1203–1212. doi:10.1098/rstb.2009.0305.
3. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Struct Dynam*. 2004;21:527–536. doi:10.1080/07391102.2004.10506946.
4. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*. 1980;8:49–62.
5. Sharp PM, Li WH. The codon Adaptation Index: a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acid Res*. 1987;15:1281–1295. doi:10.1093/nar/15.3.1281.
6. Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res*. 2003;92:1–7. doi:10.1016/S0168-1702(02)00309-X.
7. Roychoudhury S, Mukherjee D. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res*. 2010;148:31–43. doi:10.1016/j.virusres.2009.11.018.
8. Sablok G, Nayak KC, Vazquez F, Tatarinova TV. Synonymous codon usage, GC(3), and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Mol Biotechnol*. 2011;49:116–128. doi:10.1007/s12033-011-9383-9.
9. Selva KC, Nair RR, Sivaramakrishnan KG, et al. Influence of certain forces on evolution of synonymous codon usage bias in certain species of three basal orders of aquatic insects. *Mitochondr DNA*. 2012;23:447–460. doi:10.3109/19401736.2012.710203.
10. Nair RR, Nandhini MB, Monalisha E, et al. Synonymous codon usage in chloroplast genome of *Coffea arabica*. *Bioinformation*. 2012;8:1096–1104.
11. Nair RR, Nandhini MB, Sethuraman T, Doss G. Mutational pressure dictates synonymous codon usage in freshwater unicellular α—cyanobacterial

descendant *Paulinella chromatophora* and β—cyanobacterium *Synechococcus elongatus* PCC6301. *SpringerPlus*. 2013;2:492. doi:10.1186/2193-1801-2-492.

12. Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiate highly and lowly expressed genes. *Nucleic Acids Res*. 1986;14:8207–8211. doi:10.1093/nar/14.13.5125.

13. Moriyama EN, Powell JR. Codon usage bias and tRNA abundance in Drosophila. *J Mol Evol*. 1997;45:514–523. doi:10.1007/PL00006256.

14. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res*. 1982;10:7055–7074. doi:10.1093/nar/10.22.7055.

15. Stenico M, Lloyd AT, Sharp PM. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res*. 1994;22:2437–2446. doi:10.1093/nar/22.13.2437.

16. Ikemura T. Correlation between the abundance of yeast tRNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol*. 1982;158:573–579. doi:10.1016/0022-2836(82)90250-9.

17. Zama M. Codon usage and secondary structure of mRNA. *Nucleic Acids Symp Ser*. 1990;22:93–94.

18. Xia X. Maximizing transcription efficiency causes codon usage bias. *Genetics*. 1996;144:1309–1320.

19. Oresic M, Shalloway D. Specific correlations between relative synonymous codon usage and protein secondary structure. *J Mol Biol*. 1998;281:31–48. doi:10.1006/jmbi.1998.1921.

20. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol*. 2006;62:551–563. doi:10.1007/s00239-005-0221-1.

21. Kress C, Thomassin H, Grange T. Local DNA methylation in vertebrates, how could it be performed and targeted? *FEBS Lett*. 2001;494:135–140. doi:10.1016/S0014-5793(01)02328-6.

22. Beutler E, Gelbart T, Han J, Koziol JA, Beutler B. Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. *Proc Natl Acad Sci U S A*. 1989;86:192–196.

23. Breslauer KJ, Frank R, Blocker H, Marky LA. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A*. 1986;83:3746–3750.

24. Liu YS, Zhou JH, Chen HT, et al. The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. *Infect Genet Evol*. 2011;11:1168–1173.

25. Liu XS, Zhang YG, Fang YZ, Wang YL. Patterns and influencing factions of synonymous codon usage in porcine circovirus. *Virol J*. 2012;9:68. doi:10.1186/1743-422X-9-68.

26. Zhou JH, Gao ZL, Zhang J, et al. Comparative the codon usage between the three main viruses in pestivirus genus and susceptible livestock. *Virus Genes*. 2012;44:475–481. doi:10.1007/s11262-012-0731-z.

27. Pandit A, Sinha S. Differential trends in the codon usage patterns in HIV-1 genes. *PLoS ONE*. 2011;6:28889. doi:10.1371/journal.pone.0028889.

28. D'Andrea L, Pinto RM, Bosch A, Musto H, Cristina J. A detailed comparative analysis on the overall codon usage patterns in hepatitis A virus. *Virus Res*. 2011;157:19–24. doi:10.1016/j.virusres.2011.01.012.

29. Shackelton LA, Holmes EC. The evolution of large DNA viruses, combining genomic information of viruses and their hosts. *Trends Microbiol*. 2004;12:458–465. doi:10.1016/j.tim.2004.08.005.

30. Chen Y. A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: quantifying the relative importance of mutational pressure and natural selection. *Biomed Res Int*. 2013;2013:406342. doi:10.1155/2013/406342.

31. Strauss EG, Strauss JH, Levin AJ. Virus evolution. In: Fields BN, Knipe DM, Howley PM, eds. *Virology*. Philadelphia, PA: Lippincott-Raven; 1996:153–171.

32. Zhao KN, Liu WJ, Frazer IH. Codon usage bias and A+T content variation in human papillomavirus genomes. *Virus Res*. 2003;98:95–104.

33. Shchelkunov SN, Totmenin AV, Safronov PF, et al. Analysis of the monkeypox virus genome. *Virology*. 2002;297:172–194. doi:10.1006/viro.2002.1446.

34. Moss B. Poxviridae: the viruses and their replication. In: Knipe DM, Howley PM eds. *Fields Virology*. Philadelphia, PA: Lippincott; 2001:2849–2883.

35. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID. *Smallpox and Its Eradication*. Geneva, Switzerland: World Health Organization; 1988. doi:10.1163/182539189X01076.

36. Jezek Z, Fenner F. Human monkeypox. *Monogr Virol*. 1988;17:1–140.

37. Shchelkunov SN, Totmenin AV, Babkin IV, et al. Human monkeypox and smallpox viruses: genomic comparison. *FEBS Lett*. 2001;509:66–70. doi:10.1006/viro.2002.1446.

38. Brown K, Leggat PA. Human monkeypox: current state of knowledge and implications for the future. *Trop Med Infect Dis*. 2016;1:8. doi:10.3390/tropicalmed1010008.

39. Nolen L, Osadebe L, Katomba J, et al. Extended human-to-human transmission during a monkeypox outbreak in the Democratic Republic of the Congo. *Emerg Infect Dis*. 2016;22:1014–1021.

40. Khodakevich L, Manbu MD, Szczeniowski M, et al. The role of squirrels in sustaining monkeypox virus transmission. *Trop Geogr Med*. 1987;39:115–122.

41. Khodakevich L, Szczeniowski M, Manbu MD, et al. Monkeypox virus in relation to the ecological features surrounding human settlements in Bumba zone, Zaire. *Trop Geogr Med*. 1987;39:56–63.

42. Hutin YJ, Williams RJ, Malfait P, et al. Outbreak of human monkeypox, democratic republic of Congo, 1996 to 1997. *Emerg Infect Dis*. 2001;7:434–438.

43. Weaver JR, Isaacs SN. Monkeypox virus and insights into its immunomodulatory proteins. *Immunol Rev*. 2008;225:96–113.

44. Hammarlund E, Lewis MW, Carter SV, et al. Multiple diagnostic techniques identify previously vaccinated individuals with protective immunity against monkeypox. *Nat Med*. 2005;11:1005–1011.

45. Giulio DDB, Eckburg PB. Human monkeypox: an emerging zoonosis. *Lancet Infect Dis*. 2004;4:15–25.

46. Nalca A, Rimoin AW, Bavari S, Whitehouse CA. Reemergence of monkeypox: prevalence, diagnostics, and countermeasures. *Clin Infect Dis*. 2005;41:1765–1771.

47. Chen N, Li G, Liszewski MK, et al. Virulence differences between monkeypox virus isolates from West Africa and the Congo Basin. *Virology*. 2005;340:46–63. doi:10.1016/j.virol.2005.05.030.

48. Kugelman JR, Johnston SC, Mulembakani PM, et al. Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. *Emerg Infect Dis*. 2014;20:232–239.

49. Koren S, Phillippy AM. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol*. 2015;23:110–120.

50. Zhao K, Wohlhueter RM, Li Y. Finishing monkeypox genomes from short reads: assembly analysis and a neural network method. *BMC Genomics*. 2016;17:497.

51. Likos AM, Sammons SA, Olson VA, et al. A tale of two clades: monkeypox viruses. *J Gen Virol*. 2005;86:2661–2672.

52. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus genome evolution: the role of gene loss. *Viruses*. 2010;2:1933–1967.

53. Bahar MW, Graham SC, Chen RA, et al. How vaccinia virus has evolved to subvert the host immune response. *J Struct Biol*. 2011;175:127–134.

54. Trindade GDS, Emerson GL, Sammons S, et al. Serro 2 virus highlights the fundamental genomic and biological features of a natural vaccinia virus infecting humans. *Viruses*. 2016;8:328.

55. Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol Biol Rep*. 2009;36:2039–2046. doi:10.1007/s11033-008-9414-1.

56. Wright F. The "effective number of codons" used in a gene. *Gene*. 1990;87:23–29. doi:10.1016/0378-1119(90)90491-9.

57. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;5:157:105–132. doi:10.1016/0022-2836(82)90515-0.

58. Zhang J, Wang M, Liu WQ, et al. Analysis of codon usage and nucleotide composition bias in polioviruses. *Virology J*. 2011;8:146. doi:10.1186/1743-422X-8-146.

59. Hammer Q, Harper DAT, Ryan PD. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 2001;4:1–9.

60. Peden JF. *Analysis of codon usage* [PhD thesis]. Nottingham, UK: University of Nottingham; 1999.

61. Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–2739. doi:10.1093/molbev/msr121.

62. Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol*. 2013;30:1720–1728. doi:10.1093/molbev/mst064.

63. Umashankar V, Arunkumar V, Dorairaj S. ACUA: a software tool for automated codon usage analysis. *Bioinformation*. 2007;2:62–63.

64. Karumathil S, Dirisala VR, Srinadh U, Nikhil V, Kumar NS, Nair RR. Evolution of synonymous codon usage in the mitogenomes of certain species of bilaterian lineage with special reference to Chaetognatha. *Bioinform Biol Insights*. 2016;10:167–184. doi:10.4137/BBI.S38192.

65. Sueoka N. Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci U S A*. 1988;85:2653–2657.

66. Rudner R, Karkas JD, Chargaff E. Seperation of *B. subtilis* DNA into complementary strands. 3. Direct analysis. *Proc Natl Acad Sci U S A*. 1968;60:921–922.

67. Sanchez G, Bosch A, Pinto RM. Genome variability and capsid structural constraints of hepatitis A virus. *Virol*. 2003;77:452–459.

68. Pinto RM, Aragone L, Costafreda MI, Ribes E, Bosch A. Codon usage and replicative strategies of hepatitis A virus. *Virus Res*. 2007;127:158–163.

69. Grant RJ, Baldwin CD, Nalca A, et al. Application of the Ibis-T5000 pan-Orthopoxvirus assay to quantitatively detect monkeypox viral loads in clinical specimens from macaques experimentally infected with aerosolized monkeypox virus. *Am J Trop Med Hyg*. 2010;82:318–323. doi:10.4269/ajtmh.2010.09-0361.

70. Hammarlund E, Dasgupta A, Pinilla C, et al. Monkeypox virus evades antiviral CD4+ and CD8+ T cell responses by suppressing cognate T cell activation. *Proc Natl Acad Sci U S A*. 2008;105:14567–14572. doi:10.1073/pnas.0800589105.

71. Rubins KH, Hensley LE, Relman DA, Brown PO. Stunned silence: gene expression programs in human cells infected with monkeypox or vaccinia virus. *PLoS ONE*. 2011;6:e15615. doi:10.1371/journal.pone.0015615.

72. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Sys Biol*. 2009;5:311.

73. Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science*. 2004;303:1626–1632.

74. Plotkin J, Robins H, Levine AJ. Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci U S A*. 2004;101:12588–12591.

75. Trapp A, Einem JV, Hofmann H, et al. Potential of equine herpesvirus 1 as a vector for immunization. *J Virol*. 2005;54:45–54. doi:10.1128/JVI.79.9.5445-5454.2005.

76. Donofrio G, Sartori C, Ravanetti L, et al. Establishment of a bovine herpesvirus 4 based vector expressing a secreted form of the bovine viral diarrhoea virus structural glycoprotein E2 for immunization purposes. *BMC Biotechnol*. 2007;7:68. doi:10.1186/1472-6750-7-68.

77. Jazayeri M, Soleimanjahi H, Fotouhi F, Pakravan N. Comparison of intramuscular and footpad subcutaneous immunization with DNA vaccine encoding HSV-GD2 in mice. *Comp Immun Microbiol Infect Dis*. 2009;32:453–461.

78. Han JH, Choi YS, Kim WJ, et al. Codon optimization enhances protein expression of human peptide deformylase in *E. coli*. *Protein Expr Purif*. 2009;70:224–230.

79. Zhou J, Gao Z, Zhang J, et al. The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. *Infect Genet Evol*. 2013;14:105–110. doi:10.1016/j.meegid.2012.09.020.

80. Delcourt SG, Blake RD. Stacking energies in DNA. *J Biol Chem*. 1991;266:15160–15169.

81. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog*. 2008;4:e1000079. doi:10.1371/journal.ppat.1000079.

82. Woo PC, Wong BH, Huang Y, et al. Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in coronaviruses. *Virology*. 2007;369:431–442.

83. Vetsigian K, Goldenfeld N. Genome rhetoric and the emergence of compositional bias. *Proc Natl Acad Sci U S A*. 2009;106:215–220. doi:10.1073/pnas.0810122106.